Chapter 1

Preliminaries

1.1 Index notation

We follow *Einstein notation* and use regular font for both scalars and tensors, whose types are to be inferred from context. For example, a vector v expressed in terms of a basis g_i is written as

$$v = v^i g_i. \tag{1.1}$$

In most cases, we make no distinction between *vectors* and *covectors*, as a Cartesian basis is assumed unless stated otherwise. Accordingly, we also write $v = v^i g_i = v_i g^i = v_i g_i$. Similarly, the *inner product* between two geometric vectors is given by

$$(a, b) := a \cdot \overline{b} = a^i \overline{b_i} = a_i \overline{b_i}, \tag{1.2}$$

where $\cdot$ denotes *(single) contraction* and $\overline{(\,)}$ indicates complex conjugation of the subtended quantity.

1.2 Inner product and induced norm

A *set* equipped with inner product is called an *inner product space*. An inner product is a *bilinear form* (or a *sesquilinear form* for complex vectors) such that

$$(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{F}. \tag{1.3}$$

Here $\mathbb{F}$ is either a real number $\mathbb{R}$ or a complex number $\mathbb{C}$. Thus, an inner product takes two vectors in $\mathcal{V}$ and returns a scalar.

In addition, an inner product must satisfy the following properties [Oden and Demkowicz, 2017]:

- *Linearity* with respect to the first argument

$$(\alpha u + \beta v, w) = \alpha\,(u, w) + \beta\,(v, w) \quad \forall \alpha, \beta \in \mathbb{F},\ \forall u, v, w \in \mathcal{V}. \tag{1.4}$$

- *Conjugate symmetry*

$$(u, v) = \overline{(v, u)} \quad \forall u, v \in \mathcal{V}. \tag{1.5}$$

- *Positive definiteness*

$$(u,u) > 0 \quad \forall u \neq 0,\ u \in \mathcal{V}. \tag{1.6}$$

Note that an inner product is *anti-linear* with respect to the second argument, i.e.,

$$(u, \alpha v) = \overline{\alpha}\,(u,v) \quad \forall \alpha \in \mathbb{F},\ \forall u, v \in \mathcal{V}. \tag{1.7}$$

Orthogonality is defined such that

$$(u,v) = 0. \tag{1.8}$$

Any inner product space is equipped with an *induced norm* such that

$$\|u\| \coloneqq \sqrt{(u,u)}. \tag{1.9}$$

Theorem 1.2.1 (Cauchy–Schwarz Inequality) *Let $(\cdot,\cdot)$ be an inner product on a vector space $\mathcal{V}$. Then, $\forall u, v \in \mathcal{V}$,*

$$|(u,v)| \leq \|u\|\,\|v\|\,. \tag{1.10}$$

Equality holds when u and v are linearly dependent.

Triangle inequality can be proved from the Cauchy-Schwarz inequality as

$$\begin{aligned} \|u+v\|^2 &= (u+v, u+v) \\ &= (u,u) + (u,v) + (v,u) + (v,v) \\ &= (u,u) + 2\mathrm{Re}\,\{(u,v)\} + (v,v) \\ &\leq \|u\|^2 + 2\,|(u,v)| + \|v\|^2 \\ &\leq \|u\|^2 + 2\,\|u\|\,\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2\,. \end{aligned} \tag{1.11}$$

Thus, we have $\|u+v\| \leq \|u\| + \|v\|$.

