Leszek F. Demkowicz

Lecture Notes on

MATHEMATICAL METHODS IN SCIENCE AND ENGINEERING

Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Spring 2024

Contents

Preface

These lecture notes represent my second attempt to select and prepare a proper material for the first year students in our graduate CSEM* program housed in the Oden Institute at the University of Texas at Austin. The class is addressed to students who enter the program with non-math majors and it follows and builds on an earlier class covering foundations of modern mathematics, vector spaces, Lebesgue measure and integration theory, and foundations of topology and metric spaces, all with an outlook at infinite dimensional spaces but not the proper Functional Analysis yet.

This class is supposed to satisfy a number of goals essentially contradicting each other.

First of all, it is supposed to provide a quick overview of what I call *operational mathematics*: 3D calculus including curvilinear systems of coordinates and classical calculus of variations, solution of elementary ODEs and an introduction to elementary PDEs. Separation of variables and solution of linear systems of ODEs lead to the elementary spectral theory - the (not so popular) Jordan Theorem and Sturm-Liouville theory. Even the most elementary exposition to Fourier and Laplace transforms leads to the use of the Residue Theorem and the need for introducing fundamentals of complex analysis.

Secondly, the class is supposed to provide an outlook at Functional Analysis and Hilbert space methods for PDEs, i.e. distributions, energy spaces and variational formulations. And it is supposed to serve two groups of students, those that will *not* continue studying the subject, and those that *will* enroll into a two semester sequence covering those topics in depth. These contradictory goals explain the attempted style of teaching the subject. I prove only relatively simple and straightforward results delegating the proper poofs to the next, more advanced classes.

All presented material is very elementary except for a short exposition on the *Spectral Theorem for Unbounded Self-Adjoint Operators* and its connection with the separation of variables. I just could not resist demonstrating that the classical Sturm-Liouville theory and Fourier transform are parts of the same picture. Many problems have been borrowed from the excellent book of Greenberg [3].

*Computational Engineering Science and Mathematics

My special thanks go to Youguang Chen and Jiaqi Li, CSEM students and TAs for this class who did an excellent job grading homework and conducting discussion sessions. Thanks to Jonathan Zhang who helped with solutions to problems involving Matlab.

The class may or may not follow the order of exposition in the notes. In Spring 24, I presented the more advanced Section 3.6 only after discussing the more elementary examples for separation of variables from Section 6.1 that can be solved using the classical Sturm-Liouville theory. I also concluded the semester with Chapter 5 although I had to introduce some of the concepts discussed there earlier.

If you are interested in using the Lecture Notes to teach a class, please contact me to obtain the solution manual.

Leszek F. Demkowicz

Austin, Spring 2024

1

Preliminaries. Differential and Integral Calculus

1.1 Differentiation

Let X be a finite-dimensional real vector space, $\dim X = n$, equipped with an inner product $(x, y)_X$ and the corresponding Euclidean norm $\|x\|^2 = (x, x)_X$. Let e_i, $i = 1, \ldots, n$ be an orthonormal basis in X. Equivalently, you can simply think about

$$X = \mathbb{R}^n, \quad (x, y)_X = \sum_{i=1}^{n} x_i y_i, \quad \|x\|^2 = \sum_{i=1}^{n} x_i^2, \quad e_i = (0, \ldots, \underset{(i)}{1}, \ldots, 0)\,.$$

Directional and partial derivatives. Let $G \subset X$ be an open set, and $f : G \to \mathbb{R}$ a real-valued function. Let $x \in G$, and let $u \in X$ be an arbitrary non-zero vector. The *directional derivative of function f at x in the direction u* is defined as:

$$\partial_x^u f := \lim_{\epsilon \to 0} \frac{1}{\epsilon}(f(x + \epsilon u) - f(x))\,,$$

provided the limit exists. Some authors define the directional derivative using a one-sided limit,

$$\partial_x^u f := \lim_{\epsilon \to 0_+} \frac{1}{\epsilon}(f(x + \epsilon u) - f(x))\,.$$

In $\mathbb{R}^n$, the directional derivatives in the direction of canonical basis vectors are identified as partial derivatives of f,

$$\partial_i f(x) = \frac{\partial f}{\partial x_i}(x) := \partial_x^{e_i} f\,.$$

Gateaux differential and gradient. Assume that the directional derivative $\partial_x^u f$ exists for any direction u. If, additionally, function:

$$X \ni u \to \partial_x^u f \in \mathbb{R}$$

is linear, it is identified as the *Gateaux differential of function f at point x* and denoted by $d_x f$. By the definition,

$$d_x f \in X' = L(X, \mathbb{R}) \qquad d_x f(u) = \partial_x^u f\,.$$

Function f is said to be *Gateaux differentiable at x*. By the Riesz Representation Theorem, there exists a vector $v \in X$ such that

$$d_x f(u) = (v, u)_X\,.$$

Vector v defines the *gradient of function f at x*, $v = \operatorname{grad} f(x)$. If we equip $\mathbb{R}^n$ with the canonical inner product*, we get

$$\operatorname{grad} f(x) = (\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)) \,.$$

Fréchet differential. If function f is Gateaux differentiable at x and, additionally,

$$f(x+u) - f(x) - d_x f(u) = o(\|u\|) \,,$$

i.e.,

$$\frac{|f(x+u) - f(x) - d_x f(u)|}{\|u\|} \to 0 \quad \text{as} \quad \|u\| \to 0 \,,$$

$d_x f$ is called the *Fréchet derivative of function f at x*, and function f is said to be *Fréchet differentiable at x*. In a finite dimensional space X, any Gateaux differentiable function is automatically Fréchet differentiable and there is no need to differentiate between the two notions.

The derivative function. If function f is (Gateaux or, equivalently, Fréchet) differentiable at every $x \in G$, function:

$$G \ni x \to d_x f \in X'$$

is identified as the *derivative function* of f.

REMARK 1.1.1 All discussed notions generalize in a straightforward way to vector-valued functions,

$$X \supset G \ni x \to f(x) \in Y$$

where Y is another finite-dimensional vector space. In applications, typically $X = \mathbb{R}^n, Y = \mathbb{R}^m$. Most of the notions carry over also to general, infinite-dimensional Hilbert and Banach spaces (we need a normed space at minimum). In an infinite dimensional space, Gateaux and Fréchet differentiability are no longer equivalent. For most of applications that I am familiar with, Gateaux differentiability is all what you need. ∎

Higher order differentials. Differential of the derivative function is identified as the *second differential of function f*,

$$d_x^2 f := d_x f' \in L(X, X') = L(X, L(X, \mathbb{R})) \,.$$

The space $L(X, L(X, \mathbb{R}))$ is isometrically isomorphic with the space $M^2(X, \mathbb{R})$ of bilinear functionals defined on X, comp. Exercise 1.1.4. This allows us to identify the second differential as a bilinear functional on X. One can show that the second differential must be symmetric.

*Note the geometric character of the gradient as opposed to a purely algebraic character of the differential. If you change the inner product, the resulting gradient will look different.

LEMMA 1.1.1

Let function $f : X \supset G \to \mathbb{R}$ be twice differentiable at $x \in G$. Then

$$d_x^2 f(u,v) = d_x^2 f(v,u) \quad u,v \in X\,.$$

Matrix representation of the second differential in the canonical basis is identified as the *Hessian*,

$$H_{ij}(x) := d_x^2 f(e_i, e_j)\,.$$

For $X = \mathbb{R}^n$,

$$H_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x)\,.$$

The notion of second differential generalizes to an arbitrary order differentials. The k-th differential, denoted $d_x^k f$ is a symmetric k-linear functional,

$$d_x^k f \in M^k_{\text{sym}}(X,\mathbb{R}), \quad \text{i.e.} \quad d_x^k(\ldots, \underset{i}{u}, \ldots, \underset{j}{v}, \ldots) = d_x^k(\ldots, \underset{i}{v}, \ldots, \underset{j}{u}, \ldots) \quad u,v \in X\,.$$

Multiindex notation. Manipulation with higher order differentials is facilitated using the so-called *multi-index notation*. Let f be a real-valued function defined on an open set $G \subset \mathbb{R}^n$. Partial derivatives of f will be denoted by:

$$\partial^\alpha f(x) := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x)$$

where

$$\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n, \quad |\alpha| = \alpha_1 + \ldots + \alpha_n\,.$$

Multiindex representation of differentials. Let $x \in G$. The first differential of function f at x, denoted $d_x f$, is a linear functional on $\mathbb{R}^n$, $d_x f \in (\mathbb{R}^n)'$,

$$(d_x f)(y) = (d_x f)(\sum_{i=1}^n y_i e_i) = \sum_{i=1}^n \underbrace{(d_x f)(e_i)}_{=\frac{\partial f}{\partial x_i}(x)} y_i = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)\, y_i\,.$$

The second differential of function f at x, denoted $d_x^2 f$, is a bilinear, symmetric functional on $\mathbb{R}^n$, $d_x^2 f \in M^2_{\text{sym}}(\mathbb{R}^n)$,

$$\begin{aligned}(d_x^2 f)(y,y) = (d_x^2 f)(\sum_{i=1}^n y_i e_i, \sum_{j=1}^n y_j e_j) &= \sum_{i=1}^n \sum_{j=1}^n y_i y_j \underbrace{(d_x^2 f)(e_i,e_j)}_{=\frac{\partial^2 f}{\partial x_i \partial x_j}(x)} \\ &= \sum_{|\alpha|=2} \frac{2!}{\alpha_1! \cdots \alpha_n!}\, \partial^\alpha f(x)\; y_1^{\alpha_1} \cdots y_n^{\alpha_n} \\ &= \sum_{|\alpha|=2} \frac{2!}{\alpha!}\, \partial^\alpha f(x)\; y^\alpha\end{aligned}$$

where

$$y^\alpha := y_1^{\alpha_1} \cdots y_n^{\alpha_n}, \qquad \alpha! := \alpha_1! \cdots \alpha_n! \quad .$$

Notice how the multiindex notation helps us to avoid using two separate indices i and j. The k-th differential is a k-linear, symmetric functional on $\mathbb{R}^n$, $d_x^k f \in M^k_{\text{sym}}(\mathbb{R}^n)$,

$$(d_x^k f)(\underbrace{y, \ldots, y}_{k \text{ times}}) = \sum_{|\alpha|=k} \frac{k!}{\alpha!}\, \partial^\alpha f(x)\; y^\alpha \,. \tag{1.1}$$

Taylor's formula. We can write now a particular version of the Taylor formula in a very compact form, see Exercise 1.1.5,

$$f(x+y) = \sum_{k=0}^{m} \frac{1}{k!} d_x^k f(y, \ldots, y) + \frac{1}{m!} \int_0^1 (1-t)^m d_{x+ty}^{m+1} f(y, \ldots, y)\, dt \,. \tag{1.2}$$

If we define the k-th derivative of f, denoted $f^{(k)}$, as the function that for each $x \in G$, prescribes the corresponding k-th order differential at x,

$$f^{(k)} \,:\, G \ni x \to d_x^k f \in M^k_{\text{sym}}(\mathbb{R}^n) \,,$$

then we can rewrite the Taylor formula in a form resembling its 1D version,

$$f(x+y) = \sum_{k=0}^{m} \frac{1}{k!} f^{(k)}(x)(y, \ldots, y) + \frac{1}{m!} \int_0^1 (1-t)^m f^{(m+1)}(x+ty)(y, \ldots, y)\, dt \,. \tag{1.3}$$

Gradient of a vector-valued function. Let $u : X \supset G \ni x \to u(x) \in X$ be a vector-valued function. Then $d_x u \in L(X, X)$. The space $L(X, X)$ is one of possible realizations of the tensor product space $X \otimes X$, see Section 2.12 in [5]. The tensor product in this space is given by:

$$(x \otimes y)(z) = (x, z)\, y$$

where $(\cdot, \cdot)$ is the scalar product in X. The standard norm in $L(X, X)$, induced by the norm in X,

$$\|A\|_{L(X,X)} := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \,, \tag{1.4}$$

is not, in general, Euclidean (does not derive from an inner product), comp. Exercise 1.1.7. Consequently, the definition of the gradient for scalar-valued functions does not generalize in a straightforward way to vector-valued functions. In these notes, $\operatorname{grad} u$ will be simply synonymous with the differential $d_x u$.

Review Section 2.13 in [5] on multilinear algebra.

1.1.1 Necessary and Sufficient Conditions for the Existence of Local Extrema

The differentials are an indispensable tool in searching for local minima and maxima. We record now a few fundamental related results. In the following, $f : X \supset D \to \mathbb{R}$ is a real-valued function defined on an open subset D of a finite-dimensional vector space X. You can think about $X = \mathbb{R}^n$.

PROPOSITION 1.1.1 (Necessary and sufficient conditions for the existence of a local extremum)

(i) Let f be differentiable in D and possess a minimum (maximum) at $x_0 \in D$. Then

$$d_{x_0} f = 0 \,.$$

(ii) Let f be differentiable m-times in D and possess a minimum (maximum) at $x_0 \in D$. Assume, consistently with (i), that

$$d_{x_0} f = \ldots = d_{x_0}^{m-1} f = 0 \quad \text{and} \quad d_{x_0}^m f \neq 0 \,,$$

Then m must be even, and $d_{x_0}^m f$ is semi-positive (-negative) defined, i.e.,

$$d_{x_0}^m f(h, \ldots, h) \geq 0 \quad (\leq 0) \quad \forall\, h \in X \,.$$

(iii) Let f be differentiable m-times in D (with even m),

$$d_{x_0} f = \ldots = d_{x_0}^{m-1} f = 0$$

and $d_{x_0}^m f$ is positive (negative) defined, i.e.,

$$d_{x_0}^m f(h, \ldots, h) > 0 \quad (< 0) \quad \forall\, h \in X,\, h \neq 0 \,.$$

Then f possesses a local minimum (maximum) at point x_0.

The proof is delegated to Exercise 1.1.8.

1.1.2 Implicit Function Theorem

Consider the circle:

$$x^2 + y^2 = 1 \,.$$

Obviously, the circle cannot be identified globally with a graph of function but we can do it locally. For any $x_0 \in (-1, 1)$, we can identify a sufficiently small neighborhood B_{x_0} of x_0 such that for $x \in B_{x_0}$, a portion of a circle around point (x_0, y_0) on the circle, can be identified with a graph of function $y = y(x)$. For $y_0 > 0$, we have $y = \sqrt{1 - x^2}$, and for $y_0 < 0$, we have $y = -\sqrt{1 - x^2}$. We can compute then the derivative dy/dx. For instance, for $y > 0$, we have,

$$\frac{dy}{dx} = -\frac{x}{\sqrt{1 - x^2}} \,.$$

Once we know y_0 corresponding to x_0, we can compute the derivative $\frac{dy}{dx}(x_0)$ without explicitly inverting for $y(x)$. Assuming that $y = y(x)$, we have:

$$x^2 + (y(x))^2 = 1 \,.$$

Differentiating both sides in x, we obtain,

$$2x + 2y\frac{dy}{dx} = 0 \quad \Rightarrow \quad \frac{dy}{dx} = -\frac{x}{y}\,,$$

the same formula as above. Note that computation is possible only for $y \neq 0$. We cannot[†] represent y as a function of x at $x = \pm 1$. The possibility of the local inversion and the computation of the derivative of the implicitly defined function is the subject of the next classical theorem.

THEOREM 1.1.1 (Implicit Function Theorem)

Let $g(x, y) \in \mathbb{R}^m$, $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ be a C^1 function defined in a neighborhood of a point (x_0, y_0) such that

$$g(x_0, y_0) = 0\,.$$

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

$$\det\left(\frac{\partial g_k}{\partial y_i}\right) \neq 0\,,$$

then there exists a neighborhood B of point x_0 and a locally defined C^1-function $y : \mathbb{R}^n \supset B \ni x \to y(x) \in \mathbb{R}^m$ such that

$$g(x, y(x)) = 0\,.$$

The differential $d_x y(x_0)$ can be computed by differentiating the equation above,

$$\frac{\partial}{\partial x_j} g_k(x, y(x)) = \frac{\partial g_k}{\partial x_j} + \frac{\partial g_k}{\partial y_i}\frac{\partial y_i}{\partial x_j} = 0 \quad \Rightarrow \quad \frac{\partial y_i}{\partial x_j}(x_0) = -\left(\frac{\partial g_k}{\partial y_i}(x_0, y_0)\right)^{-1}\frac{\partial g_k}{\partial x_j}(x_0, y_0)\,.$$

The assumptions on regularity can be relaxed, and the theorem can be extended to infinite dimensional spaces.

1.1.3 Lagrange Multipliers

Let $f : \mathbb{R}^n \supset D \to \mathbb{R}$ be again a real-valued C^1 function defined on an open subset D of $\mathbb{R}^n$. Very often we deal with *constrained minimization problems* where the minimum (or maximum) of function f is sought not over the whole set D but a subset of D consisting of points x satisfying specific condition (constraints). In the simplest case, we ask x to satisfy an additional system of (possibly nonlinear) equations specified by an additional C^1 function:

$$g\,:\,\mathbb{R}^n \to \mathbb{R}^m, \quad m < n\,.$$

It is natural to assume that we have less constraints than unknowns so the set of potential solutions is non-empty. We seek the solution of the following constrained minimization problem:

$$\min_{x \in D,\, g(x)=0} f(x)\,.$$

[†] But we can represent x a s a function of y at those points.

We have the following fundamental result.

THEOREM 1.1.2 (Lagrange multipliers)

Let the constrained minimization problem above have a local minimum (maximum) at x_0, and let

$$\operatorname{rank}\left(\frac{\partial g_i}{\partial x_j}\right) = m\,.$$

There exists then a $\lambda_0 \in \mathbb{R}^m$ such that

$$d_{x_0} f = \lambda_0^T d_{x_0} g\,.$$

PROOF We can always reorder the components of x in such a way that $x = (z, y), z \in \mathbb{R}^m, y \in \mathbb{R}^{n-m}$, and $d_z h(x_0)$ is non-singular, i.e.,

$$\det\left(\frac{\partial g_i}{\partial z_j}\right) \neq 0\,.$$

By Theorem 1.1.1, there exists a neighborhood $N_z \times N_y$ of $x_0 = (z_0, y_0)$, and a function $h : N_y \to N_z, h(y_0) = z_0$ such that

$$g(h(y), y) = 0\,.$$

If f attains the minimum (maximum) at $x_0 = (z_0, y_0)$ then the function

$$\mathbb{R}^m \supset N_y \ni y \to f(h(y), y) \in \mathbb{R}$$

attains a minimum (maximum) at y_0. Consequently,

$$\frac{\partial f}{\partial z_i}\frac{\partial h_i}{\partial y_j} + \frac{\partial f}{\partial y_j} = 0\,. \tag{1.5}$$

By Theorem 1.1.1,

$$\frac{\partial g_k}{\partial z_i}\frac{\partial h_i}{\partial y_j} + \frac{\partial g_k}{\partial y_j} = 0 \quad \Rightarrow \quad \frac{\partial h_i}{\partial y_j} = -\left(\frac{\partial g_k}{\partial z_i}\right)^{-1}\frac{\partial g_k}{\partial y_j}\,.$$

Substituting the formula into (1.5), we obtain:

$$\frac{\partial f}{\partial y_j} - \underbrace{\frac{\partial f}{\partial z_i}\left(\frac{\partial g_k}{\partial z_i}\right)^{-1}}_{=:\lambda_k}\frac{\partial g_k}{\partial y_j}\,.$$

At the same time, trivially,

$$\frac{\partial f}{\partial z_i} = \underbrace{\frac{\partial f}{\partial z_i}\left(\frac{\partial g_k}{\partial z_i}\right)^{-1}}_{=:\lambda_k}\frac{\partial g_k}{\partial z_i}\,.$$

∎

We call λ the *Lagrange multiplier*. The usual way to execute the result is to introduce the *Lagrangian*:

$$L(x,\lambda) := f(x) - \lambda^T g(x)\,,$$

and restate the result as vanishing of differential of the Lagrangian at point (x_0,λ_0),

$$d_{(x_0,\lambda_0)}L(x,\lambda) = 0\,.$$

Equivalently, we satisfy the following system of equations at (x_0,λ_0):

$$\begin{cases} \dfrac{\partial}{\partial x_i}L(x,\lambda) = \dfrac{\partial f}{\partial x_i} - \lambda_j \dfrac{\partial g_j}{\partial x_i} = 0 & i = 1,\ldots,n \\ \dfrac{\partial}{\partial \lambda_j}L(x,\lambda) = g_j \qquad\qquad\qquad\quad = 0 & j = 1,\ldots,m\,. \end{cases}$$

Example 1.1.1

Find point $x = (x_1,x_2)$ on line $x_1 + x_2 = 1$ closest to the origin. The constrained minimization problem is:

$$\min_{x_1+x_2=1} \frac{1}{2}(x_1^2 + x_2^2)\,.$$

Note that minimizing the distance is equivalent to minimizing the half of the distance squared, and the squared (Euclidean) distance function is more regular. The Lagrangian is:

$$L(x,\lambda) = \frac{1}{2}(x_1^2 + x^2) - \lambda(x_1 + x_2 - 1)\,.$$

Differentiating in x_i, we get:

$$x_i = \lambda, \quad i = 1,2\,.$$

Substituting into the constraint equation (differentiating in λ),

$$x_1 + x_2 = 2\lambda = 1 \quad \Rightarrow \quad \lambda = \frac{1}{2}\,.$$

The solution is thus $(\frac{1}{2},\frac{1}{2})$, as expected.

□

Exercises

Exercise 1.1.1 Construct example of a function

$$\mathbb{R}^2 \ni x \to f(x) \in \mathbb{R}$$

that has all partial derivatives at some point $x = (x_1,x_2)$ but it is *not* Gateaux differentiable at this point.

(10 points)

Exercise 1.1.2 A "sanity check". Let

$$f(x) = \frac{1}{6}x_1x_2x_3, \quad x = (x_1, x_2, x_3) \in \mathbb{R}^3 .$$

Compute differentials of *all orders* at point $x_0 = (1, 1, 1)$. Represent them using the standard notation for multilinear functionals, e.g.,

$$d^3_{x_0} f(u, v, w) = \dots \qquad u, v, w \in \mathbb{R}^3 ,$$

and, for the same argument $u = v = w = y$

$$d^3_{x_0} f(y, y, y) = \dots \qquad y \in \mathbb{R}^3 ,$$

using the multiindex notation.

(5 points)

Exercise 1.1.3 Let $b \in M^2(X, \mathbb{R})$ be a bilinear form. Function:

$$X \ni x \to q(x) := b(x, x) \in \mathbb{R}$$

is termed to be a *(homogeneous) quadratic form* corresponding to bilinear form b.

- Prove that all quadratic forms form a vector space *isomorphic* to the subspace of all *symmetric* bilinear forms. In other words, construct a linear bijection between all quadratic forms and symmetric bilinear forms. This is a practical observation. For instance, if we know value of the hesian for the same argument y,

  $$d^2_x(y, y) ,$$

  we know it automatically also for different arguments u and v,

  $$d^2_x(u, v) ,$$

- Does the observation generalize to forms of arbitrary order? *Hint:* Prove first the formula for the $(m-1)$-order differential of m-order form:

  $$q(x) := a(\underbrace{x, x, \dots, x}_{m \text{ times}}) \qquad a \text{ symmetric } m\text{-linear form}$$
  $$d^{m-1}_x q(u_1, u_2, \dots, u_{m-1}) = m!\, a(x, u_1, u_2, \dots, u_{m-1}) .$$

(15 points)

Exercise 1.1.4 Let X, Z be two finite-dimensional vector spaces. Prove that the map

$$L(X, L(X, Z)) \ni A \to \{X \times X \ni (x, y) \to B(x, y) := (Ax)(y) \in Z\} \in M^2(X, Z)$$

defines a canonical isomorphism between the two spaces. Recall that norms $\|\cdot\|_X, \|\cdot\|_Z$ in spaces X, Z induce norms in the spaces of linear and bilinear functions,

$$\begin{aligned} A \in L(X,Z) \qquad & \|A\|_{L(X,Z)} = \sup_{x\neq 0} \frac{\|Ax\|_Z}{\|x\|_X} \\ B \in M^2(X,Z) \qquad & \|B\|_{M^2(X,Z)} = \sup_{x\neq 0, y\neq 0} \frac{\|B(x,y)\|_Z}{\|x\|_X \, \|y\|_X} \,. \end{aligned}$$

Prove that, with these norms, the map above is an isometry.

(5 points)

Exercise 1.1.5 Derive Taylor formula (1.3).

Outline of the proof:

(i) Use integration by parts and induction to derive Taylor formula in one dimension:

$$f(s) = \sum_{j=0}^{m} \frac{f^{(j)}(0)}{j!} s^j + \frac{s^{m+1}}{m!} \int_0^1 (1-t)^m f^{(m+1)}(ts)\, dt \,.$$

(ii) Take $f(s) := u(x+sy)$ and apply the 1D formula.

(10 points)

Exercise 1.1.6 Differentiation of product of functions. Prove the formula:

$$\partial^\alpha (fg) = \sum_{\gamma \leq \alpha} \binom{\alpha}{\gamma} \partial^\gamma f \; \partial^{\alpha-\gamma} g \,.$$

Hint: Use induction in $|\alpha|$. (10 points)

Exercise 1.1.7 Let X, Y be Hilbert real spaces. Is $L(X,Y)$ with a standard norm a Hilbert space as well ? If $Y = \mathbb{R}$, the answer is positive. By the Riesz Theorem, for every $f \in X'$, there exists a unique $x_f \in X$ such that

$$f(y) = (x_f, y) \quad y \in X \,,$$

and $\|x_f\|_X = \|f\|_{X'}$. We can use the Riesz map to transfer the inner product structure from X onto X',

$$(f,g)_{X'} := (x_f, x_g)_X \,.$$

One easily verifies the axioms for an inner product. The statement is no longer true when both spaces are of dimension greater than one. Provide a counterexample for the case of $\dim X = \dim Y = 2$.

(10 points)

Exercise 1.1.8 Prove Proposition 1.1.1. For $n = 1$, the theorem reduces to the standard 1D calculus result which you may assume to be true.

(10 points)

Exercise 1.1.9 Show that of all parallelograms with a prescribed area A, the one with the smallest perimeter L is the square. *Hint: Use Lagrange multipliers.*

(5 points)

Exercise 1.1.10 Show that of all parallelograms with a prescribed perimeter L, the one with the largest area is the square. Can you conclude the answer from Exercise 1.1.9 ?

(5 points)

1.2 Operators of Grad, Curl and Div. Curvilinear Systems of Coordinates

In this section, we will restrict ourselves to $\mathbb{R}^3$ only. Along with the already defined operator of the gradient, we recall the definitions of operators of curl and divergence.

$$\begin{aligned}
\operatorname{grad} w = \boldsymbol{\nabla} w \quad &:= (\frac{\partial w}{\partial x_1}, \frac{\partial w}{\partial x_2}, \frac{\partial w}{\partial x_3}) && = w_{,i} e_i && = \frac{\partial w}{\partial x_i} e_i && w : \mathbb{R}^3 \to \mathbb{R}\,, \\
\operatorname{curl} E = \boldsymbol{\nabla} \times E &:= (\frac{\partial E_3}{\partial x_2} - \frac{\partial E_2}{\partial x_3}, \frac{\partial E_1}{\partial x_3} - \frac{\partial E_3}{\partial x_1}, \frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2},) && = \epsilon_{ijk} \tfrac{\partial E_k}{\partial x_j} e_i && = -\frac{\partial E}{\partial x_i} \times e_i && E : \mathbb{R}^3 \to \mathbb{R}^3\,, \\
\operatorname{div} v = \boldsymbol{\nabla} \cdot v \quad &:= \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3} && = \tfrac{\partial v_i}{\partial x_i} && = \frac{\partial v}{\partial x_i} \cdot e_i && v : \mathbb{R}^3 \to \mathbb{R}^3\,.
\end{aligned}$$

Above, ϵ_{ijk} stands for the Rizzi's symbol,

$$\epsilon_{ijk} = \begin{cases} 0 & \text{if any two indices are equal,} \\ 1 & \text{if } ijk \text{ is an even permutation of } 123, \\ -1 & \text{if } ijk \text{ is an odd permutation of } 123; \end{cases}$$

we use the Einstein summation convention and e_i are the canonical basis vectors. Gradient sets scalar fields into vector fields, curl sets vector fields into vector fields, and divergence sets vector fields into scalar fields. The operations of grad and div easily extend to any n dimensions but the curl operator is strictly 3-dimensional.

Integration by parts (see the next section) reveals that

$$\begin{aligned}
(\operatorname{grad} w, v)_{L^2(G)} &= (w, -\operatorname{div} v)_{L^2(G)} + B.T. \\
(\operatorname{curl} E, F)_{L^2(G)} &= (E, \operatorname{curl} F)_{L^2(G)} + B.T.
\end{aligned}$$

where B.T. stands for "boundary terms". This proves that grad and -div operators are (formal) adjoints to each other, and the curl operator is (formally) self-adjoint.

REMARK 1.2.1 Note that contrary to the grad operator which was defined in Section 1.1 without any reference to a system of coordinates, the operators of curl and divergence are defined

in the canonical basis. A coordinates free approach leads to the theory of differential forms that holds for any space dimension n and, in fact, it can be formulated for differential manifolds. The grad, curl and div operators are then special cases of the so-called *exterior derivative*. I am tempted to show you some of this, maybe later. ∎

For completeness, we record also the formulas for the gradient and divergence of a vector field.

$$\boldsymbol{\nabla} u := \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{pmatrix} = \frac{\partial u_i}{\partial x_j} e_i \otimes e_j = \frac{\partial u}{\partial x_j} \otimes e_j \qquad u : \mathbb{R}^3 \to \mathbb{R}^3$$

$$\operatorname{div} \sigma := \begin{pmatrix} \sigma_{11,1} + \sigma_{12,2} + \sigma_{13,3} \\ \sigma_{21,1} + \sigma_{22,2} + \sigma_{23,3} \\ \sigma_{31,1} + \sigma_{32,2} + \sigma_{33,3} \end{pmatrix} \qquad \sigma : \mathbb{R}^3 \to \mathbb{R}^3 \otimes \mathbb{R}^3 ,$$

where, in the formula for $\operatorname{div} \sigma$, we use the engineering notation for derivatives. The operators of the gradient of a vector field and divergence of a tensor are again related by integration by parts formula,

$$(\boldsymbol{\nabla} u, \sigma)_{L^2(G)} = -(u, \operatorname{div} \sigma) + B.T.$$

1.2.1 Curvilinear Systems of Coordinates

Affine coordinates. Any three linearly independent (LI) vectors $a_i \in \mathbb{R}^3$, $i = 1, 2, 3$, generate an *affine system of coordinates*. Let a^j be the corresponding co-basis vectors, i.e. $(a_i, a^j) = \delta_{ij}$. Let v be a vector-valued function defined on an open set $G \subset \mathbb{R}^3$. We can represent $v(x)$ in basis a_i or the cobasis a^j,

$$v(x) = \sum_{i=1}^{3} v^i(x) a_i = v^i(x) a_i \qquad v(x) = \sum_{j=1}^{3} v_j(x) a^i = v_i(x) a^j .$$

Components $v^i(x)$ are the *contravariant* components of $v(x)$, and components $v_j(x)$ are the *covariant* components of $v(x)$ (with respect to basis a_i). If the basis is *orthonormal*, the basis and its cobasis coincide, and the contra- and covariant components of vector field $v(x)$ are identical, there is no need for distinguishing between them.

It is of practical interest to represent the operators of grad, curl and div in general affine coordinates. We will go one step further and develop the relevant representations in a more general context of *curvilinear system of coordinates* where vectors a_i (and the corresponding cobasis vectors) may vary with x.

A curvilinear system of coordinates. Any smooth (at least C^1 with bounded derivatives) bijective map

$$\mathbb{R}^3 \supset \Omega \ni \xi \to x = x(\xi) \in G \subset \mathbb{R}^3$$

defines a *curvilinear system of coordinates* in set G. Vectors

$$a_i(x) := \frac{\partial x}{\partial \xi_i}(\xi) = \frac{\partial (x_k e_k)}{\partial \xi_i}(\xi) = \frac{\partial x_k}{\partial \xi_i}(\xi)\, e_k \quad \text{where } x(\xi) = x$$

are identified as the basis vectors. Note that we are committing the usual engineering notation crime by using the same letter x for both the vector-valued function $x(\xi)$ and its values.

LEMMA 1.2.1

The cobasis vectors are given by:

$$a^j(x) = \frac{\partial \xi_j}{\partial x_l}(x)\, e_l \,.$$

The following formulas hold (and are easy to remember).

$$\begin{aligned} \boldsymbol{\nabla} w &= \tfrac{\partial w}{\partial \xi_j} a^j \,, \\ \boldsymbol{\nabla} \times E &= -\tfrac{\partial E}{\partial \xi_j} \times a^j \,, \\ \boldsymbol{\nabla} \cdot v &= \tfrac{\partial v}{\partial \xi_j} \cdot a^j \,, \\ \boldsymbol{\nabla} u &= \tfrac{\partial u}{\partial \xi_j} \otimes a^j \,. \end{aligned} \tag{1.6}$$

PROOF We have,

$$(a_i, a^j) = (\frac{\partial x_k}{\partial \xi_i} e_k, \frac{\partial \xi_j}{\partial x_l} e_l) = \frac{\partial x_k}{\partial \xi_i}\frac{\partial \xi_j}{\partial x_l} \underbrace{(e_k, e_l)}_{=\delta_{kl}} = \frac{\partial x_k}{\partial \xi_i}\frac{\partial \xi_j}{\partial x_k} = \delta_{ij} \,.$$

Next,

$$\boldsymbol{\nabla} w = \frac{\partial w}{\partial x_k}\, e_k = \frac{\partial w}{\partial \xi_i} \underbrace{\frac{\partial \xi_i}{\partial x_k}}_{=a^i} e_k \,.$$

The proof of the remaining three formulas is left for Exercise 1.2.1. ∎

Note that the formulas (1.6) do not imply whether you should use contra- or co-variant components for vectors.

Example 1.2.1 Cylindrical coordinates

We review here the classical undergraduate material on the simplest example of curvilinear coordinates - the cylindrical coordinates. We shall switch to the standard notation x, y, z (in place of x_i) for coordinates in $\mathbb{R}^3$ and (r, θ, z) for the curvilinear cylindrical coordinates (in place of ξ_j). Just for this example, we will use boldface to denote vectors. We recall the definition:

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \,. \end{cases} \tag{1.7}$$

The corresponding basis vectors are:

$$\begin{cases} \boldsymbol{a}_r = \dfrac{\partial \boldsymbol{r}}{\partial r} = \boldsymbol{e}_r \\ \boldsymbol{a}_\theta = \dfrac{\partial \boldsymbol{r}}{\partial \theta} = r\boldsymbol{e}_\theta \\ \boldsymbol{a}_z = \dfrac{\partial \boldsymbol{r}}{\partial z} = \boldsymbol{e}_z \end{cases} \tag{1.8}$$

with the unit vectors $\boldsymbol{e}_r, \boldsymbol{e}_\theta, \boldsymbol{e}_z$ given by:

$$\begin{cases} \boldsymbol{e}_r = (\cos\theta, \sin\theta, 0)^T \\ \boldsymbol{e}_\theta = (-\sin\theta, \cos\theta, 0)^T \\ \boldsymbol{e}_z = (0,0,1)^T \, . \end{cases} \tag{1.9}$$

As the system is orthogonal (the basis vectors are orthogonal), the calculation of the cobasis vectors reduces to a scaling only,

$$\begin{cases} \boldsymbol{a}^r = \boldsymbol{e}_r \\ \boldsymbol{a}^\theta = \dfrac{1}{r}\boldsymbol{e}_\theta \\ \boldsymbol{a}^z = \boldsymbol{e}_z \, . \end{cases} \tag{1.10}$$

In the case of an orthogonal curvilinear systems of coordinates, it is customary to introduce the concept of *physical components* for a vector defined as the components with respect to the *unit basis vectors* that are common for both the basis and its cobasis. We will use now general formulas (1.6) to develop specialized formulas for grad, curl and div in terms of the physical components.

Recording the derivatives of the unit vectors with respect to θ,

$$\frac{\partial \boldsymbol{e}_r}{\partial \theta} = (-\sin\theta, \cos\theta, 0)^T = \boldsymbol{e}_\theta \qquad \frac{\partial \boldsymbol{e}_\theta}{\partial \theta} = (-\cos\theta, -\sin\theta, 0)^T = -\boldsymbol{e}_r \, , \tag{1.11}$$

we specialize easily general formulas (1.6) to the cylindrical case.

$$\begin{aligned} \boldsymbol{\nabla} w &= \frac{\partial w}{\partial r}\boldsymbol{e}_r + \frac{1}{r}\frac{\partial w}{\partial \theta}\boldsymbol{e}_\theta + \frac{\partial w}{\partial z}\boldsymbol{e}_z \\ \boldsymbol{\nabla}\cdot\boldsymbol{v} &= \frac{\partial}{\partial r}(v_r\boldsymbol{e}_r + v_\theta\boldsymbol{e}_\theta + v_z\boldsymbol{e}_z)\cdot\boldsymbol{e}_r + \frac{\partial}{\partial \theta}(v_r\boldsymbol{e}_r + v_\theta\boldsymbol{e}_\theta + v_z\boldsymbol{e}_z)\cdot\frac{1}{r}\boldsymbol{e}_\theta + \frac{\partial}{\partial z}(v_r\boldsymbol{e}_r + v_\theta\boldsymbol{e}_\theta + v_z\boldsymbol{e}_z)\cdot\boldsymbol{e}_z \\ &= \frac{\partial v_r}{\partial r} + \frac{v_r}{r} + \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{\partial v_z}{\partial z} \\ \boldsymbol{\nabla}\times\boldsymbol{E} &= -\frac{\partial}{\partial r}(E_r\boldsymbol{e}_r + E_\theta\boldsymbol{e}_\theta + E_z\boldsymbol{e}_z)\times\boldsymbol{e}_r - \frac{\partial}{\partial \theta}(E_r\boldsymbol{e}_r + E_\theta\boldsymbol{e}_\theta + E_z\boldsymbol{e}_z)\times\frac{1}{r}\boldsymbol{e}_\theta - \frac{\partial}{\partial z}(E_r\boldsymbol{e}_r + E_\theta\boldsymbol{e}_\theta + E_z\boldsymbol{e}_z)\times\boldsymbol{e}_z \\ &= \left(\frac{1}{r}\frac{\partial E_z}{\partial \theta} - \frac{\partial E_\theta}{\partial z}\right)\boldsymbol{e}_r + \left(\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r}\right)\boldsymbol{e}_\theta + \left(\frac{\partial E_\theta}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \theta} + \frac{E_\theta}{r}\right)\boldsymbol{e}_z \end{aligned} \tag{1.12}$$

We use the same strategy for deriving the formula for the gradient of a vector field.

$$\begin{aligned}
\boldsymbol{\nabla u} &= \frac{\partial}{\partial r}(u_r\boldsymbol{e}_r + u_\theta\boldsymbol{e}_\theta + u_z\boldsymbol{e}_z)\otimes\boldsymbol{e}_r \\
&+\frac{\partial}{\partial\theta}(u_r\boldsymbol{e}_r + u_\theta\boldsymbol{e}_\theta + u_z\boldsymbol{e}_z)\otimes\frac{1}{r}\boldsymbol{e}_\theta \\
&+\frac{\partial}{\partial z}(u_r\boldsymbol{e}_r + u_\theta\boldsymbol{e}_\theta + u_z\boldsymbol{e}_z)\otimes\boldsymbol{e}_z \\
&= \left(\frac{\partial u_r}{\partial r}\boldsymbol{e}_r + \frac{\partial u_\theta}{\partial r}\boldsymbol{e}_\theta + \frac{\partial u_z}{\partial r}\boldsymbol{e}_z\right)\otimes\boldsymbol{e}_r \\
&+\left(\frac{\partial u_r}{\partial\theta}\boldsymbol{e}_r + u_r\boldsymbol{e}_\theta + \frac{\partial u_\theta}{\partial\theta}\boldsymbol{e}_\theta - u_\theta\boldsymbol{e}_r + \frac{\partial u_z}{\partial\theta}\boldsymbol{e}_z\right)\otimes\frac{1}{r}\boldsymbol{e}_\theta \\
&+\left(\frac{\partial u_r}{\partial z}\boldsymbol{e}_r + \frac{\partial u_\theta}{\partial z}\boldsymbol{e}_\theta + \frac{\partial u_z}{\partial z}\boldsymbol{e}_z\right)\otimes\boldsymbol{e}_z \\
&= \frac{\partial u_r}{\partial r}\boldsymbol{e}_r\otimes\boldsymbol{e}_r + \frac{\partial u_\theta}{\partial r}\boldsymbol{e}_\theta\otimes\boldsymbol{e}_r + \frac{\partial u_z}{\partial r}\boldsymbol{e}_z\otimes\boldsymbol{e}_r \\
&+\frac{1}{r}\left(\frac{\partial u_r}{\partial\theta} - u_\theta\right)\boldsymbol{e}_r\otimes\boldsymbol{e}_\theta + \frac{1}{r}\left(\frac{\partial u_\theta}{\partial\theta} + u_r\right)\boldsymbol{e}_\theta\otimes\boldsymbol{e}_\theta + \frac{1}{r}\frac{\partial u_z}{\partial\theta}\boldsymbol{e}_z\otimes\boldsymbol{e}_\theta \\
&+\frac{\partial u_r}{\partial z}\boldsymbol{e}_r\otimes\boldsymbol{e}_z + \frac{\partial u_\theta}{\partial z}\boldsymbol{e}_\theta\otimes\boldsymbol{e}_z + \frac{\partial u_z}{\partial z}\boldsymbol{e}_z\otimes\boldsymbol{e}_z
\end{aligned} \tag{1.13}$$

Utilizing the integration by parts formula,

$$\int \boldsymbol{v}\cdot\boldsymbol{\nabla}\phi\, r dr d\theta dz = -\int(\boldsymbol{\nabla}\cdot\boldsymbol{v})\phi\, r dr d\theta dz\,,$$

we can derive the formula for the divergence of a vector field in the divergence form,

$$\boldsymbol{\nabla}\cdot\boldsymbol{v} = \frac{1}{r}\frac{\partial}{\partial r}(rv_r) + \frac{1}{r}\frac{\partial v_\theta}{\partial\theta} + \frac{\partial v_z}{\partial z}\,. \tag{1.14}$$

By the same token, we can utilize the corresponding identity for tensors,

$$\int \boldsymbol{\sigma} : \boldsymbol{\nabla v}\, r dr d\theta dz = -\int(\mathbf{div}\boldsymbol{\sigma})\cdot\boldsymbol{v}\, r dr d\theta dz\,,$$

to derive the formula for the divergence of the tensor field $\boldsymbol{\sigma}$,

$$\begin{aligned}
\mathbf{div}\boldsymbol{\sigma} = &\left(\frac{1}{r}\frac{\partial}{\partial r}(r\sigma_{rr}) + \frac{1}{r}\left(\frac{\partial\sigma_{r\theta}}{\partial\theta} - \sigma_{\theta\theta}\right) + \frac{\partial\sigma_{rz}}{\partial z}\right)\boldsymbol{e}_r \\
&+\left(\frac{1}{r}\frac{\partial}{\partial r}(r\sigma_{\theta r}) + \frac{1}{r}\left(\frac{\partial\sigma_{\theta\theta}}{\partial\theta} + \sigma_{r\theta}\right) + \frac{\partial\sigma_{\theta z}}{\partial z}\right)\boldsymbol{e}_\theta \\
&+\left(\frac{1}{r}\frac{\partial}{\partial r}(r\sigma_{zr}) + \frac{1}{r}\frac{\partial\sigma_{z\theta}}{\partial\theta} + \frac{\partial\sigma_{zz}}{\partial z}\right)\boldsymbol{e}_z\,.
\end{aligned} \tag{1.15}$$

Finally, a similar exercise stemming from the formula,

$$\int \boldsymbol{E}\cdot(\boldsymbol{\nabla}\times\boldsymbol{F})\, r dr d\theta dz = \int(\boldsymbol{\nabla}\times\boldsymbol{E})\cdot\boldsymbol{F}\, r dr d\theta dz\,,$$

yields an equivalent formula for the curl in a slightly different form,

$$\boldsymbol{\nabla}\times\boldsymbol{E} = \left(\frac{1}{r}\frac{\partial E_z}{\partial\theta} - \frac{\partial E_\theta}{\partial z}\right)\boldsymbol{e}_r + \left(\frac{\partial E_r}{\partial z} - \frac{1}{r}\frac{\partial}{\partial r}(rE_z) + \frac{E_z}{r}\right)\boldsymbol{e}_\theta + \left(\frac{1}{r}\frac{\partial}{\partial r}(rE_\theta) - \frac{1}{r}\frac{\partial E_r}{\partial\theta}\right)\boldsymbol{e}_z\,. \tag{1.16}$$

□

1.2.2 Piola Transforms

Expressions (1.6) provide just a starting point for developing the actual useful formulas for grad, curl and div in a specific curvilinear system of coordinates. In Example 1.2.1 we applied them to the cylindrical coordinates, and in Exercise 1.2.2 to spherical coordinates. Both cylindrical and spherical coordinates are *orthogonal* and the difference between the contra- and co-variant components of a vector reduces to a scaling only. Hence the rationale of using *physical components* instead[‡]. For a general system of coordinates, we have to make a choice. For instance, when we compute $\mathrm{curl}\, E$, we can express vector field E in either contravariant or covariant components, and we have the same choice for the value of the operator - the $\mathrm{curl}\, E$, a total of four possible scenarios.

It turns out though that only specific choices lead to simple formulas[§]. It is convenient to expand $\boldsymbol{\nabla} w, E$ in the cobasis, and $\mathrm{curl}\, E, v$ in the (scaled) basis. The logic of using the same type of representation for $\boldsymbol{\nabla} w$ and E, and then $\mathrm{curl}\, E$ and v, comes from the structure of the differential grad-curl-div complex formed by the operators:

$$C^\infty(G) \xrightarrow{\boldsymbol{\nabla}} C^\infty(G)^3 \xrightarrow{\boldsymbol{\nabla}\times} C^\infty(G)^3 \xrightarrow{\boldsymbol{\nabla}\cdot} C^\infty(G)\,. \tag{1.17}$$

By the differential complex we mean the property that the null space of each of the operators contains the range of the previous one. In simple terms,

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla} w) = 0 \quad \text{and} \quad \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times E) = 0\,.$$

Let $x_i = x_i(\xi_j)$ be now a curvilinear system of coordinates. Let a_i and a^j be the corresponding basis and cobasis vectors, and $\mathrm{jac} = \det(\partial x_i/\partial \xi_j)$.

LEMMA 1.2.2

We have,

$$\begin{aligned} &\boldsymbol{\nabla}\times a^j = 0 && j = 1,2,3,\\ &\boldsymbol{\nabla}\cdot(\mathrm{jac}^{-1} a_i) = 0 && i = 1,2,3\,. \end{aligned}$$

PROOF

$$\boldsymbol{\nabla}\times a^l = \boldsymbol{\nabla}\times\left(\frac{\partial \xi_l}{\partial x_k} e_k\right) = \epsilon_{ijk}\xi_{l,kj} e_i = 0$$

since the product of axisymmetric (in jk) matrix ϵ_{ijk} with the symmetric matrix (in j,k) $\xi_{l,kj}$ is zero. To prove the second property, recall the definition of inverse jacobian jac^{-1} (determinant of inverse Jacobian matrix),

$$\epsilon_{ijk}\frac{\partial \xi_1}{\partial x_i}\frac{\partial \xi_2}{\partial x_j}\frac{\partial \xi_3}{\partial x_k} = \mathrm{jac}^{-1}$$

[‡]Besides, the physical components inherit the units in which we measure the vector. For instance, all physical components of a velocity vector are measured in m/s.

[§]Up to the scaling jac^{-1} factor, identical with those in a Cartesian system of coordinates.

or, more generally,

$$\epsilon_{ijk}\frac{\partial\xi_\alpha}{\partial x_i}\frac{\partial\xi_\beta}{\partial x_j}\frac{\partial\xi_\gamma}{\partial x_k}=\mathrm{jac}^{-1}\epsilon_{\alpha\beta\gamma}\,.$$

Multiplying both sides by $\partial x_l/\partial\xi_\alpha$, we get,

$$\epsilon_{ijk}\underbrace{\frac{\partial x_l}{\partial\xi_\alpha}\frac{\partial\xi_\alpha}{\partial x_i}}_{=\delta_{li}}\frac{\partial\xi_\beta}{\partial x_j}\frac{\partial\xi_\gamma}{\partial x_k}=\mathrm{jac}^{-1}\epsilon_{\alpha\beta\gamma}\frac{\partial x_l}{\partial\xi_\alpha}$$

or

$$\epsilon_{ljk}\frac{\partial\xi_\beta}{\partial x_j}\frac{\partial\xi_\gamma}{\partial x_k}=\mathrm{jac}^{-1}\epsilon_{\alpha\beta\gamma}\frac{\partial x_l}{\partial\xi_\alpha}\,.$$

In particular, differentiating both sides wrt x_l, we learn that

$$\epsilon_{\alpha\beta\gamma}\frac{\partial}{\partial x_l}(\mathrm{jac}^{-1}\frac{\partial x_l}{\partial\xi_\alpha})=\epsilon_{ljk}\frac{\partial^2\xi_\beta}{\partial x_j\partial x_l}\frac{\partial\xi_\gamma}{\partial x_k}+\epsilon_{ljk}\frac{\partial\xi_\beta}{\partial x_j}\frac{\partial\xi_\gamma}{\partial x_k\partial x_l}=0$$

as the product of a symmetric and an unsymmetric matrix must vanish. Selecting β,γ in such a way that $\epsilon_{\alpha\beta\gamma}=1$, we obtain,

$$\frac{\partial}{\partial x_l}(\mathrm{jac}^{-1}\frac{\partial x_l}{\partial\xi_\alpha})=0\quad\alpha=1,2,3\,.$$

which implies the final result,

$$\boldsymbol{\nabla}\cdot(\mathrm{jac}^{-1}a_i)=\boldsymbol{\nabla}\cdot(\mathrm{jac}^{-1}\frac{\partial x_k}{\partial\xi_i}e_k)=\frac{\partial}{\partial x_k}(\mathrm{jac}^{-1}\frac{\partial x_k}{\partial\xi_i})=0\,.$$

∎

We have already learned the formula for a gradient of a scalar-valued function $w:G\to\mathbb{R}$,

$$\boldsymbol{\nabla}w=\frac{\partial w}{\partial x_i}e_i=\frac{\partial w}{\partial\xi_j}\underbrace{\frac{\partial\xi_j}{\partial x_i}e_i}_{=a^j}=\frac{\partial w}{\partial\xi_j}\,a^j\,.$$

Gradient $\boldsymbol{\nabla}w$ of function w is naturally represented in terms of its *covariant components*. In view of the fact that gradients live in the domain of the curl operator, it is natural to extend the representation of the gradient to general vectors $E(x)$,

$$E=E_j\,a^j \tag{1.18}$$

and attempt to derive the formula for $\boldsymbol{\nabla}\times E$ in terms of the covariant components of E. We have,

$$\boldsymbol{\nabla}\times E=a^j\times\frac{\partial}{\partial\xi_j}(E_ka^k)=\frac{\partial E_k}{\partial\xi_j}a^j\times a^k+E_k\underbrace{a^j\times\frac{\partial a^k}{\partial\xi_j}}_{=\boldsymbol{\nabla}\times a^k}\,,$$

By Lemma 1.2.2, the second term vanishes. An elementary argument, comp. Exercise 1.2.3, shows that

$$a^j\times a^k=\mathrm{jac}^{-1}\epsilon_{ijk}a_i$$

where

$$\mathrm{jac} = \det\left(\frac{\partial x_i}{\partial \xi_j}\right) = \epsilon_{ijk} a_i \cdot (a_j \times a_k) = [a_1, a_2, a_3]\,.$$

Consequently,

$$\boldsymbol{\nabla} \times E = \mathrm{jac}^{-1} \epsilon_{ijk} \frac{\partial E_k}{\partial \xi_j} a_i\,.$$

Thus, the $\boldsymbol{\nabla} \times E$ is naturally given in terms of the *contravariant* components and the formula for it reduces to the elementary formula in Cartesian coordinates modulo the extra jacobian factor. The differential complex logic leads now to seek the formula for the divergence of a vector field in terms of its *scaled contravariant components* ¶

$$v = \mathrm{jac}^{-1} v^i a_i\,. \tag{1.19}$$

We have,

$$\begin{aligned}
\boldsymbol{\nabla} \cdot v &= \frac{\partial}{\partial \xi_i}\left(\mathrm{jac}^{-1} v^k a_k\right) \cdot a^i \\
&= \frac{\partial v^k}{\partial \xi_i} \mathrm{jac}^{-1} \underbrace{a_k \cdot a^i}_{=\delta_{ki}} + v^k \underbrace{\frac{\partial}{\partial \xi_i}\left(\mathrm{jac}^{-1} a_k\right) \cdot a^i}_{=\boldsymbol{\nabla}\cdot(\mathrm{jac}^{-1} a_k)} \\
&= \mathrm{jac}^{-1} \frac{\partial v^i}{\partial \xi_i}
\end{aligned}$$

since, by Lemma 1.2.2, the second term vanishes.

