Advanced Analysis

For not-so-advanced mathematicians

Štěpán Zapadlo

Table of contents

Preface

These notes follow the course **Advanced Analysis** (MAT.A0001UF) taught by Univ.-Prof. Dipl.-Ing. Dr.techn. Peter Grabner at the Institute of Analysis and Number Theory, University of Technology, Graz, as it appeared in the Winter semester 2025.

Most of the material comes from my hand-written notes from the lectures, however there are some additions to make the material (in my opinion) more understandable and/or complete. Use of additional material is always explicitly pointed out and cited. Nevertheless, I would like to thank Laura Philomena Mossböck for their LaTeX notes, which I referenced at certain points.

The typesetting was done in Quarto, and hence they come in both PDF and web versions. Feel free to use whichever feels more comfortable for you. They were put together as a preparation for the final exam, and as such are bound to contain typos and errors. I will be very happy if you report and/or fix them, either by email (stepan@zapadlo.name) or by submitting an Issue/Merge Request on the git repository.

1 Fourier Analysis

Definition 1.1 (Fourier transform)**.** Let $f \in \mathcal{L}^1(\mathbb{R})$. Then

$$\hat{f}(t) := \int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x$$

for $t \in \mathbb{R}$ is called the **Fourier transform** of f.

Remark 1.1. From $f \in \mathcal{L}^1$ follows that $\hat{f}(t)$ is defined for all $t \in \mathbb{R}$.

Let us note there are different notions one can use for the Fourier transform, namely

$$\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}x, \quad \text{or} \quad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}x,$$

though all are just normalizations and change of variables w.r.t. each other.

Before further study of this concept, a motivation for it is due. As will be clear later, the Fourier transform turns complicated operations on functions into simpler operations on the transform. Indeed, consider $f, f' \in \mathcal{L}^1(\mathbb{R})$, then

$$\begin{aligned}\int_{-\infty}^{\infty} f'(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x &= \overbrace{f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\big|_{-\infty}^{\infty}}^{0} + 2\pi\mathrm{i}t\int_{-\infty}^{\infty} f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x \\ &= 2\pi\mathrm{i}t\hat{f}(t).\end{aligned} \tag{1.1}$$

This, however, introduces the need for an "inverse" transform, where it turns out one can use

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(t)\mathrm{e}^{2\pi\mathrm{i}tx}\,\mathrm{d}t.$$

Intuitively, we decompose an $\mathcal{L}^1$-function into a collection for corresponding oscillations.

i Decay of $\mathcal{L}^1$ functions

If one recalls the computation (1.1), we used the fact that $f, f' \in \mathcal{L}^1(\mathbb{R})$ to evaluate the boundary condition $f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}|_{-\infty}^{\infty}$ to 0. As $\left|\mathrm{e}^{-2\pi\mathrm{i}tx}\right| = 1$, it poses the question whether indeed for such f it holds that $\lim_{|x|\to\infty} f(x) = 0$.
In this spirit of this question, let us take $f, f' \in \mathcal{L}^1(\mathbb{R})$ and we shall assume there exists a sequence $(x_n)_{n=1}^{\infty}$ such that $x_n \to \infty$ as $n \to \infty$ with $|f(x_n)| \to B > 0$. In other words, we assume $|f|$ either has a positive limit B in infinity or at least B is an accumulation point of

$|f|$ if it does not admit a limit. Firstly, notice that $f' \in \mathcal{L}^1(\mathbb{R})$ implies f is continuous and differentiable (but not necessarily *continuously differentiable*). Furthermore, let $B > \varepsilon > 0$ be chosen arbitrarily then by continuity of f there exists a sequence of disjunct intervals $((a_n, b_n))_{n=1}^{\infty}$ such that for all $n \in \mathbb{N}$ it holds that

- $b_n < a_{n+1} < b_{n+1}$ (equivalent with the intervals being disjunct);
- $x_n \in (a_n, b_n)$;
- $\forall x \in (a_n, b_n)$ we have that $|f(x)| \geq \varepsilon > 0$.