1.3 Function space

1.3.1 Banach space

- *Linear space*: A set X is a linear space over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$) if its elements $x, y, \ldots \in X$ satisfy the following axioms under addition and scalar multiplication by $\alpha, \beta, \ldots \in \mathbb{F}$:

  1. $x + y = y + x$ (commutativity of addition);
  2. $(x+y)+z = x+(y+z)$ (associativity of addition);
  3. There exists a zero element $0 \in X$ such that $x + 0 = x$ for all $x \in X$;
  4. For each $x \in X$, there exists an additive inverse $-x \in X$ such that $x + (-x) = 0$;
  5. $1 \cdot x = x$ for all $x \in X$ (scalar identity);

6. $\alpha(\beta x) = (\alpha\beta)x$ (associativity of scalar multiplication);
7. $(\alpha + \beta)x = \alpha x + \beta x$ (distributivity over scalar addition); *and*
8. $\alpha(x + y) = \alpha x + \alpha y$ (distributivity over vector addition).

- *Normed linear space*: A linear space X is said to be *normed* if each element $x \in X$ is assigned a nonnegative real number $\|x\|$, called the *norm* of x, such that
  1. $\|x\| = 0$ if and only if $x = 0$ (positive definiteness);
  2. $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{F}$ and $x \in X$ (absolute homogeneity); *and*
  3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$ (triangle inequality).
- *Banach space*: A normed linear space X is said to be a *Banach space* if it is complete with respect to its norm. That is, every Cauchy sequence (x_n) in X converges to a limit $x \in X$ such that $\lim_{n\to\infty} \|x_n - x\| = 0$.

A Cauchy sequence is a sequence whose elements become arbitrarily close to each other as the sequence progresses. Namely, a sequence (x_n) is Cauchy if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\|x_m - x_n\| < \varepsilon$ for all $m, n \geq N$.

- C^0 *space*: A function $u : \Omega \to \mathbb{R}$ belongs to $C^0(\Omega)$ if it is continuous on Ω, i.e.,

$$\lim_{x\to x_0} u(x) = u(x_0), \quad \forall x_0 \in \Omega. \tag{1.12}$$

  When equipped with the supremum norm,

$$\|u\|_{C^0(\Omega)} = \sup_{x\in\Omega} |u(x)|, \tag{1.13}$$

  the space $C^0(\Omega)$ is a Banach space.

- C^1 *space*: A function $u : \Omega \to \mathbb{R}$ belongs to $C^1(\Omega)$ if the function and all its first-order partial derivatives are continuous on Ω. The space $C^1(\Omega)$ is a Banach space when equipped with the C^1 norm,

$$\|u\|_{C^1(\Omega)} = \|u\|_{C^0(\Omega)} + \|\operatorname{grad} u\|_{C^0(\Omega)}. \tag{1.14}$$

  Here, $\|\operatorname{grad} u\|_{C^0(\Omega)} = \sup_{x\in\Omega} |\operatorname{grad} u(x)|$. Note that the gradient should be understood in the strong (classical) sense.

- C^k *space*: A function $u : \Omega \to \mathbb{R}$ belongs to $C^k(\Omega)$ if all partial derivatives up to order k are continuous.
- C^∞ *space*: A function $u : \Omega \to \mathbb{R}$ belongs to $C^\infty(\Omega)$ if all partial derivatives of all orders are continuous. Such functions are called *smooth* or *infinitely differentiable*.

1.3.2 Hilbert space

- *Inner product space*: A linear space X is said to be an *inner product space* if each pair of elements $x, y \in X$ is assigned a scalar $(x, y) \in \mathbb{F}$, called the *inner product* of x and y, such that
  1. $(x, x) \geq 0$ and $(x, x) = 0$ if and only if $x = 0$ (positive definiteness);
  2. $(x, y) = \overline{(y, x)}$ for all $x, y \in X$ (conjugate symmetry);

3. $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$ for all $\alpha, \beta \in \mathbb{F}$ and $x, y, z \in X$ (linearity in the first argument).

The inner product induces a norm via $\|x\| = \sqrt{(x, x)}$.

- *Hilbert space*: An inner product space X is said to be a *Hilbert space* if it is complete with respect to the norm induced by its inner product, $\|x\| = \sqrt{(x, x)}$. Equivalently, a Hilbert space is a Banach space whose norm derives from an inner product.