PROPOSITION 1.2.1

The following formulas for grad, curl and div hold in any curvilinear system of coordinates:

$$\begin{aligned}
\boldsymbol{\nabla} w &= \frac{\partial w}{\partial \xi_j} a^j \\
\boldsymbol{\nabla} \times E &= \mathrm{jac}^{-1} \epsilon_{ijk} \frac{\partial E_k}{\partial \xi_j} a_i \qquad \text{where } E = E_k a^k \\
\boldsymbol{\nabla} \cdot v &= \mathrm{jac}^{-1} \frac{\partial v^i}{\partial \xi_i} \qquad \text{where } v = \mathrm{jac}^{-1} v^i a_i\,.
\end{aligned}$$

Piola transforms. Rewriting the formulas above in the canonical coordinates, we have:

$$\begin{aligned}
\boldsymbol{\nabla} w &= \frac{\partial w}{\partial \xi_j} \frac{\partial \xi_j}{\partial x_k} e_k \\
\boldsymbol{\nabla} \times E &= \mathrm{jac}^{-1} \epsilon_{ijk} \frac{\partial E_k}{\partial \xi_j} \frac{\partial x_k}{\partial \xi_i} e_k \\
\boldsymbol{\nabla} \cdot v &= \mathrm{jac}^{-1} \frac{\partial v^i}{\partial \xi_i}\,.
\end{aligned}$$

Let $\hat{w}(\xi)$ denote the composition of map $x(\xi)$ with function $w(x)$,

$$\hat{w}(\xi_j) := w(x_i(\xi_j)) = (w \circ f)(\xi_j)$$

¶In other words, with respect to components in scaled basis $\mathrm{jac}^{-1} a_i$.

where $x = f(\xi)$ or $x_i = f_i(\xi_j)$, and we commit the usual engineering crime using the same symbol for the function $f_i = x_i$ as well its values. We shall simply write:

$$\hat{w} = w$$

with the understanding that either the right-hand side if composed with map x, or the left hand side is composed with its inverse. The formula for the gradient and the differential complex logic implies a transformation formula for the E vector,

$$E_k = \frac{\partial \xi_j}{\partial x_k} \hat{E}_j \,.$$

In the same way, the formula for the curl and the differential complex logic implies a transformation formula for the v vector,

$$v_k = \mathrm{jac}^{-1} \frac{\partial x_k}{\partial \xi_i} \hat{v}_i \,.$$

Finally, the formula for the div and the differential complex logic implies a transformation formula for a scalar-valued function f,

$$f = \mathrm{jac}^{-1} \hat{f} \,.$$

Maps:

$$w = \hat{w} \qquad E_k = \frac{\partial \xi_j}{\partial x_k} \hat{E}_j \qquad v_k = \mathrm{jac}^{-1} \frac{\partial x_k}{\partial \xi_i} \hat{v}_i \qquad f = \mathrm{jac}^{-1} \hat{f} \tag{1.20}$$

or, in a vector form,

$$w = \hat{w} \qquad E = \mathrm{Jac}^{-T} \hat{E} \qquad v = \mathrm{jac}^{-1} \mathrm{Jac}\, \hat{v} \qquad f = \mathrm{jac}^{-1} \hat{f}$$

are known as Piola transforms or *pullback maps*. The quantities with hats are functions of ξ where $x = x(\xi)$. Consistently with the differential complex structure, $\mathrm{grad}\, w$ transforms like E, $\mathrm{curl}\, E$ transforms as v, and $\mathrm{div}\, v$ transforms as f. The Piola transforms are fundamental for the construction of *parametric* Finite Elements.

We complete the discussion on the Piola transforms with the derivation of the tranformation formula for a tangent vector to a curve, and a normal vector for a surface. You may want to refresh first your knowledge about the line and the surface integrals discussed in Section 1.3.

Transformation of a tangent vector. Consider a curve in the parametric domain parametrized with

$$\xi_k = \xi_k(t), \quad t \in [0, 1] \,.$$

The image of the curve through map $x(\xi)$ is naturally parametrized with the composition of the parametrization in the parametric domain and map $x(\xi)$,

$$x_j = x_j(\xi_k(t)), \quad t \in [0, 1] \,.$$

Computing the tangent component of a E field,

$$\frac{\partial x_j}{\partial \xi_k} \frac{\partial \xi_k}{\partial t} E_j = \frac{\partial x_j}{\partial \xi_k} \frac{\partial \xi_k}{\partial t} \hat{E}_i \frac{\partial \xi_i}{\partial x_j} = \frac{\partial \xi_i}{\partial t} \hat{E}_i \,,$$

we obtain the tangent component of field $\hat{E}$ in the parametric domain. Equivalently,

$$E_t \, ds = \hat{E}_t \, ds_0 \tag{1.21}$$

where ds, ds_0 stand for the length of the tangent vectors $\frac{dx}{dt}$ and $\frac{d\xi}{dt}$ before the normalization. The Piola map preserves tangent components, and the tangential component of E along the curve in the x domain depends only upon the restriction of map $x(\xi)$ to the corresponding curve in the parametric domain.

Transformation of a normal vector. A similar result holds for the normal component of a v field. We begin again with the formula for the determinant,

$$\epsilon_{ijk} \frac{\partial x_i}{\partial \xi_\alpha} \frac{\partial x_j}{\partial \xi_\beta} \frac{\partial x_k}{\partial \xi_\gamma} = \mathrm{jac} \, \epsilon_{\alpha\beta\gamma} \, .$$

Let now $\xi_\beta(s,t)$ be a parametrization of a surface $\hat{S}$ in the parametric domain ξ. Then $x_j(\xi_\beta(s,t)$ provides a parametrization for the corresponding surface S in domain x. Multiplying the formula above by $\frac{\partial \xi_\beta}{\partial s} \frac{\partial \xi_\gamma}{\partial t}$, we obtain,

$$\epsilon_{ijk} \frac{\partial x_i}{\partial \xi_\alpha} \frac{\partial x_j}{\partial s} \frac{\partial x_k}{\partial t} = \epsilon_{ijk} \frac{\partial x_i}{\partial \xi_\alpha} \frac{\partial x_j}{\partial \xi_\beta} \frac{\partial \xi_\beta}{\partial s} \frac{\partial x_k}{\partial \xi_\gamma} \frac{\partial \xi_\gamma}{\partial t} = \mathrm{jac} \, \epsilon_{\alpha\beta\gamma} \frac{\partial \xi_\beta}{\partial s} \frac{\partial \xi_\gamma}{\partial t} \, .$$

As the cross product of two tangent vectors to a surface gives a normal to the surface, we obtain the relation between normal vectors for $\hat{S}$ and the corresponding image surface S,

$$\frac{\partial x_i}{\partial \xi_\alpha} n_i \, dS = \mathrm{jac} \, \hat{n}_\alpha \, dS_0 \, ,$$

or,

$$n_l \, dS = j \frac{\partial \xi_\alpha}{\partial x_l} \hat{n}_\alpha \, dS_0 \, .$$

where $\hat{n}, n$ are now the unit vectors and dS_0, dS denote the length of normal vectors before normalization. This implies now the relation between normal components of v and $\hat{v}$ fields in the parametric and x domains,

$$n_l H_l \, dS = \mathrm{jac} \frac{\partial \xi_\alpha}{\partial x_l} \hat{n}_\alpha \, \mathrm{jac}^{-1} \frac{\partial x_l}{\partial \xi_\beta} \hat{H}_\beta \, dS_0 = \hat{n}_\alpha \hat{H}_\alpha \, dS_0 \, . \tag{1.22}$$

Exercises

Exercise 1.2.1 Finish proof of formulas (1.6).

(5 points)

Exercise 1.2.2 Develop formulas analogous to those in Example 1.2.1 for spherical coordinates:

$$\begin{cases} x = r \sin\psi \cos\theta \\ y = r \sin\psi \sin\theta \\ z = r \cos\psi \, . \end{cases} \tag{1.23}$$

(10 points)

Exercise 1.2.3 Proof elementary formulas relating a basis a_i and its cobasis a^j in 3D:

$$a^j \times a^k = \mathrm{jac}^{-1} \epsilon^{ijk} a_i \quad a_j \times a_k = \mathrm{jac}\, \epsilon_{ijk} a^i \,.$$

(2 points)

1.3 Integration

We continue working mostly in $X = \mathbb{R}^3$.

1.3.1 Elementary Integrals

Line integral of the first kind. Let

$$(a, b) \ni \xi \to x(\xi) \in c \subset \mathbb{R}^n$$

be a C^1 parametrization of a curve (segment) c in $\mathbb{R}^n$, Let

$$\mathbb{R}^n \supset c \ni x \to \rho(x) \in \mathbb{R}$$

be a scalar-valued function defined on the curve c (think: density of mass, charge etc.). The *line integral of the first kind* is defined as follows.

$$\int_c \rho(x)\, ds := \int_a^b \rho(x(\xi)) \underbrace{|\frac{dx}{d\xi}|\, d\xi}_{=ds} \,.$$

This is a *geometrical quantity*, i.e. the value of the integral is independent of a particular parametrization of the curve, see Exercise 1.3.1.

Line integral of the second kind. Let

$$\mathbb{R}^n \supset c \ni x \to v(x) \in \mathbb{R}^n$$

be a vector-valued function defined on the curve c (think: force, current etc.). The *line integral of the second kind* is defined as follows.

$$\int_c v(x) \cdot dx := \int_a^b v(x(\xi)) \cdot \frac{dx}{d\xi}\, d\xi = \int_c v(x) \cdot t\, ds$$

where

$$t := \frac{\frac{dx}{d\xi}}{|\frac{dx}{d\xi}|}$$

is the unit vector tangent to the curve oriented consistently with the parametrization. This is again a *geometrical quantity*, i.e. the value of the integral is independent of a particular parametrization of the curve, see Exercise 1.3.1. Note that the line integral of the second type *does depend* upon the orientation of the curve.

REMARK 1.3.1 What do we mean precisely by the *orientation of curve* c ? The intuition is clear, if we define $c = \widehat{AB}$, i.e., we specify the starting point A and the ending point B, we can say that the curve is oriented from A to B. A parametrization $x(\xi)$ is then consistent with the orientation if $x(a) = A, x(b) = B$. By the way, we can always assume (explain, why ?) $a = 0, b = 1$. A more elegant way to introduce the concept of the orientation of the curve, would be to introduce an equivalence relation between parametrizations. Namely, parametrizations $x = f(\xi)$ and $x = g(\eta)$ are equivalent, if they induce a change of variable $\xi(\eta)$,

$$g(\eta) = f(\xi(\eta)) \quad \text{such that } \frac{d\xi}{d\eta} > 0\,.$$

All possible parametrizations are then partitioned into two equivalence classes which can be identified as the two possible orientations of the curve.

∎

Surface integral of the first kind. Let

$$\mathbb{R}^2 \supset G \ni (\xi_1, \xi_2) = \xi \to x(\xi) \in S \subset \mathbb{R}^3$$

be a C^1 parametrization of a surface (segment) S in $\mathbb{R}^3$, Let

$$\mathbb{R}^3 \supset S \ni x \to \rho(x) \in \mathbb{R}$$

be a scalar-valued function defined on the surface S (think: density of mass, charge etc.). The *surface integral of the first kind* is defined as follows.

$$\int_S \rho(x)\, dS := \int_G \rho(x(\xi)) \underbrace{|\frac{\partial x}{\partial \xi_1} \times \frac{\partial x}{\partial \xi_2}|\, d\xi_1 d\xi_2}_{=dS}\,.$$

This is again a *geometrical quantity*, i.e. the value of the integral is independent of a particular parametrization of the surface, see Exercise 1.3.2.

Surface integral of the second kind. Let

$$\mathbb{R}^3 \supset S \ni x \to E(x) \in \mathbb{R}^3$$

be a vector-valued function defined on the surface S (think: velocity, electric field etc.). The *surface integral of the second kind* is defined as follows.

$$\int_S E(x) \cdot dS := \int_G E(x(\xi)) \cdot (\frac{\partial x}{\partial \xi_1} \times \frac{\partial x}{\partial \xi_2})\, d\xi_1 d\xi_2 = \int_S E(x) \cdot n\, dS$$

where n is a unit normal to surface S consistent with its parametrization, i.e.,

$$n = \frac{\frac{\partial x}{\partial \xi_1} \times \frac{\partial x}{\partial \xi_2}}{|\frac{\partial x}{\partial \xi_1} \times \frac{\partial x}{\partial \xi_2}|} .$$

This is one more time a *geometrical quantity*, i.e. the value of the integral is independent of a particular parametrization of the surface, see Exercise 1.3.2. The surface integral of the second type depends upon the orientation of the surface. By the orientation of the surface we can again understand the equivalence class of parametrizations $x = f(\xi_1, \xi_2)$ corresponding to the equivalence relation as follows. We say that parametrizations $f(\xi_1, \xi_2)$ and $g(\eta_1, \eta_2)$ are equivalent, if

$$g(\eta_1, \eta_2) = f(\xi_1(\eta_1, \eta_2), \xi_2(\eta_1, \eta_2)) \quad \text{where } \det\left(\frac{\partial \xi_i}{\partial \eta_j}\right) > 0 ,$$

comp. Remark 1.3.1. Geometrically, the orientation is specified by choosing one of the two unit normals to the surface.

1.3.2 Integration by Parts. Gauss' and Stokes' Theorems

LEMMA 1.3.1 Elementary integration by parts formula

Let $G \subset \mathbb{R}^n, n = 2, 3$ be an open set. Let u, v be sufficiently regular scalar-valued functions defined on set G. The following formula holds.

$$\int_G \frac{\partial u}{\partial x_i} v = -\int_G u \frac{\partial v}{\partial x_i} + \int_{\partial G} uv\, n_i \tag{1.24}$$

where the boundary integral on the right is a line integral in $\mathbb{R}^2$, or a surface integral in $\mathbb{R}^3$ (of the first kind), and n_i stands for the i-th component of the outward unit vector n normal to boundary ∂G.

PROOF We present an elementary argument in 2D. Let us start with a special domain shown in Fig. 1.1. The 1D integration by parts formula implies,

$$\int_a^b \int_{\phi_1(x)}^{\phi_2(x)} \frac{\partial u}{\partial y}(x, y) v(x, y)\, dy\, dx = -\int_a^b \int_{\phi_1(x)}^{\phi_2(x)} u(x, y) \frac{\partial v}{\partial y}(x, y)\, dy\, dx + \int_a^b u(x, \phi_2(x)) v(x, \phi_2(x))\, dx - \int_a^b u(x, \phi_1(x)) v(x, \phi_1(x))\, dx .$$

On C_2 part of the boundary,

$$n = \frac{(-\phi_2', 1)}{\sqrt{1 + (\phi_2')^2}} .$$

Consequently,

$$\int_a^b u(x, \phi_2) v(x, \phi_2(x))\, dx = \int_a^b u(x, \phi_2) v(x, \phi_2(x)) \underbrace{\frac{1}{\sqrt{1 + (\phi_2')^2}}}_{=n_y} \underbrace{\sqrt{1 + (\phi_2')^2}\, dx}_{=ds} .$$

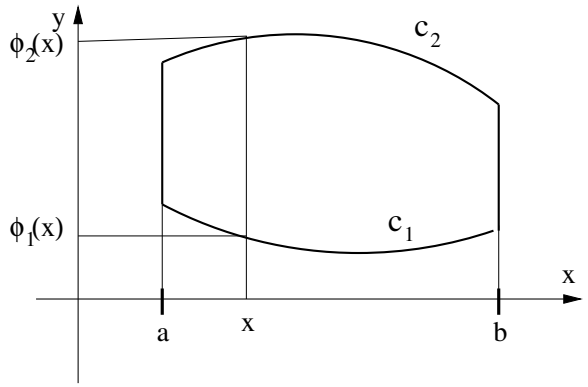

Figure 1.1
a 2D domain.

A similar relation holds on C_1 part of the boundary. On vertical parts of the boundary $n_y = 0$. As we can see, the 2D integration by parts formula is essentially a consequence of the definition of the line integral. The result for a general domain is obtained by partitioning the domain into subdomains like the one considered, and summing up the contributions corresponding to the subdomains. Line integrals over curves internal to the domain, cancel each other out. ∎

With the help of formula (1.24), we can develop now a number of integration by parts formulas for more complicated operators.

$$\begin{aligned}
\int_G \frac{\partial w}{\partial x_i} v_i \qquad &= -\int_G w \frac{\partial v_i}{\partial x_i} + \int_{\partial G} w \underbrace{v_i n_i}_{=:v_n} \\
\int_G \epsilon_{ijk} \frac{\partial E_k}{\partial x_j} F_i &= -\int_G \underbrace{\epsilon_{ijk}}_{=-\epsilon_{kji}} E_k \frac{\partial F_i}{\partial x_j} + \int_{\partial G} \epsilon_{ijk} n_j E_k F_i \\
\int_G \frac{\partial u_i}{\partial x_j} \sigma_{ij} \qquad &= -\int_G u_i \frac{\partial \sigma_{ij}}{\partial x_j} + \int_{\partial G} u_i \underbrace{\sigma_{ij} n_j}_{=:t_i}
\end{aligned}$$

i.e.,

$$\begin{aligned}
\int_G \boldsymbol{\nabla} w \cdot v \qquad &= -\int_G w \,\mathrm{div}\, v + \int_{\partial G} w v_n \\
\int_G (\boldsymbol{\nabla} \times E) \cdot F &= \int_G E \cdot (\boldsymbol{\nabla} \times F) + \int_{\partial G} (n \times E) \cdot F \qquad (1.25) \\
\int_G (\boldsymbol{\nabla} u) :: \sigma \qquad &= -\int_G u \cdot \mathrm{div}\, \sigma + \int_{\partial G} u \cdot t \,.
\end{aligned}$$

Note that the boundary term in the case of the curl operator can be rewritten in a variety of different ways,

$$\int_{\partial G} (n \times E) \cdot F = \int_{\partial G} (n \times E_t) \cdot F = \int_{\partial G} (n \times E_t) \cdot F_t = \int_{\partial G} E_t \cdot (n \times F_t) = \int_{\partial G} E \cdot (n \times F_t) = \int_{\partial G} E \cdot (n \times F)$$

where the tangent component F_t is defined as:

$$F_t = F - (F \cdot n) n = -n \times (n \times F) \,.$$

Vector $n \times F = n \times F_t$ is known as the *rotated tangent component*.

Selecting $w = 1$ in formula $(1.25)_1$, we obtain the classical *Gauss Theorem*.

THEOREM 1.3.1 Gauss' Theorem

Let v be a C^1 vector-valued field defined on a domain $G \subset \mathbb{R}^n$. Then

$$\int_G \operatorname{div} v = \int_{\partial G} v_n \,.$$

2D curl operators. The 3D curl operator gives rise to two different two-dimensional curl operators. The first one takes a vector and returns a scalar, and can be identified as the (only non-zero) third component of the 3D curl of a 2D field,

$$\boldsymbol{\nabla} \times (E_1(x_1, x_2), E_2(x_1, x_2), 0) = (0, 0, \underbrace{E_{2,1} - E_{1,2}}_{=\operatorname{curl}(E_1, E_2)}) \,.$$

The second one takes a scalar $E_3(x_1, x_2)$ and returns a vector; it can be identified with the first (the non-zero) components of the 3D curl of vector $(0, 0, E_3)$,

$$\boldsymbol{\nabla} \times (0, 0, E_3(x_1.x_2)) = (\underbrace{E_{3,2}, -E_{3,1}}_{=:\boldsymbol{\nabla} \times E_3}, 0) \,.$$

If we rotate vector (E_1, E_2) clockwise by 90 degrees, to obtain vector v,

$$v_1 = E_2, \quad v_2 = -E_1$$

the $\operatorname{curl} E$ transforms into $\operatorname{div} v$,

$$E_{2,1} - E_{1,2} = v_{1,1} + v_{2,2} \,.$$

The Gauss theorem for vector v can then be reinterpreted as the 2D version of the *Stokes' Theorem*:

$$\int_G \operatorname{curl} E = \int_G \operatorname{div} v = \int_{\partial G} (v_1 n_1 + v_2 n_2)\, ds = \int_{\partial G} (E_1(-n_2) + E_2 n_1)\, ds = \int_{\partial G} E \cdot dx$$

since tangent versor is obtained by rotating the normal versor,

$$t_1 = -n_2, \quad t_2 = n_1 \,.$$

THEOREM 1.3.2 Stokes' Theorem.

Let E be a C^1 vector-valued field defined on a surface $S \subset \mathbb{R}^3$. Then

$$\int_S (\boldsymbol{\nabla} \times E) \cdot dS = \int_{\partial S} E \cdot dx \,.$$

PROOF Let

$$\mathbb{R}^2 \supset G \ni \xi = (\xi_1, \xi_2) \to x(\xi) \in S$$

be a parametrization of surface S. Recalling the $\epsilon - \delta$ identity, see Exercise 1.3.6, we have,

$$\begin{aligned}
\int_S (\boldsymbol{\nabla} \times E) \cdot dS &= \int_G (\boldsymbol{\nabla} \times E) \cdot \left(\frac{\partial x}{\partial \xi_1} \times \frac{\partial x}{\partial \xi_2} \right) d\xi \\
&= \int_G \epsilon_{ijk} \frac{\partial E_k}{\partial x_j} \epsilon_{imn} \frac{\partial x_m}{\partial \xi_1} \frac{\partial x_n}{\partial \xi_2} \, d\xi \\
&= \int_G \left\{ \delta_{jm}\delta_{kn} \frac{\partial E_k}{\partial x_j} \frac{\partial x_m}{\partial \xi_1} \frac{\partial x_n}{\partial \xi_2} - \delta_{jn}\delta_{km} \frac{\partial E_k}{\partial x_j} \frac{\partial x_m}{\partial \xi_1} \frac{\partial x_n}{\partial \xi_2} \right\} d\xi \\
&= \int_G \left\{ \frac{\partial E_k}{\partial x_j} \frac{\partial x_j}{\partial \xi_1} \frac{\partial x_k}{\partial \xi_2} - \frac{\partial E_k}{\partial x_j} \frac{\partial x_k}{\partial \xi_1} \frac{\partial x_j}{\partial \xi_2} \right\} d\xi \\
&= \int_G \left\{ \frac{\partial E_k}{\partial \xi_1} \frac{\partial x_k}{\partial \xi_2} - \frac{\partial E_k}{\partial \xi_2} \frac{\partial x_k}{\partial \xi_1} \right\} d\xi \\
&= - \int_G E_k \frac{\partial^2 x_k}{\partial \xi_2 \partial \xi_1} \, d\xi + \int_{\partial G} E_k \frac{\partial x_k}{\partial \xi_2} \, n_1 \, ds + \int_G E_k \frac{\partial^2 x_k}{\partial \xi_1 \partial \xi_2} \, d\xi - \int_{\partial G} E_k \frac{\partial x_k}{\partial \xi_1} \, n_2 \, ds \\
&= \int_{\partial G} E_k \frac{\partial x_k}{\partial \xi_2} \, n_1 \, ds - \int_{\partial G} E_k \frac{\partial x_k}{\partial \xi_1} \, n_2 \, ds
\end{aligned}$$

Let $\xi(t)$ be a parametrization of boundary ∂G in the reference domain. Then $x(\xi(t))$ is a parametrization of boundary ∂S. Recalling that

$$n_1 = \frac{d\xi_2}{dt} \quad \text{and} \quad -n_2 = \frac{d\xi_1}{dt} \,,$$

we obtain,

$$\int_S (\boldsymbol{\nabla} \times E) \cdot dS = \int_{\partial G} E_k \{ \underbrace{\frac{\partial x_k}{\partial \xi_1} \frac{d\xi_1}{dt} + \frac{\partial x_k}{\partial \xi_2} \frac{d\xi_2}{dt}}_{= \frac{dx_k}{dt}} \} \, dt = \int_{\partial S} E \cdot dx \,.$$

Alternate proof is based on Piola transforms (1.21) and (1.22), and the already discussed 2D version of the theorem. We have,

$$\begin{aligned}
\int_S (\boldsymbol{\nabla} \times E) \cdot n \, dS &= \int_G (\hat{\boldsymbol{\nabla}} \times \hat{E}) \cdot n_0 \, dS_0 && \text{(transform (1.22))} \\
&= \int_{\partial G} \hat{E}_t ds_0 && \text{(2D version of Stokes' Theorem)} \\
&= \int_{\partial S} E_t ds && \text{(transform (1.21)) .}
\end{aligned}$$

∎

Exercises

Exercise 1.3.1 Prove that line integrals of the first and second kind are independent of a parametrization of the curve.

(5 points)

Exercise 1.3.2 Prove that surface integrals of the first and second kind are independent of a parametrization of the surface.

(10 points)

Exercise 1.3.3 Integration sanity checks. Compute the following line and surface integrals of the first kind.

- Moment of inertia of a non-homogenous arch:

$$\int_c \rho(x) x_2^2 \, ds$$

  where c is the upper part of circle centered at 0 with unit radius, and

$$\rho(x) = |x_1| \, .$$

- Mass of a non-homogeneous half-sphere:

$$\int_S \rho(x) \, dS$$

  where S is the upper part of a unit sphere centered at 0, and

$$\rho(x) = x_3 \, .$$

(5 points)

Exercise 1.3.4 Integration sanity checks. Compute the following line and surface integrals of the second kind.

- Work of a force field $F(x)$ along the half-circle $\widehat{AB}$ where $A = (1, 0), B = (-1, 0)$,

$$\int_{\widehat{AB}} F(x) \cdot dx$$

  where

$$F(x) = (1, 1) \, .$$

- Flow of a velocity field through half-sphere:

$$\int_S v(x) \cdot dS$$

  where S is the upper part of a unit sphere centered at 0 with the normal pointing upward, and

$$v(x) = (0, 0, 1) \, .$$

(5 points)

Exercise 1.3.5 Prove Lemma 1.3.1 in 3D.

(10 points)

Exercise 1.3.6 Prove $\epsilon - \delta$ identity:

$$\epsilon_{ijk}\epsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}\,.$$

Hint: With the right geometrical interpretation of the left-hand side and logical interpretation of the right-hand side, you can "see" the identity.

(10 points)

Exercise 1.3.7 A sanity check. Let V be the cylinder:

$$r \le 1, \quad 0 \le \theta < 2\pi, \quad 0 < z < 1\,.$$

Consider a vector field given in the cylindrical coordinates,

$$E = re_r + \theta e_\theta + ze_z\,.$$

Verify the Gauss' Theorem by computing the volume and surface integrals and comparing them with each other. *Hint:* Use cylindrical coordinates.

(10 points)

Exercise 1.3.8 Another sanity check. Let S be the upper part of the unit sphere ($z > 0$) oriented with vector e_r. Consider a vector field given in the spherical coordinates system,

$$E = re_r + \psi e_\psi + \theta e_\theta\,.$$

Verify the Stokes' Theorem by computing the surface and line integrals and comparing them with each other. *Hint:* Use spherical coordinates.

(10 points)

1.4 Classical Calculus of Variations

1.4.1 Classical Calculus of Variations

We refer to the classics by Gelfand and Fomin [2] for a superb exposition of the subject.

The classical calculus of variations is concerned with the solution of the constrained minimization problem:

$$\begin{cases} \text{Find } u(x), x \in [a,b], \text{ such that:} \\ u(a) = u_a \\ J(u) = \min\limits_{w(a)=u_a} J(w) \end{cases} \tag{1.26}$$

where the *cost functional* $J(w)$ is given by

$$J(w) = \int_a^b F(x, w(x), w'(x))\,dx \tag{1.27}$$

Integrand $F(x, u, u')$ may represent an arbitrary scalar-valued function of three arguments[‖] : x, u, u'. Boundary condition: $u(a) = u_a$, with u_a given, is known as the *essential BC*.

In our opening discussion we will neglect precise regularity assumptions and proceed formally. In other words, we assume whatever is necessary to make sense of the considered integrals and derivatives.

Assume now that $u(x)$ is a solution to problem (1.26). Let $v(x), x \in [a, b]$ be an arbitrary *test function*.[**] Function

$$w(x) = u(x) + \epsilon v(x)$$

satisfies the essential BC iff $v(a) = 0$, i.e., the test function must satisfy the *homogeneous essential BC*. Consider an auxiliary function,

$$f(\epsilon) := J(u + \epsilon v)$$

If functional $J(w)$ attains a minimum at u then function $f(\epsilon)$ must attain a minimum at $\epsilon = 0$ and, consequently,

$$\frac{df}{d\epsilon}(0) = 0$$

It remains to compute the derivative of function $f(\epsilon)$,

$$f(\epsilon) = J(u + \epsilon v) = \int_a^b F(x, u(x) + \epsilon v(x), u'(x) + \epsilon v'(x))\, dx$$

By Leibniz formula (see [3], p.17),

$$\frac{df}{d\epsilon}(\epsilon) = \int_a^b \frac{d}{d\epsilon} F(x, u(x) + \epsilon v(x), u'(x) + \epsilon v'(x))\, dx$$

so, utilizing the chain formula, we get

$$\frac{df}{d\epsilon}(\epsilon) = \int_a^b \left\{ \frac{\partial F}{\partial u}(u(x) + \epsilon v(x), u'(x) + \epsilon v'(x))v(x) + \frac{\partial F}{\partial u'}(u(x) + \epsilon v(x), u'(x) + \epsilon v'(x))v'(x) \right\} dx$$

Setting $\epsilon = 0$, we get

$$\frac{df}{d\epsilon}(0) = \int_a^b \left\{ \frac{\partial F}{\partial u}(u(x), u'(x))v(x) + \frac{\partial F}{\partial u'}(u(x), u'(x))v'(x) \right\} dx \tag{1.28}$$

Again, remember that u, u' in $\partial F/\partial u, \partial F/\partial u'$ denote simply the second and third arguments of F. Derivative (1.28) is identified as the *directional derivative* of functional $J(w)$ in the direction of test function $v(x)$, denoted $\partial_u^v J$. The linear operator,

$$v \to \langle (d_u J, v\rangle := \partial_u^v J = \int_a^b \left(\frac{\partial F}{\partial u}(u(x), u'(x))v(x) + \frac{\partial F}{\partial u'}(u(x), u'(x))v'(x) \right) dx \tag{1.29}$$

is the *Gâteaux differential* of $J(w)$ at u, generalizing the definition from Section 1.1 to an infinite-dimensional (function) space.

[‖] Note that, in this classical notation, x, u, u' stand for arguments of the integrand. We could have used any other three symbols, e.g., x, y, z.

[**] In the classical notation, v is replaced with δu and called the *variation* of u.

The necessary condition for u to be a minimizer reads now as follows

$$\begin{cases} u(a) = u_a \\ \langle d_u J, v\rangle = \displaystyle\int_a^b \left(\frac{\partial F}{\partial u}(x,u,u')v + \frac{\partial F}{\partial u'}(x,u,u')v' \right) dx = 0 \quad \forall v \,:\, v(a) = 0 \end{cases} \tag{1.30}$$

Integral identity (1.30) has to be satisfied for any eligible test function v, and it is identified as the *variational formulation* corresponding to the minimization problem.

In turns out that the variational formulation is equivalent to the corresponding *Euler–Lagrange* differential equation, and an additional *natural BC* at $x = b$. The key tool to derive both of them is the following Fourier's argument.

LEMMA 1.4.1 (Fourier)

Let $f \in C(\Omega)$ such that

$$\int_\Omega fv = 0$$

for every continuous test function $v \in C(\Omega)$ with compact support in Ω.

Then $f = 0$ in Ω.

PROOF The non-triviality of the result consists in the fact that we are testing only with functions v with a compact support in Ω, in particular $v = 0$ on $\partial\Omega$. If we could test with *all* continuous functions, the result would have been trivial. Indeed, taking $v = \bar{f}$, we get $\int_\Omega |f|^2 = 0$ which implies $f = 0$. Fourier's lemma is a consequence of density of continuous functions with compact support in $L^2(\Omega)$. Take a sequence of continuous functions v_n with support in Ω converging to $\bar{f}$ in the L^2-norm. By the continuity of the L^2-product, we have

$$0 = \int_\Omega f v_n \to \int_\Omega f\bar{f} = \int_\Omega |f|^2$$

and, therefore, $\|f\|_{L^2(\Omega)} = 0$ from which the result follows. ∎

In order to apply Fourier's argument, we need first to remove the derivative from the test function in the second term in (1.30). We get

$$\int_a^b \left(\frac{\partial F}{\partial u}(x,u,u') - \frac{d}{dx}\frac{\partial F}{\partial u'}(x,u,u') \right) v\,dx + \frac{\partial F}{\partial u'}(x,u(x),u'(x))v(x)\Big|_a^b = 0$$

But $v(a) = 0$, so the boundary terms reduce only to the term at $x = b$ (we do not test at $x = a$),

$$\int_a^b \left(\frac{\partial F}{\partial u}(x,u,u') - \frac{d}{dx}\frac{\partial F}{\partial u'}(x,u,u') \right) v\,dx + \frac{\partial F}{\partial u'}(b,u(b),u'(b))v(b) = 0 \tag{1.31}$$

We can follow now with the Fourier argument.

Step 1: Assume additionally that we test only with test functions with compact support in (a, b), which vanish *both* at $x = a$ and $x = b$. The boundary term in (1.31) disappears and, by Fourier's lemma, we can conclude that

$$\frac{\partial F}{\partial u}(x, u(x), u'(x)) - \frac{d}{dx}\frac{\partial F}{\partial u'}(x, u(x), u'(x)) = 0 \tag{1.32}$$

We say that *we have recovered the differential equation*.

Step 2: Once we know that the function above vanishes, the integral term in (1.31) must vanish *for any* test function v. Consequently,

$$\frac{\partial F}{\partial u'}(b, u(b), u'(b))v(b) = 0$$

for any v. Choose a test function such that $v(b) = 1$, to learn that the solution must satisfy the *natural BC* at $x = b$,

$$\frac{\partial F}{\partial u'}(b, u(b), u'(b)) = 0 \tag{1.33}$$

We have recovered the natural BC. The Euler–Lagrange equation (1.32) along with the essential and natural BCs constitute the *Euler–Lagrange Boundary-Value Problem* (E-L BVP),

$$\begin{cases} u(a) = u_a & \text{(essential BC)} \\ \dfrac{\partial F}{\partial u}(x, u, u') - \dfrac{d}{dx}\left(\dfrac{\partial F}{\partial u'}(x, u, u')\right) = 0 & \text{(Euler–Lagrange equation)} \\ \dfrac{\partial F}{\partial u'}(b, u(b), u'(b)) = 0 & \text{(natural BC)} \end{cases} \tag{1.34}$$

Neglecting the regularity issues, we can say that the E-L BVP and variational formulations are, in fact, equivalent with each other. Indeed, we have already shown that the variational formulation implies the E-L BVP. To show the converse, we multiply the E-L equation with a test function $v(x)$, integrate it over interval (a, b), and add to it the natural BC premultiplied with $v(b)$. We then integrate (back) by parts to arrive at the variational formulation. We say that the variational formulation and the E-L BVP are *formally equivalent*, formally meaning without paying attention to regularity assumptions.

Example 1.4.1 (Shortest path between a point and a line)

Given a point P and a line l in plane $\mathbb{R}^2$, we want to determine the shortest curve connecting the point with the line. By selecting a proper system of coordinates, we can always assume that $P = (0, y_0)$ and line $l : x = b$, see Fig. 1.2. It helps that we know the solution ahead of time - line: $y = y_0$. We will make a simplifying assumption that the curve is a graph of a function $y(x)$, i.e., we can parametrize the curve with x:

$$x = x,\ y = y(x) \quad x \in [0, b]\,.$$

Recalling the formula for the length of a curve, we arrive at the functional:

$$J(z) := \int_0^b \sqrt{1 + (z')^2}\, dx\,,$$

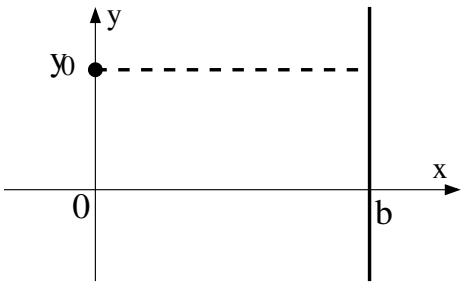

Figure 1.2
The shortest path between a point and a line.

and seek the solution of the following variational problem.

$$\begin{cases} \text{Find } y(x),\ x \in [0,b] \text{ such that:} \\ y(0) = y_0 \\ J(y) = \min\limits_{z(0)=y_0} J(z)\,. \end{cases}$$

The corresponding variational problem is:

$$\begin{cases} \text{Find } y(x),\ x \in [0.b],\ y(0) = y_0 \text{ such that:} \\ \displaystyle\int_0^b \frac{y'}{\sqrt{1+(y')^2}}\, z' = 0 \quad \forall z = z(x),\ z(0) = 0\,. \end{cases}$$

The Euler-Lagrange BVP reads as follows:

$$\begin{cases} y(0) = y_0 \\ \displaystyle -\frac{d}{dx}\left(\frac{y'}{\sqrt{1+(y')^2}}\right) = 0 \\ \displaystyle \frac{y'}{\sqrt{1+(y')^2}}(b) = 0\,. \end{cases}$$

The differential equation implies that

$$\frac{y'}{\sqrt{1+(y')^2}} = c\,,$$

and the natural BC at $x = b$ that $c = 0$. Consequently, $y' = 0$ and, by the essential BC at $x = 0$, $y(x) = y_0$, as expected. □

Example 1.4.2 (Geodesics for a sphere)

Let $r = a$ be the radius of the sphere. The task is to determine the shortest curve on a sphere connecting two points. Without loosing generality, we can assume that, in a spherical coordinates r, θ, ψ, the spherical coordinates of the points are: $r_A = a, \theta_A = 0, \psi_A = \pi/2$, $r_B = a, 0 < \theta_B \leq$

$\pi, \psi_B = \pi/2$. Parameterizing the curve on the sphere with θ, we look for a function $\psi = \psi(\theta)$, $\theta \in [0, \theta_B]$. We have,

$$\begin{aligned} x &= a\sin\psi\cos\theta & \frac{dx}{d\theta} &= a\cos\psi\frac{d\psi}{d\theta}\cos\theta - a\sin\psi\sin\theta \\ y &= a\sin\psi\sin\theta & \frac{dy}{d\theta} &= a\cos\psi\frac{d\psi}{d\theta}\sin\theta + a\sin\psi\cos\theta \\ z &= a\cos\psi & \frac{dz}{d\theta} &= -a\sin\psi\frac{d\psi}{d\theta}\,, \end{aligned}$$

and,

$$\sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 + \left(\frac{dz}{d\theta}\right)^2} = a\sqrt{\left(\frac{d\psi}{d\theta}\right)^2 + \sin^2\psi}\,.$$

The minimization problem reads thus as follows:

$$\begin{cases} \text{Find } \psi = \psi(\theta),\ \theta \in [0, \theta_B] \text{ such that} \\ \psi(0) = \psi(\theta_B) = 0 \quad \text{and} \\ a\displaystyle\int_0^{\theta_B} \sqrt{(\psi')^2 + \sin^2\psi}\,d\theta = \min_{\chi(0)=\chi(\theta_B(=0} a\int_0^{\theta_B} \sqrt{(\chi')^2 + \sin^2\chi}\,d\theta\,. \end{cases}$$

This leads to a rather complicated, nonlinear ODE for $\psi(\theta)$. Letr us try to parametrize with ψ instead. Again, without loosing generality, we can assume $\theta_A = \theta_B = 0$ and $\psi_A = 0, \psi_B > 0$. We have now,

$$\begin{aligned} x &= a\sin\psi\cos\theta & \frac{dx}{d\psi} &= a\cos\psi\cos\theta - a\sin\psi\sin\theta\theta' \\ y &= a\sin\psi\sin\theta & \frac{dy}{d\theta} &= a\cos\psi\sin\theta + a\sin\psi\cos\theta\theta' \\ z &= a\cos\psi & \frac{dz}{d\theta} &= -a\sin\psi\,, \end{aligned}$$

and,

$$\sqrt{\left(\frac{dx}{d\psi}\right)^2 + \left(\frac{dy}{d\psi}\right)^2 + \left(\frac{dz}{d\psi}\right)^2} = a\sqrt{1 + \sin^2\psi(\theta')^2}$$

where $\theta = \theta(\psi)$, $\theta(0) = \theta(\psi_B) = 0$. This time, the E-L equation is much simpler:

$$-\left[\frac{\theta'}{\sqrt{1 + \sin^2\psi(\theta')^2}}\right]' = 0\,.$$

Consequently, we have:

$$\frac{\theta'}{\sqrt{1 + \sin^2\psi(\theta')^2}} = c\,.$$

We can continue solving the equation but we can argue in an easier way. If θ is zero at both endpoints then the Mean-Value Theorem implies that derivative θ' must vanish at some intermediate point $\psi \in (0, \psi_B)$. This implies that constant $c = 0$ and, consequently, $\theta' = 0$ everywhere. The BC imply then that $\theta = 0$, as expected. The geodesics is a large circle connecting the two points.

□

Example 1.4.3 (An isoperimetric problem)

Let $y = y(x)$, $x \in [0, a]$ represent the shape of a chain of length $l > a$, fixed at the endpoints, i.e., $y(0) = y(a) = 0$. Find the shape of the chain under its own weight. Let $ds = \sqrt{1 + (y')^2}dx$. Assume unit mass density $\rho = 1$. The solution should minimize the total potential energy in the gravitational force field, i.e.,

$$J(y) = \int gy\, ds = \int_0^a gy\sqrt{1 + (y')^2}\, dx \to \min\,,$$

under the constraints:

$$y(0) = y(a) = 0 \quad \text{and} \quad \int_0^a \sqrt{1 + (y')^2}\, dx = l\,.$$

We proceed as in the finite-dimensional case by introducing the Lagrangian:

$$L(y, \lambda) := \int_0^a gy\sqrt{1 + (y')^2}\, dx - \lambda\left(\int_0^a \sqrt{1 + (y')^2}\, dx - l\right) = \int_0^a (gy - \lambda)\sqrt{1 + (y')^2}\, dx\,.$$

By Exercise 1.4.3, the corresponding E-L equation admits the first integral:

$$\frac{gy - \lambda}{\sqrt{1 + (y')^2}} = A\,.$$

Separation of variables leads to:

$$\frac{dy}{\sqrt{\left(\frac{gy-\lambda}{A}\right)^2 - 1}} = dx\,.$$

Substituting,

$$\frac{gy - \lambda}{A} = \cosh z \quad \Rightarrow \quad dy = \frac{A}{g}\sinh z\, dz\,,$$

leads to:

$$\frac{dy}{\sqrt{\left(\frac{gy-\lambda}{A}\right)^2 - 1}} = \frac{A}{g}\frac{\sinh z\, dz}{\sqrt{\cosh^2 z - 1}} = \frac{A}{g}\frac{\sinh z\, dz}{\sinh z} = \frac{A}{g}dz = dx\,.$$

This gives

$$z = \frac{g}{A}x + B \quad \Rightarrow \quad \frac{gy - \lambda}{A} = \cosh z = \cosh\left(\frac{g}{A}x + B\right)$$

and, eventually,

$$y = g^{-1}\left(\lambda + A\cosh(\frac{g}{A}x + B)\right)\,.$$

Constants A, B, λ must be determined numerically from conditions: $y(0) = y(a) = 0$ and

$$\int_0^a \sqrt{1 + (y')^2}\, dx = l\,.$$

This is the best what we can do analytically. □

1.4.2 Generalizations

The derivation of variational formulation and Euler-Lagrange BVP presented in Section 1.4.1 extends to more complicated scenarios in practically identical way. The key lies in extending the Fourier argument (lemma) to higher space dimensions and manifolds. We record now shortly a few representative examples.

Functionals depending on higher derivatives. Let $y = y(x),\ x \in [a,b]$ be a sufficiently regular function, and $F = F(x, y, y', y'')$ an integrand that depends now upon the second derivative as well. Essential BCs may involve now not only function values but first derivatives as well. It is convenient to introduce formally a *set of kinematically admissible solutions*, For instance, we may consider the following BCs:

$$W := \{z(x), x \in [a,b] \,:\, z \text{ is sufficiently regular, and } z(a) = u_a,\ z'(a) = v_a\}$$

The corresponding *space of test functions (variations)* will employ the *homogeneous* BCs:

$$V := \{z(x), x \in [a,b] \,:\, z \text{ is sufficiently regular, and } z(a) = 0,\ z'(a) = 0\}$$

Let u be a particular element of W (a *lift* of non-homogeneous BCs). Then

$$W = u + V\,,$$

i.e., set W has the structure of an *affine manifold*, see [5], Example 2.2.5. The minimization problem reads now as follows.

$$\begin{cases} \text{Find } u \in W \text{ such that:} \\ J(u) = \min_{z \in W} J(z) \end{cases}$$

where

$$J(z) = \int_a^b F(x, z(x), z'(x), z''(x))\, dx\,.$$

The corresponding variational formulation is:

$$\begin{cases} \text{Find } u \in W \text{ such that:} \\ \displaystyle\int_a^b \left(\frac{\partial F}{\partial z}(x, u(x), u'(x), u''(x))\, \delta u + \frac{\partial F}{\partial z'}(x, u(x), u'(x), u''(x))\, (\delta u)' + \frac{\partial F}{\partial z''}(x, u(x), u'(x), u''(x))\, (\delta u)'' \right) dx \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, \delta u \in V\,. \end{cases}$$

We emphasize that letters in δu cannot be separated, it is simply a two-letter symbol for a function. We can replace δu with v,

$$\begin{cases} \text{Find } u \in W \text{ such that:} \\ \displaystyle\int_a^b \left(\frac{\partial F}{\partial z}(x, u(x), u'(x), u''(x))\, v + \frac{\partial F}{\partial z'}(x, u(x), u'(x), u''(x))\, v' + \frac{\partial F}{\partial z''}(x, u(x), u'(x), u''(x))\, v'' \right) dx \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, v \in V\,. \end{cases}$$

Integration by parts (twice) and application of the Fourier argument leads to the Euler-Lagrange BVP:

$$\begin{cases} \text{Find } u(x) \text{ such that:} \\ \dfrac{\partial F}{\partial z} - \dfrac{d}{dx}\left(\dfrac{\partial F}{\partial z'}\right) + \dfrac{d^2}{dx^2}\left(\dfrac{\partial F}{\partial z''}\right) = 0 & \text{(Euler-Lagrange equation)} \\ u(a) = u_a,\ u'(a) = v_a & \text{(essential BCs)} \\ \left[\dfrac{\partial F}{\partial z'} - \dfrac{d}{dx}\left(\dfrac{\partial F}{\partial z''}\right)\right](b) = 0 & \text{(natural BC)} \\ \left[\dfrac{\partial F}{\partial z''}\right](b) = 0 & \text{(natural BC.)} \end{cases}$$

All expressions are functions of $x, u(x), u'(x), u''(x)$. We end up with a (generally nonlinear) fourth order ODE with four BCs.

Functionals depending on partial derivatives. Let $\Omega \in \mathbb{R}^n$ be a domain with boundary split into two disjoint parts. More precisely,

$$\partial\Omega = \Gamma = \overline{\Gamma_1} \cup \overline{\Gamma_2}, \quad \Gamma_1 \cap \Gamma_2 = \emptyset$$

where Γ_1, Γ_2 are (relatively) open in Γ, and the closure is understood in the topological subspace Γ. We define the *set of kinematically admissible functions:*

$$W := \{u(x),\, x \in \bar{\Omega} \,:\, u \text{ is sufficiently regular, and } u = u_0 \text{ on } \Gamma_1\}$$

with the corresponding *space of test functions:*

$$V := \{v(x),\, x \in \bar{\Omega} \,:\, v \text{ is sufficiently regular, and } v = 0 \text{ on } \Gamma_1\}\,.$$

The energy functional is:

$$J(z) = \int_\Omega F(x, u(x), \frac{\partial u}{\partial x_i}(x))\, dx = \int_\Omega F(x, u(x), u_{x_i}(x))\, dx\,.$$

The minimization problem is now:

$$\begin{cases} \text{Find } u \in W \text{ such that:} \\ J(u) = \min_{z \in W} J(z)\,. \end{cases}$$

The corresponding variational formulation is:

$$\begin{cases} \text{Find } u \in W \text{ such that:} \\ \displaystyle\int_\Omega \left(\frac{\partial F}{\partial u}\, v + \frac{\partial F}{\partial u_{x_i}}\, \frac{\partial v}{\partial x_i} \right) dx = 0 \quad \forall v \in V\,, \end{cases}$$

and the E-L BVP reads as follows:

$$\begin{cases} \text{Find } u = u(x),\, x \in \bar{\Omega} \text{ such that:} \\ \displaystyle\frac{\partial F}{\partial u} - \frac{\partial}{\partial x_i}\left(\frac{\partial F}{\partial u_{x_i}}\right) = 0 & \text{(E-L equation)} \\ u = u_0 \text{ on } \Gamma_1 & \text{(essential BC)} \\ \displaystyle\frac{\partial F}{\partial u_{x_i}} n_i = 0 \text{ on } \Gamma_2 & \text{(natural BC.)} \end{cases}$$

Again, all expressions are functions of $x, u(x)$ and all partial derivatives u_{x_i}. In general, we end up with a second order non-linear PDE accompanied with linear essential BC on Γ_1, and a nonlinear natural BC on Γ_2.

The list of possible generalizations continues. We can have higher order partial derivatives or combinations of those, e.g. the div and curl operators, we can have multiple unknowns leading to systems of ODEs or PDEs etc. The complexity of the resulting E-L BVPs is usually too overwhelming for analytical methods but we can discretize them and approximate numerically. Finite elements use the variational formulation as a starting point for discretization whereas (classical) finite difference methods are based on the E-L equations.

The discussed examples do not represent the most general cases. For instance, we can additional point values (1D) or boundary integrals (multi-D) to the functionals. In the case of *quadratic* energy functionals, the corresponding E-L BVP is linear. We will discuss this case separately using a different notation.

Exercises

Exercise 1.4.1 Determine the shortest curve connecting two arbitrary points A, B on a plane. You may assume $A = (0, 0)$ and $B = (l, 0)$. Formulate the minimization problem, and write out the corresponding variational formulation and the Euler-Lagrange BVP. Finally, solve the E-L problem,

(5 points)

Exercise 1.4.2 Find the geodesics for a cylinder. *Hint:* Look for the geodesics as a graph of function $z = z(\theta)$ or $\theta = \theta(z)$.

(10 points)

Exercise 1.4.3 Assume the integrand $F(x, y, y')$ has no explicit dependence on x, i.e., $F = F(y, y')$. Show that the Euler-Lagrange equation corresponding to the minimization problem:

$$\int_a^b F(y, y')\, dx \to \min$$

admits the first integral:

$$F - y' \frac{\partial F}{\partial y'} = \text{const}$$

(5 points)

Exercise 1.4.4 Example from 'beamology'[††]. Consider the Euler-Bernoulli beam problem shown in Fig. 1.3. The elastic energy of the beam is given by:

$$\frac{1}{2} \int_0^{3l/2} EI(w'')^2\, dx$$

where EI is the stiffness of the beam. Write down the work of the external forces, and the total potential energy functional. Specify the set (space) of kinematically admissible displacements, i.e., formulate the essential BCs, and write down the energy minimization problem. Write down then (derive ?) the corresponding variational formulation (Principle of Virtual Work). Use integration by parts and the Fourier argument to derive the corresponding E-L equation, and natural boundary and interface (at $x = l$) conditions. You may want to consult also Section 5.3.

(10 points)

[††]Terminology of Prof. Babuška.

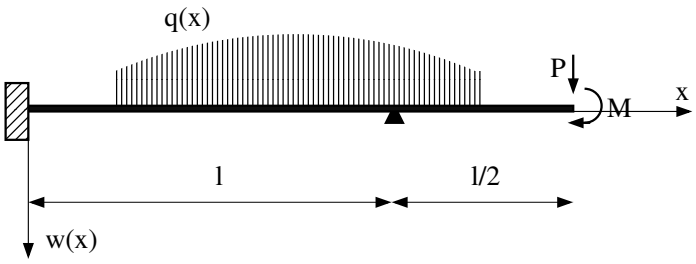

Figure 1.3
A beam problem. The beam is fixed at $x = 0$, free supported at $x = l$, subjected to a distributed load with intensity $q = q(x)$, and a concentrated force F and a concentrated moment M at $x = 3l/2$.

Exercise 1.4.5 Consider the rectangular elastic plate problem shown in Fig. 1.4 and two possible scenarios for the BCs.

(i) *Clamped plate:* $w = \frac{\partial w}{\partial n} = 0$ on Γ.

(ii) *Free supported plate:* $w = 0$ on Γ

where, as usual, n is the outward normal unit vector on boundary Γ. The elastic energy of the plate is given by:

$$\frac{1}{2}D\int_0^a\int_0^b \left[(w_{,xx}+w_{,yy})^2 - 2(1-\nu)(w_{,xx}w_{,yy}-w_{,xy}^2)\right]\, dydx$$

where $D = \frac{Eh^3}{12(1-\nu^2)}$ is the stiffness of the plate with E being the Young modulus, ν the Poisson ratio, and h the thickness of the plate. The plate is loaded with a distributed load with intensity $q = q(x, y)$. Write down the minimization problem for the total potential energy and the corresponding variational formulation (Principle of Virtual Work). Use then integration by parts and the Fourier argument to derive the corresponding Euler-Lagrange equation and natural BCs.

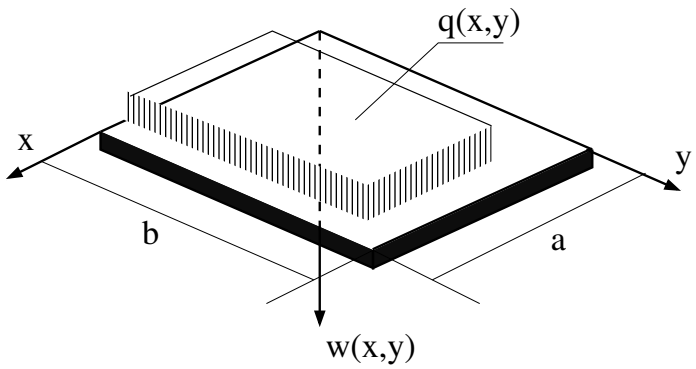

Figure 1.4
An elastic plate problem. The plate is loaded with a distributed load with intensity $q(x)$.