Taking $d_n := b_n - a_n$, we get by $f \in \mathcal{L}^1(\mathbb{R})$ that $d_n \to 0$ as $n \to \infty$ (and everywhere else the function must tend to 0).
We have assumed $|f(x_n)| \to B$, hence for any $\sigma > 0$ follows that there exists an index $N_0 \in \mathbb{N}$ such that for $k > N_0$ we have $|f(x_k) - B| < \sigma$. However, using $f \in \mathcal{L}^1(\mathbb{R})$ for any $\varepsilon > \delta > 0$ there also has to exist some $y \in \mathbb{R}$ such that $|f(x)| < \delta$ for $x \in (y, \infty) \setminus \bigcup_{n=1}^{\infty}(a_n, b_n)$.
Denote as $N > N_0$ an index for which $x_N > y$. Now, for each $k > N_0$ there also exists a $m_k \in (a_k, b_k)$ such that $|f|$ is *increasing* on (a_k, m_k) (and this m_k is maximal). This implies that (since f is differentiable)

$$\underbrace{B - \sigma - \varepsilon}_{C} \leq f(m_k) - f(a_k) = \int_{a_k}^{m_k} f'(s)\,\mathrm{d}s \leq \int_{a_k}^{m_k} |f'(s)|\,\mathrm{d}s.$$

Finally, by $f' \in \mathcal{L}^1(\mathbb{R})$

$$\infty > \int_{\mathbb{R}} |f'(s)|\,\mathrm{d}s \geq \sum_{k=N+1}^{\infty} \int_{a_k}^{m_k} |f'(s)|\,\mathrm{d}s \geq \sum_{k=N+1}^{\infty} C = \infty,$$

which is a contradiction. Therefore, in this setting of $f, f' \in \mathcal{L}^1(\mathbb{R})$ the function f must decay point-wise to 0 at infinity.
For completeness' sake, let us mention one can consider the compactly supported truncation of f and f', i.e., for $\chi_A := \chi_{[-A,A]}$ we have that $f\chi_A \xrightarrow[A\to\infty]{} f$ (and similarly for f') whilst allowing us to vanish the boundary term, without the necessity of having point-wise limits of f at infinity be zero.

Lastly, recall that Fourier series for 1-periodic functions $f \in \mathcal{L}^1(\mathbb{R}/\mathbb{Z})$ we have

$$f(x) \sim \sum_{k\in\mathbb{Z}} a_k \mathrm{e}^{2\pi \mathrm{i} kx},$$

where the coefficients a_k solve the following $\mathcal{L}^2$-problem,

$$\int_0^1 \left| f(x) - \sum_{k=-n}^{n} a_k \mathrm{e}^{2\pi \mathrm{i} kx} \right|^2 \mathrm{d}x,$$

is minimal for all n. In other words, the coefficients a_k must be chosen such that $\sum_{k=-n}^{n} a_k \mathrm{e}^{2\pi \mathrm{i} kx}$ is a projection of f to this lower dimensional subspace. We can prove convergence in the $\mathcal{L}^2$-sense, although not point-wise.

Definition 1.2 (Convolution)**.** Let $f, g \in \mathcal{L}^1(\mathbb{R})$, then

$$f * g(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\,\mathrm{d}y$$

is called the **convolution** of f and g.

Here we note that the point-wise product of $f, g \in \mathcal{L}^1(\mathbb{R})$ is not an $\mathcal{L}^1$ function, but convolution is, as will be shown by the following theorem.

Theorem 1.1. *Let $f, g \in \mathcal{L}^1(\mathbb{R})$. Then*

$$\int_{-\infty}^{\infty} |f(x-y)|\,|g(y)|\,\mathrm{d}y < \infty$$

almost everywhere on $x \in \mathbb{R}$. Furthermore, $f * g \in \mathcal{L}^1(\mathbb{R})$ and $\|f * g\|_1 \leq \|f\|_1 \cdot \|g\|_1$.