- L^2 *space*: For a domain $\Omega \subset \mathbb{R}^n$, the space $L^2(\Omega)$ is defined as

$$L^2(\Omega) = \left\{ u : \Omega \to \mathbb{R} \text{ (or } \mathbb{C}) : \int_\Omega |u|^2 \, dx < \infty \right\}. \tag{1.15}$$

The space $L^2(\Omega)$ is a Hilbert space equipped with the inner product

$$(u, v)_{L^2(\Omega)} = \int_\Omega u \cdot \overline{v} \, dx \tag{1.16}$$

and induced norm

$$\|u\|_{L^2(\Omega)} = \sqrt{(u, u)_{L^2(\Omega)}} = \left(\int_\Omega |u|^2 \, dx \right)^{1/2}. \tag{1.17}$$

- *Sobolev space of the first order*:

$$H^1(\Omega) := \left\{ u \in L^2(\Omega) \ : \ \operatorname{grad} u \in L^2(\Omega)^3 \right\}. \tag{1.18}$$

The corresponding inner product is

$$\begin{aligned} (u, v)_{H^1(\Omega)} &= (u, v) + (\operatorname{grad} u, \operatorname{grad} v) \\ &= \int_\Omega u\overline{v} + \int_\Omega \operatorname{grad} u \cdot \overline{\operatorname{grad} v} \end{aligned} \tag{1.19}$$

and the induced norm is

$$\|u\|_{H^1(\Omega)} = \sqrt{(u, u)_{H^1(\Omega)}}. \tag{1.20}$$

Note that the gradient must be understood in the weak (distributional) sense.

- *Sobolev space of the k-th order*:

$$H^k(\Omega) := \left\{ u \in L^2(\Omega) \ : \ D^\alpha u \in L^2(\Omega), \ \forall |\alpha| \le k \right\}. \tag{1.21}$$

Here, $D^\alpha = \partial^{|\alpha|} / (\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n})$ denotes partial derivative of order $|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n$ in the weak sense.

- $H(\operatorname{div})$ space:

$$H(\operatorname{div}, \Omega) = \left\{ u \in L^2(\Omega)^3 \ : \ \operatorname{div} u \in L^2(\Omega) \right\}. \tag{1.22}$$

- H (curl) space:

$$H\left(\mathrm{curl}, \Omega\right) = \left\{ u \in L^2\left(\Omega\right)^3 \ : \ \mathrm{curl}\, u \in L^2\left(\Omega\right)^3 \right\}. \tag{1.23}$$

Exercise 1.3.1 *Let $f(x) = C$ for some constant $C \neq 0$. Determine whether f belongs to each of the following function spaces on $\Omega = \mathbb{R}$:*

1. *$C^0(\mathbb{R})$ and $C^\infty(\mathbb{R})$,*
2. *$L^2(\mathbb{R})$,*
3. *$H^k(\mathbb{R})$ for $k \geq 1$.*

1.4 Differentiation

The *directional derivative* of a map $u : X \to \mathbb{R}$ at a point $x \in X$ in the direction $h \in X$ is defined as

$$d_h u(x) \coloneqq \lim_{\epsilon \to 0} \frac{u\left(x + \epsilon h\right) - u\left(x\right)}{\epsilon}. \tag{1.24}$$

This limit probes u along the single direction h and need not be linear in h in general.