(10 points)

2

Complex Analysis

2.1 Introduction

This short chapter reviews a minimum information leading to the Residue Theorem and its applications, mainly in the computation of inverse Fourier and Laplace transforms. It is by no means intended to replace a systematic study on the subject, see e.g. [3].

Complex numbers. Set $\mathbb{R}^2$ equipped with standard, component-wise addition, and the non-standard multiplication:

$$(x_1, x_2) \times (y_1, y_2) := (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1)\,,$$

forms a commutative field. We usually write $x = x_1 + i x_2$ in place of pair (x_1, x_2) where $i = (0, 1)$ is the imaginary unit, and by the first term we really understand $x_1(1, 0)$, i.e., we identify real numbers with the first axis. One has $i^2 = -1$ $(= -(1, 0))$, and the multiplication is easily remembered by recalling this property. We introduce the operation of *complex conjugate*,

$$\overline{x} = \overline{x_1 + i x_2} := x_1 - i x_2\,,$$

and identify the Euclidean norm,

$$|x|^2 = x\bar{x}$$

as the *modulus* of x. We denote the field of complex numbers with $\mathbb{C}$. In terms of topology it coincides with $\mathbb{R}^2$ equipped with the Euclidean norm and, therefore, it is a complete metric space. From now on, we will use exclusively letter z to denote a complex number.

Analytic functions and their complex extensions. Let $f : \mathbb{R} \supset D \to \mathbb{R}$ be a real-valued function of a real argument. Recall that function f is analytic in its domain D if, at every point $x_0 \in D$, the corresponding Taylor series of f at x_0 converges in a neighborhood of x_0, and the sum is equal to the value of f,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \qquad |x - x_0| < r\,.$$

The maximum (supremum really) value of $r = r(x_0)$ for which the Taylor series converges, is called the *radius of convergence at* x_0.

LEMMA 2.1.1 (Convergence of power series)

Consider a power series:

$$\sum_{n=0}^{\infty} a_n z^n \,.$$

Assume the series converges for a particular $z = z_1$, $\rho_1 := |z_1|$. Then

(i) The series is absolutely convergent for $|z| < \rho_1$,

(ii) The series is uniformly convergent for $|z| \leq \rho$ for any $\rho < \rho_1$.

Conversely, if the series diverges for a particular $z = z_1$, $\rho_1 := |z_1|$, then the series is divergent for any $|z| > \rho_1$.

PROOF If the series converges at $z = z_1$ then, in particular, $a_n z_1^n \to 0$ and, therefore, it is bounded,

$$|a_n z_1^n| < M < \infty \,.$$

Consequently,

$$|a_n z^n| = |a_n z_1^n \left(\frac{z}{z_1}\right)^n| < M \left|\frac{z}{z_1}\right|^n = M \left(\frac{\rho}{\rho_1}\right)^n, \quad \rho = |z| \,,$$

As the geometric series on the right converges, the comparison criterion implies both statements. Use proof by contradiction and the positive result to prove the negative result. ∎

We return now to our analytic function $f(x)$. Let r be the radius of convergence of the Taylor series at a point $x \in D$. By Lemma 2.1.1 we can claim that the series is converges not only for real arguments but also for any complex z with $|z - x_0| < r$. The convergent series defines a *complex extension* $f(z)$ of the original function $f(x)$. By construction, the function is equal to the sum of its Taylor series, i.e., it is analytic in z. We usually think of a *maximum extension*, i.e., a complex extension with a maximum domain of definition.

Example 2.1.1

In the case of many elementary functions, we can identify the complex extension *explicitly*. This is the case for arbitrary polynomials*, x^n extends to z^n. Recalling the Taylor expansion of e^x at $x_0 = 0$, we extend the exponential function to complex numbers,

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n \quad \Rightarrow \quad e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n \,,$$

with the radius of convergence $r(0) = \infty$. In both cases, the functions are defined on the whole real line and complex plane, and the radius of convergence $r = \infty$. Recalling definitions of hyperbolic

*Taylor expansion at 0 for an arbitrary polynomial coincides simply with the polynomial itself.

cosine and sine, we can easily extend the functions to the whole complex plane,

$$\cosh z = \frac{1}{2}(e^z + e^{-z}) \quad \Rightarrow \quad \sinh z = \frac{1}{2}(e^z - e^{-z})\,.$$

A comparison of Taylor's expansion for e^z with $z = ix$, and Taylor's expansions for the sine and cosine functions, leads to the fundamental Euler formula (comp.Exercise 2.1.1):

$$e^{ix} = \cos x + i \sin x\,. \tag{2.1}$$

This leads to formulas for the cosine and sine functions in terms of the exponentials,

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}) \quad \text{and} \quad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$$

and, in turn, to the explicit analytical extension of sine and cosine functions,

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}) \quad \Rightarrow \quad \sin z = \frac{1}{2i}(e^{iz} - e^{-iz})\,.$$

All the discussed extensions are defined on the *entire* complex plane and the radius of convergence at any point is infinite. Such beautifully regular functions are called *entire functions*. □

Multi-valued functions. Branch points and branch cuts. Consider another seemingly simple function, the logarithm: $\ln x$, $x > 0$. We represent complex argument z in terms of polar coordinates,

$$z = re^{i\theta}\,, \quad r = r(z) = |z|, \quad \theta = \theta(z) =: \arg(z)\,.$$

The $\arg(z)$ function is an example of a *multi-valued function*. It is not really a function, it is a spiral-like surface prescribing for each z an infinite number of values. In order to have a function, we have to cut out a *branch* of the surface. If we restrict ourselves to $\theta \in [0, 2\pi)$, we speak about the *principal argument function* and denote it with a capital letter $\mathrm{Arg}(z)$. In order to emphasize the multi-valuedness of $\arg(z)$, we can write:

$$\arg(z) = \mathrm{Arg}(z) + k2\pi, \quad k \in \mathbb{Z}\,.$$

Using the polar coordinates, we can now easily guess the complex extension for the logarithm function,

$$\ln z = \ln(re^{i\theta}) = \ln r + \ln ei\theta = \ln r + i\theta\,.$$

As we can see, we run into a multivalued function and we have to resort to a branch cut to make it into a proper function. We can use the branch cut used for defining the principal argument, see Fig. 2.1. The origin is identified as the *branch point*. The choice of branch cut is by no means unique. Fig 2.2 presents an alternative branch cut, this time coinciding with the negative x axis. No matter which branch cut we use, we learn that the complex extension is a discontinuous function across the branch cut. Note that, in particular, the complex extension is defined for the negative real numbers. As we will see later, the possibility of selecting different branch cuts is crucial in computing contour integrals through the Residue Theorem.

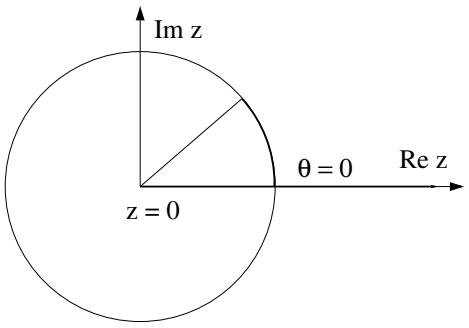

Figure 2.1
A branch cut to define $\ln(z)$.

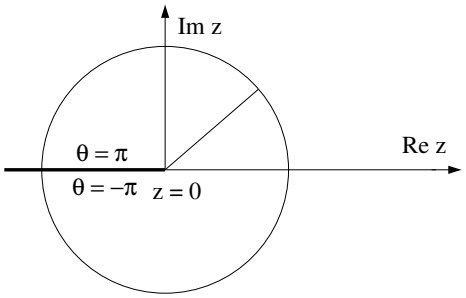

Figure 2.2
A alternative branch cut to define $\ln(z)$.

Example 2.1.2

Complex extension for $\sqrt{x}$. We use again polar coordinates,

$$\sqrt{z} = \sqrt{re^{i\theta}} = \sqrt{r}e^{i\theta/2}\,,$$

The extension is actually only *double-valued* since $e^{i\theta/2}$ can assume only two values for $\theta = \Theta + k2\pi$, one for even k, and another one for odd k. We can cut, e.g. with any half-line originating from the origin. □

Example 2.1.3

Define branch cut(s) for function,

$$f(z) = \sqrt{(z-1)(z-2)}$$

to render it single-valued. We use two systems of polar coordinates, one with the origin at $z = 1$ and another one with the origin at $z = 2$,

$$z - 1 = r_1 e^{i\theta_1} \qquad z - 2 = r_2 e^{i\theta_2}\,.$$

We define:

$$\sqrt{(z-1)(z-2)} = \sqrt{r_1 r_2 e^{i\theta_1} e^{i\theta_2}} = \sqrt{r_1 r_2} e^{i\theta_1/2} e^{i\theta_2/2}\,.$$

We can use then two individual cuts for the two argument functions shown in Fig. 2.3 where $\theta_1 \in [0, 2\pi)$, $\theta_2 \in [0, 2\pi)$. Note that the particular branch is continuous across the segment $(1, 2)$. □

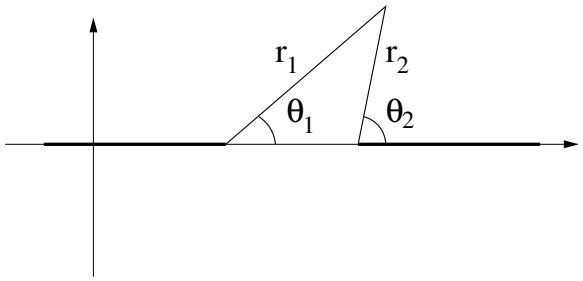

Figure 2.3
Branch cuts to define $\sqrt{(z-1)(z-2)}$.

Complex derivative. We define the complex derivative in the usual way,

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z}$$

where the division (multiplication with the inverse of Δz) is understood in the sense of complex multiplication. As Δz is an element of $\mathbb{C} = \mathbb{R}^2$, we can pass to the limit in different ways. Let,

$$f(z) = f(x, y) = u(x, y) + iv(x, y)$$

where $u(x, y)$ and $v(x, y)$ are real and imaginary parts of $f(z)$. Using a real $\Delta z = \Delta x$, we obtain,

$$f'(z) = u_{,x}(x, y) + iv_{,x}(x, y)\,.$$

Using, however, an imaginary $\Delta z = i\Delta y$, we obtain,

$$f'(z) = \frac{1}{i} u_{,y}(x, y) + v_{,y}(x, y) = v_{,y}(x, y) - iu_{,y}(x, y)\,.$$

As the two values must be equal, we obtain the following necessary conditions for the complex differentiability,

$$u_{,x}(x, y) = v_{,y}(x, y) \quad \text{and} \quad v_{,x}(x, y) = -u_{,y}(x, y)\,.$$

These are the famous *Cauchy-Riemann conditions*. Differentiating the first condition in x, the second in y, and summing them up, we obtain that the real part is a *harmonic function*, i.e., it satisfies the Laplace equation. Similarly, differentiating the first equation in y, the second in x, and subtracting the equations, we conclude that the real part is also a harmonic function. We call them the *conjugate harmonic functions* as they

correspond to one complex function $f(z)$. Cauchy-Riemann condition imply that $\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v = 0$. Since $\boldsymbol{\nabla} u$ is orthogonal to curve $u = \text{const}$ and, similarly, $\boldsymbol{\nabla} v$ is orthogonal to curve $v = \text{const}$, the two families of curves: $u = \text{const}$, and $v = \text{const}$, are orthogonal to each other. We see that complex differentiable functions are extremely regular. We record the following useful result without a proof.

THEOREM 2.1.1 (Sufficient conditions for complex differentiability)

If the Cauchy-Riemann equations are satisfied at a point $z = (x, y)$, and all partial derivatives $u_{,x}, u_{,y}, v_{,x}, v_{,y}$ are continuous at the point, then function $f(x, y) = u(x, y) + iv(x, y)$ is complex differentiable at $z = (x, y)$.

In the following sections, we will prove the following amazing result.

THEOREM 2.1.2 (Complex differentiability $\Leftrightarrow$ complex analicity)

Let $f : \mathbb{C} \supset D \to \mathbb{C}$ be a complex-valued function of complex argument z. Then f is complex differentiable in D iff f is analytic in D.

According to Theorem 2.1.2, for complex-valued functions of complex argument, it makes no sense to distinguish between spaces typical in real analysis such as $C^k(D), C^\infty(D), C^\omega(D)$. *They are identical !* Frequently, we use the name of *holomorphic functions* when talking about analytic functions of complex variable.

Exercises

Exercise 2.1.1 Prove Euler formula (2.1). *Hint:* Compare Taylor's expansions at zero for both sides.

(5 points)

Exercise 2.1.2 Prove that

$$e^{z+w} = e^z\, e^w \quad z, w \in \mathbb{C}$$

and use it to prove the more general Euler formula for an arbitrary complex argument:

$$e^z = e^{\Re z}(\cos \Im z + i \sin \Im z)\,.$$

Hint: Compare Taylor's expansions at zero for both sides.

(5 points)

Exercise 2.1.3 Discuss an alternative choice of cut shown in Fig. 2.4 where $\theta_1, \theta_2 \in [0, 2\pi)$, for function from Example 2.1.3.

(5 points)

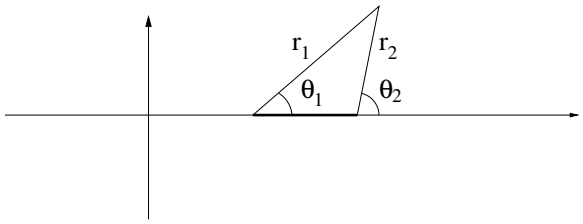

Figure 2.4
An alternative branch cut to define $\sqrt{(z-1)(z-2)}$.

Exercise 2.1.4 Prove that if $f(z)$ is complex differentiable then its complex conjugate $\overline{f(z)}$ *cannot be* complex differentiable, unless $f(z)$ is a constant function. Conclude that, for a non-constant holomorphic function $f(z)$, its modulus $|f(z)|$ is not holomorphic.

(10 points)

Exercise 2.1.5 Derive Cauchy-Riemann conditions in polar coordinates:

$$u_{,r} = \frac{v_{,\theta}}{r} \quad \text{and} \quad v_{,r} = -\frac{u_{,\theta}}{r}\,,$$

and use them to test the following functions for complex differentiability.

$$f(z) = r^2 + \theta + 2ir^2\theta, \qquad f(z) = |z|z^2\,.$$

(10 points)

Exercise 2.1.6 Interpret the complex multiplication in terms of the polar coordinates:

$$|z_1 z_2| = |z_1|\,|z_2| \quad \text{and} \quad \arg(z_1 z_2) = \arg(z_1) + \arg(z_2)\,.$$

Note that the second equation involves multi-valued functions, and explain how we should understand it. Prove then *De Moivre's theorem*:

$$z^n = r^n(\cos n\theta + i \sin n\theta)\,,$$

and explain its usefulness in computing trigonometric formulas for $\cos n\theta$ and $\sin n\theta$, for $n = 2, 3, \ldots$.

(10 points)

Exercise 2.1.7 Identify the branch points, if any, of the following functions and select two suitable branch cuts.

(a) $\sin\sqrt{z}$ (b) $(\sqrt{z})^2$ (c) $\frac{1}{\sqrt{z^2-2z}}$
(d) z^z (e) $\sqrt{z+1}+\sqrt{z-1}$ (f) $\cos\sqrt{z}$

(10 points)

Exercise 2.1.8 Find all solutions z to $f(z) = 0$, for the following functions.

(a) e^z (b) $\sin z$ (c) $\sinh z$
(d) $z^2 + 2z$ (e) $\ln z$ (f) $\cosh z^2$

(10 points)

2.2 Integration

Complex integrals. Let $c = \widehat{AB}$ be an oriented curve in the complex plane, and

$$z(t),\, t \in (a,b), \quad z(a) = A, z(b) = B\,,$$

its parametrization. We shall assume that the curve is always piece-wise C^1, i.e., its parametrization is a piece-wise C^1-function. Let $f(z)$ be a complex-valued function defined on the curve. We define the integral of f over the curve $c = \widehat{AB}$ as:

$$\int_c f(z)\,dz = \int_{\widehat{AB}} f(z)\,dz := \int_a^b f(z(t))\,z'(t)\,dt\,. \tag{2.2}$$

One can show that the integral is *independent of the parametrization*, comp. Exercise 2.2.1. Note the complex multiplication between $f(z(t))$ and $z'(t)$ which makes the integral different from the standard curve integral of a real-valued function. For a segment $c = (a,b)$ of the real line, the integral coincides with the standard 1D real integral. Finally, notice that the integral depends upon the orientation of the curve,

$$\int_{\widehat{AB}} f(z)\,dz = -\int_{\widehat{BA}} f(z)\,dz\,.$$

If c is a closed curve then the integral over c depends upon its orientation: *counterclockwise* (ccw) or *clockwise* (cw), but it is independent of the starting (= ending) point. Explain why?

Example 2.2.1

Let c be a ccw oriented circle centered at $z = a$ with radius r. Compute:

$$\int_c (z-a)^n\,dz\,,$$

for any integer $n \in \mathbb{Z}$. Parametrizing the circle in the standard way,

$$z = a + re^{i\theta}, \qquad \frac{dz}{d\theta} = ire^{i\theta}\,, \qquad \theta \in [0, 2\pi]\,,$$

we obtain,

$$\int_c (z-a)^n\,dz = \int_0^{2\pi} r^n e^{in\theta}\, ire^{i\theta}\,d\theta = ir^{n+1}\int_0^{2\pi} e^{i(n+1)\theta}\,d\theta = \begin{cases} 0 & n \neq -1 \\ 2\pi i & n = -1 \end{cases}\,.$$

□

THEOREM 2.2.1 (Cauchy)

Let function $f(z)$ be complex differentiable in a simply connected † *domain $D \subset \mathbb{C}$, and let c be a*

†Recall that an open set is simply connected if each loop contained in the domain can be shrunk to a point without leaving the domain. In 2D this means that the domain simply does not have holes in it.

closed curve (contour) contained in domain D. Then

$$\int_c f(z)\,dz = 0\,.$$

PROOF Let Ω be the domain enclosed by the curve. Let $z(t) = x(t)+iy(t)$ be a parametrization of the curve. Then

$$f(z(t))z'(t) = (u(x(t),y(t))x'(t) - v(x(t),y(t))y'(t)) + i(u(x(t),y(t))y'(t) + v(x(t),y(t))x'(t))\,,$$

and,

$$\int f(z(t))z'(t)\,dt = \int (u(x(t),y(t))x'(t) - v(x(t),y(t))y'(t))\,dt + i\int (u(x(t),y(t))y'(t) + v(x(t),y(t))x'(t))\,.$$

We have for the first integral by the Gauss' Theorem,

$$\begin{aligned}\int_c (u(x,y)x' - v(x,y)y')\,dt &= \int_c (u(x,y)\underbrace{\frac{x'}{\sqrt{(x')^2+(y')^2}}}_{=-n_y} - v(x,y)\underbrace{\frac{y'}{\sqrt{(x')^2+(y')^2}}}_{=n_x})\underbrace{\sqrt{(x')^2+(y')^2}dt}_{=ds}\\ &= \int_\Omega (-u_{,y} - v_{,x})\,dxdy = 0\,,\end{aligned}$$

since the integrand vanishes by the second Cauchy-Riemann condition. In the same way we show that the second integral vanishes as well. ∎

COROLLARY 2.2.1

Let $D \subset \mathbb{C}$ be a domain with a hole, and let c_1 and c_2 be two closed curves containing the hole, see Fig. 2.5. Let $f(z)$ be complex differentiable in D. Then its contour integrals over curves c_1 and c_2 have the same value.

We leave the proof for Exercise 2.2.2.

THEOREM 2.2.2 (Cauchy Integral Formula)

Let $f(z)$ be complex differentiable in a domain $D \subset \mathbb{C}$. Let c be a closed ccw curve contained in D. Then, for any z inside of curve c, the following formula holds.

$$f(z) = \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{\zeta - z}\,d\zeta\,. \tag{2.3}$$

Moreover, derivative $f^{(n)}(z)$ of any order n exists, and,

$$f^{(n)}(z) = \frac{n!}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta - z)^{n+1}}\,d\zeta\,. \tag{2.4}$$

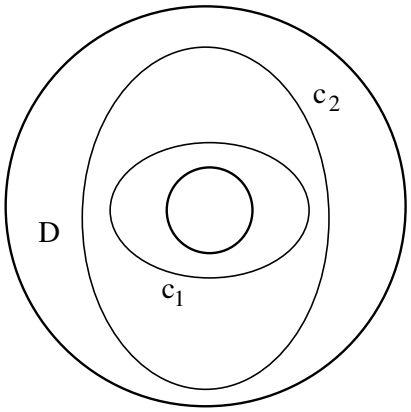

Figure 2.5
Contours around a hole.

PROOF In view of Corollary 2.2.1, we can assume that c is a ccw circle with an arbitrary radius r centered at point z. We have,

$$\frac{1}{2\pi i}\int_c \frac{f(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi i}\int_c \frac{f(z)}{\zeta - z}\, d\zeta + \frac{1}{2\pi i}\int_c \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta\,,$$

By the result from Example 2.2.1, the first integral equals $f(z)$, and we need to prove that the second integral vanishes. But f being differentiable at z must be continuous at z. Let $\epsilon > 0$ be an arbitrary small number, and let $\delta > 0$ be such that, for $|\zeta - z| \le \delta$, $|f(\zeta) - f(z)| \le \epsilon$. Choose circle c with radius $r = \delta$. We have,

$$|\frac{1}{2\pi i}\int_c \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta| \le \frac{1}{2\pi}\int_c \frac{|f(\zeta) - f(z)|}{|\zeta - z|}\, d\zeta \le \frac{1}{2\pi}\int_c \frac{\epsilon}{\delta}\, d\zeta = \epsilon\,.$$

Since ϵ was arbitrary, the integral must vanish.
Consider now the difference quotient,

$$\begin{aligned}\frac{f(z+\Delta z) - f(z)}{\Delta z} &= \frac{1}{2\pi i \Delta z}\int_c f(\zeta)\left[\frac{1}{\zeta - z - \Delta z} - \frac{1}{\zeta - z}\right] d\zeta \\ &= \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta - z - \Delta z)(\zeta - z)}\, d\zeta \\ &= \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta - z)^2}\, d\zeta + \frac{\Delta z}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta - z)^2(\zeta - z - \Delta z)}\, d\zeta\end{aligned}$$

We need to show that the second integral vanishes as $\Delta z \to 0$. Let $|\Delta z| < \frac{r}{2}$. Then $|\zeta - z - \Delta z| > \frac{r}{2}$ and we have the following bound:

$$|\frac{\Delta z}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta - z)^2(\zeta - z - \Delta z)}\, d\zeta| \le \frac{|\Delta z|}{2\pi}\int_c \frac{M}{r^2 \frac{r}{2}}\, d\zeta = |\Delta z|\frac{2M}{r^2} \to 0 \quad \text{as } \Delta z \to 0\,,$$

where M is a bound for $|f(\zeta)|$ on c (Weierstrass at work).
Proof of the general formula proceeds by induction, and we leave it for Exercise 2.2.3. ∎

Do you realize that we have just proved that any complex differentiable function is automatically C^∞ (in the complex sense) ?

Let M be a bound of $f(z)$ on a circle with radius r, centered at $z = 0$. Cauchy's formula (2.4) for $n = 1$ (first derivative) implies the bound:

$$|f'(z)| \leq \frac{M}{r}\,.$$

In particular, if f is a globally bounded *entire* function then, passing with $r \to \infty$, we conclude that $f'(z) = 0$ and, therefore, f must be a constant function. This result is known as the *Liouville Theorem*. Let now

$$P(z) = a_0 + a_1 z + \ldots + a_n z^n$$

be a polynomial of order $n > 0$. We claim that $P(z)$ must have at least one root, say z_1. Indeed, if $P(z) \neq 0$ everywhere, then $1/P(z)$ is complex differentiable everywhere, and it converges to zero as $|z| \to \infty$. Consequently, it must be globally bounded (explain, why?) and, by the Liouville Theorem, must be a constant, a contradiction. If $P(z)$ has a single root, it must have exactly n roots (counting roots of order k as k roots). Indeed, it is sufficient to divide $P(z)$ by $(z - z_1)$, and repeat the reasoning for the resulting polynomial.
You have just proved the *Fundamental Theorem of Algebra*.

Exercises

Exercise 2.2.1 Show that the value of complex integral (2.2) is independent of the parametrization.

(5 points)

Exercise 2.2.2 Prove Corollary 2.2.1.

(5 points)

Exercise 2.2.3 Use induction to prove Cauchy formula (2.4).

(10 points)

Exercise 2.2.4 Evaluate the following integrals where c is a unit ccw circle. Do not put cart ahead of horses. You are not supposed to use the Residue Theorem at this point but you can use the Cauchy Theorem 2.2.1.

$$\text{(a)} \int_c (z^2 - \sin z)\,dz \qquad \text{(b)} \int_c \frac{\sin z}{z}\,dz \qquad \text{(c)} \int_c \frac{dz}{z^2 + z + 2}$$

$$\text{(d)} \int_c \frac{\sinh z}{z}\,dz \qquad \text{(e)} \int_c \frac{\cosh z}{z}\,dz \qquad \text{(f)} \int_c \frac{dz}{z^2(z^2+3)}\,.$$

(10 points)

2.3 Taylor and Laurent Series

We start with the remarkable result.

THEOREM 2.3.1 (Complex differentiable function is analytic)

Let $f(z)$ be a complex differentiable function defined in a domain $D \subset \mathbb{C}$. Then f is analytic in D, i.e.,

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \qquad |z - z_0| < r(z_0)\,,$$

for every $z_0 \in D$.

PROOF Let c be a ccw circle with radius r, centered at z_0, contained in D, see Fig.2.6 for illustration.

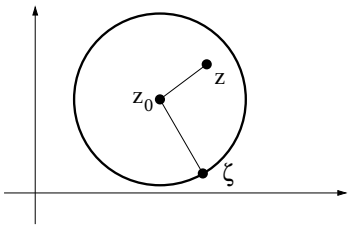

Figure 2.6
Circle used in the proof of Theorem 2.3.1.

By the Cauchy integral formula, for every z inside of c, we have,

$$f(z) = \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{\zeta - z}\,d\zeta = \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{\zeta - z_0}\left(\frac{1}{1 - \frac{z - z_0}{\zeta - z_0}}\right)\,d\zeta\,.$$

Denoting

$$t := \frac{z - z_0}{\zeta - z_0} \qquad |t| < 1\,,$$

and recalling the geometric series identity:

$$\frac{1}{1-t} = 1 + t + t^2 + \ldots + t^{n-1} + \frac{t^n}{1-t}\,, \tag{2.5}$$

we can expand the term in the parantheses to obtain:

$$f(z) = \frac{1}{2\pi i}\left\{\int_c \frac{f(\zeta)}{\zeta - z_0}\,d\zeta + (z - z_0)\int_c \frac{f(\zeta)}{(\zeta - z_0)^2}\,d\zeta + \ldots + (z - z_0)^{n-1}\int_c \frac{f(\zeta)}{(\zeta - z_0)^n}\,d\zeta\right\} + R_n$$

where

$$R_n = \frac{(z-z_0)^n}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta-z_0)^n(\zeta-z)}\,d\zeta\,.$$

By Theorem 2.2.2, the integrals in the first part are equal exactly to the terms in Taylor's series, and it is sufficient to show that the residual $R_n \to 0$ as $n \to \infty$. Denote the maximum of $|f|$ on c by M. Noting that

$$|\zeta - z| \geq r - \underbrace{|z-z_0|}_{=:\rho},$$

we estimate,

$$|R_n| = \frac{|z-z_)|^n}{2\pi}|\int_c \frac{f(\zeta)}{(\zeta-z_0)^n(\zeta-z)}\,d\zeta| \leq \frac{\rho^n}{2\pi}\frac{M}{r^n(r-\rho)}2\pi r = \frac{Mr}{r-\rho}\left(\frac{\rho}{r}\right)^n \to 0 \quad \text{as } n \to 0\,,$$

since $\rho/r < 1$. ∎

Isolated sigularities. If $f(z)$ is complex differentiable inside of a circle centered at z_0 *excluding* point z_0, we call point z_0 *an isolated singularity point*. Obviously, we cannot hope for a Taylor series expansion at z_0 but, amazingly, $f(z)$ can still be expanded into a series centered at z_0, the *Laurent series*.

THEOREM 2.3.2 (Laurent expansion)

Let $f(z)$ be complex differentiable in $D - \{z_0\}$ where D is a domain $D \subset \mathbb{C}$. Let c be a ccw circle with radius r, centered at z_0 and contained in D. The following expansion result holds.

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-z_0)^n \qquad |z-z_0| < r$$

where coefficients c_n are given by:

$$c_n = \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{(\zeta-z_0)^{n+1}}\,d\zeta\,.$$

Note that if $f(z)$ *happens* to be complex differentiable at z_0, the series reduces to the Taylor series and the integrals above coincide with terms in the Taylor series. But in general, of course, we cannot claim any relation between these integrals and derivatives of $f(z)$ at z_0 which simple may not exist.

PROOF Let c be a ccw circle with radius r, centered at z_0, contained in D. Let $z \neq z_0$ be any point inside of the circle. Introduce a second, smaller cicrle c_a with radius a such that z is in between the two circes, see Fig. 2.7 for illustration. Introducing a cut shown in the figure, we can claim the Cauchy expansion at z,

$$f(z) = \frac{1}{2\pi i}\int_d \frac{f(\zeta)}{\zeta - z}\,d\zeta$$

where d is the ccw union of the two circles and the two cuts. As the contributions over the cuts cancel each other, we end up with integrals over the two ccw circles only,

$$f(z) = \frac{1}{2\pi i}\int_c \frac{f(\zeta)}{\zeta - z}\,d\zeta - \frac{1}{2\pi i}\int_{c_a} \frac{f(\zeta)}{\zeta - z}\,d\zeta\,.$$

For the first integral, we use exactly the reasoning as in the proof of Theorem 2.3.1, the only difference being that we cannot claim the realtion of the countour integrals with the derivatives of $f(z)$. For the second integral, we use the representation:

$$\frac{1}{\zeta - z} = -\frac{1}{z - z_0}\,\frac{1}{1 - \frac{\zeta - z_0}{z - z_0}}\,. \tag{2.6}$$

Not that $|t| < 1$ for $t = \frac{\zeta - z_0}{z - z_0}$ with $\zeta \in c_a$. Using again the geometric series expansion (2.5), we use the reasoning fully analogous to that in proof of Theorem 2.3.1 to end up with the part of the Laurent expansion with negative powers. Details are left for Exercise 2.3.1. ∎

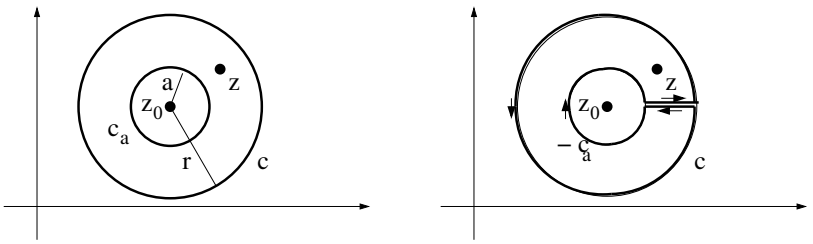

Figure 2.7
Circle used in the proof of Theorem 2.3.

Residue of function $f(z)$ at z_0 . Coefficient c_{-1} in the Laurent eaxpansion at z_0 is defined as the *residue of f at z_0*, denoted $\mathrm{res}_{z_0} f$. Note that

$$\mathrm{res}_{z_0} f = \frac{1}{2\pi i}\int_c f(\zeta)\,d\zeta\,.$$

Poles. If the Laurent series starts from a particular $-n, n > 0$, the isolated singularity is identified as *the pole of order n*. In the case of $n = 1$, we talk about a *simple pole*. If the Laurent expansion starts with $-\infty$, we talk about an *essential singularity*.

Example 2.3.1

Consider function:

$$f(z) = \frac{2}{z^2 - 1} = \frac{1}{z - 1} - \frac{1}{z + 1}\,.$$

We will play with the geometric series to come up with different Laurent and Laurent-like expansions for the function.

Case: $z_0 = -1$, $0 < |z+1| < 2$.

$$\frac{1}{z-1} = \frac{-1}{2-(z+1)} = -\frac{1}{2}\frac{1}{1-\frac{z+1}{2}} = -\sum_{n=0}^{\infty}\frac{(z+1)^n}{2^{n+1}}\,,$$

so,

$$f(z) = -\sum_{n=0}^{\infty}\frac{(z+1)^n}{2^{n+1}} - \frac{1}{z+1}\,.$$

Note that the negative part of the Laurent series reduces to the single term as the function has a simple pole at $z = -1$. Due to the singularity at $z = 1$, the radius of convergence at $z = -1$ is equal 2.

Case: $z_0 = 1$, $0 < |z-1| < 2$.

$$\frac{1}{z+1} = \frac{1}{2+(z-1)} = \frac{1}{2}\frac{1}{1+\frac{z-1}{2}} = \frac{1}{2}\frac{1}{1-\frac{1-z}{2}} = \frac{1}{2}\sum_{n=0}^{\infty}\frac{(1-z)^n}{2^n} = \sum_{n=0}^{\infty}(-1)^n\frac{(z-1)^n}{2^{n+1}}\,,$$

so,

$$f(z) = \sum_{n=0}^{\infty}(-1)^n\frac{(z-1)^n}{2^{n+1}} + \frac{1}{z-1}\,.$$

Again, we have a simple pole at $z_0 = 1$ and the radius of convergence $r = 2$.

Case: $z_0 = 2$, $1 < |z-2| < 3$. We have:

$$\frac{1}{z-1} = \frac{1}{(z-2)+1} = \frac{1}{z-2}\frac{1}{1+\frac{1}{z-2}} = \sum_{n=0}^{\infty}(-1)^n\frac{1}{(z-2)^{n+1}}$$

$$\frac{1}{z+1} = \frac{1}{3+(z-2)} = \frac{1}{3}\frac{1}{1+\frac{z-2}{3}} = \sum_{n=0}^{\infty}(-1)^n\frac{(z-2)^n}{3^{n+1}} \quad \text{so,}$$

$$f(z) = -\sum_{n=0}^{\infty}(-1)^n\frac{(z-2)^n}{3^{n+1}} + \sum_{n=0}^{\infty}(-1)^n\frac{1}{(z-2)^{n+1}}\,.$$

Note that $z_0 = 2$ is *not* an isolated singularity of the function. An examination of the proof of Theorem 2.3.2 reveals that we can construct a Laurent-like series in *any ring* centered at z_0. The domain of convergence: $1 < |z-2| < 3$ is implied by the singularities at $z = 1$ and $z = -1$. As $z_0 = 2$ is not an isolated singularity, the series above cannot be used to classify the 'singularity' at the point. For $|z-2| < 1$, the function can be expanded into its Taylor series at $z = 2$. Can you explain why the radius of convergence at $z_0 = 2$ is equal one?

Case: $|z| > 1$. We can also expand at 'infinity'. We have:

$$\frac{1}{z-1} = \frac{1}{z}\frac{1}{1-\frac{1}{z}} = \sum_{n=0}^{\infty}\frac{1}{z^{n+1}}$$

$$\frac{1}{z+1} = \frac{1}{z}\frac{1}{1+\frac{1}{z}} = \sum_{n=0}^{\infty}\frac{(-1)^n}{z^{n+1}} \quad \text{so,}$$

$$f(z) = -\sum_{n=0}^{\infty}\frac{1-(-1)^n}{z^{n+1}}\,.$$

Note again how the presence of singularities at 1 and -1 limits the domain of convergence. The expansion at ∞ is equivalent to the expansion of $z \to f(\frac{1}{z})$ at $z = 0$. □

Exercises

Exercise 2.3.1 Fill in details in the second part of the proof of Theorem 2.3.2.

(10 points)

Exercise 2.3.2 What kind of singularity (if any) does

$$f(z) := \frac{1}{2z} + \frac{1}{(2z)^2} + \frac{1}{(2z)^3} + \dots$$

have at $z = 0$?

(10 points)

Exercise 2.3.3 Find and classify all singularities of the following functions.

$$\text{(a)}\ \csc z = \frac{1}{\sin z} \qquad \text{(b)}\ \sec z = \frac{1}{\cos z} \qquad \text{(c)}\ \frac{1}{e^z - 1}$$

$$\text{(d)}\ \frac{1}{e^z + 1} \qquad \text{(e)}\ \frac{e^{-z}}{z(z^2 + 1)} \qquad \text{(f)}\ ze^{-\frac{1}{z}}$$

(10 points)

2.4 Residue Theorem

THEOREM 2.4.1 (Residue Theorem)

Let $D \subset \mathbb{C}$ be a domain, and $f(z)$ a complex-differentiable function defined on $D - \{z_1, \dots, z_N\}$, i.e., with isolated singularities at $z_1, \dots, z_N$. Let c be a closed ccw curve encirling the singular points. Then,

$$\int_c f(z)\, dz = 2\pi i \sum_{j=1}^{N} \text{res}_{z_j} f\,.$$

PROOF See Fig. 2.8 for the reasoning. Let $z \in D - \{z_1, \dots, z_N\}$ be an arbitrary point.

We encircle each point z_j with a sufficiently small circle c_j (so point z is outside of c_j) and then, similar to proof of Theorem 2.3.2, we introduce the cuts in between the circles to claim the Cauchy

integral expansion at z. The integrals over cuts cancel each other, and we end up with the result:

$$\int_c f(z)\,dz = \sum_{j=1}^{N} \int_{c_j} f(z)\,dz\,.$$

But each of the integrals over a small circle is equal exactly to the residue at the singular point. ∎

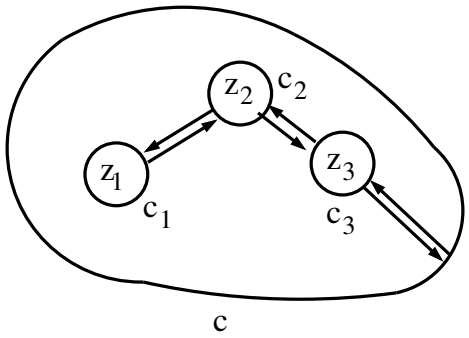

Figure 2.8
Proof of Theorem 2.4.1.

Computation of residues. In practice, we have an easy way to compute the residues. Suppose that $f(z)$ has a simple pole at z_0, i.e.,

$$f(z) = \sum_{n=-1}^{\infty} c_n (z - z_0)^n\,.$$

Multiplying both sides with $z - z_0$ and passing with $z \to z_0$, we obtain:

$$\lim_{z\to z_0} (z - z_0) f(z) = \lim_{z\to z_0} \sum_{n=-1}^{\infty} c_n (z - z_0)^{n+1} = c_{-1} + \lim_{z\to z_0} \sum_{n=0}^{\infty} c_n (z - z_0)^{n+1} = c_{-1}\,.$$

Notice that, if we make a wrong assumption about the multiplicity of the pole, i.e., we are dealing with a higher order pole or even an essential singularity, the limit will simply be infinite. Getting the infinity will indicate that our assumption about having a simple pole was wrong.

In the case of a second order (double) pole, we proceed slightly differently. We first multiply $f(z)$ with factor $(z - z_0)^2$,

$$(z - z_0)^2 f(z) = c_{-2} + c_{-1}(z - z_0) + \sum_{n=0}^{\infty} c_n (z - z_0)^{n+2}\,,$$

differentiate in z,

$$\frac{d}{dz}\left[(z - z_0)^2 f(z)\right] = c_{-1} + \sum_{n=0}^{\infty} c_n (n+2)(z - z_0)^{n+1}\,,$$

and only then pass to the limit with $z \to z_0$ to obtain:

$$c_{-1} = \lim_{z \to z_0} \frac{d}{dz} \left[(z - z_0)^2 f(z) \right] .$$

A similar procedure is used to compute the residue at a pole of order n,

$$c_{-1} = \lim_{z \to z_0} \frac{d^{n-1}}{dz^n} \left[(z - z_0)^n f(z) \right] .$$

Note again that, if are hypothesis about the order of the pole is wrong, we simply obtain the infinity which means that we have to try again.

There is no simple procedure for evaluating the residue at an essential singularity.

Example 2.4.1

Evaluate

$$\int_0^\infty \frac{x^2}{(x^2+4)^2}\, dx$$

using the Residue Theorem. We notice that the integrand is even, so

$$\int_0^\infty \frac{x^2}{(x^2+4)^2}\, dx = \frac{1}{2} \int_{-\infty}^\infty \frac{x^2}{(x^2+4)^2}\, dx .$$

The function is an L^1-function, so by the Lebesgue Theorem,

$$\int_{-\infty}^\infty \frac{x^2}{(x^2+4)^2}\, dx = \lim_{R \to \infty} \int_{-R}^R \frac{x^2}{(x^2+4)^2}\, dx .$$

We extend now segment $(-R, R)$ to the contour shown in Fig. 2.9. For sufficiently large R, we have

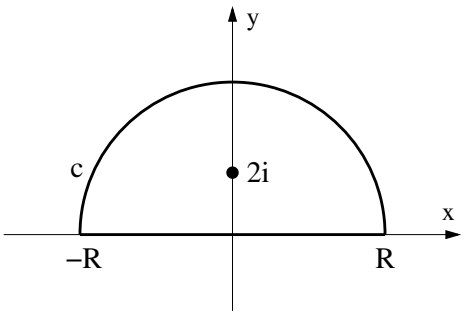

Figure 2.9
The closed contour used in Example 2.4.1

a single second order pole inside of the contour at $z = 2i$. By the Residue Theorem, the integral over the closed contour is equal to the residue at $z = 2i$. We have,

$$(z-2i)^2 \frac{z^2}{(z^2+4)^2} = \frac{z^2}{(z+2i)^2} \quad \frac{d}{dz} \frac{z^2}{(z+2i)^2} = \frac{2z(z+2i)^2 - z^2 2(z+2i)}{(z+2i)^4} = \frac{2z(z+2i) - 2z^2}{(z+2i)^3} = \frac{4iz}{(z+2i)^3}$$

so

$$\mathrm{res}_{2i} = \lim_{z\to 2i} \frac{4iz}{(z+2i)^3} = \frac{-8}{-64i} = \frac{-i}{8}\,.$$

Thus, the closed countour integral equals

$$2\pi i \frac{-i}{8} = \frac{\pi}{4}\,,$$

and, in particular, it is independent of R. We argue now that the integral over the semicircle c_R vanishes with $R\to\infty$. Indeed, the integrand is of order R^2/R^4, and the length of the semicircle is $2\pi R$ which leaves us still with an extra R in the denominator. The final value of the desired integral is thus $\frac{\pi}{8}$. Try to get the same result with standard real analysis tools, good luck ! Can you think of another contour to compute the integral ? □

The following example is a bit more complicated as it involves multi-valued functions and branch cuts.

Example 2.4.2

Compute:

$$\int_0^\infty \frac{x^{a-1}}{1+x}\,dx \qquad \text{where } 0<a<1\,.$$

First of all, note that the integrand is of order $\frac{1}{x^{2-a}}$ at ∞, and it is singular at 0 but of order $\frac{1}{x^{1-a}}$ which makes it summable (in $L^1(\mathbb{R})$). Secondly, function $z^{a-1} = r^{a-1}\theta^{a-1}$ is multi-valued (since the argument function is multi-valued). We choose the branch shown in Fig. 2.10. The choice is rather unique. The corresponding closed contour consists of segment (ϵ, R) on the side $\theta = 0$ of the branch, but it also includes the same segment on the side $\theta = 2\pi$. If we did not have a branch cut there, the two integrals would cancel each other out. As usual, we close the contour with the circle of radius R, and we will let $R\to\infty$ in order to converge to the integral over $(0,\infty)$. But, in order to cope with the singularity at $z=0$, we also have to cut out the origin with a cicrle $c_\epsilon, \epsilon < R$, and we will let $\epsilon\to 0$ as well. There is only one simple pole at $z=-1$, and the residue at the pole is:

$$\mathrm{Res}_{-1} = \lim_{z\to -1}(z+1)\frac{z^{a-a}}{1+z}) = \lim_{z\to -1} z^{a-1} = (-1)^{a-1} = (e^{i\pi})^{a-1} = e^{i(a-1)\pi}\,.$$

The integral over cicrle c_R is of order $\frac{1}{R^{1-a}}$ (one R comes from the jacobian), so it vanishes as $R\to\infty$. The integral over circle c_ϵ is of order ϵ^a (watch for the jacobian) and, therefore, it vanishes in the limit as well. The integral over the segment above the cut converges to the desired integral. We need to pay attention only to the segment on the side $\theta = 2\pi$ of the branch where

$$z^{a-1} = x^{a-1}e^{i2\pi(a-1)}\,.$$

Consequently, the integral is equal to

$$-\int_\epsilon^R \frac{x^{a-1}}{1+x}\,dx\, e^{i2\pi(a-1)}\,,$$

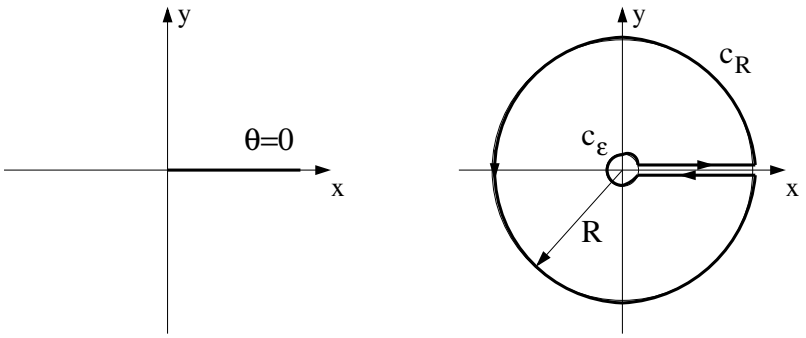

Figure 2.10
Branch cut for z^{a-1} and the closed contour used in Example 2.4.2.

and the sum of the two integrals over the straight segments is equal to:

$$(1 - e^{i2\pi(a-1)}) \int_{\epsilon}^{R} \frac{x^{a-1}}{1+x}\, dx\,.$$

Passing to the limit with $R \to \infty$ and $\epsilon \to 0$, we obtain,

$$(1 - e^{i2\pi(a-1)}) \int_{0}^{\infty} \frac{x^{a-1}}{1+x}\, dx = 2\pi i\, e^{i(a-1)\pi}\,.$$

After some algebra,

$$\int_{0}^{\infty} \frac{x^{a-1}}{1+x}\, dx = \frac{\pi}{\sin(\pi(1-a))}\,.$$

□

Exercises

Exercise 2.4.1 Verify the following integrals by means of the Residue Theorem.

(a) $\displaystyle\int_0^\infty \frac{x^2}{x^6+1}\, dx = \frac{\pi}{6}$ (b) $\displaystyle\int_0^\infty \frac{dx}{x^4+a^4} = \frac{\pi}{2\sqrt{2}a^3},\ a>0$ (c) $\displaystyle\int_0^\infty \frac{\cos x}{(x^2+1)^2}\, dx = \frac{\pi}{2e}$

(d) $\displaystyle\int_0^\pi \frac{dx}{a-\cos x} = \frac{\pi}{\sqrt{a^2-1}},\ a>1$ (e) $\displaystyle\int_{-\infty}^\infty \frac{dx}{4x^2+2x+1} = \frac{\pi}{\sqrt{3}}$ (f) $\displaystyle\int_0^{2\pi} \cos^2\theta\, d\theta = \pi$

(30 points)

Exercise 2.4.2 Evaluate

$$\int_0^\infty \frac{dx}{x^2+x+1}\,.$$

Hint: Consider the integral of

$$\frac{\ln z}{z^2+z+1}$$

and the contour used in Example 2.4.2.

(10 points)

Exercise 2.4.3 Evaluate

$$\text{(a)} \quad \int_0^\infty \frac{dx}{x^2+3x+2} \qquad \text{(b)} \quad \int_0^\infty \frac{dx}{x^2+4}$$

using techniques from previous exercises. Do you see the essential difference between the two cases ?

(10 points)

3

Spectral Analysis

3.1 Spectral Analysis in Finite Dimension

In this section we will review fundamentals of spectral analysis for linear operators defined on a finite dimensional space or, equivalently, matrices. We will consider the general complex setting first and then specialize the results to real spaces.

Eigenvalue problem. Let X be a finite-dimensional complex vector space equipped with an inner product $(\cdot,\cdot)$, and $A : X \to X$ a linear operator. For a finite dimensional space X, any linear operator is automatically continuous. A complex number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of operator A if there exists a non-zero vector $x \neq 0$ such that

$$Ax = \lambda x \quad \text{or, equivalently,} \quad (A - \lambda I)x = 0\,. \tag{3.1}$$

In particular, you can think about $X = \mathbb{C}^n$, A being a $n \times n$ complex matrix, and $(\cdot,\cdot)$ being the canonical inner product. It is easy to see that all eigenvectors x corresponding to the same eigenvalue λ, form a subspace of space X, denoted by X_λ and identified as the *eigenspace* corresponding to eigenvalue λ. In the case of a one-dimensional eigenspace, we speak sometimes about an *eigendirection*. Dimension of X_λ is identified as the *geometric multiciplicity* of eigenvalue λ.

Equation (3.1) will admit a non-zero solution iff the matrix (operator) $A - \lambda I$ is singular, i.e.

$$\det(A - \lambda I) = 0\,,$$

(Recall the definition of the determinant of a linear operator, see [5], p.184). Upon expanding the equation in λ (utilizing the multilinearity of the determinants), we obtain the so-called *characteristic equation* for operator (matrix) A,

$$(-1)^n\lambda^n + (-1)^{n-1}I_1\lambda^{n-1} + \ldots + I_n = 0 \tag{3.2}$$

where $n = \dim X$. Coefficients $I_i, i = 1,\ldots,n$, are known as the *invariants* of the characteristic equation. For a matrix A_{ij}, we have,

$$I_1 = A_{ii} = A_{11} + \ldots + A_{nn}, \qquad I_n = \det A\,,$$

In three space dimension,

$$I_2 = \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} + \begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} + \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}\,.$$

It follows from the *Fundamental Theorem of Algebra* that the algebraic characteristic equation (3.2) has exactly n roots taking into account possible multiplicities. In other words, we have

$$(-1)^n\lambda^n + (-1)^{n-1}I_1\lambda^{n-1} + \ldots + I_n = -(1)^n(\lambda-\lambda_1)^{n_1}\cdots(\lambda-\lambda_k)^{n_k}$$

where $\lambda_1, \ldots, \lambda_k$ are complex roots of the characteristic equation, and the multiciplities sum up to n,

$$n_1 + \ldots + n_k = n\,.$$

Number n_k is known as the *algebraic multiplicity* of eigenvalue λ_k. One can show that *the geometric multiplicity never exceeds the algebraic multiplicity* (a non-trivial statement) but it can be smaller, see the example below.

Note that, for a real operator (matrix), the eigenvalues and the corresponding eigenvectors may still be complex. Hence it makes little sense to consider the real setting separately. However, for a real operator,

$$(A-\lambda I)x = 0 \quad \Rightarrow \quad \overline{(A-\lambda I)x} = (A-\overline{\lambda})\overline{x} = 0$$

implies that, if λ is an eigenvalue of A, then so is its complex conjugate $\bar{\lambda}$, the corresponding eigenvectors are also complex conjugates or each other, and the corresponding eigenspaces are of the same dimension. In particular, if n is odd, then a real operator A must have at least one real eigenvalue.

Example 3.1.1

Consider

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\,.$$

Show that $\lambda = 1$ is the only eigenvalue of A with algebraic multiplicity equal two but geometrical multiplicity equal one only. □

3.1.1 Self-Adjoint Operators. Elementary Spectral Theorem

Let A be a self-adjoint operator, i.e.,

$$(Ax, y) = (x, Ay) \quad x, y \in X\,,$$

i.e., $A^* = A$. For matrices, this is equivalent for A being a *hermitian matrix* (symmetric in the real case).

Let λ be an eigenvalue of operator A and e the corresponding eigenversor, i.e., normalized eigenvector. We have,

$$\lambda = \lambda \underbrace{(e,e)}_{=1} = (\lambda e, e) = (Ae, e) = (e, Ae) = (e, \lambda e) = \bar{\lambda}(e,e) = \bar{\lambda}\,,$$

i.e. λ must be real. Note that the reasoning applies to any self-adjoint operator including the case of an infinite dimensional space X. *All eigenvalues of a self-adjoint operator are real.*

Consider now two eigenpairs (λ_i, e_i), (λ_j, e_j) where $\lambda_i \neq \lambda_j$. We have,

$$\lambda_i(e_i, e_j) = (\lambda_i e_i, e_j) = (Ae_i, e_j) = (e_i, Ae_j) = (e_i, \lambda_j e_j) = \lambda_j(e_i, e_j)$$

which implies

$$(\lambda_i - \lambda_j)(e_i, e_j) = 0 \quad \Rightarrow \quad (e_i, e_j) = 0\,.$$

Eigenvectors corresponding to different eigenvalues must be orthogonal to each other. Again, the observation applies to any general setting.

Finally, one can show that, for an eigenvalue of a self-adjoint operator, the corresponding *algebraic and geometric multiciplities are equal* (a non-trivial result). In other words, if k is the algebraic multiplicity of eigenvalue λ then $\dim X_\lambda = k$ and, in particular, one can select an orthonormal basis * for X_λ. Consequently, one always select an orthonormal basis e_i for X consisting of eigenversors of operator A. Let λ_i be the eigenvalues corresponding to eigenversors e_i. We usually organize the eigenpairs in the order of ascending eigenvalues,

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-1} \leq \lambda_n\,.$$

Let x be expanded in the eigenbasis e_i, i.e. $x = \sum_{i=1}^n x_i e_i$. Then

$$Ax = A(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n x_i Ae_i = \sum_{i=1}^n \lambda_i x_i e_i\,.$$

The action of operator A on spectral components x_i reduces to the multiplication with λ_i. This is the message behind the simplest version of the *Spectral Theorem for Self-Adjoint Operators*.