Proof. Notice that $F(x, y) = f(x-y)g(y)$ is measurable, if f, g are measurable (which, in turn, follows from $f, g \in \mathcal{L}^1$). Thus,

$$\begin{aligned}\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |F(x,y)|\,\mathrm{d}x \right) \mathrm{d}y &= \int_{\mathbb{R}} \Bigg(\overbrace{\int_{\mathbb{R}} |f(x-y)|\,\mathrm{d}x}^{\|f\|_1} |g(y)| \Bigg) \mathrm{d}y \\ &= \|f\|_1 \cdot \|g\|_1 < \infty\end{aligned}$$

by Fubini's theorem. Therefore $F \in \mathcal{L}^1(\mathbb{R}^2)$, and

$$\int_{\mathbb{R}} |F(x,y)|\,\mathrm{d}y = \int_{\mathbb{R}} |f(x-y)|\,|g(y)| \overset{\text{a.e.}}{<} \infty.$$

Finally,

$$\|f * g\|_1 = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x-y)g(y)\,\mathrm{d}y \right| \mathrm{d}x \leq \int_{\mathbb{R}} \int_{\mathbb{R}} |F(x,y)|\,\mathrm{d}y\,\mathrm{d}x = \|f\|_1 \cdot \|g\|_1 .$$

□

As a reminder we state that $(\mathcal{L}^1(\mathbb{R}), +, *)$ is a *Banach algebra*, i.e., "multiplication" (commutative, distributive, associative) must be compatible with norm, which follows from Theorem 1.1.

Remark 1.2 (Properties of the Fourier transform).

1. $g(x) = f(x)\mathrm{e}^{2\pi\mathrm{i}\alpha x}$, then $\widehat{g}(t) = \widehat{f}(t-\alpha)$;
2. $g(x) = f(x-\alpha)$, then $\widehat{g}(t) = \widehat{f}(t)\mathrm{e}^{-2\pi\mathrm{i}\alpha t}$;
3. $f, g \in \mathcal{L}^1(\mathbb{R})$, then $\widehat{f * g}(t) = \widehat{f}(t) \cdot \widehat{g}(t)$;
4. $g(x) = \overline{f(-x)}$, then $\widehat{g}(t) = \overline{\widehat{f}(t)}$;
5. $g(x) = f\left(\frac{x}{\lambda}\right)$ for $\lambda > 0$, then $\widehat{g} = \lambda\widehat{f}(\lambda t)$;

6. $g(x) = -2\pi \mathrm{i} x f(x)$ such that $g \in \mathcal{L}^1(\mathbb{R})$, then $\widehat{g}(t) = (\widehat{f})'(t)$.

Proof. Properties 1., 2., 4. and 5. should follow trivially from change of variables. Let us first show the third property. Take $f, g, \in \mathcal{L}^1(\mathbb{R})$, then

$$\begin{aligned}
\widehat{f * g}(t) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x-y) g(y) \,\mathrm{d}y \,\mathrm{e}^{-2\pi \mathrm{i} t x} \,\mathrm{d}x \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{R}} \underbrace{\int_{\mathbb{R}} f(x-y) \mathrm{e}^{-2\pi \mathrm{i} t (x-y)} \,\mathrm{d}x}_{\widehat{f}(t)} \, g(y) \mathrm{e}^{-2\pi \mathrm{i} t y} \,\mathrm{d}y \\
&= \widehat{f}(t) \cdot \widehat{g}(t).
\end{aligned}$$

Now let us turn our attention to 6., where we take $g(x) = -2\pi \mathrm{i} x f(x) \in \mathcal{L}^1(\mathbb{R})$. To prove the desired equality, consider the following

$$\frac{\widehat{f}(s) - \widehat{f}(t)}{s-t} = \int_{\mathbb{R}} f(x) \frac{\mathrm{e}^{-2\pi \mathrm{i} s x} - \mathrm{e}^{-2\pi \mathrm{i} t x}}{s-t} \,\mathrm{d}x.$$

Re-arranging $\mathrm{e}^{-2\pi \mathrm{i} s x} - \mathrm{e}^{2\pi \mathrm{i} t x}$ yields

$$\begin{aligned}
\mathrm{e}^{-2\pi \mathrm{i} s x} - \mathrm{e}^{2\pi \mathrm{i} t x} &= \mathrm{e}^{-\pi \mathrm{i} (s+t) x} \left(\mathrm{e}^{-\pi \mathrm{i} (s-t) x} - \mathrm{e}^{-\pi \mathrm{i} (t-s) x} \right) \\
&= 2\mathrm{i} \sin(-\pi (s-t) x) \mathrm{e}^{-\pi \mathrm{i} (s+t) x},
\end{aligned}$$

thus

$$\left| \frac{\mathrm{e}^{-2\pi \mathrm{i} s x} - \mathrm{e}^{-2\pi \mathrm{i} t x}}{s-t} \right| = \frac{\overbrace{\left| 2\mathrm{i} \sin(-\pi (s-t) x) \right|}^{\leq 2\pi |s-t||x|} \overbrace{\left| \mathrm{e}^{-\pi \mathrm{i} (s+t) x} \right|}^{=1}}{|s-t|} \leq 2\pi |x|.$$