If the map $h \to d_h u(x)$ is linear and bounded, then $d_h u(x)$ is called the *Gâteaux differential* of u at x in the direction h, and we write

$$d_h u(x) = du(x)[h]. \tag{1.25}$$

Theorem 1.4.1 (Riesz Representation Theorem) *Let H be a Hilbert space with inner product $(\cdot, \cdot)_H$. For every continuous linear functional $\varphi : H \to \mathbb{R}$, there exists a unique element $f \in H$ such that*

$$\varphi(x) = \left(f,\, x\right)_H, \quad \forall x \in H. \tag{1.26}$$

Moreover, $\|\varphi\|_{H^*} = \|f\|_H$. The element f is called the Riesz representation *of φ.*

The Riesz Representation Theorem guarantees the existence of a unique element $g \in X$ such that

$$du(x)[h] = \left(g,\, h\right)_X, \quad \forall h \in X. \tag{1.27}$$

This element g is called the *gradient* of u at x, denoted $\nabla u(x)$ or $\mathrm{grad}\, u(x)$. It is the unique Riesz representation of the Gâteaux derivative $du(x)\,[h]$ in X, and depends on the choice of inner product on X.

The function u is said to be *Fréchet differentiable* if

$$u\left(x + h\right) - u\left(x\right) - d_h u\left(x\right) = o\left(\|u\|\right), \tag{1.28}$$

i.e.,

$$\lim_{\|h\| \to 0} \frac{|u(x + h) - u(x) - d_h u(x)|}{\|h\|} = 0. \tag{1.29}$$

The functional $d_h u(x)$ is called the *Fréchet differential* of u at x.

Definition 1.4.1 (Weak Derivative) *Let $u \in L^2(\Omega)$. A function $v \in L^2(\Omega)$ is called the* weak derivative *of u with respect to x_i, denoted $v = \partial u / \partial x_i$, if*

$$\left(v,\, \phi\right) = -\left(u,\, \frac{\partial \phi}{\partial x_i}\right), \quad \forall \phi \in C_0^\infty(\Omega). \tag{1.30}$$

More generally, for a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ with $|\alpha| = \alpha_1 + \cdots + \alpha_n$, the weak derivative $D^\alpha u$ is the function $v \in L^2(\Omega)$ satisfying

$$\left(v,\, \phi\right) = (-1)^{|\alpha|}\left(u,\, D^\alpha \phi\right), \quad \forall \phi \in C_0^\infty(\Omega), \tag{1.31}$$

where $(\,\cdot\,,\,\cdot\,) \coloneqq (\,\cdot\,,\,\cdot\,)_{L^2(\Omega)}$.

When $u \in C^1(\Omega)$, the weak derivative coincides with the classical derivative, so the weak derivative is an extension of classical differentiation to functions that may not be pointwise differentiable. This extension is fundamental to Sobolev space theory: the Sobolev space $H^k(\Omega)$ consists precisely of those functions in $L^2(\Omega)$ whose weak derivatives up to order k also belong to $L^2(\Omega)$.

Example 1.4.1 (Weak Derivative of the Absolute Value Function) *Consider $u(x) = |x|$ on $\mathbb{R}$. For any $\phi \in C_0^\infty(\mathbb{R})$,*

$$\begin{aligned}\int_{-\infty}^{\infty} |x|\,\phi'(x)\,dx &= \int_{-\infty}^{0} (-x)\,\phi'(x)\,dx + \int_{0}^{\infty} x\,\phi'(x)\,dx \\ &= [-x\phi]_{-\infty}^{0} - \int_{-\infty}^{0} -\phi(x)\,dx + [x\phi]_0^\infty - \int_0^\infty \phi(x)\,dx \\ &= -\int_{-\infty}^{\infty} \operatorname{sgn}(x)\,\phi(x)\,dx, \end{aligned} \tag{1.32}$$

where

$$\operatorname{sgn}(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases} \tag{1.33}$$

The boundary terms vanish because ϕ has compact support and the contributions at $x = 0$ cancel. Thus, the weak derivative is $v(x) = \operatorname{sgn}(x)$.

Exercise 1.4.1 (Heaviside Step Function) *Consider the Heaviside function on $\mathbb{R}$:*

$$H(x) = \begin{cases} 0, & x \le 0, \\ 1, & x > 0. \end{cases} \tag{1.34}$$

Find the weak derivative of the Heaviside step function.