Diagonalization of hermitian matrices. Suppose, we have represented a vector x in two orthonormal bases, an *old basis* e_j, and a *new basis* e'_k,

$$x'_k e'_k = x = x_j e_j\,.$$

How do we compute the new components x'_k, provided we already know the old components x_j? Multiplying both sides of the equation above with versor e'_i (from the left), we get,

$$e'_k \underbrace{(e'_i, e'_k)}_{=\delta_{ik}} = x_j \underbrace{(e'_i, e_j)}_{=:\alpha_{ij}}$$

or, equivalently,

$$x'_i = \alpha_{ij} x_j\,.$$

Matrix α_{ij} is known as the *transformation matrix* between the old and the new coordinates. It is easy to see that the transformation matrix is *orthonormal*, i.e. $\alpha^{-1} = \alpha^* := \overline{\alpha^T}$. Consequently,

$$x_j = \overline{\alpha}_{ij} x'_i\,.$$

*E.g., via the Gram-Schmidt orthonormalization procedure.

The transformation rule for vectors implies now the corresponding transformation for the matrix representation of operator A (tensors). Indeed, we have,

$$y'_i = \alpha_{ik} y_k = \alpha_{ik} A_{kl} x_l = \alpha_{ik} A_{kl} \overline{\alpha}_{jl} x'_j = \underbrace{\alpha_{ik} \overline{\alpha}_{jl} A_{kl}}_{=A'_{ij}} x'_j \,,$$

i.e.,

$$A'_{ij} = \alpha_{ik} \overline{\alpha}_{jl} A_{kl} \,.$$

Of a particular importance is the question whether we can find a transformation matrix α_{ij} (meaning new system of coordinates) such that the matrix representation of operator A in the new system of coordinates becomes diagonal. The answer comes from the Spectral Theorem. If we set,

$$\alpha_{ij} = (e'_i, e_j)$$

where e'_i are the eigenversors of operator A (they consititute rows of the transformation matrix), the new representation of operator A will reduce to

$$A'_{ij} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \,.$$

Diagonalization of arbitrary matrices. Some of the facts for the hermitian matrices generalize to arbitrary matrices. First of all, we have to work now with arbitrary, not necessary orthonormal bases. The transformation matrix from an old basis e_j to a new basis e'_i is obtained by representing the new basis vectors in the old basis [†],

$$e'_i = \alpha_{ij} e_j \qquad \Rightarrow \qquad e_j = \alpha^{-1}_{ji} e'_i \,.$$

We have now,

$$x'_i e'_i = x = x_j e_j = \alpha^{-1}_{ji} x_j e'_i$$

which yields the transformation formula:

$$x'_i = \alpha^{-1}_{ji} x_j \,.$$

The inverse relation is:

$$x_j = \alpha_{ji} x'_i \,.$$

As before, this leads to the transformation rule for matrix representations of linear maps (second order tensors),

$$A'_{ij} = \alpha^{-1}_{ki} A_{kl} \alpha_{lj}$$

or, in the matrix form,

$$A' = \alpha^{-T} A \alpha \,.$$

If operator (matrix) A has n linearly independent eigenvectors e_i, we can use them for an (eigen)basis and, in the corresponding system of coordinates, matrix A will again be diagonalized with the eigenvalues on the

[†] The definition is consistent with the one for the orthonormal bases.

diagonal. The important question to as is whether *we can always find n linearly independent eigenvectors*? A partial answer is provided by the following lemma.

LEMMA 3.1.1

Assume operator A has n distinct eigenvalues λ_j. Then the corresponding eigenvectors e_j are linearly independent.

PROOF Let

$$\alpha_1 e_1 + \ldots + \alpha_j e_j + \ldots + \alpha_n e_n = 0\,.$$

Applying operator A k times, we get:

$$\alpha_1 \lambda_1^k e_1 + \ldots + \alpha_j \lambda_j^k e_j + \ldots + \alpha_n \lambda_n^k e_n = 0\,.$$

Assume that

$$|\lambda_j| = \max_i |\lambda_i|$$

and divide the equation above by λ_j to obtain:

$$\alpha_1 \left(\frac{\lambda_1}{\lambda_j}\right)^k e_1 + \ldots + \alpha_j e_j + \ldots + \alpha_n \left(\frac{\lambda_n}{\lambda_j}\right)^k e_n = 0\,,$$

Passing with $k \to \infty$, we get $\alpha_j = 0$. We can eliminate now the j-th term from the linear combination and keep repeating the reasoning until we prove that all coefficients must vanish. ∎

Clearly, we can anticipate trouble only in the case of eigenvalues with algebraic multiplicity greater than one. We shall study the general case in Section 3.2.

Exercises

Exercise 3.1.1 An undergraduate sanity check. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\,,$$

and determine its eigenvalues and the corresponding eigenspaces. Use the eigenvectors to form an eigenbasis diagonalizing the matrix, determine the transformation matrix from the canonical basis to the eigenbasis and double check that, in the new system of coordinates, the matrix becomes indeed diagonal.

(10 points)

3.2 Jordan Decomposition (Representation) Theorem

In this section we return to a general, *arbitrary* operator $T : X \to X$ where X is a finite-dimensional space. We have learned in Section 3.1 that any self-adjoint operator can be diagonalized using a special orthonormal basis composed of its eigenversors. Can we claim a similar result for general operators? The following exposition is based on [4].

Preliminaries. Use powers of operator T to define a sequence of null and range spaces:

$$T^0 = I, \quad T^n = \underbrace{T \circ \ldots \circ T}_{n \text{ times}}, \quad N_n := \mathcal{N}(T^n), \quad R_n = \mathcal{R}(T^n)\,.$$

We have the following simple observations.

- Sequence N_n is increasing: $N_n \subset N_{n+1}$.
  Indeed, let $x \in N_n$, i.e., $T^n x = 0$. Then $T^{n+1}x = T(T^n x) = T0 = 0$, i.e., $x \in N_{n+1}$.

- Sequence R_n is decreasing: $R_n \supset R_{n+1}$.
  Let $y \in R_{n+1}$. There exists x such that $y = T^{n+1}x$, i.e., $y = T^n(Tx)$. Consequently, there exists $x_1 (= Tx)$ such that $y = T^n x_1$ which implies that $y \in R_n$.

- We have thus two monotone sequences of null and range spaces:

$$\{0\} = N_0 \subset N_1 \subset N_2 \subset \ldots$$
$$X = R_0 \supset R_1 \supset R_2 \supset \ldots$$

  The inclusions above can be proper or can turn into equalities. We claim that once we encounter an equality, all remaining inclusions will turn into equalities as well. We show that

$$R_n = R_{n+1} \quad \Rightarrow \quad R_{n+1} = R_{n+2}\,.$$

  Let $y \in R_{n+1}$. There exists x such that $y = T^{n+1}x = TT^n x$. But $T^n x \in R_n = R_{n+1}$ and, therefore, there exists z such that $T^n x = T^{n+1}z$. Consequently, $y = TT^{n+1}z = T^{n+2}z$ which proves that $y \in R_{n+2}$. Note that the Rank and Nullity Theorem implies that the two sequences of spaces turn into constant sequences *simultaneously*. Define $m = \min\{n : R_n = R_{n+1}\}$.

REMARK 3.2.1 Note that, for a self-adjoint operator T, $m = 1$, i.e., $\mathcal{N}(T) = \mathcal{N}(T^2)$. Take $x \in \mathcal{N}(T^2)$, i.e., $T^2 x = 0$. Then $0 = (T^2 x, x) = (Tx, Tx) = \|Tx\|^2$ implies that $Tx = 0$, i.e., $x \in \mathcal{N}(T)$. ∎

Invariant subspaces. We sat that a subspace $Y \subset X$ is *invariant under transformation* $T : X \to X$, if T sets Y into Y, i.e., $TY \subset Y$. The transformation:

$$T_Y : Y \to Y, \quad T_Y x := Tx$$

is called *the part of* T *in* Y. We have the following observations.

- N_n are invariant under T.
  Let $y \in T(N_n)$. There exists $x \in N_n$ such that $y = Tx$. Then $0 = T^n x = T^{n-1}Tx = T^{n-1}y$, so $y \in N_{n-1} \subset N_n$.
- R_n are invariant under T:
  $T(R_n) = R_{n+1} \subset R_n$.
- The part T_{R_n} is singular (has a non-trivial null space) for $n < m$. Explain, why?

Decomposition of the operator. We say that operator T is *decomposed (reduced) by subspaces* $M_1, \ldots, M_s$ if M_i are invariant under T, and $X = M_1 \oplus \ldots \oplus M_s$.

Nilponent operators. We say that operator T is *nilpotent* if $T^n = 0$ for some $n \geq 1$. In the trivial case $n = 1$, the operator is simply 0. Note that T must be singular, i.e., 0 is an eigenvalue of T with $\mathcal{N}(T)$ being the corresponding eigenspace. It turns out that 0 is the only eigenvalue of T.

LEMMA 3.2.1

Operator T is nilpotent iff zero is the only eigenvalue of T.

PROOF Suppose, to the contrary, that there exists $\lambda \neq 0, x \neq 0$ such that $Tx = \lambda x$. Then $T^n x = \lambda^n x \neq 0$, *for every* n. Conversely, let m be the index after which the sequence of spaces R_n stops decreasing. If R_m is trivial then $N_m = X$ and we are done. Suppose, to the contrary, that R_m is non-trivial, and consider part T_{R_m}. Any operator must have at least one eigenvalue (explain, why?) and so does T_{R_m}. Now, for a non-trivial R_m, any eigenvalue of part T_{R_m} is also an eigenvalue of T, and T is assumed to have the zero eigenvalue only. Consequently, T_{R_m} must be singular which contradicts the fact that $R_{m+1} = R_m$. ∎

Example 3.2.1

Let $X = \mathbb{C}^2$, and

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

then $A^2 = 0$. This is perhaps the simplest example of a nilpotent operator. □

We continue now our observations on sequences N_n and R_n.

- $X = N_m \oplus R_m$.
  By the Rank and Nullity Theorem, the dimensions of N_n and R_n sum up to n. It is sufficient thus to show that $N_m \cap R_m = \{0\}$. Suppose to the contrary that there exists $y \in R_m \cap N_m, y \neq 0$. Then there exists $x \in X$ such that $y = T^m x$ which in turn implies that $x \notin N_m$, since $y \neq 0$. But, at the same time, $0 = T^m y = T^m(T^m x) = T^{2m} x$ and, therefore, $x \in N_{2m} = N_m$, a contradiction.

- Consequently, T is decomposed by spaces N_m and R_m.

- The part T_{N_m} is nilpotent.
  For $m = 1$, T_{N_m} is zero. Let $m > 1$. Suppose to the contrary that there exists a non-zero eigenvalue μ of T_{N_m} with a corresponding eigenvector $x \in N_m$. As sequence N_n is increasing, ther must be an index n such that $x \in N_n$ but $x \notin N_{n-1}$. We have:

  $$0 \neq T^{n-1}x = T^{n-1}(\mu^{-1}Tx) = \mu^{-1}T^n x = 0\,,$$

  a contradiction.

Generalized eigenspace. Let λ be an eigenvalue of map T. The subspace $X_\lambda := \mathcal{N}((T - \lambda I)^m)$ (m denotes the index from which the sequence of null spaces becomes constant) is called the *generalized eigenspace corresponding to eigenvalue* λ.

REMARK 3.2.2 Note that, according to Remark 3.2.1, for a self-adjoint operator T, the generalized eigenspace coincides with the eigenspace. ∎

Let now λ_1 be any eigenvalue of T. Introduce $X_1 := X_{\lambda_1} = \mathcal{N}((T - \lambda_1 I)^m), Y_1 := \mathcal{R}((T - \lambda_1 I)^m)$. According to the results above, $T - \lambda_1 I$ is reduced by X_1 and Y_1, and the part $(T - \lambda_1 I)_{X_1} =: D_1$ is nilpotent.

$$(T - \lambda_1 I)x = D_1 x_1 + (T - \lambda_1 I)y_1 \quad \text{where } x = x_1 + y_1, \quad x_1 \in X_1, y_1 \in Y_1\,.$$

Equivalently,

$$Tx = (\lambda_1 + D_1)x_1 + T_{Y_1} y_1\,.$$

Note that D_1 may be zero. This is, in particular, the case when T is self-adjoint. We proceed then with part T_{Y_1} introducing a second generalized eigenspace[‡] X_2 with complement Y_2 in Y_1, and continue the process until we exhaust all eigenvalues, i.e., when $Y_s = \{0\}$. Our final result reads as follows.

[‡]Note that an eigenvalue of T_{Y_1} is also an eigenvalue of T.

THEOREM 3.2.1 Jordan Decomposition Theorem

Let $\lambda_1, \ldots, \lambda_s$ denote the eigenvalues of map T, with corresponding generalized eigenspaces $X_1, \ldots, X_s$. Then

$$X = X_1 \oplus X_2 \oplus \ldots \oplus X_s$$

and the map T can be represented as

$$Tx = \sum_{j=1}^{s} \lambda_j x_j + \underbrace{\sum_{j=1}^{s} D_j x_j}_{=:Dx} \tag{3.3}$$

where x_j are components of x corresponding to the decomposition above, and operator D is nilpotent.

PROOF Note that

$$Dx = \mu x \quad \Leftrightarrow \quad D_j x_j = \mu x_j \,, \qquad j = 1, \ldots, s \,.$$

Consequently, D may have only a zero eigenvalue and, therefore, is nilpotent. ∎

Let $P_j : X \ni x \to x_j \in X_j$ be (linear) projections correponding to the decomposition. Spectral representation (3.3) cand be rewritten in an argumentless form as:

$$T = \sum_{j=1}^{s} \lambda_j P_j + D \,.$$

Matrix representation of a nilpotent. If you think in terms of matrices, we have learned that any matrix T can be block-diagonalized using the generalized eigenspaces $X_j := \mathcal{N}((T - \lambda I)^m)$ corresponding to eigenvalues λ. We will try now to come up with a clever selection of a basis for each X_j that would make the matrix representation of

$$Tx_j = \lambda_j x_j + D_j x_j \,.$$

as close as possible to a diagonal matrix. This motivates us to study the matrix representation of nilpotent D_j. It is sufficient to focus on a single generalized eigenspace X_j and the representation of D_j. Replace X_j with X and D_j with D. Assume a non-trivial case of $\dim X > 1$. Let m be such that $D^m = 0$ but $D^{m-1} \neq 0$. Let $x_1^1, \ldots, x_{p_1}^1$ be a basis in R_{m-1}. There exist $x_i^m, i = 1, \ldots, p_1$ such that $x_i^1 = D^{m-1} x_i^m$. Set $x_i^2 := D^{m-2} x_i^m$, so $D x_i^2 = x_i^1$. We claim that vectors $x_1^1, \ldots, x_{p_1}^1, x_1^2, \ldots, x_{p_1}^2$ are linearly independent. Indeed, let

$$\sum \alpha_i x_i^1 + \sum \beta_j x_i^2 = 0 \,,$$

Applying D to both sides, we get,

$$\underbrace{D(\sum \alpha_i x_i^1)}_{=0} + \sum \beta_j \underbrace{D x_j^2}_{=x_j^1} = 0$$

which implies that β_j's are zero. The linear combination reduces to the first term and all α_i's must be zero as well.

Space R_{m-2} can be decomposed into the null space of D and an algebraic complement Y,

$$R_{m-2} = \mathcal{N}(D|_{R_{m-2}}) \oplus Y \,.$$

By construction, restriction $D|_Y$ is an isomorphism onto R_{m-1}, and one can select

$$Y = \mathrm{span}\{x_j^2 \,:\, j = 1, \ldots, p_1\} \,.$$

Original vectors x_j^1 are in $\mathcal{N}(D|_{R_{m-2}})$, and they are linearly independent, but they may not span the whole null space. We complete them now with vectors $x_{p_1+1}^2, \ldots, x_{p_2}^2$ to a basis in $\mathcal{N}(D|_{R_{m-2}})$. By construction thus, vectors $x_1^1, \ldots, x_{p_1}^1, x_1^2, \ldots, x_{p_1}^2, x_{p_1+1}^2, \ldots, x_{p_2}^2$ provide a basis for R_{m-2}.

We continue in the same way until we obtain a basis for $R_0 = X$:

$$x_1^1, \ldots, x_{p_1}^1, \ldots, x_1^m, \ldots, x_{p_m}^m \,.$$

The procedure is illustrated below for $m = 3$.

$$\begin{array}{lllllllllll}
R_{m-1} & x_1^1 & \ldots & x_{p_1}^1 & & & & & & & \\
 & \downarrow & & \downarrow & & & & & & & \\
R_{m-2} & x_1^2 & \ldots & x_{p_1}^2 & x_{p_1+1}^2 & \ldots & x_{p_2}^2 & & & & \\
 & \downarrow & & \downarrow & \downarrow & & & \downarrow & & & \\
R_{m-3} & x_1^3 & \ldots & x_{p_1}^3 & x_{p_1+1}^3 & \ldots & x_{p_2}^3 & x_{p_2+1}^3 & \ldots & x_{p_3}^3 \,.
\end{array}$$

By construction, we obtain a basis for X.

If we rearrange now the vectors columnwise (by *trains* of generalized eigenvectors),

$$x_1^1, x_1^2, \ldots, x_1^n, x_2^1 \ldots \,,$$

the matrix representation of nilpotent D will look like this:

$$\begin{pmatrix}
0 & 1 & & & & & & & \\
 & 0 & 1 & & & & & & \\
 & & 0 & 1 & & & & & \\
 & & & 0 & & & & & \\
 & & & & 0 & 1 & & & \\
 & & & & & 0 & 1 & & \\
 & & & & & & 0 & & \\
 & & & & & & & 0 & 1 \\
 & & & & & & & & 0
\end{pmatrix}$$

In the picture above, we have three eigenvectors with corresponding *trains* of generalized eigenvectors. Rep-

resentation of operator T in the generalized eigenspace corresponding to eigenvalue λ will look as follows:

$$\begin{pmatrix} \lambda & 1 & & & & & & \\ & \lambda & 1 & & & & & \\ & & \lambda & 1 & & & & \\ & & & \lambda & & & & \\ & & & & \lambda & 1 & & \\ & & & & & \lambda & 1 & \\ & & & & & & \lambda & \\ & & & & & & & \lambda \end{pmatrix}$$

The matrix representation of map T in the (generalized) eigenbasis vectors consists of blocks like this,

$$\begin{pmatrix} \lambda_1 & 1 & & & & & & & \\ & \lambda_1 & 1 & & & & & & \\ & & \lambda_1 & & & & & & \\ & & & \lambda_1 & & & & & \\ & & & & \lambda_2 & 1 & & & \\ & & & & & \lambda_2 & & & \\ & & & & & & \lambda_2 & & \\ & & & & & & & \cdots & \\ & & & & & & & & \lambda_s & 1 \\ & & & & & & & & & \lambda_s \end{pmatrix}$$

and it is not diagonal but this is as close to a diagonal matrix as we can get.

REMARK 3.2.3 Note that we can conclude a posteriori that the *dimension of the generalized eigenspace matches exactly the algebraic multiciplicity of* λ. ∎

Example 3.2.2

Consider the following matrix.

$$A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

It is an upper triangular matrix so, without any computations (explain, why?) we conclude that we have a triple eigenvalue $\lambda_1 = \lambda_2 = \lambda_3 = 1$, and a single eigenvalue $\lambda_4 = 2$. The eigenproblem for $\lambda_4 = 2$ reduces to the system of equations:

$$A = \begin{pmatrix} -1 & 1 & 2 & 0 \\ 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which gives the solution:

$$(x_1, x_2, x_3, x_4)^T = (5t, 3t, t, 0)^T \quad \text{where } t \in \mathbb{R}\,.$$

As expected, the eigenspace is one-dimensional. We can select any $t \neq 0$ for a concrete eigenvector. The eigenproblem for the triple eigenspace $\lambda_1 = 1$ reduces to:

$$A = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which also gives a one-dimensional eigenspace only:

$$(x_1, x_2, x_3, x_4)^T = (t, 0, 0, 0)^T \quad \text{where } t \in \mathbb{R}\,. \tag{3.4}$$

However, this eigenvector initiates now a train of generalized eigenvectors. To get the first one, we need to solve:

$$A = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} t \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which gives

$$(x_1, x_2, x_3, x_4)^T = (s, t, 0, 0)^T \quad \text{where } s, t \in \mathbb{R}\,. \tag{3.5}$$

To get the second and the last generalized eigenvector of the train, we consider:

$$A = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s \\ t \\ 0 \\ 0 \end{pmatrix}$$

leading to:

$$(x_1, x_2, x_3, x_4)^T = (r, s - \frac{2}{3}t, \frac{1}{3}t, -\frac{1}{6}t)^T \quad \text{where } r, s, t \in \mathbb{R}\,. \tag{3.6}$$

Note that formula (3.4) represents the eigenspace, i.e. $\mathcal{N}(A - \lambda I)$, formula (3.5) represents $\mathcal{N}(A - \lambda I)^2$, and formula (3.6) represents the generalized eigenspace, i.e. $\mathcal{N}(A - \lambda I)^3$. It is important to compute the generalized eigenvectors parametrically and, if needed, fix the parameters at the end to conclude with concrete generalized eigenvectors. We can select here, e.g., $r = s = t = 1$.

□

Exercises

Exercise 3.2.1 Consider the nilpotent matrix (all unfilled entries are zero):

$$D := \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

Determine minimal index m for which $D^m = 0$. Construct explicitly the sequence of null and range spaces,

$$N_1 \subset N_2 \ldots \subset N_m = \mathbb{R}^5$$
$$R_1 \supset R_2 \ldots \supset R_m = \{0\},$$

and study the procedure of selecting the generalized eigenvectors resulting in the Jordan representation of nilpotent D.

(10 points)

Exercise 3.2.2 Consider the matrix:

$$A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Determine generalized eigenvectors of matrix A and the corresponding Jordan form.

(10 points)

3.3 Sturm-Liouville Theory

In this section, we will take our study of self-adjoint operators to infinite dimensional spaces. We are given infinite-dimensional Hilbert space X with an inner product (x, y). In practice, we will restrict ourselves to L^2-spaces, possibly with a weight. For simplicity, we will restrict ourselves to the case of linear operators $A : X \to X$.

The focus of this section is on 1D differential operators. Let $I = (a, b) \subset \mathbb{R}$, $X = L^2(I)$. We start with a set (in)formal examples.

Example 3.3.1

We consider complex-valued functions, and start with the simplest example of the first derivative operator.

$$A : L^2(I) \supset D(A) \ni u \to Au = iu' \in L^2(I)$$
$$D(A) := \{u \in L^2(I) : u \text{ is sufficiently regular, and } u(a) = 0\}.$$

Let us skip for a moment the regularity issues and focus on the boundary conditions (BC). In order to compute the L^2-adjoint of the operator, we multiply Lu (in the sense of the L^2-product) with a function $v(x)$ and integrate by parts,

$$\int_a^b \underbrace{iu'(x)}_{=Au} \overline{v(x)}\, dx = -\int_a^b iu(x)\overline{v'(x)}\, dx + (iu\bar{v})|_a^b$$
$$= \int_a^b u(x) \underbrace{\overline{iv'(x)}}_{=:A^*v} dx + (iu\bar{v})|_a^b .$$

If we disregard for a moment the boundary terms, we say that the ***formal adjoint*** of A, $A^*v = iv'$. The operator is thus ***formally self-adjoint***. Now, concerning the BCs, the term at $x = a$ vanishes for $u \in D(A)$. In order to make the term at $x = b$ vanish as well, we have no choice but assume that $v(b) = 0$,

$$D(A^*) := \{v \in L^2(I) \,:\, v \text{ is sufficiently regular, and } v(b) = 0\}\,.$$

We see that $D(A^*)$ is different from $D(A)$. The operator is thus not self-adjoint. In the end we have,

$$(Au, v) = (u, A^*v) \qquad \forall\, u \in D(A), v \in D(A^*)\,. \tag{3.7}$$

You might be wondering whether we could assume extra BCs for v, e.g.,

$$\text{(new) } D(A^*) := \{v \in L^2(I) \,:\, v \text{ is sufficiently regular, and } v(a) = v(b) = 0\}\,.$$

The extra condition at $x = 0$ is not necessary for the boundary terms to vanish since we already have $u(a) = 0$ but would it hurt if we assume it? The answer is related to the precise definition of the L^2-adjoint. Defining the adjoint involves finding out the formula for A^*v ***and identifying*** its domain $D(A^*)$ as the ***maximality*** of functions v for which the identity (3.7) holds. In other words, we look for ***minimal conditions*** which make the boundary terms vanish.

Another interesting case is when $I = \mathbb{R}$. We have no boundary then and the BC's are removed from the definitions of both $D(A)$ and $D(A^*)$. Pending a discussion on regularity assumptions, $D(A) = D(A^*)$. The operator ***is then self-adjoint !*** □

Example 3.3.2

Consider real-valued functions and a general ***diffusion-reaction*** operator,

$$Au := -(au')' + cu$$
$$D(A) := \{u \in L^2(I) \,:\, u \text{ is sufficiently regular, and } \alpha_0 u(a) + \beta_0 u'(a) = 0, \quad \alpha_1 u(b) + \beta_1 u'(b) = 0\}\,.$$

Here $a(x) > 0$, $c(x)$ are sufficiently regular coefficients [§], and $\alpha_i^2 + \beta_i^2 > 0, i = 1, 2$, i.e., at least one of the coefficients α_i, β_i is different from zero. For $\alpha \neq 0, \beta = 0$, we have a ***Dirichlet***[¶] BC, for $\alpha = 0, \beta \neq 0$, we have a ***Neumann*** [‖] BC. When both constants are different from zero, we speak of ***Cauchy*** or ***Robin*** [**]BC.

Integration by parts reveals that

$$A^*v = -(av')' + cv$$
$$D(A^*) = \{v \in L^2(I) \,:\, v \text{ is sufficiently regular, and } \alpha_0 v(a) + \beta_0 v'(a) = 0, \quad \alpha_1 v(b) + \beta_1 v'(b) = 0\}\,.$$

[§] For instance, $a \in \overline{C^1}(\bar{I}), c \in C(\bar{I})$.

[¶] After a German mathematician - Johann Peter Gustav Lejeune Dirichlet (1805 – 1859).

[‖] After a German mathematician - Carl Neumann (1832-1925).

[**] After a French mathematician - Victor Gustave Robin (1855–1897).

Thus, pending the regularity considerations, the operator is self-adjoint.

Note that for the operator including a convection term,

$$Au := -(au')' + bu' + cu\,,$$

the formal adjoint is:

$$A^*v = -(av')' - (bv)' + cv = -(av')' - bv' + (c - b')v\,.$$

There is no chance for the operator to be self-adjoint, even with a constant advection vector b. □

Example 3.3.3 (Periodic BCs)

Consider complex-valued functions and a general *diffusion-reaction* operator with periodic BC,

$$\begin{aligned} &Au := -(au')' + cu \\ &D(A) := \{u \in L^2(I) \,:\, u \text{ is sufficiently regular, and } u(a) = u(b), \quad u'(a) = u'(b)\} \end{aligned}$$

with an additional assumption on periodicity of the diffussion coefficient,

$$a(a) = a(b)\,.$$

Integration by parts reveals that the operator is self-adjoint. □

Example 3.3.4

An interesting special situation occurs when coefficient $a(x)$ vanishes at the endpoints of the interval. Let $I = (-1, 1)$. Consider the Legendre operator,

$$\begin{aligned} &Au := -((1 - x^2)u')' \\ &D(A) := \{u \in L^2(I) \,:\, u \text{ is sufficiently regular}\}\,. \end{aligned}$$

Integration by parts reveals that the operator is self-adjoint. □

Topological dual and the transpose operator. The space of all linear (antilinear) functionals defined on X that are continuous, is identified as the *topological dual* of space X and denoted by X'. By construction, the topological dual is a subspace of the algebraic dual, $X' \subset X^*$. Restriction of the *transpose operator* A^T to the topological dual is identified as the *topological transpose* and denoted by A',

$$A' = A^T|_{X'}, \qquad A'x' = x' \circ A\,.$$

Note that the topological transpose is well-defined as the composition of two continuous functions is continuous.

Continuous operators. Recall that an operator $A \in L(X, X)$ is *continuous* iff it is bounded, i.e., there exists a constant $C > 0$ such that

$$\|Ax\| \leq C\|x\| \qquad \forall\, x \in X\,,$$

see Theorem 5.1.1. The *adjoint operator* A^* is defined exactly in the same way as in the purely algebraic case,

$$A^* = R_X^{-1} \circ A' \circ R_X$$

where A' is the topological transpose operator and $R_X : X \to X'$ is the *Riesz operator*. Equivalently, we can define the adjoint operator by the identity:

$$(Ax, y) = (x, A^* y) \qquad x, y \in X\,.$$

Finally, operator A is said to be *self-adjoint* if $A^* = A$. Note that we have considered the continuous operator to be defined on the *whole* space X. This is not a coincidence. If $D(A)$ were a proper subspace of X then, by continuity, A could be extended in a unique way to closure $\overline{D(A)}$. The *Orthogonal Decomposition Theorem*, see Theorem 5.1.4, allows then to represent X as an orthogonal sum of $\overline{D(A)}$ and its orthogonal complement,

$$X = \overline{D(A)} \oplus \overline{D(A)}^{\perp}\,,$$

and we can extend (for instance, by zero) the operator to the whole X preserving the continuity constant. Consequently, without loss of generality, one can consider a continuous operator to be defined always on the whole space X.

The trouble with the spectral theory for the continuous operators is that it does not cover, at least directly, differential operators. Differential operators are simply not continuous in the L^2-norm, see Exercise 3.3.3. In order to study differential operators, we need to develop a much more sophisticated concept of *closed operators*.

We return now to the examples studied at the beginning of this section. How do we define derivatives for a L^2-functions ? The answer is: *in the sense of distributions*.

The theory of distributions, developed by a French mathematician - Laurent Schwarz in forties of 20th century, forms the backbone of the modern theory of differential equations. The theory builds heavily on the concepts of Lebesgue measure and the Lebesgue integral.

Test functions and distributions. We begin with the concept of Schwartz test functions. Let $\Omega \subset \mathbb{R}^n$ be an open set. Define,

$$C_0^\infty(\Omega) := \{u \in C^\infty(\Omega) \,:\, \operatorname{supp} u \text{ is compact, and } \operatorname{supp} u \subset \Omega\}\,.$$

By the *support* of u we understand the closure of the set consisting of all points where the function does not vanish,

$$\operatorname{supp} u := \overline{\{x \in \Omega \,:\, u(x) \neq 0\}}\,,$$

Thus, by construction, the support is closed. The definition of test functions requires it also to be *bounded* (recall Heine-Borel Theorem). The important assumption here is that this closed (and bounded) set must be contained in the *open* set Ω. This implies that test functions vanish in a neighborhood of boundary $\partial\Omega$ and, consequently, function u, along with all its derivatives, vanishes on the boundary. We can integrate by parts as many times as we wish, without obtaining any boundary terms.

Space of test functions $C_0^\infty(\Omega)$ is equipped with a sophisticated topology and turned in a *topological vector space*, denoted with a new symbol, $\mathcal{D}(\Omega)$. In prinicple, we symbol $\mathcal{D}(\Omega)$ in place of $C_0^\infty(\Omega)$ when the topology matters. The dual space, i.e., the space of all linear and continuous functionals defined on $\mathcal{D}(\Omega)$, denoted $\mathcal{D}'(\Omega)$, is the space of *Schwartz distributions*. A distribution reveals itself through its action on test functions.

Regular and irregular distributions. A function u defined on Ω, is *locally integrable*, $u \in L^1_{\rm loc}(\Omega)$ if, for every neighborhood $B(x,\epsilon)$ of $x \in \Omega$, $\int_{B(x,\epsilon)} |u|$ exists. Every function $u \in L^1_{\rm loc}(\Omega)$ generates the so-called *regular distribution*,

$$\mathcal{D}(\Omega) \ni \phi \to \langle R_u, \phi\rangle := \int_\Omega u\phi \in \mathbb{R}(\mathbb{C})\,. \tag{3.8}$$

One proves that the regular distributions are well-defined, and that $L^1_{\rm loc}(\Omega)$ is continuously embedded into $\mathcal{D}'(\Omega)$. Any distribution that is not regular, i.e., it is not defined through the n-dimensional integral above, is called an *irregular distribution*. The simplest example of an irregular distribution is the famous Dirac δ,

$$\mathcal{D}(\Omega) \ni \phi \to \langle \delta_{x_0}, \phi\rangle := \phi(x_0) \in \mathbb{R}(\mathbb{C})\,.$$

There are two non-trivial facts about the Dirac's delta: a) it is well-defined, i.e., it is continuous, and b) there is no locally integrable function that would generate it. Any integral involving the Dirac delta is purely formal, i.e., mathematically incorrect. We have tons of examples of irregular distributions. If we replace the n-dimensional integral in (3.8) with an integral over a m-dimensional manifold, $m < n$ (curve or surface integral for $n = 3$, we obtain an irregular distribution.

Distributional derivatives. Motivated with the integration by parts, we define (partial) derivative ∂_i of a distribution u by:

$$\langle \partial_i u, \phi\rangle := -\langle u, \partial_i\phi\rangle \qquad \phi \in \mathcal{D}(\Omega)\,.$$

More generally,

$$\langle \partial^\alpha u, \phi\rangle := (-1)^{|\alpha|}\langle u, \partial^\alpha\phi\rangle \qquad \phi \in \mathcal{D}(\Omega)\,.$$

One can show that the derivative of any distribution is a well-defined distribution as well. Consequently, any distribution has derivatives of any order !

Example 3.3.5

Let $I = (a, b)$. Consider a function u consisting of two branches:

$$u(x) := \begin{cases} u_1(x) & x \in (a, x_0) \\ c & x = x_0 \\ u_2(x) & x \in (x_0, b) \end{cases}$$

where $u_1 \in C^1([a, x_0])$ and $u_2 \in C^1([x_0, b])$ and c is an arbitrary number. Recall the definition of L^p functions and explain why the value at x_0 does not matter. Integration by parts leads to the formula (Exercise 3.3.4):

$$-\int_I u\phi' = \int_I v\phi + [u(x_0)]\phi(x_0) \qquad \phi \in \mathcal{D}(I) \tag{3.9}$$

where v is the pointwise derivative,

$$v(x) = \begin{cases} u_1'(x) & x \in (a, x_0) \\ u_2'(x) & x \in (x_0, b) \end{cases}$$

and $[u(x_0)]$ is the jump of function u at x_0,

$$[u(x_0)] := u_2(x_0) - u_1(x_0)\,.$$

Rewriting formula (3.9) in argument-less form, we have,

$$R_u' = R_v + [u(x_0)]\delta_{x_0}\,,$$

The moral of the story is that the distributional derivative of function u is a function (a regular distribution) iff the Dirac term vanishes, i.e., the function is globally continuous.

□

We are ready now to return to our simplest Example 3.3.1, and explain precisely the definition of the domain of the operator,

$$D(A) := \{u \in L^2(I) \,:\, u' \in L^2(I) \text{ and } u(a) = 0\}$$

where the derivative is understood *in the sense of distributions*. A few comments:

- Any function $u \in L^2(I)$ is in $L^1_{\text{loc}}(I)$ (explain, why ?) and, therefore any L^2 function generates a regular distribution.
- The main message in the definition above is that the distributional derivative of function u *is a function*. If u is given as a union of piece-wise smooth branches, it must be globally continuous. Otherwise, its derivative will include Dirac's deltas that are *not* functions. The second message in the definition that the derivative (which is a function) is assumed to be L^2-integrable. Consequently, A takes $D(A)$ back *into* $L^2(I)$.

- The totality of L^2 functions whose distributional derivative is also an L^2-function is identified as the *Sobolev* [††] *space of first order*,

$$H^1(I) := \{u \in L^2(I) \,:\, u' \in L^2(I)\}\,.$$

- One can show that elements of $H^1(I)$ are continuous functions. Hence the BC at $x = a$ makes sense.

Closed operators. So far, we have managed to make the definition of operator from Example 3.3.1 precise. We also know that the operator is not continuous. How do we develop the theory of L^2-adjoints then ? The answer is the concept of *closed operators*. Consider a Hilbert space X (in practice a (possibly weighted) $L^2(\Omega)$ space), and a general operator A defined on a *subspace* $D(A)$ of space X, identified as *the domain of the operator*,

$$X \supset D(A) \ni u \to Au \in X\,.$$

The operator A is closed iff, by definition, its graph:

$$G(A) := \{(x, y) \in L^2(\Omega) \times L^2(\Omega) \,:\, x \in D(A),\, y = Au\}$$

is a closed subspace of $L^2(\Omega) \times L^2(\Omega)$. Equivalently,

$$G(A) \ni (x_n, y_n) \to (x, y) \quad \Rightarrow \quad x \in D(A) \text{ and } y = Ax\,.$$

We will show now that the derivative is a closed operator. Take $u_n \in D(A)$ with $y_n = Au_n$. Assume $u_n \to u, y_n \to y$ in $L^2(I)$. It follows form the definition of the distributional derivative that, for each test function $\phi \in D(I)$,

$$\int_I \underbrace{iu_n'}_{=y_n} \overline{\phi} = \int_I u_n \overline{i\phi'}\,.$$

Passing to the limit with u_n and y_n, we get,

$$\int_I y\,\overline{\phi} = \int_I u\,\overline{i\phi'}\,.$$

But this implies that $y = iu'$ in the sense of distributions and, therefore, $u \in D(A)$.

Adjoint of a closed operator. The definition of the adjoint for a closed operator is much more involved than for a continuous operator. First of all, in order for a closed operator to possess an adjoint, its domain $D(A)$ *must be dense in* X. This condition is usually easily satisfied. The domain of the adjoint, $D(A^*)$, is defined as a totality (maximality) of elements v in X for which there exists another element $w \in X$ such that

$$(Au, v)_X = (u, w)_X \qquad \forall u \in D(A)\,.$$

[††] Named after a Russian mathematician - Sergei Lvovich Sobolev (1908 – 1989).

Observe that w corresponding to v is unique. Indeed, it is sufficient to argue that

$$(u, w) = 0 \quad \forall\, u \in D(A) \quad \Rightarrow \quad w = 0\,.$$

Bu this is exactly a consequence of the assumption that $D(A)$ is dense in X. Indeed, let $D(A) \ni u_n \to w$ in X. Passing to the limit in

$$\underbrace{(u_n, w)}_{\to (w,w)} = 0\,,$$

we conclude that $\|w\|^2 = 0$ and, therefore, $w = 0$, as required. Function w is identified as the value of the adjoint operator, $A^* v := w$. One can easily show, see Exercise 3.3.6, that the adjoint operator A^* is a closed operator as well. For a more constructive definition, see [5], Section 5.18.

THEOREM 3.3.1 (Sturm-Liouville)

Let $X = L^2(I)$. Operators from Examples 3.3.2, 3.3.3, and 3.3.4 are closed and self-adjoint in the sense of the closed operators theory. Each operator has a series of real eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \ldots \lambda_n \to \infty \quad \text{as} \quad n \to \infty$$

with the corresponding eigenvectors $u_i \in L^2(I)$ that form an orthogonal basis in $L^2(I)$.

Usually, we normalize the eigenvectors in the L^2-norm to obtain an *orthonormal basis*.

REMARK 3.3.1 For operators from Example 3.3.2, the theorem can be easily generalized to a weighted $L^2_w(I)$ space where $w(x) > 0$ is an analytic function in I. The operators have to include then the weight in the definition:

$$Au = \frac{1}{w(x)}[-(a(x)u')' + c(x)u]\,.$$

∎

Example 3.3.6

Consider the simplest Sturm-Liouville operator,

$$Au = -u'', \qquad D(A) = \{u \in L^2(0,l) \,:\, u'' \in L^2(0,l) \text{ and } u(0) = u(l) = 0\}\,.$$

The operator is self-adjoint and positive definite. Indeed,

$$(Au, u) = (-u'', u) = (u', u') \geq 0\,,$$

At the same time, if $\|u'\| = 0$ then $u' = 0$ (in the sense of distributions) which implies that u is constant, $u = c$, see Exercise 3.3.5. The BCs imply that $c = 0$. We know thus upfront that eigevalues

of operator A are real and positive, $\lambda = r^2$, $r \in \mathbb{R}, r \neq 0$. The eigenvalue problem reduces to:

$$u'' + r^2 u = 0 \quad \Rightarrow \quad u(x) = A\cos rx + B\sin rx\,.$$

BC at $x = 0$ implies $A = 0$, and BC at $x = l$ implies

$$B\sin rl = 0 \quad \Rightarrow \quad rl = n\frac{\pi}{2},\ n = 1, 2, \ldots\,.$$

This leads to the sequence of eigenpairs:

$$\lambda_n = \Big(\frac{n\pi}{2l}\Big)^2, \quad e_n(x) = B_n \sin\frac{n\pi x}{2l}, \quad n = 1, 2, \ldots\,.$$

Note that $-n$ gives exatly the same eigenvalue and eigenvector as n, so the sequence is numbered with $n \in \mathbb{N}$ only. With

$$\int_0^l \sin^2\Big(\frac{n\pi x}{2l}\Big)\,dx = \frac{1}{2}\int_0^l \Big(1 - \cos\frac{n\pi x}{l}\Big)\,dx = \frac{l}{2}$$

normalization of e_n leads to $B_n = \sqrt{\frac{2}{l}}$. Sturm-Liouville Theorem implies that functions

$$e_n = \sqrt{\frac{2}{l}}\sin\frac{n\pi x}{2l}, \quad n = 1, 2, \ldots$$

provide an orthonormal basis for $L^2(0, l)$. Expansion

$$u(x) = \frac{2}{l}\sum_{i=1}^{\infty}\left(\int_0^l u(s)\sin\frac{n\pi s}{2l}\,ds\right)\sin\frac{n\pi x}{2l}$$

represents the classical sine series.

□

Example 3.3.7

Consider the Sturm-Liouville operator with periodic BCs,

$$Au = -u'', \qquad D(A) = \{u \in L^2(0, l)\,:\, u'' \in L^2(0, l) \text{ and } u(0) = u(l),\, u'(0) = u'(l)\}\,.$$

The operator is self-adjoint and semi-positive definite. Indeed,

$$(Au, u) = (-u'', u) = (u', u') \geq 0\,.$$

We know thus upfront that eigevalues of operator A are real and non-negative, $\lambda = r^2$, $r \in \mathbb{R}$. The eigenvalue problem reduces to:

$$u'' + r^2 u = 0 \quad \Rightarrow \quad u(x) = Ae^{irx} + Be^{-irx}\,.$$

The periodic BCs are automatically satisfied if rl is a multiple of 2π,

$$rl = n2\pi \quad \Rightarrow \quad r = r_n := \frac{n2\pi}{l}\,.$$

We have thus a series of eigenvalues $\lambda_n = r_n^2$. Except for $n = 0$, each eigenspace is two-dimensional, spanned by $e^{ir_n x}$ and $e^{-ir_n x}$. Normalizing the eigenvectors, we obtain

$$\frac{1}{\sqrt{l}} e^{ir_n x}, \frac{1}{\sqrt{l}} e^{-ir_n x}\,.$$

Expansion

$$u(x) = \frac{1}{l} \sum_{n=-\infty}^{\infty} \left(\int_0^l u(s) \overline{e^{ir_n x}}\, ds \right) e^{ir_n x}$$

represents the classical (complex) Fourier series. Replacing the exponentials with cosines and sines, we obtain the standard (real) Fourier series. □

Example 3.3.8 (Operator with piecewise constant coefficients)

Consider Sturm-Liouville operator:

$$Au = -(au')', \quad D(A) = \{u \in L^2(-\pi, \pi) \,:\, Au \in L^2(-\pi, \pi), \quad u(-\pi) = u(\pi) = 0\}$$

where diffusion coefficient $a(x)$ is piecewise constant,

$$a(x) = \begin{cases} 4 & |x| < \frac{\pi}{2} \\ 1 & \text{otherwise}\,. \end{cases}$$

Integrability conditions on u imply that u must satisfy the following interface conditions at $x = \pm\frac{\pi}{2}$,

$$[u] = [au'] = 0\,.$$

As A is positive definite (explain, why ?), we can assume $\lambda = k^2, k > 0$. For $|x| < \frac{\pi}{2}$, the solution of the differential equation: $4u'' + k^2 u = 0$ is:

$$u = A \cos \frac{k}{2} x + B \sin \frac{k}{2} x\,.$$

It is convenient to consider odd and even eigenvectors separately, see Exercise 3.3.9. Let us worry about the scaling coefficients afterwards and consider first the odd function,

$$u = \sin \frac{k}{2} x\,.$$

It is sufficient to determine $u(x)$ for positive x and take the odd extension to the whole interval afterwards. For $x \in (\frac{\pi}{2}, \pi)$, we have $u'' + k^2 u = 0$, and the solution once again is a combination of sine and cosine functions. Due to the Dirichlet BC at $x = \pi$, it is convenient to introduce a phase and represent the solution in the form:

$$u = A \cos k(x - \pi) + B \sin k(x - \pi)\,.$$

The BC at $x = \pi$ eliminates the first term, and we are left with $u = B \sin k(x - \pi)$. The interface conditions at $x = \frac{\pi}{2}$ lead to two equations:

$$\begin{aligned} [u(\tfrac{\pi}{2})] = 0 \quad &\Rightarrow \qquad \sin \tfrac{k\pi}{4} = B \sin k(-\tfrac{\pi}{2}) = -B \sin \tfrac{k\pi}{2} \\ [(au')(\tfrac{\pi}{2})] = 0 \quad &\Rightarrow \quad 2k \cos \tfrac{k\pi}{4} = Bk \cos \tfrac{k\pi}{2}\,. \end{aligned}$$

Diving side-wise the equations, we obtain:

$$2\cot\frac{k\pi}{4} = -\cot\frac{k\pi}{2}\,.$$

But

$$\cot 2\alpha = \frac{\cos^2\alpha - \sin^2\alpha}{2\sin\alpha\cos\alpha} = \frac{1}{2}[\cot\alpha - \tan\alpha]$$

which leads to the equation:

$$5\cot\frac{k\pi}{4} = \tan\frac{k\pi}{4}\,.$$

Plot functions $\tan x$ and $3\cot x$ to convince yourself that the equation admits an infinite number of solutions $k = k_n > 0$, $n = 1, 2, \ldots$.

Determination of even eigenmodes proceeds along the same lines, see Exercise 3.3.10. Note that all eigenvalues are double eigenvalues as for each λ we have the odd and even eigenvectors, comp. Exercise 3.3.9.

□

Exercises

Exercise 3.3.1 Proceed formally to identify the (formal) adjoint with its domain.

(i) Convection-reaction operator:

$$Au = b(x)u' + c(x)u, \qquad b(x) > 0, c(x) \geq 0, \quad x \in (0,1)$$

with different domains defined as follows.

(a) $D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, u(0) = 0\}$

(b) $D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, u(1) = 0\}$

(c) $D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, u(0) = 0 \text{ and } u(1) = 0\}$

(d) $D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\,$ with no BCs $\}$

(ii) Diffusion-reaction operator:

$$Au = -(a(x)u')' + c(x)u, \qquad a(x) > 0, c(x) \geq 0, \quad x \in (0,1)$$

with periodic BCs:

$$D(A) = \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, u(0) = u(1) \text{ and } a(0)u'(0) = a(1)u'(1)\}\,.$$

(10 points)

Exercise 3.3.2 Show that operator

$$Au = -u''$$
$$D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, 2u(0) - u(1) - 4u'(1) = 0 \text{ and } u(0) - 2u'(1) = 0\}$$

has no eigenvalues. By contrast, show that for operator:

$$Au = -u''$$
$$D(A) := \{u \in L^2(0,1) \,:\, Au \in L^2(0,1),\, u(0) - u(1) = 0 \text{ and } u'(0) + u'(1) = 0\}$$

every complex number λ is an eigenvalue.

(10 points)

Exercise 3.3.3 Let $I = (0,1)$. Consider the simplest differential operator,

$$Au = u'$$
$$D(A) = \{u \in L^2(I) \,:\, Au \in L^2(I)\} = H^1(I)\,.$$

Demonstrate by means of a counterexample that the operator is not bounded in the L^2-norm.

(5 points)

Exercise 3.3.4 Derive formula (3.9).

(5 points)

Exercise 3.3.5 An exercise giving you a taste of distributions. Assume that $u \in \mathcal{D}'(a,b)$ and $u' = 0$, i.e.,

$$\langle u', \phi\rangle := -\langle u, \phi'\rangle = 0 \quad \forall\, \phi \in C_0^\infty(a,b)\,.$$

Prove that u must be a constant, i.e., a regular distribution generated by a constant function,

$$\langle u, \phi\rangle = \int_a^b c\,\phi \quad \forall\, \phi \in C_0^\infty(a,b)\,.$$

Hint: Let $\psi \in C_0^\infty(a,b)$ be a test function with an average value of one. Let $\phi \in C_0^\infty(a,b)$ be an arbitrary test function. Start by proving that function

$$\chi(x) := \int_a^x (\phi(s) - (\int_a^b \phi)\psi(s)\,ds)\,ds\,,$$

is a test function as well.

(10 points)

Exercise 3.3.6 Use the definition of the adjoint of a closed operator to demonstrate that the adjoint operator is closed, too.

(5 points)

Exercise 3.3.7 Consider the Sturm-Liouville operator:

$$Au = -u'', \quad D(A) = \{u \in L^2(0,1) \, : \, u'' \in L^2(0,1), \quad u(0) = 0, \quad u(1) - 2u'(1) = 0\}\,.$$

Show that the eigenfunctions are of the form $\sin\sqrt{\lambda_n}x$, where the eigenvalues λ_n are solutions to the transcendental equation:

$$\tan\sqrt{\lambda} = 2\sqrt{\lambda}\,.$$

Argue that

$$\lambda_n = (2n-1)^2\frac{\pi^2}{4} \quad \text{as } n \to \infty\,.$$

(10 points)

Exercise 3.3.8 Find the eigenpairs of the Sturm-Liouville operator:

$$Au = -u'', \quad D(A) = \{u \in L^2(0, \frac{\pi}{2}) \, : \, u'' \in L^2(0, \frac{\pi}{2}), \quad u'(0) = u(\frac{\pi}{2}) = 0\}.$$

Expand function

$$u(x) = \begin{cases} 0 & x \in (0, \frac{\pi}{4}) \\ 1 & x \in (\frac{\pi}{4}, \frac{\pi}{2}) \end{cases}$$

in the eigenbasis, and plot its spectral approximations u_n obtained with $n = 4, 16, 64$ terms.

(10 points)

Exercise 3.3.9 Consider the Sturm-Liouville operator,

$$Au = -(au')' + cu, \quad D(A) = \{u \in L^2(-l,l) \, : \, Au \in L^2(-l,l),\, u(-l) = u(l) = 0\}$$

where the diffusion and reaction coefficients are even functions, i.e.,

$$a(-x) = a(x), \quad c(-x) = c(x)\,.$$

Prove that if $(\lambda, u(x))$ is an eigenpair for operator A then so is $(\lambda, u(-x))$. Conclude that, If the eigenvector is neither even nor odd, the even and odd parts of function u ($\frac{1}{2}(u(x) + u(-x))$, $\frac{1}{2}(u(x) - u(-x))$) must be eigenvectors corresponding to λ as well. One can search then from the very beginning separately for even and odd eigenvectors which simplifies greatly the algebra. The eigenspace is then at least two-dimensional. Note that, if the original eigenvector is even (odd) to begin with, then the search for the odd (even) eigenvector will simply fail.

(10 points)

Exercise 3.3.10 Finish Example 3.3.8 by finding even eigenvectors.

(10 points)

Exercise 3.3.11 Consider the Sturm- Liouville operator,

$$Au = -(a(x)u')', \quad D(A) = \{u \in L^2(-10,10) \,:\, Au \in L^2(-10,10) \quad u(-10) = u(10) = 0\}\,,$$

where

$$a(x) = \begin{cases} 2 & |x| < 1 \\ 1 & \text{otherwise}\,. \end{cases}$$

Determine the eigenpairs of the operator. Plot a few selected eigenvectors. You may want to use Matlab or Mathematica to solve this problem. *Hint:* The integrability conditions lead to interface conditions at $x = \pm 1$:

$$[u] = [au] = 0\,.$$

(20 points)

3.4 Fourier Transform

L-periodic functions. A measurable function $u : \mathbb{R} \to \mathbb{C}$ is L-periodic if

$$u(x + kL) = u(x) \qquad x \in \mathbb{R},\, k \in \mathbb{Z}\,.$$

Restricting ourselves to functions that are L^2-integrable on interval $(0, L)$, we can equip the space with the inner product:

$$(u, v) = \int_{(0,L)} u(x)\overline{v(x)}\, dx\,,$$

to obtain a Hilbert space $L^2_{per}(\mathbb{R})$. One can show then that Laplace operator is a well-defined, closed and self-adjoint operator from (a dense subspace of) $L^2_{per}(\mathbb{R})$ into itself. The eigenvalues of the Laplace operator are real and non-negative. The eigenvectors are given by

$$\phi_k(x) = L^{-1/2} e^{i2\pi \frac{kx}{L}}$$

with $k \in \mathbb{Z}$, and the corresponding eigenvalues

$$\lambda_k = \left(\frac{2\pi|k|}{L}\right)^2\,.$$

Note that the eigenvectors have been normalized to form an orthonormal system. It follows from the Sturm-Liouville Theorem that eigenvectors ϕ_k form a complete orthonormal system, i.e.,

$$u(x) = \sum_{k\in\mathbb{Z}} (u, \phi_k)_{L^2_{per}(\mathbb{R})} \phi_k(x) = \frac{1}{L} \sum_{k\in\mathbb{Z}} \hat{u}_L(\frac{k}{L}) e^{i2\pi \frac{kx}{L}}$$

where

$$\hat{u}_L(\xi) = \int_{(0,L)} u(x) e^{-i2\pi\xi x}\, dx\,.$$

Let $u \in L^2(\mathbb{R})$ with a compact support in some $(-L/2, L/2)$. Consider its L-periodic extension and the corresponding frequency content illustrated in Fig. 3.1. Elementary calculations show that if we consider the original function with compact support in interval $(-L, L)$ and only then consider its $2L$-periodic extension (i.e. double the value of period L), the corresponding representation will consist of old frequencies (and identical values for them) and new frequencies in between the old ones. If we continue the process, we expect the frequencies to fill the entire real line. This is exactly the intuition behind the definition of the Fourier transform.