Hence

$$\int_{\mathbb{R}} \left| f(x) \right| \left| \frac{\mathrm{e}^{-2\pi \mathrm{i} s x} - \mathrm{e}^{-2\pi \mathrm{i} t x}}{s-t} \right| \mathrm{d}x \leq \int_{\mathbb{R}} \underbrace{\left| f(x) \right| \cdot 2\pi \left| x \right|}_{|g(x)|} \mathrm{d}x < \infty,$$

where $g \in \mathcal{L}^1$ by our assumption. In other words, $\frac{\widehat{f}(s) - \widehat{f}(t)}{s-t}$ is “dominated” by g, see Theorem 4.2, and we may write

$$\begin{aligned}
(\widehat{f})'(t) &= \lim_{s \to t} \frac{\widehat{f}(s) - \widehat{f}(t)}{s-t} \\
&= \int_{\mathbb{R}} f(x) \lim_{s \to t} \frac{2\mathrm{i} \sin(\pi (s-t) x) \mathrm{e}^{-\pi \mathrm{i} (s+t) x}}{s-t} \,\mathrm{d}x \\
&= \int_{\mathbb{R}} -2\pi \mathrm{i} x f(x) \mathrm{e}^{-2\pi \mathrm{i} t x} \,\mathrm{d}x \\
&= \mathcal{F}(-2\pi \mathrm{i} x f(x)),
\end{aligned}$$

where we recalled $\lim_{x \to 0} \frac{\sin(x)}{x} = 1$. ☐

Now we shall discuss how to recover f back from $\widehat{f}$.

Lemma 1.1. *Let* $1 \leq p < \infty$ *and* $f \in \mathscr{L}^p(\mathbb{R})$. *The map*

$$y \mapsto f(\cdot - y) =: f_y,$$

for $y \in \mathbb{R}$, *is uniformly continuous.*

Proof. Let $f \in \mathscr{L}^p(\mathbb{R})$, and $\varepsilon > 0$. Then (by density of $C_c(\mathbb{R})$ in $\mathscr{L}^p$) there exists $g \in C_c(\mathbb{R})$ with $\|f - g\|_p < \varepsilon$. Hence there also exists A such that $\operatorname{supp} g \subseteq [-A, A]$. From compact support and continuity follows *uniform continuity*, i.e., there exists $0 < \delta < A$ such that

$$|s - t| < \delta \implies |g(s) - g(t)| < (3A)^{-\frac{1}{p}} \varepsilon. \tag{1.2}$$

Let us use g as a proxy for dealing with f_y. Using (1.2),

$$\overbrace{\int_{\mathbb{R}} |g(x-s) - g(x-t)|^p \, \mathrm{d}x}^{\text{integral over finite set as } g \in C_c(\mathbb{R})} = \|g_s - g_t\|_p^p < \frac{1}{3A} \varepsilon^p (2A + \delta) < \varepsilon^p.$$

Finally, we can use triangle inequality to obtain

$$\|f_s - f_t\|_p \leq \|f_s - g_s\|_p + \|g_s - g_t\|_p + \|g_t - f_t\| < 3\varepsilon.$$

$\square$

Theorem 1.2 (Riemann-Lebesgue Lemma)**.** *Let* $f \in \mathscr{L}^1(R)$. *Then* $\widehat{f} \in C_0(\mathbb{R})$[1] *and* $\|\widehat{f}\|_\infty \leq \|f\|_1$.

Proof. Surely,

$$\left|\widehat{f}(t)\right| \leq \int_{\mathbb{R}} \Big|f(x) \overbrace{\mathrm{e}^{-2\pi \mathrm{i} t x}}^{\text{factor of } 1}\Big| \, \mathrm{d}x = \|f\|_1.$$

Moreover, this inequality also holds for supremum, which proves the second statement of the theorem.