1.5 Integration by parts

Integration by parts in one dimension reads

$$\int_0^1 \frac{du}{dx} v = [uv]_0^1 - \int_0^1 u \frac{dv}{dx}. \tag{1.35}$$

The multi-dimensional generalization of the above equation reads

$$\int_\Omega \frac{\partial u}{\partial x^i} v = \int_{\partial\Omega} u v n_i - \int_\Omega u \frac{\partial v}{\partial x^i}. \tag{1.36}$$

In the above, $\Omega \in \mathbb{R}^N$ is the domain in N dimension and its boundary is denoted by $\partial\Omega$, and n_i is the i-th component of the outward normal vector on the boundary.

Integration by parts for divergence and curl operators are (respectively)

$$\int_\Omega (\operatorname{div} u)\, v = \int_{\partial\Omega} (u \cdot n)\, v - \int_\Omega u \cdot \operatorname{grad} v \quad \text{and} \tag{1.37}$$

$$\int_\Omega \operatorname{curl} E \cdot F = \int_{\partial\Omega} (n \times E) \cdot F + \int_\Omega E \cdot \operatorname{curl} F. \tag{1.38}$$

In the above, u, v, E, and F are vector-valued functions and n is unit outward normal vector. Integration by parts involves with the divergence of a *two tensor*, $A \in \mathbb{R}^{N\times N}$, reads

$$\int_\Omega (\operatorname{div} A) \cdot v = \int_{\partial\Omega} (An) \cdot v - \int_\Omega A : \operatorname{grad} v, \tag{1.39}$$

where : denotes *double contraction* such that $A : B = A_{ij}B^{ij}$.

1.6 Implicit function theorem

The Implicit Function Theorem is a fundamental result in nonlinear analysis with wide-ranging applications, including the derivation of first-order optimality conditions via Lagrange multipliers and sensitivity analysis of PDE-constrained optimization problems. We state the theorem first in the finite-dimensional setting and then in the general Banach space setting.

Theorem 1.6.1 (Implicit Function Theorem) *Let $U \subset \mathbb{R}^n \times \mathbb{R}^m$ be an open set and let $g : U \to \mathbb{R}^m$ be a C^1 function. Suppose there exists a point $(x_0, y_0) \in U$ such that*

$$g(x_0, y_0) = 0. \tag{1.40}$$

If the partial Jacobian of g with respect to y, evaluated at (x_0, y_0), is non-singular, i.e.,

$$\det\left(\frac{\partial g_i}{\partial y_j}(x_0, y_0)\right) \neq 0, \quad i, j = 1, 2, \ldots, m, \tag{1.41}$$

then there exist open neighborhoods $B \subset \mathbb{R}^n$ of x_0 and $V \subset \mathbb{R}^m$ of y_0, and a unique C^1 function $\varphi : B \to V$ such that $\varphi(x_0) = y_0$ and

$$g(x, \varphi(x)) = 0, \quad \forall x \in B. \tag{1.42}$$

Moreover, the Jacobian of φ is given by

$$D\varphi(x) = -\left[D_y\, g(x, \varphi(x))\right]^{-1} D_x\, g(x, \varphi(x)), \quad \forall x \in B, \tag{1.43}$$

or equivalently in component form,

$$\frac{\partial \varphi_i}{\partial x_j}(x) = -\sum_{k=1}^{m} F^{-1}_{ik}(x, \varphi(x)) \frac{\partial g_k}{\partial x_j}(x, \varphi(x)), \quad \forall x \in B, \tag{1.44}$$

where $F_{ij} := \partial g_i / \partial y_j$.

$D_y\, g(x_0, y_0)$ denotes the Fréchet derivative of g with respect to y, evaluated at (x_0, y_0).

Example 1.6.1 (Unit Circle) *Consider the unit circle defined implicitly by*

$$g(x, y) = x^2 + y^2 - 1 = 0. \tag{1.45}$$

This is a classic example because the circle cannot be expressed as a single global function $y = \varphi(x)$ over all of $\mathbb{R}$, yet the implicit function theorem (IFT) guarantees a local function exists near any point where the hypotheses are satisfied.