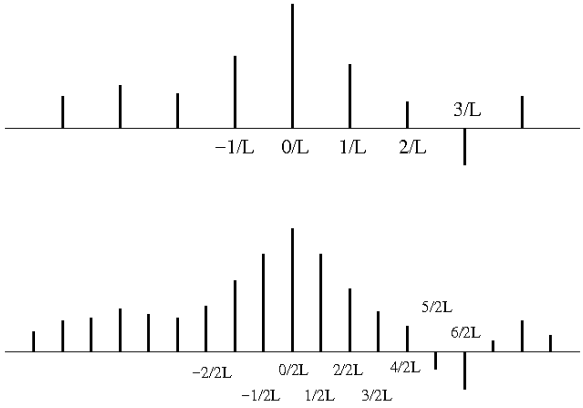

Figure 3.1
Change of frequency spectrum from L to $2L$.

Classical Fourier transform. Let $u \in L^1(\mathbb{R})$. We define its Fourier transform $\mathcal{F}u = \hat{u}$ by

$$(\mathcal{F}u)(\xi) = \hat{u}(\xi) := \int_{\mathbb{R}} u(x) e^{-i2\pi\xi x}\, dx\,.$$

The formal (at this point) L^2-adjoint of the Fourier transform is equal to:

$$(\mathcal{F}^* u)(\xi) := \int_{\mathbb{R}} u(x) e^{i2\pi\xi x}\, dx\,.$$

We expect operator $\mathcal{F}$ to be invertible with the inverse equal to its adjoint, $\mathcal{F}^{-1} = \mathcal{F}^*$. Hölder inequality implies that the Fourier transform is well-defined. We have the following classical result.

THEOREM 3.4.1

Let $u, \hat{u} \in L^1(\mathbb{R}^n)$ and u is continuous at a point x. Then

$$u(x) = (\mathcal{F}^* \hat{u})(x)\,.$$

PROOF See [1], Section 2.5, if you are interested. ∎

REMARK 3.4.1 Variable ξ, the argument of the Fourier transform, is identified as the *frequency*. The *angular frequency* is defined as

$$\omega = 2\pi\xi\,.$$

Most of applied math texts will define the Fourier transform as a function of the angular frequency rather than the frequency,

$$(\mathcal{F}_{x\to\omega}u) = \hat{u}(\omega) := \int_{\mathbb{R}} u(x)e^{-i\omega x}\,dx\,.$$

The inverse Fourier transform does no longer coincide with the adjoint as it must then be scaled with the 2π factor,

$$(\mathcal{F}^{-1}_{\omega\to x}\hat{u})(x) := \frac{1}{2\pi}(\mathcal{F}^{*}_{\omega\to x}\hat{u})(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{u}(\omega)e^{i\omega x}\,d\omega\,.$$

Frequently, in order to staisfy the Plancheral identity discussed next, we distribute the $1/2\pi$ factor between the two transforms evenly defining:

$$(\mathcal{F}_{x\to\omega}f)(\omega) := \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)e^{-i\omega x}\,dx \qquad (\mathcal{F}^{-1}_{\omega\to x}\hat{f})(x) := \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega x}\,d\omega\,.$$

Confusing... ∎

Extension to L^2-functions. The classical Fourier transform, defined for $L^1(\mathbb{R})$ functions, can be extended to space $L^2(\mathbb{R})$. Let $u, v \in C_0^\infty(\mathbb{R})$. Theorem 3.4.1 implies that

$$(\mathcal{F}u, \mathcal{F}v)_{L^2(\mathbb{R})} = (\mathcal{F}^*\mathcal{F}u, \mathcal{F}v)_{L^2(\mathbb{R})} = (u, v)_{L^2(\mathbb{R})}\,.$$

Consequently, the Fourier transform is an L^2-isometry from $C_0^\infty(\mathbb{R}) \cap L^2(\mathbb{R})$ into itself. As $C_0^\infty(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, the transform can be extended in a unique way to a map defined on the whole L^2-space preserving the norm, see Exercise 3.4.2.

THEOREM 3.4.2 (Plancherel)

Fourier transform is an isometry from $L^2(\mathbb{R}^n)$ into itself, i.e.,

$$(\mathcal{F}u, \mathcal{F}v)_{L^2} = (u, v)_{L^2} \qquad u, v \in L^2(\mathbb{R}^n)\,.$$

Additionally, $\mathcal{F}^*$ represents the L^2-adjoint of $\mathcal{F}$ and $\mathcal{F}^{-1} = \mathcal{F}^*$ which proves that $\mathcal{F}$ is a surjection.

For a more explicit construction of Fourier transform for L^2-functions, see Exercise 3.4.3.

Action of Fourier transform on derivatives. Elementary calculations show that

$$\begin{aligned} &\mathcal{F}_{x\to\xi}\left(\partial^\alpha \phi(x)\right)(\xi) = (i2\pi\xi)^\alpha \hat{\phi}(\xi) \quad \text{or} \\ &\mathcal{F}_{x\to\omega}\left(\partial^\alpha \phi(x)\right)(\omega) = (i\omega)^\alpha \hat{\phi}(\omega)\,. \end{aligned} \tag{3.10}$$

This in turn implies that

$$\mathcal{F}_{x\to\xi}\left((-i2\pi x)^\alpha \phi(x)\right)(\xi) = \partial^\alpha \hat{\phi}(\xi)\,. \tag{3.11}$$

The properties above make Fourier transform an ideal tool for solving ODEs or PDEs with constant coefficients.

Fourier transform of distributions. Without going into details, we can extend the definition of the Fourier transform to a class of distributions called *tempered* distributions. The identity for functions u, v:

$$\int_{\mathbb{R}} \mathcal{F}u\, v = \int_{\mathbb{R}} u\, \mathcal{F}u$$

suggests defining the Fourier transform of distribution $u \in \mathcal{D}'(\mathbb{R})$ by:

$$\langle \mathcal{F}u, \phi\rangle := \langle u, \mathcal{F}\phi\rangle\,.$$

In particular for Dirac's delta δ, we obtain:

$$\langle \mathcal{F}\delta, \phi\rangle := \langle \delta, \int_{\mathbb{R}} \phi(x) e^{-i\omega x}\, dx\rangle = (\int_{\mathbb{R}} \phi(x) e^{-i\cdot x}\, dx)(0) = \int_{\mathbb{R}} \phi(x)\, dx = \int_{\mathbb{R}} 1\phi(x)\, dx = \langle R_1, \phi\rangle\,.$$

Fourier transform of Dirac's delta is the regular distribution generated by the unity function. Informally, $\mathcal{F}\delta = 1$.

Example 3.4.1 (Wave equation)

Consider the wave equation:

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = 0\,,$$

where $u = u(x,t)$, $x \in \Omega$, $t \in \mathbb{R}$, accompanied with some BC on $\partial\Omega$. Fourier transforming in time we obtain the Helmholtz equation.

$$-\omega^2 \hat{u} - \Delta \hat{u} = 0\,.$$

The same result can be obtained assuming the ansatz: $u(x,t) = e^{i\omega t}\hat{u}(x)$. Function $\hat{u}(x)$ is called the *phasor*. Note that $u(x,t)$ is real-valued but its Fourier transform, the phasor, is always a complex-valued function. By solving the wave equation in the *frequency domain*, we commit ourselves automatically to complex-valued functions. Once we solve the problem in the frequency domain, we use the inverse Fourier transform to obtain the solution in the *time domain*,

$$u(x,t) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{u}(x,\omega) e^{ix\omega}\, d\omega\,.$$

Note the consistency of sign in the exponential defining the inverse Fourier transform and the exponential in the ansatz. Many authors use the $u(x,t) = e^{-i\omega t}\hat{u}(x)$ which gives the same Helmholtz problem[‡‡]. The analogy with the Fourier transform is then lost.

□

Example 3.4.2 (Infinite beam on an elastic foundation)

The simplest model for a rail (or a long foundation) would be the Euler-Bernoulli beam on an elastic (Winkler) foundation:

$$EIu'''' + ku = f\,.$$

Here E is the Young modulus, and I is the moment of inertia; product EI is identified as the *stiffness* of the beam. Coefficient $k > 0$ represents the reaction of the foundation (elastic springs), and $f = f(x)$ is the *load*. The equation is defined on the whole real line. Fourier transforming in x, we obtain,

$$EI\omega^4\hat{u}(\omega) + k\hat{u}(\omega) = \hat{f}(\omega)\,,$$

which yields the solution in the Fourier domain,

$$\hat{u}(\omega) = \frac{\hat{f}(\omega)}{EI\omega^4 + k}\,.$$

Variable ω does no longer have the interpretation of an angular frequency but nevertheless we use the same symbol. For a particular case of $f(x) = \delta_0$, we obtain,

$$\hat{u}(\omega) = \frac{1}{EI\omega^4 + k} = \frac{1}{EI}\frac{1}{\omega^4 + \alpha^4}$$

where $\alpha^4 = k/EI$. To obtain the final solution, we have to compute the inverse Fourier transform,

$$u(x) = \frac{1}{2\pi EI}\int_{\mathbb{R}} \frac{e^{i\omega x}}{\omega^4 + \alpha^4}\,d\omega\,.$$

The Residue Theorem comes in handy. First, we represent the denominator as:

$$\omega^4 + \alpha^4 = (\omega - \omega_1)(\omega - \omega_2)(\omega - \omega_3)(\omega - \omega_4) \quad \omega_k = e^{i(2k-1)\frac{\pi}{4}},\ k = 1,2,3,4\,,$$

to learn that the function has four simple poles, two above and two below the x axis. If we employ the contour from Fig. 2.9, we need to compute the residues at ω_1 and ω_2. We have,

$$(\omega_1 - \omega_2)(\omega_1 - \omega_3)(\omega_1 - \omega_4) = \sqrt{2}(\sqrt{2} + i\sqrt{2})i\sqrt{2}\alpha^3 = 2\sqrt{2}(i-1)\alpha^3$$
$$(\omega_2 - \omega_1)(\omega_2 - \omega_3)(\omega_2 - \omega_4) = -\sqrt{2}i\sqrt{2}(-\sqrt{2} + i\sqrt{2})\alpha^3 = 2\sqrt{2}(i+1)\alpha^3$$

and,

$$e^{i\omega_1 x} = e^{i\alpha(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2})x} = e^{-\alpha\frac{\sqrt{2}}{2}x}e^{i\alpha\frac{\sqrt{2}}{2}x}$$
$$e^{i\omega_2 x} = e^{i\alpha(-\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2})x} = e^{-\alpha\frac{\sqrt{2}}{2}x}e^{-i\alpha\frac{\sqrt{2}}{2}x}\,.$$

[‡‡] But different sign in the impedance BC and definition of incoming/outgoing waves.

We have now,

$$\begin{aligned}
2\pi i(\mathrm{Res}_{z_1} + \mathrm{Res}_{z_2}) &= \frac{i}{EI}\left[\frac{e^{i\omega_1 x}}{(\omega_1-\omega_2)(\omega_1-\omega_3)(\omega_1-\omega_4)} + \frac{e^{i\omega_2 x}}{(\omega_2-\omega_1)(\omega_2-\omega_3)(\omega_2-\omega_4)}\right] \\
&= \frac{i}{EI}\frac{1}{2\sqrt{2}\alpha^3} e^{-\alpha\frac{\sqrt{2}}{2}x}[\underbrace{\frac{1}{i-1}}_{=-\frac{i+1}{2}} e^{i\alpha\frac{\sqrt{2}}{2}x} + \underbrace{\frac{1}{i+1}}_{=-\frac{i-1}{2}} e^{-i\alpha\frac{\sqrt{2}}{2}x}] \\
&= \frac{1}{2\sqrt{2}\alpha^3 EI} e^{-\alpha\frac{\sqrt{2}}{2}x}\left[\cos(\alpha\frac{\sqrt{2}}{2}x) + \sin(\alpha\frac{\sqrt{2}}{2}x)\right]
\end{aligned}$$

Note that the integral over the semicircle c_R vanishes in the limit (explain, why?). As expected, the solution decays exponentially away from the load point and oscilates. If we position the unit load at an arbitrary point ξ, the solution is obtained by changing the coordinate from x to $x-\xi$, and interpreted as the *Green function* for the problem,

$$G(x,\xi) = G(x-\xi) = \frac{1}{2\sqrt{2}\alpha^3 EI} e^{-\alpha\frac{\sqrt{2}}{2}(x-\xi)}\left[\cos(\alpha\frac{\sqrt{2}}{2}(x-\xi)) + \sin(\alpha\frac{\sqrt{2}}{2}(x-\xi))\right].$$

By superposition, the solution corresponding to a general load $f(x)$ is then:

$$u(x) = \int_{\mathbb{R}} G(x,\xi) f(\xi)\, d\xi\,.$$

Alternatively, we can Fourier transform $f(x)$ and apply the inverse Fourier transform via the Residue Theorem. □

Fourier convolution. Let u, v denote complex-valued functions defined on the whole $\mathbb{R}$. The *convolution* of functions u and v is defined as:

$$(u * v)(x) := \int_{\mathbb{R}} u(x-y)v(y)\, dy\,.$$

We are implicitly assuming that product $u(x-\cdot)v(\cdot) \in L^1(\mathbb{R})$. A simple change of variables shows that convolution is symmetric,

$$\begin{aligned}
(u * v)(x) &= \int_{\mathbb{R}} u(x-y)v(y)\, dy \\
&= \int_{\mathbb{R}} u(z)v(x-z)\, dz \qquad (z = x-y) \\
&= (v * u)(x)\,.
\end{aligned}$$

More precisely, if either the left- or right-hand side is well defined, the so is the other side, and they are equal. The definition can be extended to three functions,

$$u * v * w := (u * v) * w\,.$$

As the operation is associative (Exercise 3.4.5), i.e.,

$$(u * v) * w = u * (v * w)$$

we are justified to use the notation without any parentheses indicating the order of computing the convolutions. By induction, the notion extends to any finite number of functions.

The following theorem is a sample result formulating sufficient conditions for the convolution to be well defined, and its continuity properties.

THEOREM 3.4.3

Let

$$\frac{1}{p}+\frac{1}{q}=1+\frac{1}{r}, \qquad p,q,r\in[1,\infty]\,,$$

and $u\in L^p(\mathbb{R})$, $v\in L^q(\mathbb{R})$. *Then* $u*v$ *exists a.e. and*

$$\|u*v\|_{L^r}\le\|u\|_{L^p}\,\|v\|_{L^q}\,.$$

PROOF See [1]. ∎

Fourier transform of convolutions. Let $u,v\in L^1(\mathbb{R})$. We have:

$$\begin{aligned}
\mathcal{F}_{x\to\xi}((u*v)(x))(\xi) &= \int_{\mathbb{R}} e^{-i2\pi\xi x}\int_{\mathbb{R}^n} u(x-y)v(y)\,dy\,dx\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-i2\pi\xi x}u(x-y)\,dx\,v(y)\,dy \qquad \text{(Fubini)}\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-i2\pi\xi(z+y)}u(z)\,dz\,v(y)\,dy \qquad (x-y=z)\\
&= \int_{\mathbb{R}} e^{-i2\pi\xi y}\hat{u}(\xi)v(y)\,dy\\
&= \hat{u}(\xi)\hat{v}(\xi)\,,
\end{aligned}$$

i.e., Fourier transform sets convolutions into products.

Exercises

Exercise 3.4.1 Compute the Fourier transform for the following functions.

$$\text{(a) } f(x)=e^{-\alpha|x|} \qquad \text{(b) } f(x)=\begin{cases}0 & x<0\\ e^{-\alpha x} & x>0\end{cases}$$

where $\alpha>0$.

(10 points)

Exercise 3.4.2 Let X,Y be Banach spaces, and $\mathcal{X}$ a dense subspace of X. Let A be a bounded linear operator defined on the $\mathcal{X}$ into Y,

$$\|Ax\|_Y\le C\|x\|_X \qquad x\in\mathcal{X}\,.$$

Prove that operator A admits a *unique* extension to the whole space X preserving the bound above. *Hint:* Let $x \in X$. Consider a sequence $x_n \in \mathcal{X}$ converging to x. Argue that Ax_n is Cauchy in Y and use completeness of Y to conclude the existence of a limit y. Define $Ax := y$, and prove that A is linear and preserves the bound.

(10 points)

Exercise 3.4.3 If you do not like the extension argument used in the text to define the Fourier transform for L^2-functions, here is another way to get there. Let $u \in L^2(\mathbb{R}^n)$. Take $N > 0$ and define

$$u_N(x) := \begin{cases} u(x) & \text{for } |x| < N \\ 0 & \text{otherwise} \end{cases}$$

Explain why $u_N \in L^1(\mathbb{R}^n)$, and define:

$$\underbrace{\mathcal{F}u}_{\text{new}} := \lim_{N\to\infty} \underbrace{\mathcal{F}u_N}_{\text{classical}}$$

where the limit is understood in the L^2 sense. Prove that the limit exists and show that the new definition delivers the same result as the two definitions discussed in the text. (10 points)

Exercise 3.4.4 Prove identities (3.10) and (3.11).

(5 points)

Exercise 3.4.5 Associativity of convolutions. Prove that

$$(u * v) * w = u * (v * w)$$

for $u, v, w \in L^1(\mathbb{R})$. (5 points)

Exercise 3.4.6 Take function

$$f(x) = \begin{cases} 1 & |x| < 1 \\ 0 & |x| > 1 \end{cases}$$

and compute its Fourier transform $\hat{f}(\omega)$. Compute then the inverse Fourier transform of $\hat{f}(\omega)$ and verify whether, indeed, it coincides with the original function.

(10 points)

3.5 Laplace Transform

One-sided Fourier Transform. In the case of an initial boundary-value problem,

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - \Delta u = 0 & x \in \Omega, \quad t > 0 \\ u(x,t) = 0 & x \in \partial\Omega, \ t > 0 \\ u(x,0) = u_0(x) & x \in \Omega \\ \dfrac{\partial u}{\partial t}(x,0) = v_0 & x \in \Omega\,. \end{cases} \tag{3.12}$$

it is more convenient to use a *one-sided Fourier transform* defined as (we use the same symbol):

$$\hat{u}(\omega) := \int_0^\infty u(t) e^{-i\omega t}\, dt\,.$$

It is easy to see that the one-sided Fourier transform of $u(t), t > 0$ coincides with the regular Fourier transform applied to the zero extension of $u(t)$ to the whole real line,

$$\text{(new)}\ \hat{u}(\omega) = \text{(old)}\ \hat{U}(\omega) \quad \text{where} \quad U(t) := \begin{cases} 0 & t < 0 \\ u(t) & t > 0 \end{cases}\,.$$

The one-sided Fourier derivative behaves slightly different when applied to derivative $u'(t)$,

$$\begin{aligned} \widehat{u'}(\omega) &= \int_0^\infty u'(x) e^{-i\omega x}\, dx \\ &= i\omega \int_0^\infty u(x) e^{-i\omega x}\, dx + [u(x) e^{-i\omega x}]|_0^\infty \\ &= i\omega\, \hat{u}(\omega) - u(0_+)\,. \end{aligned} \tag{3.13}$$

Note that $u \in L^1(0,\infty)$ implies that u vanishes at ∞. Alternatively, we can obtain the formula by applying the standard Fourier transform to the distributional derivative of the extension $U(t)$, comp. Exercise 3.5.1.

Similarly,

$$\widehat{u''}(\omega) = -\omega^2 \hat{u}(\omega) - u'(0_+) - i\omega u(0_+)\,.$$

Applying the one-sided Fourier transform (with respect to time) to the IBVP $(3.12)_1$, and taking into the account the initial conditions, we obtain the following problem in the frequency domain:

$$-\omega^2 \hat{u} - \Delta \hat{u} = v_0 + i\omega u_0\,.$$

The IC data enter now as a load in the Helmholtz problem.

Let us try now to solve the simple IVBV for the vibrating string from Example 6.1.7. Applying the one-sided Fourier transform to wave equation $(3.12)_1$, we obtain the following BVP for the Fourier transformed solution:

$$-\frac{d^2 \hat{u}}{dx^2} - \omega^2 \hat{u} = i\omega \sin \frac{\pi x}{l}\,,$$

The general solution to this equation is:

$$\hat{u}(x) = A\cos\omega x + B\sin\omega x + \frac{i\omega l^2}{\pi^2 - (\omega l)^2}\sin\frac{\pi x}{l}\,.$$

$\hat{u}(0) = 0$ implies $A = 0$ but $\hat{u}(l) = 0$ leads to the equation:

$$B\sin\omega l = 0\,,$$

If $\omega l = n\pi$, coefficient B is undetermined. Something went wrong...

What went wrong was the assumption that we are dealing with L^1 (in time) solution. We are not. A spring without any damping mechanism will vibrate forever with non-diminishing amplitude. Clearly the solution is not summable in time. Fortunately, we have a remedy for the problem.

Laplace transform. We relax our assumption on the function being transformed and request that it is at most of the exponential growth:

$$|f(t)| \le Ce^{\alpha t} \quad t > 0\,.$$

In particular, the condition is satisfied by functions that remain bounded (choose $\alpha = 0$) but not in $L^1(0,\infty)$, functions that may grow linearly in time (resonans), or may even grow exponentially in time but with the limited exponent α. Before we apply the (one-sided) Fourier transform, we simply multiply the function with a negative exponential $e^{-\gamma t}$, with $\gamma > \alpha$,

$$\int_0^\infty e^{-\gamma t} f(t) e^{-i\omega t}\,dt = \int_0^\infty f(t) e^{-(\gamma + i\omega)t}\,dt\,.$$

The formula is begging for introducing a complex argument $s = \gamma + i\omega$, and defining the new transform as a function of complex rather than real variable,

$$(\mathcal{L}f)(s) = \bar{f}(s) := \int_0^\infty f(t)e^{-st}\,dt\,.$$

The inversion formula for the Fourier transform,

$$f(t)e^{-\gamma t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f(t)e^{-\gamma t}}\, e^{i\omega t}\,d\omega\,,$$

and the definition of complex integral, imply the inversion formula for the Laplace transform:

$$f(t) = \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} \bar{f}(s)e^{st}\,ds\,.$$

Note that the integration path is not unique, see Exercise 3.5.2. In numerical approximations of the inverse Laplace transform, finding an optimal path of the integration is one of the main tasks.

LEMMA 3.5.1

The following properties hold:

$$\mathcal{L}(f')(s) = s\bar{f}(s) - f(0)$$
$$\mathcal{L}(f'')(s) = s^2\bar{f}(s) - sf(0) - f'(0)$$

PROOF See Exercise 3.5.3. ∎

We return now to the vibrating string example. Laplace transforming the equation, and building the initial conditions in, we obtain the following problem in the Laplace domain.

$$\begin{cases} s^2\bar{u} - \bar{u}'' = su_0 = s\sin\frac{\pi x}{l} & x \in (0,l) \\ \bar{u}(0) = \bar{u}(l) = 0 \end{cases}$$

The general solution of the ODE (in x, parametrized with s) is:

$$\bar{u} = A\cosh sx + B\sinh sx + \frac{s}{s^2 + (\frac{\pi}{l})^2}\sin\frac{\pi x}{l}\,.$$

BC: $\bar{u}(0) = 0$ implies $A = 0$, and BC: $\bar{u}(l) = 0$ implies $B = 0$. Note that with $\Re s > 0$, $\sinh sl \neq 0$. We need to compute the inverse Laplace transform:

$$\mathcal{L}^{-1}\left(\frac{s}{s^2 + (\frac{\pi}{l})^2}\right) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{se^{st}}{s^2 + (\frac{\pi}{l})^2}\,ds\,.$$

We will use the contour shown in Fig. 3.2. It consists of four parts. The integral over the vertical path converges to the inverse Laplace transform,

$$\lim_{R\to\infty}\frac{1}{2\pi i}\int_{\gamma-iR}^{\gamma+iR} \frac{se^{st}}{s^2 + (\frac{\pi}{l})^2}\,ds = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{se^{st}}{s^2 + (\frac{\pi}{l})^2}\,ds\,.$$

Parametrizing c_1,

$$s = x + iR, \quad x \in (0,\gamma)\,,$$

for sufficiently large R, we can estimate the integral as follows:

$$|\frac{1}{2\pi i}\int_{c_1} \frac{se^{st}}{s^2 + (\frac{\pi}{l})^2}\,ds| \leq c\int_0^{\gamma} \frac{Re^{\gamma t}}{R^2}\,dx \to 0 \quad \text{as } R\to\infty \qquad c > 0\,.$$

In the similar way, we show that the integral over c_2 vanishes in the limit as well. Estimating the integral over the semicircular arch is more delicate. We start with the parametrization:

$$s = -R\sin\theta + iR\cos\theta \quad \theta \in (0,\pi)\,,$$

and (for sufficiently large R) estimate as follows:

$$|\frac{1}{2\pi i}\int_{c_1} \frac{se^{st}}{s^2 + (\frac{\pi}{l})^2}\,ds| \leq c\int_0^{\pi} \frac{Re^{-R\sin\theta}}{R^2}\,Rd\theta\,.$$

The simple argument with powers of R does not work anymore but we can apply the Lebesgue Theorem. Indeed, the integrand converges pointwise to zero, and we can choose simply a constant for a dominating function. Decomposing the denominator,

$$s^2 + (\frac{\pi}{l})^2 = (s - s_1)(s - s_2) \quad s_1 = \frac{\pi}{l}i,\ s_2 = -\frac{\pi}{l}i\,,$$

we see that we have two simple poles. It remains to compute the residues at s_1 and s_2.

$$\mathrm{Res}_{s_1} = \lim_{s\to s_1}(s-s_1)\frac{se^{st}}{(s-s_1)(s-s_2)} = \lim_{s\to s_1}\frac{se^{st}}{s-s_2} = \frac{s_1e^{s_1t}}{s_1-s_2} = e^{s_1t}\,.$$

Similarly,

$$\mathrm{Res}_{s_2} = \lim_{s\to s_2}(s-s_2)\frac{se^{st}}{(s-s_1)(s-s_2)} = \lim_{s\to s_2}\frac{se^{st}}{s-s_1} = \frac{s_2e^{s_1t}}{s_2-s_1} = e^{s_2t}\,.$$

Finally,

$$\frac{1}{2\pi i}\int_c \frac{se^{st}}{s^2+(\frac{\pi}{l})^2}\,ds = \mathrm{Res}_{s_1}+\mathrm{Res}_{s_2} = \frac{1}{2}e^{i\frac{\pi}{l}t}+\frac{1}{2}e^{-i\frac{\pi}{l}t} = \cos\frac{\pi}{l}t\,.$$

We have recovered the solution obtained with the separation of variables,

$$u = \cos\frac{\pi}{l}t\sin\frac{\pi}{l}x\,.$$

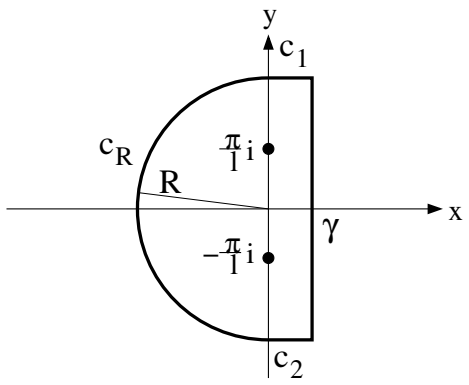

Figure 3.2
Closed contour to compute the inverse Laplace transform.

Example 3.5.1

We will solve now the parabolic problem studied in Example 6.1.9 using the Laplace transform in time. The problem reads as follows.

$$\begin{cases} \dfrac{\partial u}{\partial t} - \alpha^2\dfrac{\partial^2 u}{\partial x^2} = 0 & x\in(0,l),\ t>0 \\ u(0,t) = 1 & t>0 \\ u(l,t) = 0 & t>0 \\ u(x,0) = 0 & x\in(0,l)\,. \end{cases}$$

Note that the constant BC at $x = 0$ prevents the use of the (one-sided) Fourier transform. Laplace transforming the heat equation, we get,

$$s\bar{u} - \alpha^2 \bar{u}_{,xx} = 0\,.$$

Laplace transforming the BCs:

$$\bar{u}(0) = \int_0^\infty 1e^{-st}\,dt = -\frac{1}{s}e^{-st}|_0^\infty = \frac{1}{s} \quad \Re s > 0$$
$$\bar{u}(l) = 0\,.$$

Solving the ODE, we get:

$$\bar{u}(x) = C\cosh(\frac{\sqrt{s}}{\alpha}(x-l)) + D\sinh(\frac{\sqrt{s}}{\alpha}(x-l))\,.$$

BC: $\bar{u}(l) = 0$ implies $C = 0$. BC: $\bar{u}(0) = \frac{1}{s}$ implies:

$$\bar{u}(x) = -\frac{1}{s\sinh(\frac{\sqrt{s}}{\alpha}l)}\sinh(\frac{\sqrt{s}}{\alpha}(x-l))\,.$$

Please take time to double check that the solution satisfies the ODE and the BCs. We compute now the inverse Laplace transform:

$$u(x,t) = -\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+\infty} \frac{\sinh(\frac{\sqrt{s}}{\alpha}(x-l))e^{st}}{s\sinh(\frac{\sqrt{s}}{\alpha}l)}\,ds\,.$$

Function $\sqrt{s}$ is double-valued and we need to select a branch cut. However, the integrand is an ***even*** function of $\sqrt{s}$ and, consequently, its value is ***independent*** of the selection of the branch cut as it is simply a continuous function of s except for simple poles at $s_0 = 0$ and $s_n = -\frac{n^2\pi^2\alpha^2}{l^2}$, $n = 1, 2, \ldots$. We can work thus with the standard Laplace contour shown in Fig. 3.3. We converge with $R \to \infty$

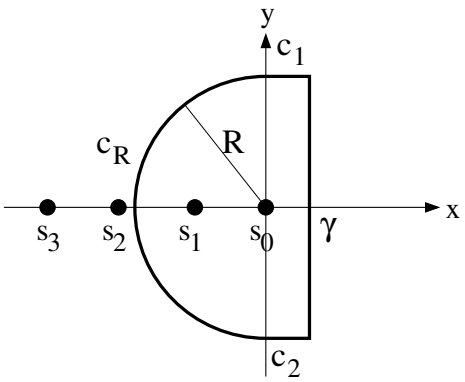

Figure 3.3
Closed contour to compute the inverse Laplace transform in Example 3.5.1.

through a sequence of discrete values R_n selected in such a way that the contour falls in between

s_{n+1} and s_n. Thus, for $R = R_n$, the contour contains poles $s_0, \ldots, s_n$, and the integral equals to the sum of the residues at those poles. Converging with $n \to \infty$, we obtain an ***inifinite series*** of residues at $s_0, s_1, \ldots$. We leave to the reader to show that the integrals over parts c_1, c_2, c_R vanish in the limit, and the remaining part of the integral delivers the inverse Laplace transform, see Exercise 3.5.6. It remains to compute the residues. Computation of the residue at $s_0 = 0$ leads to the indefinite symbol $0/0$, and we have to resort to the d'Hospital rule to obtain:

$$\mathrm{Res}_{s_0} = \lim_{s\to 0} \frac{\sinh(\frac{\sqrt{s}}{\alpha}(x-l))}{\sinh(\frac{\sqrt{s}}{\alpha}l)} \lim_{s\to 0} e^{st} \stackrel{H}{=} \lim_{s\to 0} \frac{\frac{x-l}{\alpha}\frac{1}{2\sqrt{s}}\cosh(\frac{\sqrt{s}}{\alpha}(x-l))}{\frac{l}{\alpha}\frac{1}{2\sqrt{s}}\cosh(\frac{\sqrt{s}}{\alpha}l)} = \frac{x-l}{l}\,.$$

We start the computation of the residue at s_n by computing $\sqrt{s_n}$,

$$\sqrt{s_n} = i\frac{n\pi\alpha}{l} \quad \Rightarrow \quad \frac{\sqrt{s_n}}{\alpha}l = in\pi\,.$$

Note that we have used a particular branch of $\sqrt{s}$ to compute the square roots. It is critical that the same branch is used in all the computations. We have now,

$$\mathrm{Res}_{s_0} = \lim_{s\to s_n} \frac{s-s_n}{\sinh(\frac{\sqrt{s}}{\alpha}l)} \lim_{s\to s_n} \sinh(\frac{\sqrt{s}}{\alpha}(x-l)) \lim_{s\to s_n} \frac{e^{st}}{s}\,.$$

The first limit is computed using again the d'Hospital rule:

$$\lim_{s\to s_n} \frac{s-s_n}{\sinh(\frac{\sqrt{s}}{\alpha}l)} \stackrel{H}{=} \lim_{s\to s_n} \frac{1}{\frac{l}{2\alpha\sqrt{s}}\cosh(\frac{\sqrt{s}}{\alpha}l)} = \frac{2\alpha\sqrt{s_n}}{l\cosh(\frac{\sqrt{s_n}}{\alpha}l)} = (-1)^n\frac{2\alpha^2 in\pi}{l^2}$$

since

$$\cosh(\frac{\sqrt{s_n}}{\alpha}l) = \cosh(in\pi) = \cos(n\pi) = (-1)^n\,.$$

The second limit is:

$$\lim_{s\to s_n} \sinh(\frac{\sqrt{s}}{\alpha}(x-l)) = \sinh(i\frac{n\pi}{l}(x-l)) = i\sin(\frac{n\pi}{l}(x-l))\,.$$

and the third one is simply $e^{s_n t}/s_n$. Our final result reads:

$$u(x,t) = (1-\frac{x}{l}) + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\,\sin(\frac{n\pi}{l}(x-l))\,e^{-\frac{n^2\pi^2\alpha^2}{l^2}t}\,.$$

□

Exercises

Exercise 3.5.1 Prove formula (3.13) using the zero extension $U(t)$ and standard Fourier transform. You will need to use the definitions of Fourier transform of a distribution, and the distributional derivative.

(10 points)

Exercise 3.5.2 Possible paths of integration in the definition of the inverse Laplace transform. Fig. 3.4 shows four different paths of integration: c_1 is a vertical line passing through γ_0, c_2 replaces a middle part of the path with a semicircle, c_3 with an alternative polygonal path, and c_4 is simply another vertical line passing through $\gamma > \gamma_0$. Assume that function $f(z)$ is complex-differentiable for $\Re z > \gamma_0$ and explain why the paths c_2 and c_3 will yield the same value of the integral as path c_1. Under what additional decay assumption of $|f(z)|$ for $|\Im z| \to \infty$, path c_4 will yield the same result as path c_1 ?

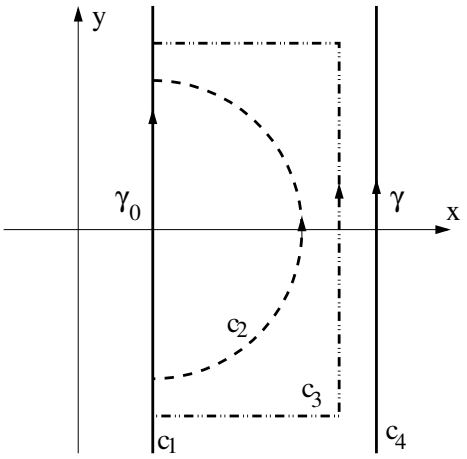

Figure 3.4
Possible paths of integration in the inverse Laplace transform.

(10 points)

Exercise 3.5.3 Prove Lemma 3.5.1

(10 points)

Exercise 3.5.4 Find the inverse Laplace transform of the following functions.

$$\text{(a) } \frac{1}{s^2+a^2} \qquad \text{(b) } \frac{s}{s^2+a^2}$$

$$\text{(c) } \frac{1}{s^3} \qquad \text{(d) } \frac{e^{-as}}{s}$$

where $a > 0$.

(20 points)

Exercise 3.5.5 Consider the following initial-value problem.

$$\begin{cases} \ddot{x} + \dot{x} = -\delta(t-2) \\ x(0) = 1,\ \dot{x}(0) = 0 \end{cases}$$

where δ denotes the Dirac's delta "function".

1. Define precisely delta functional and reinterpret its action in terms of appropriate jump conditions for $\dot{x}$ and x at $t = 2$.
2. Solve the problem using elementary means.
3. Define the Laplace transform. Apply it to both sides of the equation and find the solution in the Laplace domain.
4. Use the Residue Theorem to compute the inverse Laplace transform of the solution in the Laplace domain and compare it with the solution obtained using the elementary calculus.

(20 points)

Exercise 3.5.6 Complete Example 3.5.1 by showing that the contour integral over parts c_1, c_2, c_R vanishes in the limit $R = R_n \to \infty$.

(20 points)

3.6 Spectral Theorem for Unbounded Self-Adjoint Operators

In this section, we will discuss a generalization of Theorem 3.3.1 to a larger class of self-adjoint (bounded and) unbounded operators. Let X be a Hilbert space. Consider a general unbounded operator,

$$A : X \supset D(A) \ni u \to Au \in X\,.$$

Recall that we always assume that $D(A)$ is dense in X, otherwise we cannot define the adjoint A^*. In practice, $X = L^2_w(\Omega)$ where $\Omega \subset \mathbb{R}^n$ is an arbitrary domain, and $w(x) > 0$ denotes a smooth *weight* on Ω.

Consider $\lambda \in \mathbb{C}$ and shifted operator $A - \lambda I$ defined on $D(A)$. If the operator is *not* injective, i.e. there exists a non-zero $u \in D(A)$ such that $(A - \lambda I)u = 0$, λ is an *eigenvalue* of operator A, and u the corresponding *eigenvector*. Eigenvectors u corresponding to eigenvalue λ (plus the zero vector) form an *eigenspace* X_λ. The collection of all eigenvalues of operator A is identified as the *point* or *discrete spectrum* of operator A.

If operator $A - \lambda I$ is injective then it is invertible with the inverse $(A - \lambda I)^{-1}$ defined on the range of $A - \lambda I$. For a finite-dimensional space X, the Rank and Nullity Theorem implies that the inverse is defined on the whole space X, and it is automatically continuous. For an infinite-dimensional Hilbert space X, things can go wrong.

The range of $A - \lambda I$ may not be *dense* in X. We say then that λ belongs to the *residual spectrum* of operator A. If the inverse is defined on a dense subset of X, and *it is bounded*, it admits a unique extension to the whole space X. We say that λ belongs to the *resolvent set* of operator A, and call $R_\lambda := (A - \lambda I)^{-1}$ the *resolvent* of A at λ. If $(A - \lambda I)^{-1}$ is unbounded, we say that λ is an element of the *continuous spectrum* of A. A general operator may have discrete, residual, and continuous spectrum.

It turns out that a self-adjoint operator cannot have a residual spectrum. But it may have both discrete and continuous spectrum. Instead of attempting to formulate, even in an informal way, the *Spectral Theorem for Unbounded Self-Adjoint Operators*, we shall study a number of elementary examples in $L^2(\mathbb{R})$,

$$A : L^2(\mathbb{R}) \supset D(A) \ni u \to Au \in L^2(\mathbb{R})\,.$$

Example 3.6.1 First Derivative Operator

Consider

$$Au = i\frac{du}{dx}, \quad D(A) = \{u \in L^2(\mathbb{R}) \,:\, Au \in L^2(\mathbb{R})\} = H^1(\mathbb{R})\,.$$

The operator is self-adjoint (show it). The eigenproblem

$$i\frac{du}{dx} = \lambda u$$

leads to $\lambda = k \in \mathbb{R}$ and $u(x) = e^{-ikx}$. But function e^{-ikx} is not in $L^2(\mathbb{R})$ so it does not qualify for an eigenvector. The operator has no point spectrum but its continuous spectrum consists of the entire real line (see [5], Example 6.8.1 and Exercise 6.8.1). Normalized functions:

$$E(k,x) := \frac{1}{\sqrt{2\pi}} e^{-ikx}, \quad k \in \mathbb{R}\,,$$

called the *scattering* or *radiation modes* define the so-called *resolution of identity*. Namely, for every $u \in L^2(\mathbb{R})$, we have a representation:

$$u = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{u}(k) e^{-ikx}\, dk \tag{3.14}$$

where $\hat{u} \in L^2(\mathbb{R})$ is given by:

$$\hat{u}(k) = (u, E(\cdot,x))_{L^2(\mathbb{R})} = \int_{-\infty}^{\infty} u(x) \frac{1}{\sqrt{2\pi}} e^{ikx}\, dx\,.$$

The radiation modes define thus an integral transform:

$$\mathcal{E} \,: L^2(\mathbb{R}) \ni u \to \hat{u} \in L^2(\mathbb{R})\,.$$

Its L^2-adjoint coincides with its inverse: $\mathcal{E}^* = \mathcal{E}^{-1}$. We easily recognize that, modulo the sign in the exponent and scaling constants, transform $\mathcal{E}$ coincides with the Fourier transform. The spectral representation of the operator is given by:

$$i\frac{du}{dx} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k\, \hat{u}(k) e^{-ikx}\, dk\,.$$

□

I have not found in a literature a general mathematical term for them.

Please note an analogy with Sturm-Liouville operators. There we have a sequence of orthogonal eigenmodes which, after normalization, produce an orthonormal basis e_i for space $L^2_w(I)$ where I is a bounded interval in $\mathbb{R}$. This means that, for every $u \in L^2_w(I)$, we have the decomposition,

$$u = \sum_{i=1}^{\infty} u_i e_i$$

where the spectral components are computed by:

$$u_i = (u, e_i) = \int_I w(x) u(x) \overline{e_i(x)}\, dx\,.$$

So we can also think about two transforms. The direct transform operates from $L^2_w(I)$ into the ℓ^2, and the inverse one goes from ℓ^2 back to $L^2_w(I)$. The analogy is essentially the same as between Fourier transform and Fourier series. There are, of course, tones of subtle details that I am sweeping under the carpet. Let me mention one - the normalization of eigenmodes vs radiation modes. The normalization of eigenmodes is straightforward and it needs no comment. But *how do we normalize the radiation modes ?* Where in Example 3.6.1 does the $1/\sqrt{2\pi}$ normalizing factor come from ? Let $E(k,x)$ be a solution of the ODE defining the eigenvalue problem. Let ϕ be a test function. Compute first its 'forward' transform,

$$\hat{\phi}(x) = \int_{-\infty}^{\infty} \phi(k')\, \overline{E(x,k')}\, dk'\,.$$

Next, compute the 'inverse' transform. You should recover the original function premultiplied with a coefficient $N(k)$,

$$\int_{-\infty}^{\infty} \hat{\phi}(x)\, E(x,k)\, dx = N(k)\, \phi(k)\,.$$

The value of $N(k)$ should be real positive and independent of test function ϕ.

If we now normalize wave functions,

$$E(x,k) \to N^{-1/2}(k) E(x,k)\,,$$

after the forward and inverse transformations, we recover the original function. Indeed,

$$\int_{-\infty}^{\infty} \phi(k')\, N^{-1/2}(k') \overline{E(x,k')}\, dk' = \widehat{\phi N^{-1/2}} \quad \text{and} \quad \int_{-\infty}^{\infty} \widehat{\phi N^{-1/2}}(x)\, E(x,k)\, dx = N(k)\phi(k) N^{-1/2}(k)\,.$$

Dividing both sides by $N^{1/2}(k)$, we get

$$\int_{-\infty}^{\infty} \widehat{\phi N^{-1/2}}(x)\, N^{-1/2}(k) E(x,k)\, dx = \phi(k)\,.$$

In Example 3.6.1, if we start with $E(k,x) = e^{-ikx}$, we obtain $N(k) = \frac{1}{2\pi}$. In particular, in this *simplest possible* example, the normalizing factor is a number, and it is independent of k.

This normalizing condition is expressed in the engineering literature in a rather informal way as:

$$\int_{\mathbb{R}} E(k,x) \overline{E(k',x)}\, dx = \delta(k-k')\,.$$

Remember that radiation modes $E(k,x)$ are not L^2-functions; the condition above cannot be understood in the sense of the Lebesgue integral. However, if we test the equation above with a test function $\phi(k')$, we obtain:

$$\begin{aligned}\int_{\mathbb{R}} \phi(k') \int_{\mathbb{R}} E(k,x)\overline{E(k',x)}\,dx\,dk' &= \int_{\mathbb{R}} E(k,x) \underbrace{\int_{\mathbb{R}} \phi(k')\overline{E(k',x)}\,dk'}_{=\hat{\phi}(x)}\,dx && \text{(Fubini)}\\ &= \int_{\mathbb{R}} \phi(k')\delta(k-k')\,dk' && \text{(definition of Dirac's delta)}\\ &= \phi(k)\,,\end{aligned}$$

recovering the right meaning of the normalization condition.

Example 3.6.2 1D Laplacian

Consider

$$Au = -\frac{d^2u}{dx^2}, \quad D(A) = \{u \in L^2(\mathbb{R}) \,:\, Au \in L^2(\mathbb{R})\} = H^2(\mathbb{R})\,.$$

The operator is self-adjoint and positive-definite, see Exercise 3.6.1. The eigenproblem

$$-\frac{d^2u}{dx^2} = \lambda u, \quad \lambda > 0$$

leads to a continuous spectrum $\lambda \in (0,\infty)$ and the corresponding radiation modes $u(x) = e^{\pm ikx}$ where $k = \sqrt{\lambda}, k > 0$. Note that there are two linearly independent radiation modes for each eigenvalue λ. The resolution of identity coincides with (3.14), and the spectral representation of the operator is given by:

$$-\frac{d^2u}{dx^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k^2\, \hat{u}(k) e^{-ikx}\,dk\,.$$

Note that the spectral representation confirms that

$$-\frac{d^2}{dx^2}u = (i\frac{d}{dx})^2 u\,.$$

□

Example 3.6.3 1D Helmholtz Operator

Consider

$$Au = -\frac{d^2u}{dx^2} - \omega^2 u, \quad D(A) = \{u \in L^2(\mathbb{R}) \,:\, Au \in L^2(\mathbb{R})\} = H^2(\mathbb{R})\,.$$

The operator as a sum of two self-adjoint operators is self-adjoint. The continuous spectrum of the operator is obtained by shifting the spectrum of the Laplace operator, $\lambda \in (-\omega^2, \infty)$. The generalized eigenvectors are $u(x) = e^{\pm ikx}$ where $k = \sqrt{\lambda + \omega^2}, k > 0$. The resolution of identity coincides with (3.14), and the spectral representation of the operator is given by:

$$-\frac{d^2u}{dx^2} - \omega^2 u = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (k^2 - \omega^2)\, \hat{u}(k) e^{-ikx}\,dk\,.$$

□

Example 3.6.4 Three Layer Open Waveguide

In this example, we will encounter for the first time an operator with *both* discrete and continuous spectrum. Consider an open acoustical waveguide occupying the half-space: $x \in \mathbb{R}, z > 0$, governed by the equation:

$$-\frac{\partial^2 A}{\partial x^2} - \frac{\partial^2 A}{\partial z^2} - k_0^2 n^2(x) A = 0$$

where $n(x)$ is defined in Fig. 3.5. Separation of variables, $A = u(x)v(z)$ leads to

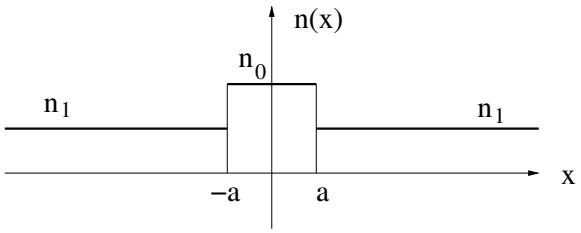

Figure 3.5
A three layer open waveguide.

$$-u''v - uv'' - k_0^2 n^2 uv = 0$$

or, equivalently,

$$-\frac{u''}{u} - k_0^2 n^2 = \frac{v''}{v} = \lambda$$

and, in turn, to the spectral analysis of the operator

$$Au = -u'' - k_0^2 n^2 u, \quad D(A) = \{u \in L^2(\mathbb{R}) \,:\, Au \in L^2(\mathbb{R})\} = H^2(\mathbb{R})\,.$$

The only difference with the previous example is the presence of variable coefficient in the zero order term. Because of that, however, the spectrum of the operator is no longer a simple shift of the Laplace operator spectrum, and the corresponding eigenvectors (and generalized eigenvectors) are very different.

The operator is clearly self-adjoint which implies that the separation constant λ must be real. We will assume $\lambda = -\beta^2$. For $\lambda < 0$, β is real and we obtain propagating modes with

$$v(z) = e^{\pm i\beta z}\,.$$

As usual, a radiation condition for $z \to \infty$ eliminates the mode representing a wave coming from infinity. Looking for the eigenvectors in the form:

$$u(x) = e^{rx}\,,$$

we arrive at the characteristic equation:

$$r^2 + k_0^2 n^2 - \beta^2 = 0\,.$$

Case: $k_0^2 n_1^2 - \beta^2 < 0$ and $k_0^2 n_0^2 - \beta^2 > 0$. Define:

$$k = k_c = (k_0^2 n_0^2 - \beta^2)^{1/2} \quad \text{and} \quad \alpha = (\beta^2 - k_0^2 n_1^2)^{1/2}\,.$$

Restricting ourselves to symmetric eigenmodes, we obtain:

$$u(x) = \begin{cases} C_2 \cos(kx) & x < a \\ C_1 e^{-\alpha x} + D_1 e^{\alpha x} & x > a\,. \end{cases}$$

The L^2-integrability eliminates $e^{\alpha x}$ and we end up with

$$u(x) = \begin{cases} C_2 \cos(kx) & x < a \\ C_1 e^{-\alpha x} & x > a\,. \end{cases}$$

The continuity conditions at $x = a$ lead to the equations:

$$\begin{aligned} C_2 \cos(ka) &= \;\; C_1 e^{-\alpha a} \\ k\, C_2 \sin(ka) &= \alpha C_1 e^{-\alpha a}\,. \end{aligned}$$

Requesting for a non-trivial solution, we obtain the dispersion relation:

$$k \tan ka = \alpha = ((k_0^2(n_0^2 - n_1^2) - k^2)^{1/2}\,. \tag{3.15}$$

We can also look for antisymmetric (odd) eigenfunctions. We briefly mention the other two cases, one needs to consider.

Case: $k_0^2 n_1^2 - \beta^2 < 0$ and $k_0^2 n_0^2 - \beta^2 < 0$. Defining,

$$\gamma = (\beta^2 - k_0^2 n_0^2)^{1/2} \quad \text{and} \quad \alpha = (\beta^2 - k_0^2 n_1^2)^{1/2}\,,$$

we obtain (still looking for symmetric solutions),

$$u(x) = \begin{cases} C_2 \cosh(\gamma x) & x < a \\ C_1 e^{-\alpha x} & x > a\,. \end{cases}$$

As before, the continuity conditions lead to the dispersion relation

$$\gamma \tanh(\gamma a) = -\alpha = (\gamma^2 + k_0^2(n_0^2 - n_1^2))^{1/2}$$

which clearly has no real solution (the left-hand side function is positive whereas the right-hand side is negative).

Case: $k_0^2 n_1^2 - \beta^2 > 0$ and, therefore, $k_0^2 n_0^2 - \beta^2 > 0$ as well. Define:

$$k_c = (k_0^2 n_0^2 - \beta^2)^{1/2} \quad \text{and} \quad k = k_x = (k_0^2 n_1^2 - \beta^2)^{1/2}\,.$$

We obtain two families of solutions. The first family consists of even solutions:

$$u_{e,k}(x) = \begin{cases} C_2 \cos k_c x & x < a \\ C_1 \cos kx + D_1 \sin kx & x > a\,. \end{cases}$$

Selecting $C_2 = 1$, the continuity conditions at $x = a$,

$$\begin{pmatrix} \cos ka & \sin ka \\ -k \sin ka & k \cos ka \end{pmatrix} \begin{pmatrix} C_1 \\ D_1 \end{pmatrix} = \begin{pmatrix} \cos k_c a \\ -k_c \sin k_c a \end{pmatrix}$$

lead to

$$C_1 = \cos(k_c - k)a + \frac{k_c - k}{k} \sin k_c a \sin ka \quad D_1 = -\sin(k_c - k)a - \frac{k_c - k}{k} \sin k_c a \cos ka\,.$$

The second family consists of odd solutions,

$$u_{o,k}(x) = \begin{cases} \sin k_c x & x < a \\ C_1 \cos kx + D_1 \sin kx & x > a\,. \end{cases}$$

with analogous formulas for C_1, D_1. Thus, for each $k > 0$, we have two radiating modes. Typically, we reparametrize the family with a single $k \in \mathbb{R}$. For instance, we may define:

$$u_{c,k} = \begin{cases} u_{e,k} + iu_{o,k} & k > 0 \\ u_{e,k} - iu_{o,k} & k < 0\,. \end{cases}$$

In the central region, the redefined radiation modes coincide simply with $e^{ik_c x}$.