Now, let us focus on proving $\widehat{f}$ is continuous. Let $t_n \to t$ be a sequence, then

$$\widehat{f}(t_n) - \widehat{f}(t) = \int_{\mathbb{R}} f(x) \left(\mathrm{e}^{-2\pi \mathrm{i} t_n x} - \mathrm{e}^{-2\pi \mathrm{i} t x}\right) \mathrm{d}x.$$

Since $\left|\mathrm{e}^{-2\pi \mathrm{i} t_n x} - \mathrm{e}^{-2\pi \mathrm{i} t x}\right| \leq 2$ (both are complex with modulus 1), we get

$$\left|\widehat{f}(t_n) - \widehat{f}(t)\right| \leq 2 |f(x)|,$$

[1] $C_0(\mathbb{R})$ denotes continuous functions tending to 0.

which by dominated convergence Theorem 4.2 yields

$$\lim_{n\to\infty}\left(\widehat{f}(t_n)-\widehat{f}(t)\right)=0.$$

Lastly, we shall prove that $\widehat{f}$ decays to zero at infinity. Note that this result is called the *Riemann-Lebesgue lemma*. By definition and Remark 1.2 (2.),

$$\begin{aligned}
\widehat{f}(t) &= \int_{\mathbb{R}} f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x\\
&= \overbrace{\mathrm{e}^{\pi\mathrm{i}-\pi\mathrm{i}}}^{=1}\int_{\mathbb{R}} f(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x\\
&= -\int_{\mathbb{R}} f(x)\mathrm{e}^{-2\pi\mathrm{i}t\left(x+\frac{1}{2t}\right)}\,\mathrm{d}x\\
&= -\int_{\mathbb{R}} f\left(x-\frac{1}{2t}\right)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x.
\end{aligned}$$

Thus,

$$2\widehat{f}(t)=\int_{\mathbb{R}}\left(f(x)-f\left(x-\frac{1}{2t}\right)\right)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x,$$

which, in turn, implies

$$2\left|\widehat{f}(t)\right|\le\left\|f-f_{\frac{1}{2t}}\right\|_1.$$

By Lemma 1.1 we know $y\mapsto f_y$ is uniformly continuous, hence for $|t|\to\infty$ we obtain

$$\lim_{|t|\to\infty}\widehat{f}(t)=0.$$

$\square$

Example 1.1 (Gaussian)**.** Let us consider $h(x)=\mathrm{e}^{-\pi x^2}$, to which we want to find $\widehat{h}$ by the means of complex analysis. Firstly, recall that $f(z)=\mathrm{e}^{-\pi z^2}$ is an entire function. Hence, we may employ Cauchy's integral theorem 4.9, i.e.,

$$\oint_C f(z)\,\mathrm{d}z=0=\int_{-R}^{R}\mathrm{e}^{-\pi x^2}\,\mathrm{d}x+\int_{R}^{R+\mathrm{i}t}f(z)\,\mathrm{d}z-\int_{-R}^{R}\mathrm{e}^{-\pi(x+\mathrm{i}t)^2}-\int_{-R}^{-R+\mathrm{i}t}f(z)\,\mathrm{d}z,$$

where C is shown on the diagram Figure 1.1.

Examining integrals along the "vertical" parts of C produces

$$\left|\int_{R}^{R+\mathrm{i}t}f(z)\,\mathrm{d}z\right|=\int_0^t\left|\mathrm{e}^{-\pi\overbrace{(R+\mathrm{i}y)^2}^{(R^2-y^2)}}\right|\mathrm{d}y\le t\mathrm{e}^{-\pi(R^2-t^2)}\xrightarrow[R\to\infty]{}0.$$

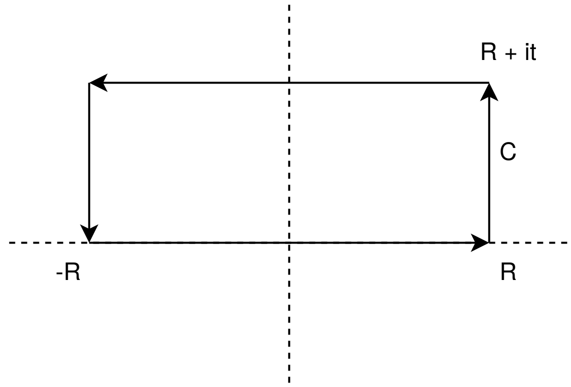

Figure 1.1: Closed curve C for the use of Cauchy's theorem with h.