Here $g : \mathbb{R}^2 \to \mathbb{R}$, so $n = 1$ and $m = 1$ in Theorem 1.6.1. The partial derivative with respect to y is

$$\frac{\partial g}{\partial y}(x, y) = 2y. \tag{1.46}$$

The IFT applies at any point (x_0, y_0) on the circle where

$$\frac{\partial g}{\partial y}(x_0, y_0) = 2y_0 \neq 0, \tag{1.47}$$

i.e., at any point except $(\pm 1, 0)$.

Take the point $(x_0, y_0) = (0, 1)$. Since $\partial g/\partial y = 2 \neq 0$ there, the IFT guarantees the existence of a neighborhood $B \subset \mathbb{R}$ of $x_0 = 0$ and a unique C^1 function $\varphi : B \to \mathbb{R}$ such that $\varphi(0) = 1$ and

$$g(x, \varphi(x)) = x^2 + \varphi(x)^2 - 1 = 0, \quad \forall x \in B. \tag{1.48}$$

Explicitly, this function is the upper semicircle

$$\varphi(x) = \sqrt{1 - x^2}, \quad x \in (-1, 1). \tag{1.49}$$

Similarly, near $(x_0, y_0) = (0, -1)$, the IFT gives the lower semicircle $\varphi(x) = -\sqrt{1 - x^2}$.

The IFT formula gives the derivative of φ:

$$\varphi'(x) = -\left[\frac{\partial g}{\partial y}(x, \varphi(x))\right]^{-1} \frac{\partial g}{\partial x}(x, \varphi(x)) = -\frac{2x}{2\varphi(x)} = -\frac{x}{\varphi(x)}, \tag{1.50}$$

which for $\varphi(x) = \sqrt{1 - x^2}$ gives

$$\varphi'(x) = -\frac{x}{\sqrt{1 - x^2}}. \tag{1.51}$$

At $x_0 = 0$, we obtain $\varphi'(0) = 0$, reflecting the horizontal tangent at the top of the circle.

At the points $(\pm 1, 0)$, the condition $\partial g/\partial y \neq 0$ fails and the IFT does not apply. Indeed, the circle has a vertical tangent at these points and cannot be expressed as a C^1 function of x near $x = \pm 1$.

The finite-dimensional theorem extends to Banach spaces, with the non-singularity of the partial Jacobian replaced by the bounded invertibility of the partial Fréchet derivative. This generalization is essential for applications to PDE-constrained optimization, where the state variable lives in an infinite-dimensional function space such as a Sobolev space.

Theorem 1.6.2 (Implicit Function Theorem in Banach Spaces) *Let X, Y, Z be Banach spaces, $U \subset X \times Y$ open, and $g : U \to Z$ a C^1 map. Suppose $(x_0, y_0) \in U$ satisfies $g(x_0, y_0) = 0$ and the partial Fréchet derivative*

$$A := D_y\, g(x_0, y_0) : Y \to Z \tag{1.52}$$

is a Banach space isomorphism, i.e., A is a bounded bijection with bounded inverse $A^{-1} : Z \to Y$. Then there exist open neighborhoods $B \subset X$ of x_0 and $V \subset Y$ of y_0, and a unique C^1 map $\varphi : B \to V$ such that $\varphi(x_0) = y_0$,

An isomorphism between two Banach spaces is a bounded bijective linear map whose inverse is also bounded.

$$g(x,\, \varphi(x)) = 0, \quad \forall x \in B, \tag{1.53}$$

and

$$D\varphi(x) = -\left[D_y\, g(x, \varphi(x))\right]^{-1} D_x\, g(x, \varphi(x)), \quad \forall x \in B. \tag{1.54}$$