Thus, for $\lambda = -\beta^2 > -k_0^2 n_1^2$, we obtain a continuous spectrum. The resolution of identity takes the form:

$$u(x) = \sum_i u_i u_{d,i}(x) + \int_{-\infty}^{+\infty} \hat{u}(k) u_{c,k}(x) dk$$

where the sum extends over the discrete eigenmodes $u_{d,i}$ (we have determined a symmetric one) and $u_{c,k}$ are the radiation modes. The two contributions above are L^2-orthogonal to each other. To determine the spectral components u_i, we multiply with an eigenmode $\overline{u_{d,i}(x)}$ and integrate over the real line to obtain:

$$u_i \int_{-\infty}^{\infty} |u_{d,i}(x)|^2\, dx = \int_{-\infty}^{\infty} u(x) \overline{u_{d,i}(x)}\, dx\,.$$

If the eigenmodes have been normalized, the coefficient on the left equals one. Once we have determined spectral components u_i, we multiply both sides with a radiation mode $\overline{u_{c,k}(x)}$, and integrate over the real line to obtain:

$$\hat{u}(k) N(k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{u}(k') u_{c,k'}(x) dk'\, \overline{u_{c,k}(x)}\, dx = \int_{-\infty}^{\infty} [u(x) - \sum_i u_i u_{d,i}(x)] \overline{u_{c,k}(x)}\, dx\,.$$

Again, for normalized radiation modes $N(k) = 1$.

Once the spectral components are known, the action of operator A on u takes the form:

$$Au = \sum_i \underbrace{(k_i^2 - k_0^2 n_0^2)}_{=\lambda_i = -\beta_i^2} u_i u_{d,i}(x) + \int_{-\infty}^{+\infty} \underbrace{(k^2 - k_0^2 n_0^2)}_{=\lambda = -\beta^2} \hat{u}(k) u_{c,k}(x) dk\,.$$

Note that all discrete modes are propagating modes as $\lambda_i = -\beta_i^2 < 0$. Concerning the radiation modes, for $|k| < k_0 n_0$ we have real β, and we obtain propagating modes in z, whereas for $|k| > k_0 n_0$, β is purely imaginary, and we obtain evanescent modes. Both types of radiation modes loose energy from the central part of the waveguide, the propagating modes by radiating it out, and the evanescent modes by decaying exponentially.

Normalization. It is relatively easy to normalize the eigenmodes. For instance, for the symmetric eigenmode, we have determined, we have $C_1 = e^{\alpha a} \cos kaC_2$. Assuming $C_2 = 1$, we compute the L^2-norm of the mode:

$$\int_{-\infty}^{\infty} |u(x)|^2\, dx = 2\int_0^{\infty} |u(x)|^2\, dx = 2\int_0^a \cos^2 kx\, dx + 2\int_a^{\infty} e^{2\alpha(a-x)} \cos^2 ka\, dx = a + \frac{1}{2k}\sin 2ka$$

where k is one of the roots of dispersion relation (3.15). Computation of factors $N(k)$ is much more cumbersome and I could not do it analytically. □

Example 3.6.5 W-Type Open Waveguide

Consider now a waveguide with $n(x)$ defined in Fig. 3.6. As in Example 3.6.4, we assume that the

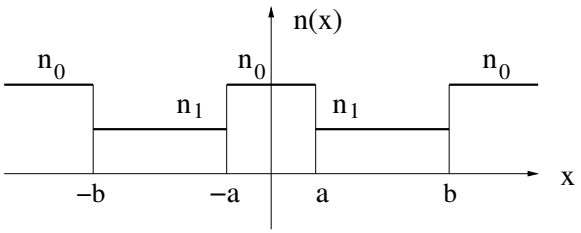

Figure 3.6
W-type waveguide.

(real) eigenvalue $\lambda = -\beta^2$.

Case: $k_0^2 n_1^2 - \beta^2 < 0$ and $k_0^2 n_0^2 - \beta^2 > 0$. Define:

$$k = (k_0^2 n_0^2 - \beta^2)^{1/2} \quad \text{and} \quad \alpha = (\beta^2 - k_0^2 n_1^2)^{1/2}\,.$$

Restricting ourselves to symmetric eigenmodes, we obtain:

$$u(x) = \begin{cases} C_2 \cos(kx) & x < a \\ C_1 e^{-\alpha x} + D_1 e^{\alpha x} & a < x < b \\ C_0 e^{-ikx} + D_0 e^{ikx} & x > b\,. \end{cases}$$

Note that in the most outer region, we have two propagating modes. The L^2-integrability condition eliminates both of them, i.e., $C_0 = D_0 = 0$. The continuity conditions at $x = b$ imply then that

$C_1 = D_1 = 0$ as well and, finally, the continuity conditions at $x = a$ imply that the solution must be trivial. Concluding, the solution above can only represent a radiation mode. We set $C_2 = 1$. Solving the continuity condition at $x = a$,

$$\begin{aligned} \cos(ka) &= C_1 e^{-\alpha a} + D_1 e^{\alpha a} \\ -k \sin ka &= -C_1 \alpha e^{-\alpha a} + D_1 \alpha e^{\alpha a} \end{aligned}$$

for C_1, D_1, we obtain,

$$\begin{aligned} C_1 &= \tfrac{1}{2\alpha}(\quad \alpha e^{\alpha a} \cos ka + \quad k e^{\alpha a} \sin ka) \\ D_1 &= \tfrac{1}{2\alpha}(\, \alpha e^{-\alpha a} \cos ka - k e^{-\alpha a} \sin ka) \end{aligned} \,. \tag{3.16}$$

Similarly, considering the continuity conditions at $x = b$,

$$\begin{pmatrix} e^{-\alpha b} & e^{\alpha b} \\ -\alpha e^{-\alpha b} & \alpha e^{\alpha b} \end{pmatrix} \begin{pmatrix} C_1 \\ D_1 \end{pmatrix} = \begin{pmatrix} e^{-ikb} & e^{ikb} \\ -ike^{-ikb} & ike^{ikb} \end{pmatrix} \begin{pmatrix} C_0 \\ D_0 \end{pmatrix}$$

and solving them for C_1, D_1, we obtain:

$$\begin{aligned} \begin{pmatrix} C_1 \\ D_1 \end{pmatrix} &= \tfrac{1}{2\alpha} \begin{pmatrix} \alpha e^{\alpha b} & -e^{\alpha b} \\ \alpha e^{-\alpha b} & e^{-\alpha b} \end{pmatrix} \begin{pmatrix} e^{-ikb} & e^{ikb} \\ -ike^{-ikb} & ike^{ikb} \end{pmatrix} \begin{pmatrix} C_0 \\ D_0 \end{pmatrix} \\ &= \tfrac{1}{2\alpha} \begin{pmatrix} e^{\alpha b}(\alpha + ik)e^{-ikb} & e^{\alpha b}(\alpha - ik)e^{ikb} \\ e^{-\alpha b}(\alpha - ik)e^{-ikb} & e^{-\alpha b}(\alpha + ik)e^{ikb} \end{pmatrix} \begin{pmatrix} C_0 \\ D_0 \end{pmatrix} \end{aligned} \tag{3.17}$$

Comparing formulas (3.16) and (3.17), we obtain a system of equations for C_0, D_0,

$$\begin{pmatrix} (\alpha + ik)e^{-ikb} & (\alpha - ik)e^{ikb} \\ (\alpha - ik)e^{-ikb} & (\alpha + ik)e^{ikb} \end{pmatrix} \begin{pmatrix} C_0 \\ D_0 \end{pmatrix} = \begin{pmatrix} e^{-\alpha(b-a)}(\alpha \cos ka + k \sin ka) \\ e^{\alpha(b-a)}(\alpha \cos ka - k \sin ka) \end{pmatrix} \tag{3.18}$$

with the determinant of the matrix on the left equal to $4i\alpha k \neq 0$.

An alternative derivation of the radiation modes. As we can see, coming up with explicit formulas for the coefficients C_0, D_0, C_1, D_1 is quite cumbersome. We can try to represent the radiation mode in a different form:

$$u(x) = \begin{cases} \cos kx & x < a \\ C_1 e^{-\alpha x} + C_2 e^{\alpha x} & a < x < b \\ C_0 \cos(kx + \phi) & x > b \,. \end{cases}$$

Note that the representation is purely real. We are interested in computing constant C_0 explicitly (as a function of k) so we can normalize the mode by dividing with C_0. This is not the exact normalization but it produces modes that are equivalent with the (properly) normalized modes *uniformly in* k. In particular, we are interested in zeros of C_0 which make the solution blow up in the central region. You may treat it simply as an example of a derivation of the radiation modes using an alternative representation.

Continuity conditions at $x = b$ lead to the expression for C_1, D_1 in terms of C_0 and phase ϕ,

$$\begin{pmatrix} C_1 \\ D_1 \end{pmatrix} = \frac{1}{2\alpha} \begin{pmatrix} \alpha e^{\alpha b} & -e^{\alpha b} \\ \alpha e^{-\alpha b} & e^{-\alpha b} \end{pmatrix} \begin{pmatrix} \cos(kb + \phi) \\ -k \sin(kb + \phi) \end{pmatrix} C_0 \,.$$

Comparing with formulas (3.16), we get a system of equations for C_0 and ϕ,

$$[\alpha\cos(kb+\phi)+k\sin(kb+\phi)]C_0 = e^{-\alpha(b-a)}[\alpha\cos ka + k\sin ka]$$
$$[\alpha\cos(kb+\phi)-k\sin(kb+\phi)]C_0 = \quad e^{\alpha(b-a)}[\alpha\cos ka - k\sin ka]$$

Adding and subtracting the two equations, we get,

$$\cos(kb+\phi)C_0 = \left\{e^{-\alpha(b-a)}[\alpha\cos ka + k\sin ka] + e^{\alpha(b-a)}[\alpha\cos ka - k\sin ka]\right\}\tfrac{1}{2\alpha}$$
$$\sin(kb+\phi)C_0 = \left\{e^{-\alpha(b-a)}[\alpha\cos ka + k\sin ka] - e^{\alpha(b-a)}[\alpha\cos ka - k\sin ka]\right\}\tfrac{1}{2k}$$

Finally, squaring and adding both equations we eliminate the phase ϕ and get a formula for C_0^2,

$$\begin{aligned}4\alpha^2k^2C_0^2 &= (k^2+\alpha^2)\{e^{-2\alpha(b-a)}[\alpha\cos ka + k\sin ka]^2 + e^{2\alpha(b-a)}[\alpha\cos ka - k\sin ka]^2\}\\ &\quad +2(k^2-\alpha^2)(\alpha^2\cos^2 ka - k^2\sin^2 ka)\\ &= \cos^2 ka\{(k^2+\alpha^2)\{e^{-2\alpha(b-a)}[\alpha + k\tan ka]^2 + e^{2\alpha(b-a)}[\alpha - k\tan ka]^2\} + 2(k^2-\alpha^2)(\alpha^2 - k^2\tan^2 ka)\}\,.\end{aligned}$$

Denoting,

$$\kappa := k/\alpha \qquad \gamma := e^{-2\alpha(b-a)}\,, \qquad \delta := e^{2\alpha(b-a)} \qquad z = k\tan ka/\alpha$$

(note that $\gamma\delta = 1$) we have,

$$\begin{aligned}&(k^2+\alpha^2)\{e^{-2\alpha(b-a)}[\alpha + k\tan ka]^2 + e^{2\alpha(b-a)}[\alpha - k\tan ka]^2\} + 2(k^2-\alpha^2)(\alpha^2 - k^2\tan^2 ka)\\ &= \alpha^4\{(\kappa^2+1)(\gamma(1+z)^2 + \delta(1-z)^2) + 2(\kappa^2-1)(1-z^2)\}\,.\end{aligned}$$

Focusing on the quadratic (in z) term:

$$\begin{aligned}A &= (\kappa^2+1)(\gamma+\delta) - 2(\kappa^2-1)\\ B &= 2(\kappa^2+1)(\gamma-\delta)\\ C &= (\kappa^2+1)(\gamma+\delta) + 2(\kappa^2-1)\\ \Delta = B^2 - 4AC &= -64\kappa^2\end{aligned}$$

$$z_{1,2} = \frac{-B\pm\sqrt{\Delta}}{2A} = \frac{-2(\kappa^2+1)(\gamma-\delta)\pm 8i\kappa}{2(\kappa^2+1)(\gamma+\delta) - 4(\kappa^2-1)}\,.$$

Dividing numerator and denominator by δ, we obtain:

$$z_{1,2} = \frac{-B\pm\sqrt{\Delta}}{2A} = \frac{-2(\kappa^2+1)(e^{-4\alpha(b-a)}-1)\pm 8i\kappa e^{-2\alpha(b-a)}}{2(\kappa^2+1)(e^{-4\alpha(b-a)}+1) - 4(\kappa^2-1)e^{-2\alpha(b-a)}}\,. \tag{3.19}$$

The coefficient C_0^2 vanishes thus at solutions to the equations:

$$\frac{k\tan ka}{\alpha} = z_{1,2}\,. \tag{3.20}$$

As expected, for $b\to\infty$, $z_1, z_2 \to 1$, and the two dispersion relations converge to

$$k\tan ka = \alpha\,,$$

the dispersion relation defining the eigenmodes in Example 3.6.4.

One can show (nor so quickly) that the two remaining scenarios:
Case: $k_0^2n_1^2 - \beta^2 < 0$ and $k_0^2n_0^2 - \beta^2 < 0$, and **Case:** $k_0^2n_1^2 - \beta^2 > 0$ and $k_0^2n_0^2 - \beta^2 > 0$, lead also to radiation modes only. The operator has only a continuous spectrum. □

A summary. The purpose of this Section was to demonstrate that the solution of PDEs stated on the whole real line via the Fourier transform (see Example 6.1.10) is based on the same principle as the separation of variables based on Sturm-Liouville theory for problems defined on a bounded interval. The key to the story is the general *Spectral Theory for (Unbounded) Self-Adjoint Operators*. In either case, the separation of variables leads to a BVP for a self-adjoint operator. Dependent upon the spectrum, we end up with the representation of the solution in terms of a series, or an integral, or both, like in Example 3.6.4.

Exercises

Exercise 3.6.1 Demonstrate that the 1D Laplace operator considered in Example 3.6.2 is self-adjoint and positive definite. Start by showing that the domain of the operator indeed coincides with space $H^2(\mathbb{R})$. *Hint:* Use Fourier transform.

(10 points)

Exercise 3.6.2 Let $\mathcal{E}$ be the transform defined by the *normalized* radiation modes from Example 3.6.4,

$$\mathcal{E} \,:\, V^\perp \ni u \to \hat{u} \in L^2(\mathbb{R}), \qquad \hat{u}(k) := \int_{\mathbb{R}} u(x)\overline{u_{c,k}(x)}\, dx$$

where $V^\perp$ denotes the orthogonal complement of subspace V spanned by the finite family of eigenmodes

$$V := \mathrm{span}\{u_{d,i}\}$$
$$V^\perp := \{u \in L^2(\mathbb{R}) \,:\, (u, u_{d,i}) = 0 \quad \forall i\}.$$

Let $\mathcal{E}^{-1}$ be its inverse:

$$\mathcal{E}^{-1} \,:\, L^2(\mathbb{R}) \ni \hat{u} \to u \in V^\perp, \qquad u(x) := \int_{\mathbb{R}} \hat{u}(k) u_{c,k}(x)\, dk\,.$$

Argue why both maps preserve the L^2-orthogonaliy, i.e.,

$$\hat{u}_1, \hat{u}_2 \in L^2(\mathbb{R}), \quad (\hat{u}_1, \hat{u}_2) = 0 \quad \Rightarrow \quad (u_1, u_2) = 0$$
$$u_1, u_2 \in L^2(\mathbb{R})^\perp, (u_1, u_2) = 0 \quad \Rightarrow \quad (\hat{u}_1, \hat{u}_2) = 0\,.$$

(10 points)

Exercise 3.6.3 Analyze **Case:** $k_0^2 n_1^2 - \beta^2 < 0$ and $k_0^2 n_0^2 - \beta^2 < 0$ in Example 3.6.5 for the existence of eigenvalues (propagating modes).

(10 points)

Exercise 3.6.4 Show that Example 3.6.4 is equivalent to the *finite well* problem for the Schrödinger equation.

(10 points)

Exercise 3.6.5 Consider the operator

$$Au = -\frac{d^2u}{dx^2}, \quad D(A) := \{u \in L^2(0,\infty) \,:\, Au \in L^2(0,\infty),\, u'(0) = 0\}\,.$$

Show that the operator is self-adjoint and has a continuous spectrum

$$\{k^2 \,:\, k \in (0,\infty)\}$$

with the corresponding (generalized) eigenmodes:

$$e_k(x) = \frac{2}{\sqrt{2\pi}} \cos kx\,.$$

Deduce the partition of unity, and the spectral representation of the operator

$$u(x) = \int_0^\infty \tilde{u}(k) e_k(x)\, dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty \tilde{u}(k) \cos kx\, dx$$
$$(Au)(x) = -u''(x) = \int_0^\infty k^2 \tilde{u}(k) e_k(x)\, dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty k^2 \tilde{u}(k) \cos kx\, dx$$

where

$$\tilde{u}(k) = \frac{2}{\sqrt{2\pi}} \int_0^\infty u(x) \cos kx\, dx\,.$$

Hint: The only non-trivial part of the problem is the scaling of the modes. (10 points)

4

Ordinary Differential Equations

4.1 Systems of First Order Equations

By identifying derivatives as new unknowns the Initial Value Problem (IVP) for any explicit system of ODEs of arbitrary order can be turned into the IVP for an explicit system of first order ODEs:

$$\begin{cases} \dot{x}(t) = f(x,t) & t \in (0,T) \\ x(0) = x_0 \end{cases}$$

where we seek a vector-valued function $x : (0,T) \ni t \to x(t) \in \Omega \subset \mathbb{C}^n$, $f(x,t) : \Omega \times (0,T) \to \mathbb{C}^n$ is a flux function, and $x_0 \in \Omega$ is an initial condition data. As usual, $\Omega \subset \mathbb{C}^n$ is a domain (= open, connected set). Classical analysis using Chebyshev spaces leads to the assumpion that the flux function is Lipschitz continuous in x uniformly in t. The solution lives then in space:

$$C^1((0,T);\mathbb{C}^n) \cap C^0([0,T];\mathbb{C}^n)\,.$$

Banach Contractive Map argument leads to the *local existence and uniqueness result*, i.e. there exists $T > 0$ such that the solution exists and it us unique, see [5], Section 4.10.

Of a particular importance is the linear case when the flux function is linear in x, i.e.,

$$f(x,t) = A(t)x + b(t)\,, \qquad A(t) \in L(\mathbb{C}^n, \mathbb{C}^n), \quad b(t) \in \mathbb{C}^n\,.$$

Uniform Lipschitz continuity of f in x translates into the assumption:

$$\|A(t)\| \leq M \qquad t > 0$$

where the operator norm for $A(t)$ is induced by a particular norm used for $\mathbb{C}^n$. For the canonical L^2-norm, this translates into a bound for the largest characteristic value of $A(t)$. With this assumption in place, the value of maximum time T for which the solution exists is only a function of constant M. We can use value $x(M)$ to restart the problem and continue the solution to $2T$ and so on. This leads to the result that the solution to the linear problem exists and it is unique for any $t > 0$.

Special case with constant $A(t) = A$. We will look first for a general solution of the homogeneous system:

$$\dot{x}(t) = Ax\,.$$

By the Jordan Decomposition Theorem, there exists an eigenbasis for $\mathbb{C}^n$ consisting of generalized eigenvectors. Let $\lambda \in \mathbb{C}$ be an eigenvalue of operator A, with corresponding eigenvector e_0. Let $e = e_1, \ldots, e_k$ be the possible corresponding Jordan train of generalized eigenvectors, i.e.,

$$Ae_0 = \lambda e_0 \quad Ae_j = \lambda e_j + e_{j-1}, \quad j = 1, \ldots, k\,.$$

We seek a solution in the form:

$$x = c_0(t)e_0 + \sum_{j=1}^{k} c_j(t)e_j\,.$$

Comparing

$$\dot{x} = \dot{c}_0 e_0 + \sum_{j=1}^{k} \dot{c}_j(t)e_j$$

with

$$Ax = \lambda c_0 e_0 + \sum_{j=1}^{k} c_j(\lambda e_j + e_{j-1})\,,$$

we obtain a semi-decoupled system of ODEs for coefficients $c_0, c_1, \ldots, c_k$,

$$\dot{c}_0 = \lambda c_0 + c_1 \quad \dot{c}_1 = \lambda c_1 + c_2 \quad \ldots \quad \dot{c}_k = \lambda c_k\,.$$

We obtain:

$$c_k = e^{\lambda t}C_k \quad c_{k-1} = e^{\lambda t}(C_{k-1} + C_k t) \quad \ldots \quad c_0 = e^{\lambda t}(C_0 + C_1 t + \ldots + C_k \frac{t^k}{k!})$$

where $C_0, C_1, \ldots, C_k$ are arbitrary integration constants. The corresponding solution is thus,

$$\begin{aligned} x(t) &= e^{\lambda t}\left[\left(C_0 + C_1 t + \ldots + C_k \frac{t^k}{k!}\right)e_0 + \left(C_1 + C_2 t + \ldots + C_k \frac{t^{k-1}}{(k-1)!}\right)e_1 + \ldots + C_k e_k\right] \\ &= e^{\lambda t}\left[C_0 e_0 + C_1\left(te_0 + e_1\right) + \ldots + C_k\left(\frac{t^k}{k!}e_0 + \frac{t^{k-1}}{(k-1)!}e_1 + \ldots + e_k\right)\right]\,. \end{aligned}$$

The ultimate (general) solution consists of terms like this for each Jordan train of eigenvectors corresponding to an eigenvalue λ. Remember that we may have multiple trains corresponding to the same λ.

Exercises

Exercise 4.1.1 Consider the matrix:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}\,.$$

- Determine generalized eigenvectors of matrix A and the corresponding Jordan form.
- Use the Jordan form to determine general solution for the system of ODEs:

$$\dot{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{u}\,.$$

(10 points)

Exercise 4.1.2 Determine the general solution to the system of five ODEs:

$$\dot{x} = Ax$$

where

$$A := \begin{pmatrix} 2 & 1 & & & \\ & 2 & 1 & & \\ & & 2 & & \\ & & & 2 & 1 \\ & & & & 2 \end{pmatrix}.$$

(10 points)

Exercise 4.1.3 Determine the general solution to the system of four ODEs:

$$\dot{x} = Ax$$

with matrix A from Example 3.2.2.

(10 points)

4.2 Standard Solution Techniques

In this section we review some very standard methods of solving particular ODEs analytically. I have made a very personal choice of techniques that I have found useful in my own academic research and career. The selection is by no means exhaustive, see [3] for a much bigger selection.

4.2.1 A Single ODE of First Order

A general ODE may be given in an *implicit form*:

$$F(x, y, y') = 0$$

where we are looking for a function $y = y(x)$, and F is a given function of three variables. If we can solve the equation above for y', we get

$$y' = f(x, y)$$

where f is an appropriate function of two variables. We talk then about an *explicit ODE*. Most of the time, we deal with explicit ODEs.

Separation of variables. This, perhaps the most known technique, applies to the situation when $f(x, y) = \phi(x)\psi(y)$. We *separate the variables* rewriting the equation in the form:

$$\psi^{-1}(y)\, dy = \phi(x)\, dx$$

and integrate both sides,

$$\int \psi^{-1}(y)\, dy = \int \phi(x)\, dx + C$$

where by the integrals on both sides we mean primitive functions of ψ^{-1} and $\phi(x)$ (indefinite integrals), and C is an arbitrary *integration constants*. We get thus a one-parameter family of solutions, the *general solution* (to be discussed in a moment). If the ODE is accompanied with an IC,

$$y(x_0) = x_0\,,$$

we incorporate the IC into the integration process using *definite integrals*,

$$\int_{x_0}^{x} \psi^{-1}(y)\, dy = \int_{y_0}^{y} \phi(x)\, dx\,,$$

and solve for $y = y(x)$.

Example 4.2.1

Solve the IVP:

$$\begin{cases} \dfrac{dy}{dx} = \dfrac{x^3}{y^2} \quad x > 0 \\ y(0) = 1\,. \end{cases}$$

We have,

$$\int_0^y y^2\, dy = \int_1^x x^3\, dx$$

which gives,

$$\frac{1}{3}y^3|_1^y = \frac{1}{4}x^4|_0^x$$

or,

$$y(x) = \left[\frac{3}{4}x^4 + 1\right]^{1/3}\,.$$

□

Variation of a constant. Consider a general non-homogeneous linear equation with general *variable* coefficients:

$$y' + m(x)y = n(x)\,.$$

Recall that *the general solution of a linear ODE equals the general solution of the (corresponding) homogeneous ODE, and a particular solution (of the original one)* * and look first for a general solution of the

*My teachers made me memorize this statement when I was an engineering undergraduate.

homogeneous equation:

$$y' + m(x)y = 0\,.$$

Separation of variables leads to the solution:

$$y = e^{-\int m(x)+c} = Ce^{-\int m(x)}$$

where $\int m$ denotes a primitive of $m(x)$ (indefinite integral), and $C = e^c$. We can look now for a particular solution of the non-homogeneous equation in the form:

$$y = C(x)e^{-\int m(x)}$$

where C is no longer constant but an unknown function, hence the name of the technique. This leads to the equation:

$$C'e^{-\int m(x)} = n(x)$$

and the formula for $C(x)$:

$$C(x) = \int e^{\int m(x)} n(x) + D$$

where again the indefinite integral is known up to an additive integration constant D. The ultimate solution is:

$$y(x) = (\int e^{\int m(x)} n(x) + D)e^{-\int m(x)}\,.$$

As depressing as it is, to my best knowledge, this is the only linear ODE with variable coefficients for which we have a general solution in the closed form.

Exact differentials. Integrating factors. Sometimes, an ODE is given in the form:

$$P(x,y)dx + Q(x,y)dy = 0 \qquad (x,y) \in G \subset \mathbb{R}^2\,.$$

If it represents a *differential* of a function $u(x,y)$, i.e.,

$$P(x,y) = \frac{\partial u}{\partial x} \qquad Q(x,y) = \frac{\partial u}{\partial y}\,,$$

then the corresponding general solution is of the form

$$u(x,y) = C$$

as vanishing of the differential $d_x u$ in G is equivalent to u being constant. The equation above defines implicitly a function $y(x)$ but it may also define a function $x(y)$, and the determination of such functions may be possible only *locally* (think about u representing a circle). By determining $u(x,y)$ we do not have to make a decision which of the two variables will be an independent and which will be a dependent variable. The problem of finding u is equivalent to the problem of finding a scalar potential of vector-valued function (P,Q). The necessary (and sufficient, if domain G is simply connected) condition is that $\mathrm{curl}(P,Q) = 0$, i.e.

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}\,. \tag{4.1}$$

If the condition is not satisfied, we still may be lucky finding the so-called *integrating factor*, i.e., a function $\phi(x, y)$ such that

$$\operatorname{curl}(\phi P, \phi Q) = 0\,.$$

In general, finding ϕ is as difficult as solving for u. In some (academic?) examples, we may be lucky finding an integrating factor that depends only on one of the variables, i.e., $\mu = \mu(x)$ or $\mu = \mu(y)$.

Envelopes and singular solutions. If the flux function does not satisfy the Lipschitz condition, we may have multiple solutions to an IVP for an ODE. Such a situation occurs when the general solution of an implicit equations $F(x, y, y') = 0$, represented in the form of a one-parameter family,

$$\psi(x, y, c) = 0$$

admits an *envelope*. By the envelope we mean a curve that is tangent to all curves in the family. If it exists, it must satisfy the equation:

$$\frac{\partial \psi}{\partial c}(x, y, c) = 0\,.$$

Eliminating c from the system above, we obtain an equation for the envelope. The envelope solves the original ODE as well. Consequently, for any point common to the envelope and one of particular solutions used for an initial condition, we have an example of an IVP with multiple solutions.

Example 4.2.2

Consider the equation

$$\frac{dy}{dx} = y^{1/2}\,.$$

Note that the flux function is not Lipschitz continuous (in y) in a vicinity of $y = 0$. We can separate the variables to get a (the?) general solution:

$$y = \left(\frac{x}{2} + c\right)^2\,. \tag{4.2}$$

Differentiating in c, we obtain,

$$0 = 2\left(\frac{x}{2} + c\right)\,.$$

Eliminating c, we get the envelope:

$$y = 0\,.$$

Note that the envelope *satisfies the equation as well, and it is not included in family (4.2)*. Some authors would define the general solution as a one-parameter family of solutions. The parameter is then fixed by satisfying an IC. If we add an IC in the form: $y(x_0) = y_0$ where $x_0 \in \mathbb{R}$, and $y_0 \geq 0$, we can determine a unique C for which the IC is satisfied. Family (4.2) is then the general solution understood in this sense. Some other authors, though, would request the general solution to include *all particular solutions* and, in this sense, we cannot claim family (4.2) as the general solution since it has missed the envelope.

Finally, note that for $y_0 = 0$, we have two solutions: one belonging to the family (4.2) and the envelope: $y = 0$. This example illustrates thus the importance of the assumption that the flux is Lipschitz continuous in y to guarantee the uniqueness of the solution. □

4.2.2 A Single ODE of Higher Order

Cauchy-Euler equation. Determine general solution of the equation:

$$a_0 x^n y^{(n)} + a_1 x^{n-1} y^{(n-1)} + \ldots + a_{n-1} x y' + a_n y = 0$$

where $a_0, a_1, \ldots, a_n$ are constants. Note that the exponent in x^i matches exactly the derivative order (i) in each term. Seeking the solution in the form $y = x^r$, we obtain factor x^r in each term and, consequently, a characteristic algebraic equation for r,

$$r(r-1)\ldots(r-n)a_0 + r\ldots(r-(n-1))a_1 + \ldots + r a_{n-1} + a_n = 0$$

that can be solved for r. In the case of multiple roots, we use *variation of a constant* method to find missing solutions.

Note that the Cauchy-Euler equation can be reduced to the equation with constant coefficients by changing the *independent* variable: $x = e^t$.

Example 4.2.3

Solve:

$$x^2 y'' + x y' - y = 0\,.$$

Using the ansatz $y = x^r$, we get:

$$r(r-1) + r - 1 = r^2 - 1 = 0\,.$$

This leads to two linearly independent solutions: $y_1 = x$, $y_2 = x^{-1}$, and the general solution:

$$y = Ax + Bx^{-1}, \quad A, B \in \mathbb{R}\,.$$

□

4.2.3 Analytical Solutions and the Frobenius Method

We shall restrict ourselves to a second order homogeneous linear ODE with variable coefficients:

$$p(x)y'' + q(x)y' + n(x)y = 0\,.$$

The main motivation of the presentation here is to introduce Bessel and Legendre functions. We follow closely [3]. Diving by $p(x)$, we obtain the explicit form of the equation,

$$y'' + \frac{q(x)}{p(x)}y + \frac{n(x)}{p(x)}y = 0\,.$$

Assume that coefficients $q(x)/p(x)$ and $n(x)/p(x)$ are analytic in a neighborhood of a point x_0, say for $|x - x_0| < r_0$. Then, by the Cauchy-Kovalevskaya Theorem, the problem admits a general, two-parameter analytical solution with the radius of convergence at x_0 greater of equal to r_0. We can solve effectively the problem by seeking the solution in the form of its Taylor series.

Example 4.2.4

Use the Taylor expansion at $x = 0$ to find the general solution of the equation:

$$y'' + xy' - y = 0\,.$$

We have:

$$\begin{aligned} y &= \sum_{k=0}^{\infty} c_k x^k \\ y' &= \sum_{k=1}^{\infty} k c_k x^{k-1} \qquad\qquad xy' = \sum_{k=1}^{\infty} k c_k x^k \\ y'' &= \sum_{k=2}^{\infty} k(k-1) c_k x^{k-2} \quad \text{or} \quad y'' = \sum_{k=0}^{\infty} (k+2)(k+1) c_{k+2} x^k\,. \end{aligned}$$

Substituting in the equation, we get,

$$\sum_{k=0}^{\infty} (k+2)(k+1) c_{k+2} x^k + \sum_{k=1}^{\infty} k c_k x^k - \sum_{k=0}^{\infty} c_k x^k = 0\,.$$

Note that the summation in the second term starts from $k = 1$. Equating coefficients corresponding to different powers x^k to zero, we obtain:

$$\begin{aligned} k = 0 \quad & 2c_2 - c_0 = 0 && \Rightarrow c_2 = \tfrac{1}{2} c_0 \\ k = 1 \quad & 6c_3 + c_1 - c_1 = 0 && \Rightarrow c_3 = 0 \\ k > 1 \quad & (k+1)(k+1) c_{k+2} + (k-1) c_k = 0 && \Rightarrow c_{k+2} = -\tfrac{k-1}{(k+2)(k+1)} c_k\,. \end{aligned}$$

We obtain thus two solutions y_1 and y_2. Function y_1 corresponds to arbitrary constant c_0 and it contains only even powers of x^k:

$$c_0 \in \mathbb{R}, \quad c_2 = \frac{1}{2} c_0 \quad c_{k+2} = -\frac{k-1}{(k+2)(k+1)} c_k \quad \text{for } k > 1\,.$$

y_2 involves a single term only corresponding to arbitrary $c_1 \in \mathbb{R}$, $y_2 = c_1 x$. Note that the solution y_1 is given in a recursion form that can be implemented on a computer. □

The situation is more interesting and complicated if coefficients $q(x)/p(x)$ and $n(x)/p(x)$ are singular at x_0. If

$$\frac{q(x)}{p(x)}(x - x_0) \quad \text{and} \quad \frac{r(x)}{p(x)}(x - x_0)^2$$

are analytic though, we call the singular point x_0 a *regular singular point* and look for the solution in a modified form:

$$y(x) = (x - x_0)^\alpha (\sum_{k=0}^{\infty} c_k (x - x_0)^k)$$

where α is an unknown constant to be determined. This the *method of Frobenius*. We are guaranteed to get at least one (possibly two) linearly independent solutions.

Example 4.2.5 Bessel functions

Equation:

$$x^2 y'' + xy' + (x^2 - \nu^2) y = 0 \qquad \nu \in \mathbb{N}$$

results from separation of variables for 2D problems in cylindrical coordinates, and it is known as the ***Bessel equation*** of order ν. Clearly, $x = 0$ is a regular singular point. We shall discuss the case $\nu = 0$, i.e., the equation:

$$xy'' + y' + xy = 0\,.$$

We have,

$$y = \sum_{k=0}^{\infty} c_k x^{\alpha+k} \qquad xy = \sum_{k=0}^{\infty} c_k x^{\alpha+k+1} = \sum_{k=2}^{\infty} c_{k-2} x^{\alpha+k-1}$$

$$y' = \sum_{k=0}^{\infty} (\alpha + k) c_k x^{\alpha+k-1}$$

$$y'' = \sum_{k=0}^{\infty} (\alpha + k)(\alpha + k - 1) c_k x^{\alpha+k-2} \qquad xy'' = \sum_{k=0}^{\infty} (\alpha + k)(\alpha + k - 1) c_k x^{\alpha+k-1}$$

provided all exponents: $(\alpha + k) \neq 0$. Substituting into the equation we get,

$$\sum_{k=0}^{\infty} (\alpha + k)(\alpha + k - 1) c_k x^{\alpha+k-1} + \sum_{k=0}^{\infty} (\alpha + k) c_k x^{\alpha+k-1} + \sum_{k=2}^{\infty} c_{k-2} x^{\alpha+k-1} = 0$$

which results in the relations:

$$k = 0 \quad (\alpha(\alpha - 1) + \alpha) c_0 = 0 \qquad \Rightarrow \alpha = 0 \quad \text{or} \quad c_0 = 0$$

$$k = 1 \quad ((\alpha + 1)\alpha + \alpha + 1) c_1 = 0 \qquad \Rightarrow \alpha = -1 \quad \text{or} \quad c_1 = 0$$

$$k > 1 \quad ((\alpha + k)(\alpha + k - 1) + (\alpha + k)) c_k + c_{k-2} = 0 \qquad \Rightarrow c_k = -\tfrac{1}{(\alpha+k)^2} c_{k-2}\,.$$

If we choose $\alpha = 0$ to satisfy the first equation, we obtain $c_1 = 0$ from the second equation. The third equation generates then a recursion formula that defines a function y_1 with only even order terms:

$$y_1 = c_0 (1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 4^2} - \ldots)\,.$$

For $c_0 = 1$, this defines ***Bessel function of the first kind and order (ν) zero***, denoted $J_0(x)$.

Choosing $c_0 = 0$ and $\alpha = -1$ leads to the same solution. Indeed, we have:

$$y = \sum_{k=1}^{\infty} c_k x^{k-1} = \sum_{k=0}^{\infty} c_{k+1} x^k$$

which, modulo renaming the coefficients, is exactly the previous case for $\alpha = 0$ that we have already examined. We have thus a situation where the Frobenius method delivers just one solution.

Construction of a second linearly independent solution is done by using the variation of constant method, seeking

$$y = A(x) J_0(x) \,.$$

This leads to the equation for $A(x)$,

$$xA'' J_0 + A'(2xJ_0' + J_0) = 0 \,.$$

Substituting $B = A'$, we can use separation of variables (a non-trivial case) leading to a semi-closed formula for A:

$$B = \frac{1}{xJ_0^2} \quad \Rightarrow \quad A = \int \frac{dx}{xJ_0^2} \,.$$

For the rest of the exposition, we refer to [3]. After a couple of extra steps, we arrive at the definition of the *Neumann function of order zero*, denoted $N_0(x)$. This is not the end of the story. In order to secure a particular behavior at infinity, we define the *Bessel function of the second kind and order zero* as:

$$Y_0(x) := \frac{2}{\pi}(N_o(x) + (\gamma - \ln 2) J_0(x))$$

where $\gamma \approx= 0.577$ is Euler's constant. Ufff.....

The things to remember are:

- J_0 is analytic whereas Y_0 is singular at $x = 0$.
- At infinity J_0 and Y_0 behave as the cosine and sine functions. More precisely,

$$J_0(x) \sim \frac{\cos(x - \frac{\pi}{4})}{\sqrt{\frac{\pi x}{2}}} \qquad Y_0(x) \sim \frac{\sin(x - \frac{\pi}{4})}{\sqrt{\frac{\pi x}{2}}} \,.$$

  Note the decay rate $x^{-1/2}$ at infinity.

The asymptotic behavior at infinity leads to the definition of the corresponding *Hankel functions of the first and second kind of order zero*:

$$H_0^{(1)}(x) = J_0(x) + iY_0(x) \qquad H_0^{(2)}(x) = J_0(x) - iY_0(x)$$

that exhibit now wave-like behavior at infinity,

$$H_0^{(1)}(x) \sim \frac{e^{i(x - \frac{\pi}{4})}}{\sqrt{\frac{\pi x}{2}}} \qquad H_0^{(2)}(x) \sim \frac{e^{-i(x - \frac{\pi}{4})}}{\sqrt{\frac{\pi x}{2}}} \,.$$

In a similar way we construct Bessel and Hankel functions of arbitrary order ν.

□

Example 4.2.6 Legendre functions and polynomials

We ran into the Legendre functions when studying the eigenvalue problem for the Legendre operator:

$$-[(1-x^2)y']' = \lambda y, \qquad x \in I =: (-1,1)\,. \tag{4.3}$$

The operator is self-adjoint in $L^2(I)$. Due to vanishing of the leading coefficient $(1-x^2)$ at the end-points of the interval, no BCs are imposed, just the energy condition: $y, Ay \in L^2(I)$. The operator is also positive semi-definite, so $\lambda \in [0,\infty)$. We begin by rewriting the Legendre equation (4.3) in the form more suitable for the Frobenius method,

$$(1-x^2)y'' - 2xy' + \lambda y = 0\,.$$

Note that any $x = \pm 1$ are regular singular points, and any other point $x_0 \in I$ is a regular point. We will expand the solution around the regular point $x = 0$.

$$\begin{aligned}
y &= \sum_{k=0}^{\infty} c_k x^k & \lambda y &= \sum_{k=0}^{\infty} \lambda c_k x^k \\
y' &= \sum_{k=1}^{\infty} k c_k x^{k-1} & -2xy' &= -2\sum_{k=1}^{\infty} k c_k x^k \\
y'' &= \sum_{k=2}^{\infty} k(k-1) c_k x^{k-2} = \sum_{k=0}^{\infty} (k+2)(k+1) c_{k+2} x^k & -x^2 y'' &= -\sum_{k=2}^{\infty} k(k-1) c_k x^k\,.
\end{aligned}$$

Substituting into the equation, we get the relations:

$$\begin{aligned}
k=0 \quad & 2c_2 + \lambda c_0 = 0 && \Rightarrow c_2 = -\tfrac{\lambda}{2} c_0 \\
k=1 \quad & 6c_3 + (\lambda - 2)c_1 = 0 && \Rightarrow c_3 = -\tfrac{\lambda-2}{6} c_1 \\
k>1 \quad & (k+2)(k+1)c_{k+2} + [\lambda - k(k+1)]c_k = 0 && \Rightarrow c_{k+2} = -\tfrac{\lambda - k(k+1)}{(k+2)(k+1)} c_k\,.
\end{aligned}$$

We obtain two solutions y_1, y_2 corresponding to pairs $c_0 = 1, c_1 = 0$ and $c_0 = 0, c_1 = 1$,

$$\begin{aligned}
y_1 &= c_0 + c_2 x^2 + c_4 x^4 + \dots \\
y_2 &= c_1 x + c_3 x^3 + c_5 x^5 + \dots
\end{aligned}$$

Notice that if $\lambda = \lambda_n := n(n+1)$, $n = 1, 2, \dots$, one of the series will terminate. If n is even, the y_1 series at some point terminates, and y_1 is simply a polynomial. Similarly, if n is odd, the y_2 series terminates. Thus, for any natural number n, we have two solutions: a polynomial solution $P_n(x)$ and a second solution $Q_n(x)$ represented with an infinite series. Functions $P_n(x)$ are the *Legendre polynomials* or *Legendre functions of the first kind and degree* n, functions $Q_n(x)$ are *Legendre functions of the second kind and degree* n. For any polynomial y, Ay is a polynomial as well and, therefore, (trivially) both $y, Ay \in L^2(I)$. The Legendre polynomials are thus eigenvectors of the Legendre

operator. It is much less simple to show that values λ_n form the whole spectrum of the Legendre operator and, consequently, the (normalized) Legendre polynomials form an orthonormal basis for $L^2(I)$.

Note finally that the Legendre polynomials can be computed by using the simple recursion in k. They are indispensable in the construction of higher order shape functions in the Finite Element (FE) method.

□

Exercises

Exercise 4.2.1 Use separation of variables to find a general solution:

$$(a) \qquad \frac{dy}{dx} = \frac{ky \ln x}{x} \qquad k > 0$$

$$(b) \qquad \frac{dy}{dx} = (y^3 - y^2)e^x \,.$$

(2 points)

Exercise 4.2.2 Find a general solution.

$$y' + \frac{y}{x} = x^2 \,.$$

(5 points)

Exercise 4.2.3 Find a general solution for the equations below. If necessary, use an integrating factor.

$$(a) \quad (x + 2y)\,dx + (y + 2x)\,dy = 0$$

$$(b) \quad 3y\,dx + dy = 0 \,.$$

(10 points)

Exercise 4.2.4 Solve by any means. Consult other sources, if necessary.

$$(a) \qquad xy' + 2y = 4x^2 \,, \quad y(0) = 0$$

$$(b) \qquad y' = -xy + y^{1/2} \,, \quad y(0) = 0$$

$$(c) \qquad y' = -xy + y^{1/2} \,, \quad y(0) = 1$$

$$(d) \qquad y = \ln y' \,, \qquad\qquad y(2) = 0 \,.$$

(20 points)

Exercise 4.2.5 Determine general solutions for the equations:

$$(a) \qquad y'' + xy' - y = 0$$

$$(b) \qquad x^3y'' + xy' - y = 0 \,.$$

Hint: Use experience from (a) to solve (b). (5 points)

Exercise 4.2.6 Obtain the general solution in terms of elementary or Bessel functions:

$$\begin{array}{ll} (a) & xy'' + y' + kxy = 0 \\ (b) & x^2y'' + xy' + (k^2x^2 - \frac{1}{9})y = 0 \\ (c) & y'' + x^4y = 0 \end{array}$$

(10 points)

4.3 Phase Portraits and Lyapunov Stability

In this section we are interested in studying a second order, possibly nonlinear, *autonomous* equation:

$$\ddot{x} = f(x, \dot{x})\,.$$

The equation is called *autonomous* if the flux function does not depend explicitly upon t. The interest in the equation originates from studying the motion of a single particle under action of various force fields. The equation can always be represented as a first order system for the position $x(t)$ and velocity $y(t) = \dot{x}(t)$,

$$\dot{x} = y \quad \dot{y} = f(x, y)\,.$$

Instead of studying dependence of x and y in time t, we will pay more attention to the particle *trajectories* in the x, y plane. For instance, consider a simple harmonic oscillator,

$$m\ddot{x} + kx = 0$$

or,

$$\begin{cases} \dfrac{dx}{dt} = y \\ \dfrac{dy}{dt} = -\dfrac{kx}{m}\,. \end{cases}$$

Eliminating formally dt, we obtain the equation:

$$my\,dy + kx\,dx = 0$$

which represents an exact differential,

$$\frac{m}{2}y^2 + \frac{k}{2}x^2 = C$$

where C is an integration constant. The equation above represents the *conservation of total energy*, the first terms represents the kinetic energy and the second part represents the potential energy of the spring. The trajectory in the phase plane x, y is an ellipse. We call it the *phase portrait* of the solution.

We will study a slightly more general system of equations:

$$\begin{cases} \dot{x} = P(x, y) \\ \dot{y} = Q(x, y)\,. \end{cases} \tag{4.4}$$

Eliminating dt, we can rewrite the system as a single equation:

$$\frac{dy}{dx} = \frac{Q(x,y)}{P(x,y)} \quad \text{or} \quad \frac{dx}{dy} = \frac{P(x,y)}{Q(x,y)}\,.$$

If $P(x_0, y_0) \neq 0$ then we can solve for function $y = y(x)$ in a vicinity of point (x_0, y_0). Similarly, if $Q(x_0, y_0) \neq 0$, we can determine $x = x(y)$ in a vicinity of point $(x_0.y_0)$. The interesting scenario is when both $P(x_0, y_0) = Q(x_0, y_0) = 0$, i.e. (x_0, y_0) is an *equilibrium point* called frequently also a *singular point*.. We shall study now the phase portraits in a vicinity of *equilibrium points* and the stability of the corresponding solutions.

Stability in the sense of Lyapunov. We say that solution $(x_0(t), y_0(t))$ of system (6.6) is *stable in the sense of Lyapunov* if, for every $\epsilon > 0$ and time t_0, there exists $\delta = \delta(\epsilon, t_0)$ such that, for any solution $(x(t), y(t))$ originating from a δ-neighborhood of $(x_0(t_0), y_0(t_0))$ at t_0, the solution stays within the ϵ-neighborhood of $(x_0(t), y_0(t))$, *for all times* $t > t_0$. In other words,

$$d((x(t_0), y(t_0)), (x_0(t_0), y_0(t_0))) < \delta \qquad \Rightarrow \qquad d((x(t), y(t)), (x_0(t), y_0(t))) < \epsilon \quad \forall\, t > t_0\,.$$

Otherwise, the solution is said to be *unstable*. If, *additionally*, $d((x(t), y(t)), (x_0(t), y_0(t))) \to 0$ as $t \to \infty$, the solution is said to be *asymptotically stable*. Note that, contrary to other terminologies, the asymptotical stability is a stronger condition than just stability.

Example 4.3.1

Consider the simplest linear differential equation,

$$\dot{x} = \lambda x$$

with the solution:

$$x = Ce^{\lambda t}\,.$$

For $\lambda < 0$ the solution is asymptotically stable, for $\lambda > 0$ is unstable, and for $\lambda = 0$, it is stable but not asymptotically stable. □

Phase portraits for a linear system. We will study first the linear system with constant coefficients:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}\,.$$

If determinant $ad - bc = 0$ then the system has infinitely many *non-isolated* equilibrium points, the case of no interest, so we assume: $ad - bc \neq 0$. There exists then a unique equilibrium point x_0, y_0. By switching to new variables: $x - x_0, y - y_0$, we can reduce our study to the case when $e = f = 0$, i.e., the equilibrium point is at the origin, $x = y = 0$. The characteristic equation for the coefficient matrix is:

$$\lambda^2 - (a + d)\lambda + ad - bc = 0$$

with $\Delta = (a - d)^2 + 4bc$. We have three possible cases.

- Case: $\Delta > 0$. We have two distinct non-zero real eigenvalues $\lambda \neq \mu$ and, therefore, two linearly independent eigenvectors. In the eigensystem of coordinates (not necessarily orthogonal), the system of ODEs reduces to

$$\dot{\xi} = \lambda\xi, \qquad \dot{\eta} = \mu\eta\,.$$

which leads to the solution $\xi = C_1 e^{\lambda t}, \eta = C_2 e^{\mu t}$. Clearly, the equilibrium point will be stable if both $\lambda, \mu < 0$, otherwise it is unstable. Eliminating t, we obtain the equation:

$$(C_1^{-1}\xi)^\mu = (C_2^{-1}\eta)^\lambda \quad \text{or} \quad \eta = C\xi^{\frac{\mu}{\lambda}}\,.$$

If $\mu/\lambda > 0$ we obtain the so-called *node*, otherwise we have the *saddle*. Note that the saddle is always unstable (one of the eigenvalues must be negative), whereas the node may be stable or unstable. See Fig.4.1 for an illustration.

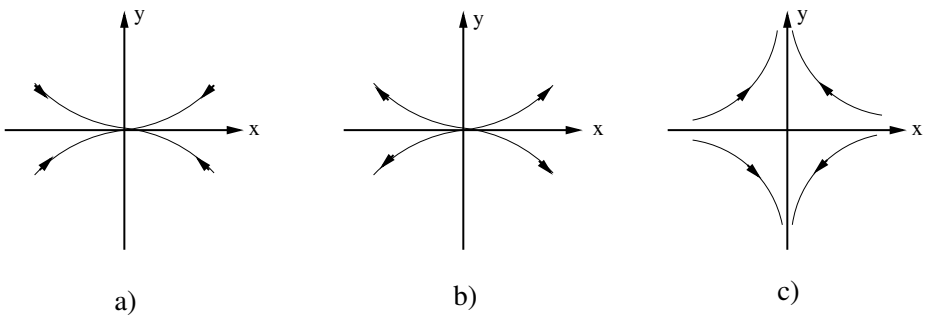

Figure 4.1
Phase portraits: a) a stable node, b) an unstable node, c) an (always) unstable saddle.

- Case: $\Delta < 0$. We have two distinct, complex conjugate eigenvalues $\lambda + i\mu$ and $\lambda - i\mu$ with corresponding eigenvectors $\boldsymbol{\alpha} + i\boldsymbol{\beta}$ and $\boldsymbol{\alpha} - i\boldsymbol{\beta}$. In the complex eigensystem of coordinates the solution is:

$$\xi = e^{(\lambda+i\mu)t} = e^{\lambda t}(\cos\mu t + \sin\mu t) \qquad \eta = e^{(\lambda-i\mu)t} = e^{\lambda t}(\cos\mu t - \sin\mu t)\,.$$

The general solution is:

$$\begin{aligned}\boldsymbol{x} &= C_1 e^{(\lambda+i\mu)t}(\boldsymbol{\alpha} + i\boldsymbol{\beta}) + C_2 e^{(\lambda-i\mu)t}(\boldsymbol{\alpha} - i\boldsymbol{\beta}) \\ &= e^{\lambda t}[\underbrace{(C_1 + C_2)}_{=:D_1}(\boldsymbol{\alpha}\cos\mu t - \boldsymbol{\beta}\sin\mu t) + \underbrace{i(C_1 - C_2)}_{=:D_2}(\boldsymbol{\beta}\cos\mu t + \boldsymbol{\alpha}\sin\mu t)] \\ &= e^{\lambda t}[(D_1\boldsymbol{\alpha} + D_2\boldsymbol{\beta})\cos\mu t + (-D_1\boldsymbol{\beta} + D_2\boldsymbol{\alpha})\sin\mu t]\,.\end{aligned}$$

For $\lambda = 0$, this equation represents a (possibly skewed) ellipse and we classify the equilibrium point as the *center*. Indeed, $\boldsymbol{a}_1 := D_1\boldsymbol{\alpha} + D_2\boldsymbol{\beta}$ and $\boldsymbol{a}_2 := -D_1\boldsymbol{\beta} + D_2\boldsymbol{\alpha}$ represent two linearly independent vectors, and the solution is then:

$$\boldsymbol{x} = \cos\mu t\boldsymbol{a}_1 + \sin\mu t\boldsymbol{a}_2\,.$$

Therefore, in contravariant affine coordinates, we have a circle or, more precisely, an ellipse, since $\boldsymbol{a}_1, \boldsymbol{a}_2$ may have different lengths. The center is always stable but not asymptotically stable. For $\lambda \neq 0$, we classify the equilibrium point as the *focus*. Dependent upon the sign of λ, the focus may be stable or unstable. See Fig.4.2 for an illustration.