Moreover, for $R \to \infty$, we also obtain

$$0 = \overbrace{\int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2} \,\mathrm{d}x}^{=1 \text{ (akin to normal dist.)}} - \int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2} \mathrm{e}^{-2\pi x t} \mathrm{e}^{\pi t^2} \,\mathrm{d}x$$

$$\Downarrow$$

$$\widehat{h}(t) = \int_{\mathbb{R}} \mathrm{e}^{-\pi x^2} \mathrm{e}^{-2\pi \mathrm{i} x t} \,\mathrm{d}x = \mathrm{e}^{-\pi t^2}.$$

Finally, by the scaling properties of Fourier transform, see Remark 1.2, we can morph h into h_λ,

$$h_\lambda(x) = \frac{1}{\sqrt{\lambda}} \mathrm{e}^{-\pi \frac{x^2}{\lambda}} \quad \& \quad \widehat{h}_\lambda(t) = \mathrm{e}^{-\pi \lambda t^2}, \tag{1.3}$$

which will become handy for us later.

Remark 1.3. An attentive reader might realize that h_λ with λ representing the time is the **heat kernel**. In other words, $f * h_\lambda$ is the solution to the *heat equation*, i.e.,

$$c u_x = u_{xx} \quad \& \quad u(0, x) = f(x).$$

Lemma 1.2. *Let $f \in \mathcal{L}^1(\mathbb{R})$. Then*

$$f * h_\lambda(x) = \int_{\mathbb{R}} \widehat{f}(t) \mathrm{e}^{-\pi \lambda t^2} \mathrm{e}^{2\pi i t x} \,\mathrm{d}t.$$

Note that for $\lambda \to 0$, we basically recover f or the inverse Fourier transform of $\widehat{f}$ on the right side, respectively.

Proof. By the assumption and Example 1.1, one obtains right away

$$f * h_\lambda(x) = \int_{\mathbb{R}} f(x-y) h_\lambda(y)\,\mathrm{d}y = \int_{\mathbb{R}} f(x-y) \int_{\mathbb{R}} \mathrm{e}^{-\pi\lambda t^2} \mathrm{e}^{2\pi \mathrm{i} t y}\,\mathrm{d}t\,\mathrm{d}y$$
$$\overset{\text{Fubini}}{=} \int \mathrm{e}^{-\pi\lambda t^2} \underbrace{\int_{\mathbb{R}} f(x-y) \mathrm{e}^{2\pi \mathrm{i} t y}\,\mathrm{d}y}_{\mathrm{e}^{2\pi \mathrm{i} t x}\widehat{f}(t)}\,\mathrm{d}t.$$

□

Theorem 1.3. *Let $g \in \mathcal{L}^\infty(\mathbb{R})$ be continuous in (fixed) $x \in \mathbb{R}$. Then*

$$\lim_{\lambda \to 0} g * h_\lambda(x) = g(x).$$

Proof. Using the fact that $\int_{\mathbb{R}} h_\lambda(y)\,\mathrm{d}y = 1$, let us examine

$$\begin{aligned} g * h_\lambda(x) - g(x) &= \int_{\mathbb{R}} (g(x-y) - g(x))\, h_\lambda(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} (g(x-y) - g(x)) \frac{1}{\sqrt{\lambda}} \mathrm{e}^{-\pi \frac{y^2}{\lambda}}\,\mathrm{d}y \\ &\overset{s=\frac{y}{\sqrt{\lambda}}}{=} \int_{\mathbb{R}} (g(x - s\sqrt{\lambda}) - g(x)) \underbrace{\mathrm{e}^{-\pi s^2}}_{\text{bounded}}\,\mathrm{d}s. \end{aligned}$$

As $g \in \mathcal{L}^\infty(\mathbb{R})$, we have $\left|g\left(x - s\sqrt{\lambda}\right) - g(x)\right| \le 2\,\|g\|_\infty$. Thus, there exists majorant of the integrand for the dominated convergence theorem, see Theorem 4.2, and we may pass $\lim_{\lambda\to 0}$ inside of the integral. Finally, by continuity of g we obtain the desired results. □

Theorem 1.4. *Let $1 \le p < \infty$ and $f \in \mathcal{L}^p(\mathbb{R})$. Then*

$$\lim_{\lambda \to 0} \left\| f * h_\lambda - f \right\|_p = 0.$$

Proof. One might observe that $h_\lambda \in \mathcal{L}^q(\mathbb{R})$, such that $\frac{1}{p} + \frac{1}{q} = 1$, and that $f * h_\lambda(x)$ is defined for every x, i.e., the integral is convergent for every x.

Similarly to the proof of Theorem 1.3, we might use the fact that h_λ is a probability measure and compute

$$f * h_\lambda(x) - f(x) = \int_{\mathbb{R}} (f(x-y) - f(x))\, h_\lambda(y)\,\mathrm{d}y.$$

Now by Jensen's inequality, see Proposition 4.1, we get

$$\left| f * h_\lambda(x) - f(x) \right|^p \le \int_{\mathbb{R}} \left| f(x-y) - f(x) \right|^p h_\lambda(y)\,\mathrm{d}y.$$

Integrating both sides w.r.t. to x yields

$$\left\| f * h_\lambda - f \right\|_p^p \leq \int_{\mathbb{R}} \left\| f_y - f \right\|_p^p h_\lambda(y)\, \mathrm{d}y.$$

Since both $y \mapsto f_y$ and norm are continuous, $\left\| f_y - f \right\|_p^p$ is continuous in y and bounded, we can apply Theorem 1.3 to obtain

$$\lim_{\lambda \to 0} \left\| f * h_\lambda - f \right\| = 0.$$

□

Remark 1.4. The main difference between Theorem 1.3 and Theorem 1.4 can be seen in the kind of convergence both theorems describe. While Theorem 1.3 gives us point-wise convergence for bounded and continuous functions, Theorem 1.4 provides an alternative view with $\mathcal{L}^p$-convergence for appropriate functions.

Lastly, Lemma 1.2 simply characterizes the same situation for the special case of $\mathcal{L}^1(\mathbb{R})$, without going into the details on convergence when $\lambda \to 0$.

Let us now turn our attention to the relationship of Fourier transform and it's inverse. Namely, we shall characterize the case when we recover the original function f almost everywhere, with the caveat of needing $\widehat{f} \in \mathcal{L}^1(\mathbb{R})$. However, there are prominent examples of f such that $\widehat{f} \notin \mathcal{L}^1(\mathbb{R})$ — we shall address this issue later.

Theorem 1.5. *If $f, \widehat{f} \in \mathcal{L}^1(\mathbb{R})$, then*

$$g(x) = \int_{\mathbb{R}} \widehat{f}(t) \mathrm{e}^{2\pi i t x}\, \mathrm{d}t, \tag{1.4}$$

with $g \in C_0(\mathbb{R})$, such that $g(x) = f(x)$ almost everywhere for $x \in \mathbb{R}$.

Proof. By Lemma 1.2, we directly have

$$f * h_\lambda(x) = \int_{\mathbb{R}} \widehat{f}(t) \mathrm{e}^{-\pi\lambda t^2} \overbrace{\mathrm{e}^{2\pi i t x}}^{\text{modulus } 1}\, \mathrm{d}t, \tag{1.5}$$

i.e., the integrand is bounded by $\left|\widehat{f}(t)\right|$ and $\widehat{f} \in \mathcal{L}^1(\mathbb{R})$. Hence, we can use the dominated convergence theorem 4.2, yielding

$$g(x) = \lim_{\lambda \to 0} f * h_\lambda(x).$$

Recalling (1.4) and the Definition 1.1 of the Fourier transform (evaluated here at $-x$) of $\widehat{f}$, we get $g \in C_0(\mathbb{R})$ by Theorem 1.2.