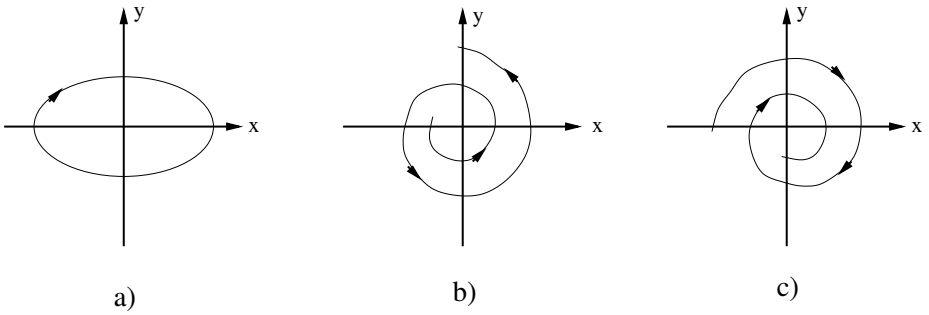

Figure 4.2
Phase portraits: a) an (always) stable center, b) an unstable focus, c) a stable focus.

- Case: $\Delta = 0$. We have a double real eigenvalue λ. If the system admits two linearly independent eigenvectors, the reasoning is exactly the same as in the first case with $\lambda = \mu$. We obtain a *node* that may be stable or not, dependent upon the sign of λ. The more interesting case is when we have only one eigenvector. Recalling the Jordan Theorem, we employ the corresponding generalized eigenvector to form an eigensystem in which the system matrix takes the form:

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

which leads to the solution:

$$\xi = (C_1 t + C_2) e^{\lambda t} \qquad \eta = C_1 e^{\lambda t} .$$

Eliminating time t, we obtain the equation:

$$\xi = (C_1 \lambda^{-1} \ln(C_2^{-1}\eta) + C_2) C_2^{-1} \eta$$

which is classified again as the *node* as the logarithmic term $\ln \eta$ is dominated by linear term η, as $\eta \to 0$. Dependent upon the sign of λ, the node may be stable or unstable.

Non-linear case. Without losing generality, assume again that the equilibrium point coincides with the origin, i.e. $P(0,0) = Q(0,0) = 0$. We compute then the differential of the vector-valued function (P, Q) at $(0,0)$ and consider the corresponding linearized problem. We have the following fundamental result.

THEOREM 4.3.1 (Poincaré)

The singularities of the linearized system are identical with the singularities of the original system except for the case of the center, i.e., when $\Delta < 0$ and $\lambda = 0$ (or, equivalently, $a+d=0$). Dependent upon the higher order terms, we may have then (still) a center or a focus.

The practical moral of the story is that in the case of the center, we have to continue with the analysis of the nonlinear problem.

Lyapunov's Function. Assume again that the system

$$\dot{x} = P(x,y) \qquad \dot{y} = Q(x,y)$$

has an isolated singularity at $(0,0)$. Function $V(x,y)$, defined in a neighborhood of $(0,0)$ is called a *Lyapunov function* if

- $V(0,0) = 0$ but $V > 0$ for $(x,y) \neq (0,0)$.
- $$\frac{\partial V}{\partial x}P(x,y) + \frac{\partial V}{\partial y}Q(x,y) \leq 0\,. \tag{4.5}$$

Note that the expression above represents the time derivative of $V(x,y)$. Indeed,

$$\frac{d}{dt}(V(x(t),y(t)) = \frac{\partial V}{\partial x}\dot{x} + \frac{\partial V}{\partial y}\dot{y} = \frac{\partial V}{\partial x}P(x,y) + \frac{\partial V}{\partial y}Q(x,y)\,.$$

THEOREM 4.3.2 (Lyapunov's First Theorem)

If there exists a Lyapunov function then the equilibrium point is stable. Additionally, if function $V(x,y)$ satisfies a strict inequality in (4.5) outside of the origin, i.e.

$$\frac{\partial V}{\partial x}P(x,y) + \frac{\partial V}{\partial y}Q(x,y) < 0 \qquad \text{for } (x,y) \neq (0,0)$$

then the equilibrium point is asymptotically stable.

We also have a negative result.

THEOREM 4.3.3 (Lyapunov's Second Theorem)

If there exists a C^1 function $W(x,y)$ defined in a neighborhood of the origin such that,

- $W(0,0) = 0$.
- *For each neighborhood $B(\mathbf{0},r)$ of zero, there exists a point $(x,y) \in B(\mathbf{0},r)$ where $W(x,y) > 0$.*

- $$\frac{\partial W}{\partial x}P(x,y) + \frac{\partial W}{\partial y}Q(x,y) > 0 \qquad \text{for } (x,y) \neq (0,0)$$

  in some neighborhood of $(0,0)$.

then the equilibrium point is unstable.

Example 4.3.2

Consider motion of a point in a conservative force field $f(x)$ with the corresponding potential (energy) $W(x)$,

$$\ddot{x} = f(x) \qquad f(x) = -W'(x)\,.$$

Without losing generality (scalar potentials are determined up to a constant), assume $W(0) = 0$. Assume also that $W > 0$ outside of the origin. The corresponding first order system with $y = \dot{x}$ is:

$$\dot{x} = y \quad \dot{y} = f(x)\,.$$

Function

$$V(x,y) = W(x) + \frac{1}{2}y^2$$

representing the total energy, is a Lyapunov function. Indeed,

$$\frac{\partial V}{\partial x}P(x,y) + \frac{\partial V}{\partial y}Q(x,y) = W'(x)y + yf(x) = 0\,,$$

The equilibrium point is thus stable. The condition above represents simply the conservation of total energy. □

Exercises

Exercise 4.3.1 Sketch the phase portrait for each of the equations below and the equilibrium points.

(a) $\ddot{x} + \frac{g}{l}\sin x = 0$

(b) $\ddot{x} + x^2 = 0$

(c) $\ddot{x} + \dot{x} + x - x^3 = 0$

(d) $\ddot{x} + c\dot{x} + \frac{g}{l}\sin x = 0$

(20 points)

Exercise 4.3.2 Locate and classify all singularities of the system below.

$$\dot{x} = x + y - 1\,, \quad \dot{y} = y - x^2 + 1\,.$$

(10 points)

Exercise 4.3.3 Consider the Volterra problem:

$$\dot{x} = (\alpha - \gamma y)x\,, \quad \dot{y} = -(\beta - \delta x)y$$

where α, γ, β and δ are positive constants. Find and classify the singular points, show that the axes coincide with trajectories, and sketch the phase portrait for $x, y \geq 0$.

(10 points)

Exercise 4.3.4 Show that $V = x^2 + y^2$ is a suitable Lyapunov function for the equation:

$$\ddot{x} + \epsilon\dot{x}^3 + x = 0 \qquad \text{where } \epsilon > 0$$

and use it to draw conclusions about the stability of the equilibrium point at the origin.

(5 points)

Exercise 4.3.5 Seek V or W of the form ax^2+by^2 and draw conclusions about the stability of the equilibrium point at the origin.

$$\dot{x} = x^3 + y^3, \quad \dot{y} = xy^2 - 2x^2y - y^3\,.$$

(10 points)

5

Elements of Theory of Hilbert Spaces

5.1 Preliminaries

Hilbert space. Any vector space equipped with an inner product (an inner product or pre-Hilbert space) that generates a norm and, in turn, a metric that is complete, is called a *Hilbert space*. Every Hilbert space is automatically a Banach space (a complete normed space). Recall elementary examples of Banach spaces: $l^p, L^p(\Omega)$, $p \in [1, \infty]$. For $p = 2$, we have Hilbert spaces. A common example is also a *weighted L^2_w-space* with the inner product:

$$(u, v) = \int_\Omega w(x) u(x) \overline{v(x)}\, dx$$

where $w > 0$ is the weight. We arrive naturally at weighted space in curvilinear coordinates, e.g., cylindrical or spherical coordinates.

Continuous (bounded) maps. Let X, Y be two Banach spaces. A linear map $A \in L(X, Y)$ is continuous iff it is bounded.

THEOREM 5.1.1

Let X, Y be two normed spaces, and let $A : X \to Y$ be a linear map. The following conditions are equivalent to each other.

(i) A is continuous.

(ii) A is continuous at 0.

(iii) A is bounded, i.e., there exists a constant $C > 0$ such that

$$\|Ax\|_Y \le C\|x\|_X \qquad \forall\, x \in X\,.$$

PROOF (i) $\Rightarrow$ (ii) obvious.
(ii) $\Rightarrow$ (iii). Continuity at 0 is equivalent to the $\epsilon - \delta$ condition:

$$\forall\, \epsilon \quad \exists\, \delta \quad : \|x\|_X \le \delta \quad \Rightarrow \quad \|Ax\|_Y \le \epsilon\,.$$

Choose $\epsilon = 1$. We have, for every $x \neq 0$,

$$\|\delta \frac{x}{\|x\|_X}\|_X = \delta$$

and, therefore,

$$\|A\,\delta \frac{x}{\|x\|_X}\|_Y \leq 1\,.$$

But this is equivalent to:

$$\|Ax\|_Y \leq \frac{1}{\delta}\,\|x\|_X\,.$$

(iii) $\Rightarrow$ (i). We have

$$\|A(x_n - x)\|_Y \leq C\|x_n - x\|_X\,,$$

Consequently, if $x_n \to x$, i.e., $\|x_n - x\|_X \to 0$, then $Ax_n \to Ax$. ∎

Space $\mathcal{L}(X, Y)$. The linear and continuous maps from a normed space X into a normed space Y, form a subspace of $L(X, Y)$, denoted $\mathcal{L}(X, Y)$. Most of the time, we equip $\mathcal{L}(X, Y)$ with the norm induced by norms in X and Y.

THEOREM 5.1.2

The following expression(s) define a norm in $\mathcal{L}(X, Y)$.

$$\|A\|_{\mathcal{L}(X,Y)} := \sup_{x\neq 0} \frac{\|Ax\|_Y}{\|x\|_X} = \sup_{\|x\|_X \leq 1} \|Ax\|_Y = \sup_{\|x\|_X = 1} \|Ax\|_Y = \inf\{M\,:\,\|Ax\|_Y \leq M\|x\|_X, \quad x \in X\}\,.$$

We leave the proof for Exercise 5.1.1. The theorem implies immediately the useful inequality:

$$\|Ax\| \leq \|A\|_{\mathcal{L}(X,Y)}\,\|x\|_X\,.$$

Recall that, if Y is complete then so is $\mathcal{L}(X, Y)$ (X needs not be complete), see [5], Section 4.8.

Unitary map (isometry). Let $T : X \to Y$ be a linear map from a Hilbert space X into a Hilbert space Y. The following conditions are equivalent to each other, see Exercise 5.1.2.

- Map T preserves inner product, i.e.,

$$(Tx_1, Tx_2)_Y = (x_1, x_2)_X \qquad x_1, x_2 \in X\,. \tag{5.1}$$

- Map T is an isometry, i.e., it preserves the norm,

$$\|Tx\|_Y = \|x\|_X \qquad x \in X\,. \tag{5.2}$$

Such a map is called a *unitary map*. Recall that every isometry is injective.

Topological dual. The space of all linear (antilinear) functionals defined on X that are continuous, is identified as the *topological dual* of space X and denoted by X'. By construction, the topological dual is a subspace of the algebraic dual, $X' \subset X^*$ and it is always complete since $\mathbb{R}$ amd $\mathbb{C}$ are complete. The topological dual is a Banach space. The Riesz Theorem discussed in Section 5.2 demonstrates that, for a Hilbert space X, its dual X' is also a Hilbert space.

Let $p \in [1, \infty]$, and let $u \in L^q(\Omega)$ where $1/p + 1/q = 1$. Hölder inequality implies that function u generates a (anti)linear and continuous functional on $L^p(\Omega)$,

$$|\int_\Omega u\bar{v}| \leq \|u\|_{L^q(\Omega)} \, \|v\|_{L^p(\Omega)} \, .$$

One can show (non-trivial) that the map:

$$R \, : \, L^q(\Omega) \ni u \to Ru := \{L^p(\Omega) \ni v \to \int_\Omega u\bar{v} \in \mathbb{R}(\mathbb{C})\} \in (L^p(\Omega)'$$

is a unitary map.

THEOREM 5.1.3 (Representation Theorem for Duals to $L^p(\Omega)$)

Let $p \in [1, \infty)$. Then map R above is a surjection, i.e., for each $f \in (L^p(\Omega))'$ there exists a unique $u_f \in L^q(\Omega)$ such that

$$\int_\Omega u_f \bar{v} = f(v) \qquad v \in L^p(\Omega) \, .$$

In other words, R is an isometric isomorphism.

See [5], Section 5.12, for the proof. The same result holds for spaces l^p.

Topological transpose. Restriction of the *transpose operator* A^T to the topological dual is identified as the *topological transpose* and denoted by A',

$$A' = A^T|_{X'}, \qquad A'x' = x' \circ A \, .$$

Note that the topological transpose is well-defined as the composition of two continuous functions is continuous.

Orthogonality. Orthogonal complement. Vectors $x, y \in X$ are *orthogonal* if $(x, y)_X = 0$. Let $M \subset X$ be a subspace. In the same way as on the purely algebraic level, we define the concept of the *orthogonal complement*,

$$M^\perp := \{x \in X \, : \, (x, y) = 0 \quad \forall \, y \in M\} \, .$$

It is easy to show that $M^\perp$ is always closed.

THEOREM 5.1.4 (Orthogonal Decomposition Theorem)

Let X be a Hilbert space, and $M \subset X$ a closed subspace of X. Then

$$X = M \oplus M^{\perp}$$

i.e., each $x \in X$ can be uniquely decomposed into orthogonal components $m \in M$, $n \in M^{\perp}$,

$$x = m + n, \quad m \in M,\, n \in M^{\perp}\,.$$

See [5], Section 6.2, for the proof.

Orthogonal projection. Let $M \subset X$ be a closed subspace of X, and $x = m + n$ the orthogonal decomposition of x. The linear projection $P_M : X \to M, Px := m$, in the direction of orthogonal complement $M^{\perp}$ is called the *orthogonal projection* of X onto M. It follows from the Pythagoras Theorem,

$$\|x\|^2 = \|m+n\|^2 = (m+n, m+n) = \|m\|^2 + \|n\|^2$$

that

$$\|m\| = \|P_M x\| \le \|x\| \quad \Rightarrow \quad \|P_M\| \le 1\,.$$

But, at the same time, $P_M m = m$, $m \in M$. Consequently, $\|P_M\| = 1$. Finally, observe that $m = P_M x$ realizes the distance between x and M, i.e., it represents the closest element in M to x,

$$\|x - P_M x\| = \|x - m\| = \min_{m_1 \in M} \|x - m_1\|\,.$$

Indeed,

$$\|x - m_1\|^2 = \| \underbrace{x - m}_{=n \in M^{\perp}} + \underbrace{m - m_1}_{\in M} \|^2 = \|x-m\|^2 + \|m - m_1\|^2 \ge \|x-m\|^2\,.$$

This leads to the variational characterization of $m = P_M x$, comp. Exercise 5.1.3,

$$\begin{cases} m \in M \\ (x - m, \delta m) = 0 \quad \forall\, \delta m \in M\,. \end{cases} \tag{5.3}$$

Orthonormal bases. A basis in X whose elements are orthogonal to each other, is called an *orthogonal basis.* It is easy to show that orthogonal vectors must be linearly independent. Consequently, any maximal subset $\mathcal{B}$ of orthogonal vectors in X,

$$(x, y) = 0 \quad \forall\, y \in \mathcal{B} \quad \Rightarrow \quad x = 0\,.$$

represents an orthogonal basis in X. If the elements of the basis have been normalized, we call it an *orthonormal basis*. We are interested only in spaces which have *countable* orthonormal bases. This is the case for *separable* * Hilbert spaces including spaces $L^2(\Omega), \ell^2, L^2_w(\Omega)$.

*A normed space is *separable* if it has a countable dense subset.

Let M denote the linear span of vectors $e_1, e_2, \ldots$ forming the basis,

$$M = \text{span } \{e_1, e_2, \ldots\} \, .$$

The definition of the orthonormal basis implies that the orthogonal complement of M reduces to the zero vector

$$M^{\perp} = \{0\}$$

which (see Exercise 5.1.4) implies that

$$\overline{M} = \left(M^{\perp}\right)^{\perp} = \{0\}^{\perp} = X \, .$$

Thus M is dense in the space X. Consequently, for any vector $x \in X$ there exists a sequence $x_n \in M$ converging to x, $x_n \to x$.

In particular, since any finite-dimensional space is automatically closed, we immediately see that the existence of a finite orthonormal basis implies that the space X is finite-dimensional. Orthonormal bases then constitute a special subclass of usual (Hamel) bases in a finite-dimensional Hilbert (Euclidean) space.

Let $X_n = \text{span } \{e_1, \ldots, e_n\}$ denote now the span of first n vectors from the basis and let P_n be the corresponding orthogonal projection onto X_n. We claim that

$$P_n x \to x, \quad \text{for every } x \in X \, .$$

Indeed, let $x_n \in M$ be a sequence converging to x. Pick an arbitrary $\varepsilon > 0$ and select an element $y \in M$ such that

$$\|y - x\| < \frac{\varepsilon}{2}$$

Let $N = N(y)$ be an index such that $y \in X_N$. We have then for every $n \geq N$,

$$\begin{aligned} \|P_n x - x\| &\leq \|P_n x - P_n y\| + \|P_n y - x\| \\ &\leq \|P_n\| \, \|x - y\| + \|y - x\| \\ &\leq \|x - y\| + \|y - x\| \leq 2\frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

since $\|P_n\| = 1$ and $P_n y = y$ for $n \geq N$.

Let $P_n x = \sum_{i=1}^{n} x_i e_i$. It follows from the orthogonality condition,

$$(P_n x - x, e_j) = 0 \quad j = 1, 2, \ldots, n \, ,$$

that $x_j = (x, e_j)$. In particular, components x_i are independent of n. Consequently,

$$\sum_{i=1}^{\infty} x_i e_i := \lim_{n \to \infty} \sum_{i=1}^{n} x_i e_i = \lim_{n \to \infty} P_n x = x$$

The coefficients x_i can be viewed as the *components* of x with respect to the orthonormal basis $\{e_i\}$.

Example 5.1.1

Vectors

$$e_k = \big(0, \ldots, 1_{(k)}, \ldots, 0\big)$$

form a (canonical) orthonormal basis in $\mathbb{C}^n$ with the canonical scalar product. □

Example 5.1.2

Vectors

$$e_k = \left(0, \ldots, \underset{(k)}{1}, \ldots, 0, \ldots\right)$$

form a (canonical) orthonormal basis in ℓ^2.

Indeed, let $x \in \ell^2$, $x = (x_1, x_2, x_3, \ldots)$. Then

$$(x, e_k) = x_k$$

and, therefore, trivially $(x, e_k) = 0$, $k = 1, 2, \ldots$ implies that $x = 0$. Also, since

$$x = \sum_{i=1}^{\infty} x_i e_i$$

numbers x_i are interpreted as components of x with respect to the canonical basis. □

Example 5.1.3

Functions

$$e_k(x) = e^{2\pi i k x}, \quad k = 0, \pm 1, \pm 2, \ldots$$

form an orthonormal basis in $L^2(0, 1)$. See [5], Example 6.3.3, for an elementary proof, and Example 3.3.7 for the connection with Sturm-Liouville theory. □

Because of this example, series

$$\sum_{i=1}^{\infty} (x, e_i) e_i \,,$$

for any orthonormal basis, is frequently called a *generalized Fourier series*.

THEOREM 5.1.5 (Characterization of orthonormal bases)

The following conditions are equivalent to each other.

(i) $\{e_n\}_{n=1}^{\infty}$ *is an orthonormal basis, i.e., it is a maximal set of orthonormal vectors.*

(ii) $x = \displaystyle\sum_{n=1}^{\infty} (x, e_n)\, e_n \qquad \forall\ x \in X.$

(iii) $(x,y) = \sum_{n=1}^{\infty} (x,e_n)\,\overline{(y,e_n)}$.

(iv) $\|x\|^2 = \sum_{n=1}^{\infty} |(x,e_n)|^2$.

PROOF (i)⇒(ii). This has already been shown.

(ii)⇒(iii). Let $x_N = \sum_{j=1}^{N}(x,e_j)e_j$, $y_N = \sum_{j=1}^{N}(y,e_j)e_j$. Use orthogonality of e_i to learn that

$$(x_N, y_N) = \sum_{i=1}^{N} x_i\overline{y}_i = \sum_{i=1}^{N}(x,e_i)\,\overline{(y,e_i)} \to \sum_{i=1}^{\infty}(x,e_i)\,\overline{(y,e_i)}$$

(iii)⇒(iv). Substitute $y = x$.

(iv)⇒(i). Suppose, to the contrary, the $\{e_1, e_2, \ldots\}$ can be extended with a vector $x \neq 0$ to a bigger orthonormal family. Then x is orthogonal with each e_i and, by property (iv), $\|x\| = 0$. So $x = 0$, a contradiction. ∎

Exercises

Exercise 5.1.1 Prove Theorem 5.1.2.

(5 points)

Exercise 5.1.2 Prove that conditions (5.1) and (5.2) are equivalent. *Hint:* Recall the *polarization formula*.

(5 points)

Exercise 5.1.3 Prove that $m = P_M x$, $m \in M$ iff m satisfies the variational problem (5.3).

(5 points)

Exercise 5.1.4 Let M be a subspace of a Hilbert space V. Prove that

$$\overline{M} = (M^{\perp})^{\perp}$$

(10 points)

5.2 Riesz Representation Theorem, Topological Transpose and Adjoint of a Continuous Operator

Representation Theorem 5.1.3 for $p = 2$ is a special case of the fundamental result of Riesz [†] valid for all Hilbert spaces.

Recall the definition of Riesz operator,

$$R_X : X \to X', \qquad \langle R_X x, y\rangle := (x, y)_X \,.$$

It follows from the Cauchy-Schwarz inequality and definition of inner product that R_X is an isometry (unitary map).

THEOREM 5.2.1 (Riesz Representation Theorem)

Riesz operator is a surjection, i.e., for every $f \in X'$, there exists a unique $x_f \in X$ such that

$$\langle f, y\rangle = (x_f, y)_X \qquad \forall\, y \in X \,,$$

Consequently, the Riesz operator provides a (canonical) isometric isomorphism between space X and its topological dual X'.

PROOF See [5], Theorem 6.4.1. ∎

Topological Transpose and Adjoint of a Continuous operator. We have already discussed these concepts in Section 3.3. Let $A : X \to Y$ be a continuous operator from a Hilbert space X to a Hilbert space Y. The algebraic transpose operator,

$$A^T : Y^* \to X^*, \quad A^T y^* = y^* \circ A$$

restricted to the topogical dual Y', is identified as the *topological transpose* of operator A, denoted $A' := A^T|_{Y'}$. Note that A' takes values in X' since the composition of two continuous maps is continuous. The *adjoint operator* $A^* : Y \to X$ is defined then exactly in the same way as in the purely algebraic case,

$$A^* = R_X^{-1} \circ A' \circ R_Y$$

where R_X, R_Y are the Riesz operators for spaces X and Y. Equivalently, we can define the adjoint operator by the identity:

$$(Ax, y)_Y = (x, A^* y)_X \qquad x \in X,\, y \in Y \,.$$

[†] Frigyes Riesz (1880 – 1956) was a Hungarian mathematician.

Finally, operator A is said to be *self-adjoint* if $X = Y$ and $A^* = A$. Note that we have considered the continuous operator to be defined on the *whole* space X. This is not a coincidence. If $D(A)$ were a proper subspace of X then, by continuity, A could be extended in a unique way to closure $\overline{D(A)}$. The *Orthogonal Decomposition Theorem* allows then to represent X as an orthogonal sum of $\overline{D(A)}$ and its orthogonal complement,

$$X = \overline{D(A)} \oplus \overline{D(A)}^{\perp},$$

and we can extend (for instance, by zero) the operator to the whole X preserving the continuity constant. Consequently, without loss of generality, one can consider a continuous operator to be defined always on the whole space X.

Example 5.2.1

Let $X = L^2(I), I = (0,1) \subset \mathbb{R}$. Consider the integral operator,

$$(Af)(x) := \int_I K(x,y) f(y)\, dy \qquad f \in L^2(I)$$

where kernel $K \in L^2(I^2)$. It is easy to show that the operator is well-defined and continuous, comp. Exercise 5.2.1.

Its adjoint is also an integral operator generated by the *adjoint kernel*:

$$K^*(x,y) := \overline{K(y,x)}.$$

The operator is thus self-adjoint iff $K^*(x,y) = K(x,y)$. □

Exercises

Exercise 5.2.1 Consider the integral operator from Example 5.2.1. Use Cauchy-Schwarz inequality to prove that the operator is well-defined (i.e., it takes $L^2(I)$ into itself) and continuous. Derive then the formula for its L^2-adjoint.

(5 points)

5.3 Variational Problems

This section provides a short introduction to the theory of variational problems laying down foundations for the Galerkin method and Finite Elements. We begin with variational problems equivalent to minimization problems and then move on to a more general class of variational problems that do not have a minimization

of energy principle behind them. We recite the Lax-Milgram and Babuška-Nečas theorems and outline the connection with Banach Closed Range Theorem.

5.3.1 Problems Stemming from Minimization

Let X be a real Hilbert 'energy space'. Consider a quadratic energy functional,

$$J(x) = \frac{1}{2}b(x,x) - l(x)$$

where $b(x,y)$ is a continuous (i.e. bounded) functional defined on $X \times X$, and $l \in X'$, i.e., it is a linear and continuous functional defined on X,

$$\begin{aligned} |b(x,y)| &\leq M\|x\|_X \, \|y\|_X \qquad x,y \in X \\ |l(x)| &\leq C\|x\|_X \qquad\qquad x \in X\,. \end{aligned} \tag{5.4}$$

Let $u_0 \in X$ be a given element, and Let $V \subset X$ be a *subspace* of X. Consider the abstract minimization problem:

$$\min_{w \in u_0 + V} J(w)\,. \tag{5.5}$$

Let $u \in u_0 + V$ be a solution to the minimization problem. Let $v \in V$ be an arbitrary test function. Define an auxiliary function:

$$f(\epsilon) = J(u + \epsilon v)\,.$$

If u minimizes $J(w)$ then f attains a minimum at $\epsilon = 0$ and, consequently, $f'(0) = 0$. The derivative $f'(0)$ equals the Gateaux derivative of functional J at u:

$$f'(0) = (d_u J)(v) = \langle d_u J, v\rangle = \frac{1}{2}(b(u,v) + b(v,u)) - l(v) = 0\,.$$

If we *assume additionally* that b is symmetric [‡], we conclude that the necessary condition for u to be a minimizer is that u satisfies the following *abstract variational problem*:

$$\begin{cases} u \in u_0 + V \\ b(u,v) = l(v) \qquad v \in V\,. \end{cases} \tag{5.6}$$

Under what additional conditions the minimization and variational problems are equivalent ?

Let u be a solution of the variational problem (5.6). Let $w = u + v$, $v \in V$ be an arbitrary element of affine space $u_0 + V$. We have:

$$J(w) = J(u+v) = \frac{1}{2}b(u+v,u+v) - l(u+v) = \underbrace{\frac{1}{2}b(u,u) - l(u)}_{=J(u)} + \underbrace{b(u,v) = l(v)}_{=0} + \frac{1}{2}b(v,v) = J(u) + \frac{1}{2}b(v,v)\,.$$

Consequently, if b is *positive-definite* on V, i.e.,

$$b(v,v) > 0 \quad \forall\, v \in V, v \neq 0\,, \tag{5.7}$$

[‡]Note that functional $J(w)$ sees only the symmetric part of $b(u,v)$ so we can replace an original $b(u,v)$ with its symmetric part.

u is *the unique* solution of the minimization problem (5.5). Summarizing, if $b(x,y)$ is symmetric and positive-definite, the minimization and variational problems are equivalent. The equivalence of the two problems does not prove that either one of them is well-posed, i.e. the solution exists and it is unique. To prove this, we upgrade condition (5.7) to a stronger, *V-coercivity condition:*

$$b(v,v) \geq \alpha \|v\|_X^2 \quad v \in V\,. \tag{5.8}$$

With continuity condition $(5.4)_1$, and the coercivity condition, $b(u,v)$ can be identified as an *equivalent inner product on* V. Representing $u = u_0 + w$, $w \in V$, we can rephrase the variational problem as:

$$\begin{cases} w \in V \\ b(w,v) = \underbrace{l(v) - b(u_0,v)}_{=:l_{\mathrm{mod}}(v)} \quad v \in V \end{cases} \tag{5.9}$$

where the *modified load functional* l_{mod} is continuous. The well-posedness of the variational problem follows now directly from the *Riesz Representation Theorem.*

REMARK 5.3.1 If we replace V in the minimization problem with a finite-dimensional subspace $V_h \subset V$, we obtain the *Ritz method*. If we do the same in the variational problem, we obtain the *(Bubnov -) Galerkin method*. In either case, we obtain the same *approximate solution* $u_h \in u_0 + V_h$, see Exercise 5.3.3. ∎

Example 5.3.1

Consider the elastic bar problem shown in Fig. 5.1. The elastic energy of the bar is given by $\frac{1}{2}b(u,u)$ where

$$b(u,v) = \int_0^l EAu'v'\,dx\,.$$

The load is:

$$l(v) = \int_0^l \rho Agv\,dx + Fv(l)\,,$$

and the functional $J(u) = \frac{1}{2}b(u,u) - l(u)$ represents the *total potential energy* of the bar. The energy space is $H^1(0,l)$, $u_0 = 0$, and the test space is

$$V := \{v \in H^1(0,l) \,:\, v(0) = 0\}\,.$$

Continuity of the bilinear form follows immediately from the Cauchy-Schwarz inequality,

$$|b(u,v)| = |\int_0^l EAu'v'\,dx| \leq EA\left(\int_0^l (u')^2\,dx\right)^{1/2}\left(\int_0^l (v')^2\,dx\right)^{1/2} \leq EA\|u\|_{H^1(0,l)}\,\|v\|_{H^1(0,l)}\,.$$

The continuity constant $M \leq EA$.

In order to show the continuity of the linear form, we need first to establish the embedding[§]:

$$H^1(0,l) \hookrightarrow C([0,l])\,. \tag{5.10}$$

Let $u \in C^\infty([0,l])$, and $\bar{u} = \frac{1}{l}\int_0^l u(x)\,dx$ be the average value of u. Cauchy-Schwarz implies:

$$\frac{1}{l}|\int_0^l u(x)\,dx| = \frac{1}{l}|\int_0^l 1\,u(x)\,dx| \leq \frac{1}{l}\left(\int_0^l 1\,dx\right)^{1/2}\left(\int_0^l u^2\,dx\right)^{1/2} \leq \frac{1}{\sqrt{l}}\|u\|_{H^1(0,l)}\,.$$

The function $u-\bar{u}$, as a function with zero average, must vanish at some point $x_0 \in [0,l]$. Consequently,

$$|u(x)-\bar{u}| = |\int_{x_0}^x (u-\bar{u})'(s)\,ds| = |\int_{x_0}^x u'(s)\,ds| \leq \left(\int_{x_0}^x ds\right)^{1/2}\left(\int_{x_0}^x (u')^2\,ds\right)^{1/2} \leq \sqrt{l}\|u\,\|u\|_{H^1(0,l)}\,.$$

Summing things up, for any $x \in [0,l]$, we have:

$$|u(x)\| \leq |u(x)-\bar{u}| + |\bar{u}| \leq \left(\sqrt{l}+\frac{1}{\sqrt{l}}\right)\|u\|_{H^1(0,l)}\,.$$

The continuity of l over the whole $H^1(0,l)$ follows now from the density of functions $C^\infty([0,l])$ in $H^1(0,l)$ (a technical result). This implies now the continuity of the linear form,

$$|l(v)| \leq \underbrace{\left(\rho A g\sqrt{l} + P(\sqrt{l}+\frac{1}{\sqrt{l}})\right)}_{=C}\|v\|_{H^1(0,l)}\,.$$

Embedding (5.10) is also necessary to justify the BC in the definition of the test space which, by the continuity of trace $v(0)$, is now a closed subspace of Hilbert space $H^1(0,l)$ and, therefore, it is Hilbert (complete), too.

Finally, in order to conclude the well-posedness of the minimization and variational problems, we need to establish the V-coercivity result. We have:

$$\int_0^l (v(x))^2\,dx = \int_0^l \left(\int_0^x v'(s)\,ds\right)^2 dx \leq \int_0^l \int_0^x 1\,ds\,\underbrace{\int_0^x (v'(s))^2\,ds}_{\leq \|v\|^2_{H^1(0,l)}}\,dx \leq \frac{l^2}{2}\|v\|^2_{H^1(0,l)}\,.$$

The (optimal) coercivity constant can thus be estimated as $\alpha \geq \frac{2}{l^2}$. We actually have shown that, for functions v vanishing at zero, the L^2-norm of v can be estimated by the H^1-seminorm of v,

$$\|v\|_{L^2(0,l)} \leq \frac{l}{\sqrt{2}}\|v'\|_{L^2(0,l)} \qquad v \in H^1(0,l),\, v(0)=0\,.$$

The result represents an elementary case of the *Poincaré inequality* that holds in multiple space dimension as well. As we can see, it takes some work to check the assumptions for proving well-posedness of the variational formulation. □

[§]One can prove a stronger result: $H^1(0,l) \hookrightarrow C^\alpha([0,l])$ where $C^\alpha([0,l])$ denotes the Hölder-continuous functions with exponent $\alpha \in (0, \frac{1}{2})$.

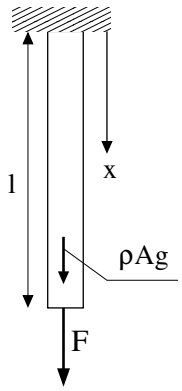

Figure 5.1
Elastic bar with stiffness EA and mass density (per unit length) ρA, loaded with its own weight and a force F.

5.3.2 Non-symmetric Coercive Problems

Once we have learned the equivalence of the variational formulation and the BVP, we realize that the entire procedure based on the integration by parts and the Fourier argument may be repeated for problems where the bilinear form is not necessary symmetric, i.e., BVPs that do not result from minimization of energy. The proof of equivalence is purely formal as we have not paid much attention to regularity issues. Once we develop the variational formulation though, we put the mathematician's hat on, and ask the questions about the well-posedness of the variational problem: *Does a solution exist? Is it unique ? Does it depend continuously upon the data?* In the previous section, we have studied a bit the issue for symmetric (hermitian) and coercive forms. We will discuss now generalizations of these results to forms that are not necessarily symmetric (hermitian).

THEOREM 5.3.1 (Lax-Milgram)

Let X be a Hilbert space with a closed subspace V and let $b : X \times X \to \mathbb{R}$ (or $\mathbb{C}$) be a sesquilinear continuous form that is V-coercive. Then, for every antilinear and continuous functional l on V, $l \in V'$, and element $\tilde{u}_0 \in X$, there exists a unique solution to the abstract variational problem

$$\begin{cases} \text{Find } u \in \tilde{u}_0 + V \text{ such that} \\ b(u,v) = l(v) \qquad \forall v \in V\,. \end{cases}$$

The solution u depends continuously on the data,

$$\|u\| \leq \frac{1}{\alpha}\|l\|_{V'} + \left(\frac{M}{\alpha} + 1\right)\|\tilde{u}_0\|_X\,.$$

The proof follows immediately from a more general result of Babuška and Nečas discussed next.

Example 5.3.2

The bilinear form in the variational formulation may be coercive even though it is not symmetric. The simplest example is provided by the diffusion-convection-reaction (with constant convection) problem in 1D:

$$\begin{cases} -(au')' + bu' + cu = f \quad \text{in } (0,1) \\ u(0) = 0 \\ a(1)u'(1) = g \end{cases}$$

Here a, b, c are the diffusion, convection and reaction coefficients, and b is constant. We assume:

$$0 < a_0 \leq a(x) \leq a_\infty < \infty \qquad 0 \leq c(x) \leq c_\infty < \infty \,.$$

Right-hand side $f(x)$ and g in the Neumann BC represent the load. Multiplying the equation with a test function $v(x)$, integrating over interval $(0,1)$, and integrating by parts the diffusion term, we obtain:

$$\int_0^1 [au'v' + bu'v + cuv]\, dx - au'v|_0^1 = \int_0^1 fv\, dx \,.$$

Assuming $v(0) = 0$, and building the Neumann BC into the formulation, we obtain,

$$\underbrace{\int_0^1 [au'v' + bu'v + cuv]\, dx}_{=:b(u,v)} = \underbrace{\int_0^1 fv\, dx + gv(1)}_{=:l(v)} \quad v(0) = 0 \,.$$

Due to the presence of the convection term, the bilinear form is not symmetric, $b(u,v) \neq b(v,u)$. We identity $H^1(0,1)$ as the energy space. The test space is:

$$V = \{v \in H^1(0,1) \,:\, v(0) = 1\},.$$

The essential (Dirichlet) BC is homogeneous and we do not need to lift the Dirichlet data ($u_0 = 0$); the variational problem is:

$$\begin{cases} u \in V \\ b(u,v) = l(v) \qquad \forall\, v \in V \,. \end{cases}$$

Continuity of forms $b(u,v)$ and $l(v)$ is proved exactly the same way as in Example 5.3.1. Before we address the coercivity condition, let us look at the contribution in $b(v,v)$ from the convection term:

$$\int_0^1 bv'v\, dx = \int_0^1 b\left(\frac{v^2}{2}\right)'\, dx = b\frac{v^2}{2}|_0^1 = \frac{b}{2}v(b)^2 \geq 0 \qquad v \in V \,.$$

Consequently,

$$b(v,v) = \int_0^1 [a(v')^2 + bv'v + cv^2]\, dx \geq a_0 \int_0^1 (v')^2\, dx \,,$$

and the proof of the coercivity follows from the Poincaré inequality discussed in Example 5.3.1. The variational problem is well-posed. ∎

5.3.3 General Variational Problems

I will give you now a glimpse into the powerful Banach theory. We are concerned with a general linear equation,

$$\begin{cases} x \in X \\ Bx = f\,. \end{cases} \tag{5.11}$$

Here X and Y are two Banach spaces, $B : X \to Y$ is a general continuous operator, and $f \in Y$. *Under what conditions on B and f, the problem is well-posed?* Assume that the problem has a solution x. Multiplying both sides of the equation with an element $y' \in Y'$ (in the sense of the duality pairing), we obtain:

$$_Y\langle Bx, y'\rangle_{Y'} = {}_X\langle x, B'y'\rangle_{X'} = {}_Y\langle f, y'\rangle_{Y'} \qquad \forall\, y' \in Y'\,.$$

Consequently, f *must* satisfy the *compatibility condition:*

$$\langle f, y'\rangle = 0 \quad \forall y' \in \mathcal{N}(B') \quad \Leftrightarrow \quad f \in \mathcal{N}(B')^{\perp} := \{y \in Y \,:\, \langle y, y'\rangle = 0 \quad \forall\, y' \in \mathcal{N}(B')\}\,.$$

Additionally, if x is to depend continuously upon the data f, we must have:

$$\|x\|_X \leq C\|f\|_Y = C\|Bx\|_Y.$$

In other words, the operator must be *bounded below:*

$$\|Bx\|_Y \geq \gamma\|x\|_X \qquad x \in X,\, \gamma > 0\,. \tag{5.12}$$

Note that the boundness below condition implies that the operator is injective, i.e., the solution must be unique. It turns out that these two conditions are not only necessary but also sufficient for the well-posedness of linear equation (5.11).

THEOREM 5.3.2 (Banach's Closed Range Theorem)

Let $B : X \to Y$ be a linear, continuous and injective operator from a Banach space X into a Banach space Y, and $B' : Y' \to X'$ denote its topological transpose. The following conditions are equivalent to each other.

(i) The operator is bounded below, i.e., condition (5.12) holds.

(ii) Range $\mathcal{R}(B)$ is closed in Y.

(iii) For every $f \in \mathcal{N}(B')^{\perp}$, there exists a unique solution x of problem (5.11) that depends continuously upon data f.

PROOF See [5], Section 5.17. ∎

THEOREM 5.3.3 (Babuška–Nečas)

Assume spaces U, V are Banach spaces with V being reflexive, and forms $b(u,v), l(v)$ are continuous. Additionally, let form $b(u,v)$ satisfy the inf-sup condition:

$$\inf_{u\neq 0}\sup_{v\neq 0}\frac{|b(u,v)|}{\|u\|_U\|v\|_V} =: \gamma > 0 \quad \Leftrightarrow \quad \sup_{v\neq 0}\frac{|b(u,v)|}{\|v\|_V} \geq \gamma\|u\|_U \tag{5.13}$$

and let $l(v)$ satisfy the compatibility condition:

$$l(v) = 0 \quad v \in V_0 \tag{5.14}$$

where

$$V_0 := \{v \in V \,:\, b(u,v) = 0 \quad \forall u \in U\}\,. \tag{5.15}$$

Then, the variational problem:

$$\begin{cases} u \in U \\ b(u,v) = l(v) \quad v \in V\,, \end{cases} \tag{5.16}$$

is well posed, i.e., there exists a unique solution u that depends continuously upon the data:[¶]

$$\|u\|_U \leq \gamma^{-1}\|l\|_{V'} = \gamma^{-1}\sup_{v\neq 0}\frac{|l(v)|}{\|v\|_V}\,. \tag{5.17}$$

PROOF The proof of Babuška–Nečas Theorem is a direct reinterpretation of Banach Closed Range Theorem. Indeed, the sesquilinear form $b(u,v)$ induces two linear and continuous operators,

$$B : U \to V', \quad \langle Bu, v\rangle_{V'\times V} = b(u,v)$$
$$B_1 : V \to U', \quad \langle B_1 v, u\rangle_{U'\times U} = \overline{b(u,v)}$$

Space V_0 is the null space of operator B_1. The inf-sup condition says that operator B is bounded below. The transpose operator B' goes from bidual V'' into dual U'. However, in the reflexive setting, spaces V and V'' are isometrically isomorphic and, therefore, operator B_1 can be identified with the transpose of B. Consequently, space V_0 can be identified as the null space of the transpose, and we arrive at the scenario covered by the Closed Range Theorem. ∎

COROLLARY 5.3.1

In the case of non-homogeneous essential boundary condition, the final stability estimate looks as follows,

$$\begin{aligned} \|u\|_X = \|\tilde{u}_0 + w\|_X &\leq \|\tilde{u}_0\|_X + \|w\|_X \\ &\leq \|\tilde{u}_0\|_X + \frac{1}{\gamma}\left(\|l\|_{V'} + M\|\tilde{u}_0\|_X\,\|v\|_V\right) \\ &\leq \frac{1}{\gamma}\|l\|_{V'} + \left(1 + \frac{M}{\gamma}\right)\|\tilde{u}_0\|_X \end{aligned} \tag{5.18}$$

[¶]The problem is stable. It goes without saying that γ is the best constant we can have.

We may argue that, from the operator theory (the classical Functional Analysis) point of view, the nature of a variational problem lies in the fact that the operator takes values in a dual space (dual to the test space). We may want to mention that neither Closed Range Theorem nor Babuška–Nečas Theorem are constructive. They do not tell us how to prove the inf-sup condition (boundness below of operator B); they only tell us that we must have it.

Babuška–Nečas Theorem implies immediately Lax–Milgram Theorem. Indeed, with $U = V$, and the coercivity assumption, we have:

$$\sup_{v\in V} \frac{|b(u,v)|}{\|v\|_V} \geq \frac{|b(u,u)|}{\|u\|_V} \geq \alpha \|u\|_V \,,$$

with the coercivity constant α providing a lower bound for the inf-sup constant γ.

Babuška–Nečas Theorem provides a foundation for variational problems with *unsymmetric variational setting*, i.e., when trial space U is different from test space V. The following example is perhaps the simplest example illustrating the need for such a theory.

Example 5.3.3 (A convection-reaction problem)

We are looking for a solution $u = u(x)$ to the 1D convection-reaction equation:

$$bu' + cu = f \quad \text{in } (0,1)$$

where $b = b(x)$ is a convection coefficient, $c = c(x) \geq 0$ is a reaction coefficient, and $f = f(x)$ represents the load. We will make appropriate assumptions on the data as we check assumptions of the Babuška–Nečas Theorem. The problem is accompanied by BCs. We will discuss four possible scenarios. For simplicity, we will consider only homogeneous BCs.

Case 1. Inflow BC: $u(0) = 0$.

Case 2. Outflow BC: $u(1) = 0$.

Case 3. Inflow and outflow BC: $u(0) = u(1) = 0$.

Case 4. No BC.

We focus on Case 1 BC and will only comment on the remaining cases. There are two possible variational formulations.

Strong variational formulation. We multiply the DE with a test function $v = v(x)$, integrate over interval $(0,1)$, and leave it alone. The resulting bilinear form is:

$$b(u,v) = \int_0^1 (bu'v + cuv)\, dx \,.$$

The functional setting emerges naturally from the continuity assumption and the Cauchy-Schwarz inequality. We assume that the advection and reaction coefficients are bounded,

$$|b(x)| \le b_\infty < \infty, \qquad c(x) \le c_\infty < \infty\,. \tag{5.19}$$

If a norm $\|\cdot\|$ does not come with an index for the space, it will always denote the L^2-norm. We have,

$$|b(u,v)| \le b_\infty \|u'\|\,\|v\| + c_\infty \|u\|\,\|v\| = (b_\infty\|u'\| + c_\infty\|u\|)\,\|v\| \le \underbrace{\sqrt{b_\infty^2 + c_\infty^2}}_{=:M}\,\|u\|_{H^1}\,\|v\|_{L^2}\,.$$

This leads to the definition of trial and test spaces:

$$U = \{u \in H^1(0,1) \,:\, u(0) = 0\} \qquad V = L^2(0,1)\,.$$

We already know that $H^1(0,1)$ is embedded continuously into $C([0,1])$, so the inflow BC makes sense and U, as a closed subspace of a Hilbert space, is itself a Hilbert space. To assure the continuity of the linear functional, we can assume $f \in L^2(0,1)$,

$$|l(v)| = |\int_0^1 fv| \le \underbrace{\|f\|}_{=:C}\,\|v\|_{L^2}\,.$$

To prove the inf-sup condition, we will try $v = u'$. We have then,

$$b(u,v) = \int_0^1 \left[b(u')^2 + \int_0^1 c\left(\frac{u^2}{2}\right)'\right] dx = \int_0^1 b(u')^2\,dx - \frac{1}{2}\int_0^1 c'u^2\,dx + \frac{1}{2}c(1)u(1)^2\,.$$

The last term is non-negative so it is harmless. Let us focus on the first two terms. With the assumption,

$$0 < b_0 \le b(x)\,, \tag{5.20}$$

the first term is bounded below by $b_0\|u'\|^2$ and, by the Poincaré inequality, by $b_0\alpha\|u\|^2$ as well, where $\alpha > 0$ is the Poincaré constant. Assuming for instance,

$$b_0\alpha - c' \ge 0 \quad \Leftrightarrow \quad c' \le \alpha b_0\,, \tag{5.21}$$

we obtain,

$$b(u,v) \ge \frac{1}{2}b_0\|u'\|^2\,.$$

The inf-sup condition now easily follows. For $v = u'$, we have

$$\|v\|_{L^2} \le \|u\|_{H^1} \quad \Rightarrow \quad \frac{1}{\|v\|_{L^2}} \ge \frac{1}{\|u\|_{H^1}}\,,$$

and,

$$\sup_{v\in L^2} \frac{b(u,v)}{\|v\|_{L^2}} \ge \frac{b(u,u')}{\|u'\|_{L^2}} \ge \frac{1}{2}b_0\alpha\frac{\|u\|_{H^1}^2}{\|u\|_{H^1}} = \frac{1}{2}b_0\alpha\|u\|_{H^1}\,.$$

Done. It remains to check on the compatibility conditions for the right-hand side f. We claim that

$$V_0 = \{v \in L^2(0,1) \,:\, b(u,v) = 0 \quad \forall u \in U\} = \{0\}\,.$$

Indeed, let $u = \int_0^x v(s)\,ds$, i.e., $u' = v$, where $v \in V_0$. The proof of the inf-sup condition implies that $\|u\|_{H^1} = 0$ and, therefore, $v = 0$ as well. Summarizing, with assumptions (5.19),(5.20), (5.21), and any $f \in L^2(0,1)$, the problem is well-posed. We emphasize that these are just ***sufficient conditions*** that we have examined; one can come up with less restrictive and more general assumptions.

Analysis of Case 2 BCs is very similar.

Analysis of Case 3 BC leads to a non-trivial V_0 and a compatibility condition for f, see Exercise 5.3.5.

Case 4 BC leads to a non-unique solution. We reconcile the non-uniqueness with the Babuška-Nečas Theorem[‖] by looking for a solution in the quotient space:

$$U = H^1(0,1)/U_0$$

where U_0 is the one-dimensional space spanned by a non-zero solution u_0 of the equation that you may obtain e.g. by separation of variables. As a finite-dimensional subspace of U, U_0 is automatically closed. One of the fundamental results from Functional Analysis is that the quotient space obtained by dividing a Banach (Hilbert) space with a closed subspace, remains complete and, therefore, it is itself a Banach (Hilbert space). The norm in a quotient space is a ***minimum energy extension norm:***

$$\|[u]\|_{U/U_0} := \inf_{w \in [u]} \|w\|_U\,.$$

The problem can now be stated in the quotient space U and $V = L^2(0,1)$. The bilinear form is defined essentially in the same way,

$$b([u],v) := \int_0^1 (au' + cu)v\,dx$$

where $u \in [u]$ is a representative from the equivalence class. Please note that the value of the bilinear form is independent of the choice of the representative and so the bilinear form is indeed well-defined. The rest of the analysis follows now in the standard way.

Weak variational formulation. This time we integrate by parts to obtain:

$$-\int_0^1 u(-(bv)' + cv)\,dx + (buv)|_0^1 = \int_0^1 fv\,.$$

[‖] The inf-sup condition implies uniqueness.

Focusing again on Case 1 BCs, we build the BC into the formulation and eliminate the boundary term at $x = 1$ by assuming $v(1) = 0$. We obtain:

$$\begin{aligned} U &:= L^2(0,1) \\ V &:= \{v \in H^1(0,1) \,:\, v(1) = 0\} \\ b(u,v) &:= \int_0^1 u(-(bv)' + cv)\,dx = \int_0^1 [-buv' + (c - b')uv]\,dx \\ l(v) &:= \int_0^1 fv\,dx\,. \end{aligned}$$

The roles of u and v have now changed. Following the same procedure as before we can come with sufficient conditions for the related (dual) variational problem:

$$\begin{cases} v \in V \\ b(w,v) = (g,w) \quad w \in U \end{cases}$$

to be well-posed. Note that, for v, the dual formulation is a strong variational formulation. Once we establish this, we can use the well-posedness of the dual problem to prove the inf-sup condition for the weak formulation. We select $g = u$ and take the corresponding solution v of the dual problem. By the well-posedness of the strong dual problem,

$$\|v\|_V \le C\|u\| \quad \Rightarrow \quad \frac{1}{\|v\|_V} \ge \frac{1}{C}\frac{1}{\|u\|}\,.$$

Testing in the dual problem with $w = u$, we obtain,

$$\|u\|^2 = (u,u) = b(u,v)\,.$$

Consequently,

$$\sup_{z\in V} \frac{b(u,z)}{\|z\|_V} \ge \frac{b(u,v)}{\|v\|_V} = \frac{\|u\|^2}{\|v\|_V} \ge \frac{1}{C}\frac{\|u\|^2}{\|u\|} = \frac{1}{C}\|u\|\,.$$

The rest of the analysis is straightforward.

Case 2,3, and 4 BC are analyzed similarly.

□

Exercises

Exercise 5.3.1 Replace u_0 in (5.9) with *any element* $\bar{u}_0 \in u_0 + V$. The solution w will obviously change into some other $\bar{w}$. Show that the ultimate solution $u = \bar{u}_0 + \bar{w} = u_0 + w$ will remain the same, though.

(2 points)

Exercise 5.3.2 Generalize the results on the equivalence of minimization and variational problems to the complex case. The bilinear symmetric form is replaced with a sesquilinear hermitian form, i.e.,

$$b(u,v) = \overline{b(v,u)} \qquad u, v \in X\,,$$

and the energy functional is redefined as:

$$J(u) = \frac{1}{2}b(u,u) - \Re l(u)\,.$$

Hint: Recall the fundamental property of linear and antilinear functionals defined on a complex vector space:

$$l(v) = 0 \quad \Leftrightarrow \quad \Re l(v) = 0\,.$$

(10 points)

Exercise 5.3.3 Argue why the Ritz and Galerkin methods are equivalent, i.e., they deliver the same approximate solution.

(5 points)

Exercise 5.3.4 Consider a slight modification of the bar problem from Example 5.3.1, presented in Fig. 5.2. Force F has now been replaced with a (linear) spring support at the end of the bar. Formulate the minimization problem (the elastic energy of the bar has now to be complemented with the elastic energy of the spring), the corresponding variational formulation and the Euler-Lagrange BVP. Prove that the variational problem is well posed.

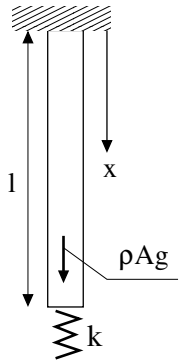

Figure 5.2
Elastic bar with stiffness EA and mass density (per unit length) ρA, supported with a spring with stiffness k at $x = l$, and loaded with its own weight.

(10 points)

Exercise 5.3.5 Provide a complete well-posedness analysis for the strong variational formulation of convection-reaction problem discussed in Example 5.3.3 and Case 3 BC.

(10 points)

6

Elementary Theory of Partial Differential Equations

6.1 Preliminaries

Any exposition on Partial Differential Equations (PDEs) starts with the classification of a single, second order PDE in two space dimensions,

$$a_{11}u_{,11} + 2a_{12}u_{,12} + a_{22}u_{,22} = F$$

where coefficients $a_{11}, a_{12}, u_{,22}$ may depend upon x. The type of equation is governed by the determinant of the symmetric matrix ($a_{12} = a_{21}$):

$$a(x) := \begin{pmatrix} a_{11}(x) & a_{12}(x) \\ a_{21}(x) & a_{22}(x) \end{pmatrix}$$

If $\det a(x) > 0$, the equation is said to be *elliptic* at x, if $\det a(x) < 0$, the equation is *hyperbolic* at x, for $\det a(x) = 0$, the equation is *parabolic* at x. If matrix $a = a(x_1, x_2)$ is only a function of independent variables x_1, x_2, and $F = F(x_1, x_2, u, u_{,1}, u_{,2})$ depends possibly not only upon x_1, x_2 but also the solution u and its first derivatives, the equation is said to be *quasilinear* [*]. If the dependence of F on $u, u_{,1}, u_{,2}$ is linear, we are dealing with a linear PDE.

Note that the classification depends only upon the coefficients in front of the second derivatives. As

$$a_{1,2}u_{,12} + a_{2,1}u_{,21} = (a_{1,2} + a_{2,1})u_{,12}\,,$$

it makes little sense to consider a nonsymmetric matrix since only the sum of the off-diagonal terms enters the equation.

Example 6.1.1

Poisson equation and its homogeneous version ($f = 0$) - the Laplace equation,

$$-\Delta u := -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f\,,$$

are elliptic. The wave equation,

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = f$$

[*] Some authors extend the definition to the case when matrix a may also depend upon u and its first derivatives. The classification depends then not only upon x but also the solution $u(x)$. We talk about the solution being elliptic, hyperbolic or parabolic.

is hyperbolic, and the heat (diffusion) equation:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = f\,,$$

is parabolic. □

The classification extends to more space dimensions. The general *diffusion-convection-reaction equation* (summation convention at work) in $\mathbb{R}^n$,

$$\underbrace{-\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right)}_{=:Au} + b_j\frac{\partial u}{\partial x_j} + cu = f$$

with a symmetric, positive-definite diffusion matrix a_{ij} is *elliptic*. If we add time dependence to the coefficients and solution u, the heat equation in $\mathbb{R}^n$,

$$\frac{\partial u}{\partial t} - Au = f\,,$$

is *parabolic*, and the n-dimensional wave equation,

$$\frac{\partial^2 u}{\partial t^2} - Au = f\,,$$

is *hyperbolic*. In most of physical applications, the diffusion operator A above is given in the divergence form [†].

The PDEs are accompanied with *boundary conditions* (BCs) or *initial conditions* (ICs) resulting in *boundary-value problems* (BVP), *initial-value problems* (IVPs) and *initial-boundary-value problems* (IBVPs).

6.1.1 Separation of Variables. Elliptic Examples

The methodology applies to all types of (single) PDEs provided the domain in which the equation is being solved is *separable*, i.e., it can be represented as a Cartesian product of lower dimensional domains (possibly in curvilinear coordinates). We will study the technique through a large collection of examples.

Example 6.1.2

Consider the following Dirichlet BVP for the Laplace equation on a square domain,

$$\begin{cases} -\Delta u = 0 & \text{in } \Omega = (0,a)^2 \\ u = g & \text{on } \Gamma = \partial\Omega\,. \end{cases}$$

We look for the solution in the form of the tensor product,

$$u(x,y) = X(x)Y(y)\,,$$

[†] Ready for integration by parts.

Substituting into the equation and separating the variables we obtain,

$$-X''Y - XY'' = 0 \quad \Rightarrow \quad -\frac{X''}{X} = \frac{Y''}{Y} = \lambda$$

where λ is a separation constant. This leads to separate equations in x and y,

$$-X'' = \lambda X \quad \text{and} \quad Y'' = \lambda Y\,.$$

The critical part of the methodology is to obtain a Sturm-Liouville problem in one of the variables. This calls for homogeneous BCs. We can accomplish this by splitting the boundary data g into parts g_1, g_2, g_3, g_4 corresponding to the four edges, and considering non-homogenous BCs on one edge only [‡], see Fig. 6.1. The final solution can then be obtained by superposition of the four,

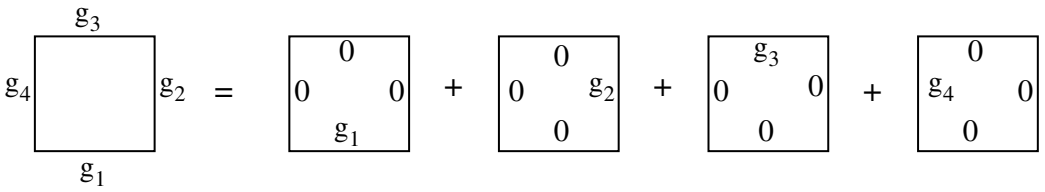

Figure 6.1
Splitting a Dirichlet problem into the superposition of four problems with non-homogeneous BC on one edge only

one-edge problems. Consider for instance the non-homogeneous data on the first edge only,

$$\begin{cases} u(x,0) = g_1(x) & x \in (0,a) \\ u(a,y) = 0 & y \in (0,a) \\ u(x,a) = 0 & x \in (0,a) \\ u(0,y) = 0 & y \in (0,a) \end{cases}$$

The BCs on u translate into homogeneous Dirichlet BCs on X which leads to the Sturm-LIouville problem:

$$\begin{cases} -X'' = \lambda X \\ X(0) = X(a) = 0 \end{cases}$$

We can say now much more about the separation constant, it must be real and positive definite, $\lambda = k^2, k > 0$. This leads to the eigenvector,

$$X(x) = A\cos kx + B\sin kx\,.$$

The BCs imply now $A = 0$ and

$$ka = n\pi \qquad \Rightarrow \quad k = k_n = \frac{n\pi}{a}\,.$$

[‡] Or two opposite edges.

By the Sturm Liouvillle Theorem, after normalization,

$$\int_0^a \sin^2(kx)\,dx = \frac{1}{2}\int_0^a (1-\cos 2kx)\,dx = \frac{a}{2}\,,$$

functions $\sqrt{\frac{2}{a}}\sin k_n x$ provide and orthonormal basis for $L^2(0,a)$. Consequently, any function $g_1(x)$ can be expanded into the series,

$$g_1(x) = \frac{2}{a}\sum_{n=1}^{\infty} \underbrace{\int_0^a g_1(s)\sin k_n s\,ds}_{=:g_{1n}} \sin k_n x\,.$$

Once we know the separation constant, we can solve now for $Y(y)$,

$$Y(y) = Ae^{k_n y} + Be^{-k_n y}\,.$$

The homogenous BC at $y=a$ suggests a different representation,

$$Y(y) = A\cosh k_n(y-a) + B\sinh k_n(y-a) \quad \Rightarrow \quad A = 0\,,$$

The constant B comes now from requesting $Y(0)=1$,

$$B = B_n = -\sinh^{-1} k_n a\,.$$

The final solution to the first edge problem has the form:

$$u_1 = -\frac{2}{a}\sum_{n=1}^{\infty} \frac{g_{1n}}{\sinh k_n a}\,\sin k_n x\,\sinh k_n(y-a)$$

where

$$k_n = \frac{n\pi}{a}, \qquad g_{1n} = \int_0^a g_1(x)\sin k_n x\,dx\,.$$

□

REMARK 6.1.1 A necessary condition for the solution u to have a finite energy ($u \in H^1(\Omega)$) is that the boundary data g must be continuous. The splitting of g discussed above does not preserve this continuity. Consequently, solutions $u_1,\ldots,u_4$ will be less regular than u which, in particular, will affect the convergence of the series. A better splitting of boundary data g starts with taking out from g its *vertex interpolant*. For instance, for the first edge,

$$g_{1v} = g_1(0)(1-\frac{x}{a}) + g_1(a)\frac{x}{a}\,.$$

The union of vertex interpolants admits a standard extension to the whole domain using the square element vertex shape functions,

$$u_I = u(0,0)(1-\frac{x}{a})(1-\frac{y}{a}) + u(a,0)\frac{x}{a}(1-\frac{y}{a}) + u(a,a)\frac{x}{a}\frac{y}{a} + u(0,a)(1-\frac{x}{a})\frac{y}{a}\,.$$

Note that this bilinear function satisfies the Laplace equation. We can now decompose the BC data for $u - u_I$ edge-wise, and look for the final solution as the superposition of solutions to one-edge problems,

$$u = u_I + u_1 + u_2 + u_3 + u_4 \,.$$

Contrary to the previous split, functions u_i enjoy now continuous BC data. ∎

Example 6.1.3

Solve the Laplace equation in a circular domain,

$$\Omega = \{x \,:\, |x| < a\}$$

with the Dirichlet BC:

$$u(a,\theta) = g(\theta) \quad \theta \in (0, 2\pi) \,.$$

The domain is not separable in Cartesian coordinates but it is separable in polar coordinates,

$$\Omega = \{(r,\theta) \,: 0 < r < a,\, 0 < \theta \leq 2\pi\} \,.$$

We recall the formula for the gradient in polar coordinates,

$$\boldsymbol{\nabla} u = \frac{\partial u}{\partial r} e_r + \frac{1}{r}\frac{\partial u}{\partial \theta} e_\theta \,,$$

and use integration by parts to develop quickly the formula for the Laplacian,

$$\int_\Omega \boldsymbol{\nabla} u \, \boldsymbol{\nabla} v = \int\int \left(\frac{\partial u}{\partial r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial u}{\partial \theta}\frac{\partial v}{\partial \theta} \right) r dr d\theta = -\int\int \underbrace{\left(\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial u}{\partial r}) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} \right)}_{=\Delta u} v \, r dr d\theta + B.T.$$

Assuming $u = R(r)\Theta(\theta)$ leads to:

$$-\frac{\Theta''}{\Theta} = \frac{r(rR')'}{R} = \lambda \,.$$

Operator $-\Theta''$ with periodic BC: $\Theta(0) = \Theta(2\pi)$, $\Theta'(0) = \Theta'(2\pi)$, is a Sturm-Liouville semi-positive operator, comp. Example 3.3.7. We can seek thus the separation constant in the form: $\lambda = k^2$, $k \geq 0$. This leads to $k = 0, 1, 2 \ldots$ and the solution:

$$\Theta = A cosk\theta + B \sin k\theta \,.$$

Turning to the equation for R, we get:

$$r(rR')' - k^2 R = 0 \,.$$

For $k = 0$, this leads to the solution:

$$R = C \ln r + D \,.$$

The first term leads to the solution $u = c \ln r$ and has infinite energy ($u \notin H^1(\Omega)$), see Exercise 6.1.1. The second term leads to the constant solution. For $k > 0$, we obtain a Cauchy-Euler equation in r which leads to the solution:

$$R = Cr^k + Dr^{-k} .$$

The second term leads again to infinite-energy solutions and must be eliminated. The first term leads to the ultimate solution in the form:

$$u = c + \sum_{k=1}^{\infty} r^k (A_k \cos k\theta + B_k \sin k\theta) .$$

The BC leads to:

$$c + \sum_{k=1}^{\infty} a^k (A_k \cos k\theta + B_k \sin k\theta) = g(\theta) .$$

The main point is that the Sturm-Liouville Theorem guarantees that $1, \cos k\theta, \sin k\theta$, $k = 1, 2, \ldots$, provides an orthogonal basis for $L^2(0, 2\pi)$. This leads to the formulas for the coefficients c, A_n, B_n:

$$c = \frac{1}{2\pi} \int_0^{2\pi} g(\theta)\, d\theta, \quad A_k = \frac{\pi}{a^k} \int_0^{2\pi} g(\theta) \cos k\theta \, d\theta, \quad B_k = \frac{\pi}{a^k} \int_0^{2\pi} g(\theta) \sin k\theta \, d\theta .$$

□

Example 6.1.4

Consider an exterior domain:

$$\Omega = \{x \,:\, |x| > a\} ,$$

and Helmholtz equation:

$$-\Delta u - \omega^2 u = 0$$

with a Neumann BC at $r = a$,

$$\frac{\partial u}{\partial n} = -\frac{\partial u}{\partial r} = g ,$$

and the requirement that u is *outgoing at infinity*. Recall that the Helmholtz equation is obtained from the wave equation,

$$\frac{\partial^2 U}{\partial t^2} - \Delta U = 0$$

by assuming the ansatz:

$$U(x,t) = \Re(u(x)e^{\omega t}) \quad \text{or} \quad U(x,t) = \Re(u(x)e^{-\omega t})$$

where ω is the angular frequency, and $u(x)$ is the new, complex-valued unknown, the *phasor*. Note that both formulas above lead to the same Helmholtz equation. The choice of the sign in front of the frequency is critical though for distinguishing between outgoing and incoming waves as we will see in a moment. We shall work with the first ansatz with the plus sign. Note also that the phasor is *complex-valued*, solving the Helmholtz equations involves always complex-valued solutions.

Separation of variables: $u = R(r)\Theta(\theta)$ leads to

$$-\frac{1}{r}(rR')'\Theta - \frac{1}{r^2}\Theta'' - \omega^2 R\Theta = 0$$

and,

$$-\frac{\Theta''}{\Theta} = \frac{r(rR')'}{R} + \omega^2 r^2 = \lambda\,.$$

As in the previous example, we can assume $\lambda = k^2$, $k \geq 0$ and the periodic BC in θ lead to the solution:

$$\Theta = \begin{cases} A_0 & k = 0 \\ A_k \cos k\theta + B_k \sin k\theta & k = 1, 2, \dots \end{cases}.$$

Since the solution is complex-valued, it is more convenient to work with the complex Fourier series,

$$\Theta = \sum_{k=-\infty}^{\infty} A_k e^{ik\theta}\,.$$

With $\lambda = k^2$, we obtain,

$$r^2 R'' + rR' + (\omega^2 r^2 - k^2)R = 0\,.$$

Changing the dependent variable, $\rho = \omega r$, we obtain the Bessel equation:

$$\rho^2 \frac{d^2R}{d\rho^2} + \rho\frac{dR}{d\rho} + (\rho^2 - k^2) = 0\,,$$

and the general solution,

$$R = \sum_{k=0}^{\infty}(A_k H_k^{(1)}(\omega r) + B_k H_k^{(2)}(\omega r))\,.$$

This is where the radiation condition comes in. Recalling the asymptotic behavior of Hankel functions at ∞, we have,

$$H_k^{(1)}(\omega r)e^{i\omega t} \sim \frac{e^{i(\omega r - \frac{\pi}{4})}}{\sqrt{\frac{\pi\omega r}{2}}}e^{i\omega t} = \frac{e^{i(\omega(r+t) - \frac{\pi}{4})}}{\sqrt{\frac{\pi\omega r}{2}}}$$

and

$$H_k^{(2)}(\omega r)e^{i\omega t} \sim \frac{e^{i(-\omega r - \frac{\pi}{4})}}{\sqrt{\frac{\pi\omega r}{2}}}e^{i\omega t} = \frac{e^{i(\omega(-r+t) - \frac{\pi}{4})}}{\sqrt{\frac{\pi\omega r}{2}}}\,.$$

The first function represents a wave coming from infinity whereas the second one a wave outgoing to infinity [§]. Consequently, we set $A_k = 0$.

The ultimate solution looks as follows,

$$u = \sum_{k=-\infty}^{\infty} c_k H_k^{(2)}(\omega r)e^{ik\theta}\,.$$

The Neumann BC at $r = a$ leads to the equation:

$$-\omega \sum_{k=-\infty}^{\infty} c_k (H_k^{(2)})'(\omega a)e^{ik\theta} = g(\theta) \qquad \theta \in (0, 2\pi)\,.$$

[§] Signs in front of time t and radius r have to be opposite for the outgoing wave.

Again, the main point is that functions $e^{ik\theta}, k \in \mathbb{Z}$, provide an orthogonal basis for $L^2(0, 2\pi)$. Multiplying the equation above with $e^{-ik\theta} = \overline{e^{ik\theta}}$, and integrating over $(0, 2\pi)$, we get:

$$c_k = -\frac{1}{2\pi\omega(H_k^{(2)})'(\omega a)} \int_0^{2\pi} g(\theta) e^{-ik\theta}\, d\theta\,.$$

□

So far, we have discussed how to solve homogeneous PDEs with non-homogeneous BCs. Can we extend the presented methodology to non-homogeneous PDEs ? Consider for instance the Poisson equation with homogeneous Dirichlet BC,

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma = \partial\Omega\,. \end{cases}$$

If we are lucky, we may be able to find *a particular solution* u_f to the Poisson equation (with no BC imposed), i.e. $-\Delta u_f = f$. We can seek then solution u in the form: $u = v + u_f$ where component v satisfies the Laplace equation and a *non-homogeneous* BC involving the particular solution u_f:

$$u = v + u_f = 0 \quad \Rightarrow \quad v = -u_f \quad \text{on } \Gamma\,.$$

Thus, we have managed to trade a non-homogeneous right-hand side for a non-homogenous BC.

A more universal technology is based on the general Spectral Theorem that guarantees that eigenvectors of a self-adjoint operator provide an orthogonal basis for space $L^2(\Omega)$ where Ω is the multi-dimensional domain. This leads to the task of solving multi-dimensional eigenvalue problems. If domain Ω is separable, the separation of variables comes handy again.

Example 6.1.5

Let $\Omega = (0, a) \times (0, b)$. Solve the two-dimenional eigenvalue problem for the Laplace operator,

$$\begin{cases} -\Delta e = \mu e & \text{in } \Omega \\ e = 0 & \text{on } \Gamma = \partial\Omega\,. \end{cases}$$

Separation of variables, $e = X(x)Y(y)$, leads to:

$$-X''Y - XY'' = \mu XY \quad \Rightarrow \quad -\frac{X''}{X} = \frac{Y''}{Y} + \mu = \lambda\,.$$

As $-X''$ with Dirichlet BCs is self-adjoint and positive definite, $\lambda = k^2$, $k \geq 0$. This leads to:

$$k = k_n = \frac{n\pi}{a}, \qquad X_n = A_n \sin\frac{n\pi x}{a} \quad n = 1, 2, \ldots,\,.$$

In turn, we obtain,

$$-\frac{Y''}{Y} = \mu - \left(\frac{n\pi}{a}\right)^2\,.$$

But operator $-Y''$ with Dirichlet BCs is self-adjoint and positive-definite as well, so we must have again,

$$\mu - \left(\frac{n\pi}{a}\right)^2 = k^2 \,.$$

This leads to

$$k = k_m = \frac{m\pi}{b} \qquad Y_m = \sin\frac{m\pi y}{b} \,.$$

The ultimate eigenpairs are:

$$\mu = \mu_{nm} = \left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 \qquad e = e_{nm} = C_{nm} \sin\frac{n\pi x}{a} \sin\frac{m\pi y}{b} \,.$$

If we expand now the right-hand side in the eigenfunctions, and look for the solution u to the Poisson problem in the same form,

$$f = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} f_{nm} e_{nm}, \qquad u = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} u_{nm} e_{nm} \,,$$

applying the Laplace operator to u and comparing with the spectral representation for f, we obtain explicit formulas for the components u_{nm} of solution u in the spectral basis,

$$u_{nm} = \frac{f_{nm}}{\mu_{nm}} \,.$$

The main point here is that the eigenvectors e_{nm} of the Laplace operator provide an orthogonal basis for $L^2(\Omega)$. As usual, components f_{nm} are obtained by multiplying the spectral representation of f with eigenvectors e_{nm} and integrating over Ω,

$$f_{nm} \underbrace{\int_0^a \sin^2\frac{n\pi x}{a}\,dx \int_0^b \sin^2\frac{n\pi y}{b}\,dy}_{=\frac{ab}{4}} = \int_0^a\int_0^b f(x,y)\sin\frac{n\pi x}{a}\sin\frac{m\pi y}{b}\,dxdy \,.$$

□

REMARK 6.1.2 Augmenting the Laplace operator with a shift $-\omega^2 I$, results simply in a shift of eigenvalues. Consequently, solution of the Helmholtz equation,

$$-\Delta u - \omega^2 u = f \,,$$

with Dirichlet BC, leads to a spectral representation of the solution with components:

$$u_{nm} = \frac{f_{nm}}{\mu_{nm} - \omega^2} \,.$$

▮

Example 6.1.6

This is a somehow more difficult example related to problems we have encountered when studying waveguide problems. We study the Laplace operator in a conical domain:

$$D := \{x \in \mathbb{R}^2 \,;\, a < r < b,\, 0 < \theta < \alpha\}.$$

We will consider both the Laplace equation (with different BCs), and the Laplace eigenvalue problem: $-\Delta u = \mu\, u$ with homogeneous Dirichlet BC.

Laplace eigenvalue problem. We know that the Laplace operator in any domain with homogeneous Dirichlet BCs is self-adjoint and positive definite, so we can assume that $\mu = \nu^2$, $\nu > 0$. The separation of variables ansatz: $u = R(r)\Theta(\theta)$ leads to the equation:

$$-\frac{1}{r}(rR')'\Theta - \frac{1}{r^2}R\Theta'' = \nu^2 R\Theta$$

and, after separating the variables, to:

$$\frac{r(rR')'}{R} + \nu^2 r^2 = -\frac{\Theta''}{\Theta} = \lambda$$

where λ is the separation constant. Both operators in r and θ are self-adjoint so, in principle, we can start with either one, but the operator in θ is easier and it involves *only* constant λ. As the operator $-\Theta''$ with homogeneous Dirichlet BCs is self-adjoint in $L^2(0,\alpha)$ and positive definite, we can assume again $\lambda = \beta^2$, $\beta > 0$. The solution is:

$$\Theta = A\cos\beta\theta + B\sin\beta\theta\,.$$

The BCs lead to $A = 0$ and,

$$\beta\alpha = n\pi \quad \Rightarrow \quad \beta = \beta_n = \frac{n\pi}{\alpha} \quad n = 1, 2, \dots\,.$$

This, in turn, leads to another eigenvalue problem in r,

$$-r(rR')' - \nu^2 r^2 R = \beta_n^2 R\,,$$

with the operator being self-adjoint in the weighted space $L^2_{1/r}(a,b)$. The problem leads to the Bessel equation:

$$r^2R'' + rR' + (\nu^2 r^2 - \beta_n^2)R = 0\,,$$

and the solution:

$$R = CJ_{\beta_n}(\nu r) + DY_{\beta_n}(\nu r)\,.$$

Homogeneous Dirichlet BCs for $r = a$ and $r = b$ lead to a transcendental equation for ν:

$$\begin{vmatrix} J_{\beta_n}(\nu a) & Y_{\beta_n}(\nu a) \\ J_{\beta_n}(\nu b) & Y_{\beta_n}(\nu b) \end{vmatrix} = J_{\beta_n}(\nu a)Y_{\beta_n}(\nu b) - J_{\beta_n}(\nu b)Y_{\beta_n}(\nu a) = 0\,. \tag{6.1}$$

The eigenvalues $\mu = \nu^2$ depend upon index n and an additional index m corresponding to the m-th root of the equation above. It is possible to consider the case $a = 0$. The BC at $r = a$ is replaced with the $L^2_{1/r}$-integrability condition for R which eliminates Y_{β_n}, and leads to the characteristic equation:

$$J_{\beta_n}(\nu b) = 0 \quad \Rightarrow \quad \nu = \nu_{n,m} = \frac{r_{m,n}}{b} \quad m = 1, 2\dots \tag{6.2}$$

where $r_{m,n}$ is the m-th root of Bessel function J_{β_n} . The ultimate eigenvectors of the Laplace operator (no normalization) are, for $a > 0$,

$$u_{nm} = \left(J_{\beta_n}(\nu_{n,m} r) - \frac{Y_{\beta_n}(\nu_{n,m} a)}{J_{\beta_n}(\nu_{n,m} a)} Y_{\beta_n}(\nu_{n,m} r) \right) \sin \beta_n \theta \quad n, m = 1, 2 \dots$$

and, for $a = 0$,

$$u_{nm} = J_{\beta_n}(\nu_{n,m} r) \sin \beta_n \theta \quad n, m = 1, 2 \dots .$$

As we said above, in principle, we can start with the problem in r:

$$-r(rR')' - \nu^2 r^2 R = \beta^2 R .$$

The condition assuring the existence of a non-trivial solution will lead again to the transcendental equation (6.1) or (6.2), involving now an *unknown constant* β and, therefore, practically non-solvable until we consider the operator in θ leading to β_n. So, the order in which we solve the 1D Sturm-Liouville problems, does matter.

Laplace equation with homogeneous BCs for $\theta = 0, \alpha$. The homogeneous BCs imply a Sturm-Liouville problem in θ. As above, we obtain

$$\Theta = \sin \beta_n \theta, \quad \beta_n = \frac{n\pi}{\alpha} \quad n = 1, 2 \dots .k$$

The corresponding equation in r now is the Cauchy-Euler equation:

$$r(rR')' = \beta_n^2 R$$

with the solution:

$$R = R_n = C_n r^{\beta_n} + D_n r^{-\beta_n}$$

and the ultimate solution:

$$u = \sum_{n=1}^{\infty} (C_n r^{\beta_n} + D_n r^{-\beta_n}) \sin \beta_n \theta .$$

Constants C_n and D_n are determined from BCs at $r = a$ and $r = b$. In the case of $a = 0$, the finite energy condition $u \in H^1(D)$ implies $D_n = 0$.

Laplace equation with homogeneous BCs for $r = a, b$. We have again,

$$-\frac{r(rR')'}{R} = \frac{\Theta''}{\Theta} = \lambda$$

but, this time, we need to consider the Sturm-Liouville problem in r. The operator is self-adjoint in $L^2_{1/r}(a, b)$ and positive-definite, and we can assume $\lambda = k^2, k > 0$. We still have the Cauchy-Euler equation in r but with more complicated solutions:

$$R = r^{\pm ik} = e^{\pm ik \ln r} = \cos(k \ln r) \pm i \sin(k \ln r) .$$

Alternatively, we can represent the general solution as:

$$R = C\cos(k\ln r) + D\sin(k\ln r)\,.$$

The condition for the existence of non-trivial solutions with the homogeneous BCs is:

$$\begin{vmatrix} \cos(k\ln a) & \sin(k\ln a) \\ \cos(k\ln b) & \sin(k\ln b) \end{vmatrix} = \sin(k\ln\frac{b}{a}) = 0 \quad \Rightarrow \quad k = k_n = \frac{n\pi}{\ln\frac{b}{a}}\,.$$

The ultimate solution is:

$$u = \sum_{n=1}^{\infty}\left(\cos(k_n\ln r) - \frac{\cos(k_n\ln a)}{\sin(k_n a)}\sin(k_n\ln r)\right)(A_n e^{k_n\theta} + B_n e^{-k_n\theta})$$

with constants A_n and B_n determined from BCs on $\theta = 0$ and $\theta = \alpha$.

With either $a = 0$ or $b = \infty$, we run into a continuous spectrum. Indeed, with substitution $t = \ln r$, we have:

$$\int_a^b |e^{ik\ln r}|^2\frac{1}{r}\,dr = \int_{\ln a}^{\ln b}|e^{ikt}|^2\,dt\,,$$

and, if $a = 0$ or $b = \infty$, the function is not $L^2_{1/r}$-integrable. For the case $r \in (0,\infty)$, the spectral transform is equivalent to the standard Fourier transform. Indeed, defining

$$\hat{\phi}(r) := \int_{-\infty}^{\infty}\phi(k')e^{ik'\ln r}\,dk'\,,$$

we obtain,

$$\begin{aligned} N(k)\phi(k) &= \int_0^{\infty}\hat{\phi}(r)e^{-ik\ln r}\,\frac{dr}{r} \\ &= \int_0^{\infty}\int_{-\infty}^{\infty}\phi(k')e^{ik'\ln r}\,dk'\,e^{-ik\ln r}\,\frac{dr}{r} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(k')e^{ik't}\,dk'\,e^{-ikt}\,dt \end{aligned}$$

which gives $N(k) = 1/\sqrt{2\pi}$. The solution to the Laplace equation combines functions

$$\frac{1}{\sqrt{2\pi}}e^{ik\ln r}(A(k)e^{k\theta} + B(k)e^{-k\theta})\,.$$

For instance, in the case of a homogeneous BC for $\theta = 0$, we obtain

$$u(r,\theta) = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{ik\ln r}A(k)\sinh k\theta\,dk\,.$$

The unknown $A(k)$ is obtained by matching the spectral transform of Dirichlet BC data for $\theta = \alpha$,

$$A(k)\sinh k\alpha = \frac{1}{\sqrt{2\pi}}\int_0^{\infty}u_\alpha(r)e^{-ik\ln r}\,\frac{dr}{r}\,.$$

□

6.1.2 Separation of Variables. Hyperbolic Examples

Example 6.1.7 (Vibrating string problem)

We solve the following IVBP.

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - \dfrac{\partial^2 u}{\partial x^2} = 0 & x \in (0,l),\ t > 0 \\ u(0,t) = u(l,t) = 0 & t > 0 \\ u(x,0) = u_0 & x \in (0,l) \\ \dfrac{\partial u}{\partial t}(x,0) = v_0 & x \in (0,l)\,. \end{cases}$$

Assuming: $u = X(x)T(t)$ leads to:

$$-\frac{T''}{T} = -\frac{X''}{X} = \lambda\,.$$

As the problem in x is a Sturm-Liouvillle problem, and the operator is self-adjoint and positive-definite, we can assume for the start that $\lambda = k^2, k > 0$. This leads to:

$$X = A\cos kx + B\sin kx\,.$$

The BCs imply that $A = 0$ and:

$$kl = n\pi \quad \Rightarrow \quad k = k_n = \frac{n\pi}{l}, \quad n = 1, 2, \ldots$$

This, in turn, leads to:

$$T = C\cos k_n t + D\sin k_n t\,,$$

and the final solution is:

$$u(x,t) = \sum_{n=1}^{\infty} \left[C_n \sin(\frac{n\pi}{l}x)\cos(\frac{n\pi}{l}t) + D_n \sin(\frac{n\pi}{l}x)\sin(\frac{n\pi}{l}t) \right]$$

where constants C_n and D_n are determined from ICs. If we assume now specific IC data: $u_0(x) = \sin\frac{\pi x}{l}$, $v_0 = 0$, we obtain the simple solution:

$$u(x,t) = \sin(\frac{\pi}{l}x)\cos(\frac{\pi}{l}t)\,.$$

□

6.1.3 Separation of Variables. Parabolic Examples

Example 6.1.8 (Heat conduction I)

We solve the following IVBP.

$$\begin{cases} \dfrac{\partial u}{\partial t} - \alpha^2 \dfrac{\partial^2 u}{\partial x^2} = 0 & x \in (0,l),\ t > 0 \\ u(0,t) = u(l,t) = 0 & t > 0 \\ u(x,0) = u_0 & x \in (0,l)\,. \end{cases}$$

where $\alpha > 0$. Separating the variables: $u = X(x)T(t)$, we obtain,

$$-\frac{1}{\alpha^2}\frac{T'}{T} = -\frac{X''}{X} = \lambda\,.$$

As usual, self-adjointness in $L^2(0,l)$ and positive definitness of $Au = -u''$ with the homogenous Dirichlet BCs, implies that $\lambda = k^2$, $k > 0$. The Sturm-Liouville problem in x leads to the solutions:

$$X_n = \sin k_n x, \quad k_n = \frac{n\pi}{l}, \quad n = 1, 2, \ldots\,.$$

In turn, the ODE in t gives:

$$T_n = e^{-\alpha^2 k_n^2 t}\,.$$

By superposition, the ultimate solution is:

$$u = \sum_{n=1}^{\infty} c_n \sin k_n x e^{-\alpha^2 k_n^2 t}\,.$$

Constants c_n are determined from the IC:

$$u(x,0) = \sum_{n=1}^{\infty} c_n \sin k_n x = u_0(x), \quad x \in (0,l)\,.$$

As usual, it is critical that $\sin k_n x$ provide an orthogonal basis for $L^2(0,l)$. Multiplying both sides of the equation above with $\sin k_n x$, integrating over $(0,l)$, and using the L^2-orthogonality condition, leads to formulas for the coefficients c_n in terms of u_0,

$$c_n = \frac{2}{l}\int_0^l u_0(x) \sin k_n x\, dx\,.$$

To make the example more concrete, assume a constant IC data: $u_0 = 1$. This gives:

$$c_n = \frac{2}{l}\int_0^1 \sin\frac{n\pi x}{l}\, dx = \begin{cases} \frac{4}{n\pi} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

and the ultimate solution is of the form:

$$u(x,t) = \sum_{\text{odd } n=1}^{\infty} \frac{4}{n\pi} \sin\left(\frac{n\pi}{l}x\right) e^{-\frac{n^2\pi^2\alpha^2}{l^2}t}\,.$$

□

REMARK 6.1.3 Observe that, with growing time t, the higher order harmonics are dying out much faster than the leading term, and the solution converges quickly to:

$$\frac{4}{\pi}\sin\left(\frac{\pi}{l}x\right)e^{-\frac{\pi^2\alpha^2}{l^2}t}\,.$$

Observe also that, even if the initial data u_0 is very localized[¶], solution $u(x,t)\neq 0\,\forall x\in(0,l)$, for arbitrary small time t. The speed of propagation of information for a parabolic problem is infinite. ∎

Example 6.1.9 (Heat conduction II)

How do we proceed if we have a non-homogeneus BC? Consider, for instance, the problem:

$$\begin{cases}\dfrac{\partial u}{\partial t}-\alpha^2\dfrac{\partial^2 u}{\partial x^2}=0 & x\in(0,l),\ t>0\\ u(0,t)=1 & t>0\\ u(l,t)=0 & t>0\\ u(x,0)=0 & x\in(0,l)\,.\end{cases}$$

If we can lift the BC data with a solution to the homogeneous PDE, we are lucky; we can trade the non-homogeneous BC for a non-homogeneous IC. In our case, we can lift the BC data simply with $1-\frac{x}{l}$. Note that the lift vanishes at $x=l$ and it satisfies the heat equation. We look then for the solution in the form:

$$u(x,t)=1-\frac{x}{l}+v(x,t)\,.$$

The new unknown $v(x,t)$ must satisfy the homogeneous heat equation, homogeneous BCs, and a non-homogeneous IC:

$$u(x,0)=0\quad\Rightarrow\quad v(x,0)=\frac{x}{l}-1\,.$$

We can determine then v like in Example 6.1.8, see Exercise 6.1.5. For an alternate technique based on the use of Laplace transform in time, see Example 3.5.1.

□

Example 6.1.10 (Heat conduction III)

We return one more time to the heat problem. This time, the problem is set up on the entire real axis. You can think about a heat diffusion in an infinite rod.

$$\begin{cases}\dfrac{\partial u}{\partial t}-\alpha^2\dfrac{\partial^2 u}{\partial x^2}=0 & x\in\mathbb{R},\ t>0\\ u(x,0)=u_0 & x\in\mathbb{R}\end{cases}$$

[¶]Non-zero in a small subinterval $(x_0-\epsilon,x_0+\epsilon)\subset(0,l)$.

where x_0 is an IC data. Similarly as in Example 3.4.2, the set up on the entire real line is inviting the use of Fourier transform. Transforming the equation and IC in x, we obtain the IVP for the transformed solution $\hat{u}(\omega, x)$:

$$\begin{cases} \hat{u}_{,t} + \alpha^2\omega^2\hat{u} = 0 \quad t > 0 \\ \hat{u}(0) = \hat{u}_0\,. \end{cases}$$

The solution in the Fourier domain is:

$$\hat{u}(\omega, t) = \hat{u}_0(\omega)e^{-\alpha^2\omega^2 t}\,.$$

To compute the inverse Fourier transform, we need a concrete choice of IC data u_0. Heading for a Green function approach, assume $u_0 = \delta_0$. As $\mathcal{F}\delta_0 = 1$, we obtain simply $\hat{u} = e^{-\alpha^2\omega^2 t}$. We compute now the inverse Fourier transform,

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha^2\omega^2 t}e^{i\omega x}\,d\omega\,.$$

We first complete the exponent to a complete square,

$$-\alpha^2\omega^2 t + i\omega x = -\alpha^2 t(\omega - c)^2 + \alpha^2 t c^2 \quad \Rightarrow \quad c = \frac{i\omega x}{2\alpha^2 t}\,.$$

This leads to:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha^2\omega^2 t}e^{i\omega x}\,d\omega = \frac{1}{2\pi}e^{-\frac{x^2}{4\alpha^2 t}}\int_{-\infty}^{\infty} e^{-\alpha^2 t(\omega-c)^2}\,d\omega = \frac{1}{2\pi\alpha\sqrt{t}}e^{-\frac{x^2}{4\alpha^2 t}}\int_{-\infty}^{\infty} e^{-\tau^2}\,d\tau\,.$$

But, see Exercise 6.1.6,

$$\int_{-\infty}^{\infty} e^{-\tau^2}\,d\tau = \sqrt{\pi}\,,$$

so, our Green-like function is given by:

$$U(x,t) := u(x,t) = \frac{1}{2\alpha\sqrt{\pi t}}e^{-\frac{x^2}{4\alpha^2 t}}\,.$$

If we position the delta function at $x = \xi$,

$$u_0(x) = \delta(x - \xi)\,,$$

it is a matter of shifting the coordinate x to obtain,

$$U_\xi(x,t) := U(x-\xi) = \frac{1}{2\alpha\sqrt{\pi t}}e^{-\frac{(x-\xi)^2}{4\alpha^2 t}}\,.$$

By superposition, solution for a general IC data $u_0(\xi)$ is given by:

$$u(x,t) = \int_{\infty}^{\infty} u_0(\xi)\,U(x-\xi)\,d\xi = \frac{1}{2\alpha\sqrt{\pi t}}\int_{\infty}^{\infty} u_0(\xi)\,e^{-\frac{(x-\xi)^2}{4\alpha^2 t}}\,d\xi\,.$$

□

Exercises

Exercise 6.1.1 Let $u = \ln r$, and D be a unit ball in $\mathbb{R}^2$, $D = B(0,1)$. Show that $\|u\|_{H^1(D)} = \infty$.

(5 points)

Exercise 6.1.2 Vibrating membrane problem. Recall the equation governing (free) motion of an elastic membrane:

$$w_{,tt} - \alpha^2 \Delta w = 0$$

where w is the deflection of the membrane, and $\alpha^2 = \frac{T}{\rho}$, with ρ denoting the density (per unit area) of the membrane, and T its tension. The membrane occupies a domain Ω, and it is fixed on its boundary, $w = 0$ on $\Gamma = \partial\Omega$. Use ansatz $w := u(x)e^{i\omega t}$ to turn the equation into the Helmholtz equation,

$$-\Delta u = \left(\frac{\omega}{\alpha}\right)^2 u\,,$$

but treat ω not as a given parameter but an unknown, i.e., study the corresponding eigenvalue problem for (λ, u). The eigenvector u is identified as a *resonant mode* of the membrane, and ω the corresponding *resonant frequency*.

- Consider first a square membrane $\Omega = (0,a)^2$. Adapt solution from Example 6.1.5 to obtain resonant frequencies ω_i for the membrane.
- Repeat the calculations for a circular membrane: $\Omega = \{|x| < a\}$.
- Use the methodology explained in Example 6.1.5 to deduce the solution to the *forced vibrations* problem:

  $$-\Delta u - \omega^2 u = f$$

  where ω is now a forcing frequency, different from any of the resonant frequencies, and function $f \in L^2(\Omega)$ represents a time-harmonic load.
- Comparing the resonant frequencies for the square and circular membranes, can you explain why drums are circular and not rectangular ?

(10 points)

Exercise 6.1.3 Solve the following boundary-value problems.

(i) $-\Delta u = 0$ in $\Omega = \{a < r < b\}$ with BCs: $u(a,\theta) = f(\theta)$, $u(b,\theta) = g(\theta)$.

(ii) $-\Delta u = 1$ in $\Omega = \{r < 1\}$ with BC: $u(1,\theta) = 0$.

(10 points)

Exercise 6.1.4 Solve the BVPs in Fig. 6.2 for the Laplace operator in the conical domain:

$$\Omega = \{(r,\theta) \,:\, 0 < \theta < \alpha,\, r < a\}\,.$$

Here u_n stands for the outward normal derivative.

(25 points)

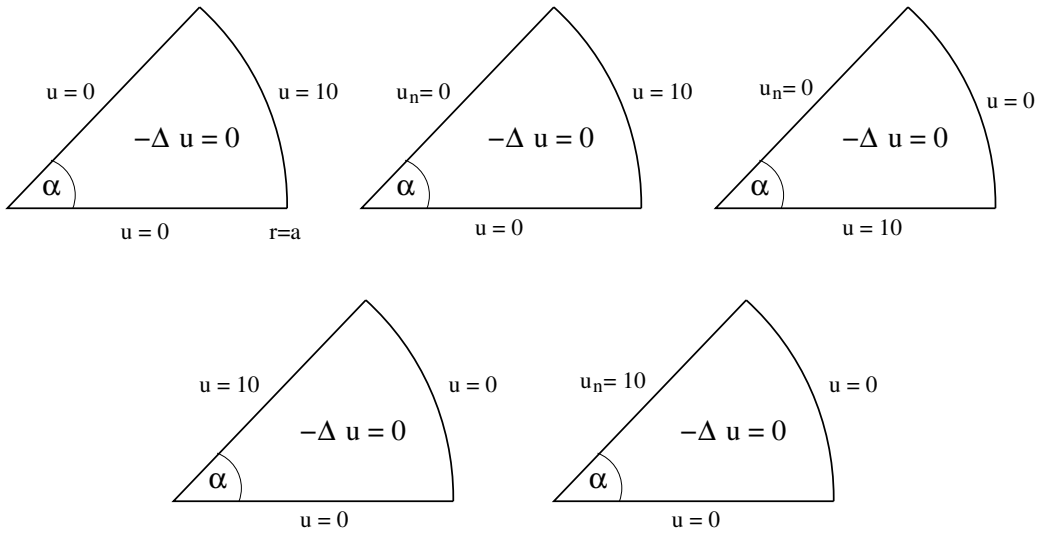

Figure 6.2
Laplace equation in a conical domain.

Exercise 6.1.5 Complete Example 6.1.9.

(5 points)

Exercise 6.1.6 Show that

$$\int_{-\infty}^{\infty} e^{-t^2}\, dt = \sqrt{\pi}\,,$$

Hint: Use

$$\left(\int_{-\infty}^{\infty} e^{-t^2}\, dt\right)^2 = \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\, dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2}\, rdrd\theta\,.$$

(1 point)

6.2 Solution of a Linear PDE of First Order. Characteristics

We discuss the solution of a single, linear, homogeneous PDE of first order with variable coefficients:

$$\sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} = 0\,. \tag{6.3}$$

The unknown, $u = u(x) = u(x_1, \ldots, x_n)$ with $x \in \mathbb{R}^n$, and the coefficients $b_i(x)$ are assumed to be Lipschitz continuous.

Characteristics. Graph of function $x = x(t), t \in (a, b)$ is called *characteristics* of (6.3) if $x(t)$ satisfies the system of autonomous ODEs:

$$\frac{dx}{dt} = b(x)\,. \tag{6.4}$$

Any solution $u(x)$ of (6.3) must be constant along its characteristics. Indeed,

$$\frac{d}{dt}u(x(t)) = \frac{\partial u}{\partial x_i}\frac{dx_i}{dt} = \frac{\partial u}{\partial x_i}b_i(x) = 0 \qquad \text{(summation convention in use.)}$$

The characteristics as curves are unique but their representation as a function of parameter t is not. If we switch from parameter t to a parameter s, $t = t(s)$, we obtain a slightly different system of ODEs:

$$\frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = b(x)\frac{dt}{ds}\,,$$

and yet,

$$\frac{d}{ds}u(x(s)) = \frac{\partial u}{\partial x_i}\frac{dx_i}{dt}\frac{dt}{ds} = (\frac{\partial u}{\partial x_i}b_i(x))\frac{dt}{ds} = 0\,.$$

The uniqueness of the characteristics is restored if we rewrite system (6.4) without parameter t as:

$$\frac{dx_1}{b_1(x)} = \frac{dx_2}{b_2(x)} = \ldots = \frac{dx_n}{b_n(x)}\,. \tag{6.5}$$

Upon eliminating t, we end up with the system of $n-1$ ODEs. We can choose any of the variables as an independent variable or pursue the solution in the form of an exact differential, comp. Section 4.2. Solving system (6.5) may be much easier than solving the original system (6.4).

Example 6.2.1

Determine characteristics for the equation:

$$y\frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} = 0\,.$$

We have:

$$\frac{dx}{y} = -\frac{dy}{x} \quad \Rightarrow \quad 2xdx = -2ydy$$

which leads to $x^2 + y^2 = C$. The characteristics are circles centered at 0. Solving for characteristics in the parametric form,

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

leads to the characteristic equation:

$$-\lambda^2 - 1 = 0$$

and two eigenpairs:

$$\lambda_1 = i,\ d_1 = d = (i, -1) \quad \lambda_2 = -i,\ d_2 = \bar{d} = (-i, -1)\,.$$

The solution is:

$$\begin{aligned}(x,y) &= C_1 d e^{it} + C_2 \bar{d} e^{-it} \\ &= C_1 d(\cos t = I \sin t) + C_2 \bar{d}(\cos t - i \sin t) \\ &= (C_1 d + C_2 \bar{d}) \cos t + i(C_1 d - c_2 \bar{d}) \sin t\end{aligned}$$

This is hardly recognizable as a parametrization for a circle. But

$$C_1 d = C_2 \bar{d} = \Re d \underbrace{(C_1 + C_2)}_{=:D_1} = \Im d \underbrace{i(C_1 - C_2)}_{=:D_2}$$

and, by the same token,

$$i(C_1 d - C_2 \bar{d}) = \Re d D_2 - \Im d D_1 \,.$$

This leads to the solution in the form:

$$\begin{cases} x = D_2 \cos t - D_1 \sin t \\ y = D_1 \cos t + D_2 \sin t \,. \end{cases}$$

Introducing a phase θ such that

$$\cos\theta = \frac{D_2}{D}, \quad \sin\theta = \frac{D_1}{D}, \quad D = \sqrt{D_1^2 + D_2^2}\,,$$

we can rewrite the parametrization as:

$$x = D\cos(t+\theta), \quad y = D\sin(t+\theta)$$

which now can be recognized as a parametrization for a circle.

□

Prime (first) integral. Consider a general system of first order ODEs:

$$\frac{dx}{dt} = b(x,t)\,. \tag{6.6}$$

A function $f(x,t)$ is called a prime (first) integral of system (6.6) if

$$f(x(t),t) = \text{const} \tag{6.7}$$

for any solution $x(t)$ of the system.

Example 6.2.2

Consider motion of a single particle in a conservative force field,

$$m\ddot{x} = F(x) = -\boldsymbol{\nabla} V(x) \quad x \in \mathbb{R}^3$$

where $V(x)$ is the potential (energy) corresponding[‖] to $F(x)$. Introduce velocity v to convert the problem to a system of six equations of first order:

$$\begin{cases} \dot{x} = v \\ \dot{v} = -\frac{1}{m}\boldsymbol{\nabla} V(x)\,. \end{cases} \tag{6.8}$$

The total energy (sum of the kinetic and potential energies):

$$E(x,v) = \frac{1}{2}m|v|^2 + V(x)$$

is the prime (first) integral of system (6.8). Indeed,

$$\frac{d}{dt}E(x(t),v(t)) = mv\cdot\frac{dv}{dt} + \boldsymbol{\nabla} V(x)\cdot\frac{dx}{dt} = -v\cdot\boldsymbol{\nabla} V(x) + \boldsymbol{\nabla} V(x)\cdot v = 0\,.$$

For another example of a prime (first) integral, see Exercise 1.4.3. □

If we know precisely n prime integrals $f_i, i = 1,\dots,n$ for system (6.6), the problem of finding a general solution to (6.6) can be reduced to the solution of a (nonlinear, in general,) system of algebraic equations:

$$f_i(x,t) = c_i \quad i = 1,\dots,n\,.$$

Such a method for solving the system of ODEs is known as the *method of prime integrals*. The following is a baby example of such a methodology.

Example 6.2.3

Consider the system of two ODEs:

$$\begin{cases} \dot{x} = y + 1 \\ \dot{y} = x\,. \end{cases}$$

Adding the equations, we obtain:

$$\frac{d}{dt}(x+y) = x + y + 1 \quad \Rightarrow \quad \frac{d}{dt}(x+y+1) = x + y + 1\,.$$

Consequently,

$$x + y + 1 = C_1 e^t\,,$$

i.e. $e^{-t}(x+y+1) = C_1$ is a prime integral for the system. Similarly, subtracting the equations, we get:

$$\frac{d}{dt}(x-y) = -x + y + 1 = -(x-y-1) \quad \Rightarrow \quad \frac{d}{dt}(x-y-1) = -(x-y-1)\,.$$

Consequently,

$$x - y - 1 = C_2 e^{-t}\,,$$

[‖] Recall that the potential is unique up to an additive constant.

i.e. $e^t(x-y-1)=C_2$ is a prime integral for the system as well. Solving the algebraic system:

$$\begin{cases} x+y+1=C_1e^t \\ x-y-1=C_2e^{-t} \end{cases}$$

for x and y, we obtain the general solution of the original system of ODEs:

$$\begin{aligned} x &= \tfrac{1}{2}(C_1e^t+C_2e^{-t}) \\ y &= \tfrac{1}{2}(C_1e^t-C_2e^{-t})-1\,. \end{aligned}$$

□

Prime integrals are also useful for finding a general solution to (6.3).

THEOREM 6.2.1

Let $\phi_1(x),\ldots,\phi_{n-1}(x)$ be $n-1$ prime integrals of system (6.4). A general solution to the PDE (6.3) is of the form:

$$u=F(\phi_1(x),\ldots,\phi_{n-1}(x)) \tag{6.9}$$

where $F=F(y_1,\ldots,y_{n-1})$ is an arbitrary function of $n-1$ variables.

PROOF

$$\sum_i \frac{\partial u}{\partial x_i}b_i=\sum_i\sum_j \frac{\partial F}{\partial y_j}\frac{\partial \phi_j}{\partial x_i}b_i(x)=\sum_j \frac{\partial F}{\partial y_j}\underbrace{\left(\sum_i \frac{\partial \phi_j}{\partial x_i}b_i(x)\right)}_{=0}=0\,.$$

■

We finish with another baby example.

Example 6.2.4

Find a general solution of:

$$\frac{\partial u}{\partial x}+2\frac{\partial u}{\partial y}+3\frac{\partial u}{\partial z}=0\,.$$

Determine then a particular solution satisfying the initial condition (IC) at $z=0$:

$$u(x,y,0)=x^2+y^2\,.$$

We have:

$$2\,dx=dy \quad \Rightarrow \quad 2x=y+A$$

and

$$3\,dy=2\,dz \quad \Rightarrow \quad 3y=2z+B$$

so $A = 2x - y$ and $B = 3y - 2z$ are two possible prime integrals. A general solution can thus be represented in the form:

$$u(x, y) = F(2x - y, 3y - 2z)$$

where F is an arbitrary function of two variables. The IC implies

$$x^2 + y^2 = F(2x - y, 3y)\,.$$

Solving for $\xi = 2x - y, \eta = 3y$, we obtain:

$$F(\xi, \eta) = (\frac{1}{2}\xi + \frac{1}{6}\eta)^2 + (\frac{1}{3}\eta)^2\,.$$

The particular solution is thus:

$$u(x, y) = (\frac{1}{2}(2x - y) + \frac{1}{6}(3y - 2z))^2 + (\frac{1}{3}(3y - 2z))^2\,. \tag{6.10}$$

If the goal is to find just the solution to the initial-value problem, we can proceed differently. Solution of the characteristics equation is:

$$x = t + a, \quad y = 2t + b, \quad z = 3t + c\,.$$

Let (x_0, y_0, z_0) be coordinates of a point in $\mathbb{R}^3$. If we impose the initial condition for the characteristics: $x(0) = x_0, y(0) = y_0, z(0) = z_0$, we get:

$$x = t + x_0, \quad y = 2t + y_0, \quad z = 3t + z_0\,.$$

The characteristics intersect the $z = 0$ plane at $t = -\frac{1}{3}z_0$, i.e., at the point: $x = x_0 - \frac{1}{3}z_0, y = y_0 - \frac{2}{3}z_0$. As the solution has to be constant along the characteristics, we obtain:

$$u(x_0, y_0, z_0) = (x_0 - \frac{1}{3}z_0)^2 + (y_0 - \frac{2}{3}z_0)^2$$

or, dropping the zero index,

$$u(x, y, z) = (x - \frac{1}{3}z)^2 + (y - \frac{2}{3}z)^2\,.$$

This coincides with (6.10). □

Exercises

Exercise 6.2.1 Consider the equation:

$$y\frac{\partial u}{\partial x} + x\frac{\partial u}{\partial y} = 0\,.$$

Use the method of prime integrals to find the general solution of the equation. Complement then the equation with the initial condition:

$$u(x, 0) = x^4\,,$$

and find the solution to the initial-value problem. Use characteristics to verify the solution.

(10 points)

Exercise 6.2.2 Consider the equation:

$$3\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + 2\frac{\partial u}{\partial z} = 0\,.$$

Use the method of prime integrals to find the general solution of the equation. Complement then the equation with the initial condition:

$$u(x, y, 0) = x + 2y\,,$$

and find the solution to the initial-value problem. Use characteristics to verify the solution.

(10 points)

7

References

[1] L. Demkowicz. Lecture notes on Energy Spaces. Technical Report 13, ICES, 2018.

[2] I.M. Gelfand and S.V. Fomin. *Calculus of Variations*. Dover, 2000.

[3] M. Greenberg. *Foundations of Applied Mathematics*. Prentice Hall, Englewood Cliffs, N.J. 07632, 1978.

[4] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, 1966.

[5] J.T. Oden and L.F. Demkowicz. *Applied Functional Analysis for Science and Engineering*. Chapman & Hall/CRC Press, Boca Raton, 2018. Third edition.