Finally, from Theorem 1.4 we know that $\lim_{\lambda \to 0} \left\| f * h_\lambda - f \right\|_p$ and by Theorem 4.3 (which, in turn, relies on Lemma 4.1), we have

$$\underbrace{\lim_{n \to \infty} f * h_{\lambda_n}(x)}_{g(x)} \overset{\text{a.e.}}{=} f(x).$$

□

Corollary 1.1. *Let $f \in \mathcal{L}^1(\mathbb{R})$ and $\widehat{f} \equiv 0$. Then $f(x) \overset{a.e.}{=} 0$.*

Proof. Follows immediately from Theorem 1.5. □

So far, we know that

$$\mathcal{F} : \mathcal{L}^1(\mathbb{R}) \to C_0(\mathbb{R}),$$

and even so, $\mathcal{F}$ is still not surjective. For example, one can define Fourier transform for measures and find measures, which do not have density. Moreover, the inversion formula works only for $\mathcal{L}^1(\mathbb{R}) \cap C_0(\mathbb{R})$. Therefore, let us now focus on finding a suitable spaces X, Y such that $\mathcal{F} : X \to Y$ is *bijective*.

1.1 Fourier Transform on $\mathcal{L}^2$

Firstly, let us note $\mathcal{L}^1 \cap \mathcal{L}^2$ is dense in $\mathcal{L}^2$ (and also in $\mathcal{L}^1$). Moreover, we already have the Fourier transform defined on $\mathcal{L}^1 \cap \mathcal{L}^2$ and our goal will be to extend it to $\mathcal{L}^2$. Without getting ahead of ourselves too much, we shall spoil that we can understand $\widehat{f}$ as an element in $\mathcal{L}^2$, which is captured by the following theorem.

Theorem 1.6 (Plancherel's)**.** *There exists a map $\mathcal{F} : \mathcal{L}^2(\mathbb{R}) \to \mathcal{L}^2(\mathbb{R})$ with the following properties:*

1. *for $f \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$ it holds $\mathcal{F}f = \widehat{f}$;*
2. *$\|f\|_2 = \|\mathcal{F}f\|_2$, i.e., $\mathcal{F}$ is an isometry;*
3. *if $\varphi_A(t) = \int_{-A}^{A} f(x) \mathrm{e}^{-2\pi i t x} \, \mathrm{d}x$ and $\psi_A(x) = \int_{-A}^{A} \mathcal{F}f(t) \mathrm{e}^{2\pi i t x} \, \mathrm{d}t$, then*

$$\lim_{A \to \infty} \|\mathcal{F}f - \varphi_A\|_2 = 0 \quad \& \quad \lim_{A \to \infty} \|f - \psi_A\|_2 = 0,$$

i.e., both integral transforms "work almost everywhere".

Proof. We shall show the existence and properties of $\mathcal{F}$ as results of its behavior on $\mathcal{L}^1 \cap \mathcal{L}^2$, e.g., by employing density arguments. In the end, we will give a concrete $\mathcal{L}^1$-proxy for $\mathcal{F} : \mathcal{L}^2 \to \mathcal{L}^2$ in the limit sense.

Denote $f^*(x) := \overline{f(-x)}$ for $f \in \mathcal{L}^1 \cap \mathcal{L}^2$ and $g := f * f^*$, then by Theorem 1.1 $g \in \mathcal{L}^1$. As

$$g(x) = \int_{\mathbb{R}} f(x - y) \overline{f(-y)} \, \mathrm{d}y = \int_{\mathbb{R}} f(x + y) \overline{f(y)} \, \mathrm{d}y,$$

we can interpret g as a scalar product of f and shifted f_{-x}. From this immediately follows that $g(0) = \|f\|_2^2$ and that $g(x) = \langle f_{-x}, f \rangle_{\mathcal{L}^2}$ is uniformly continuous, since $x \to f_{-x}$ is continuous by Lemma 1.1. By Cauchy-Schwarz inequality 4.4, we get

$$|g(x)| \le \|f\|_2^2,$$

thus $g(x)$ is bounded.

Now, recalling the scaled gaussian "mollifier" h_λ from (1.3), we may invoke Lemma 1.2 in order to obtain

$$g * h_\lambda(0) = \int_{\mathbb{R}} \widehat{g}(t) \mathrm{e}^{-\pi\lambda t^2}\,\mathrm{d}t.$$

By Remark 1.2 (3. & 4.), $\widehat{g}(t) = \widehat{f}(t) \cdot \overline{\widehat{f}(t)} = \left|\widehat{f}(t)\right|^2$, therefore

$$g * h_\lambda(0) = \int_{\mathbb{R}} \left|\widehat{f}(t)\right|^2 \mathrm{e}^{-\pi\lambda t^2}\,\mathrm{d}t,$$

with $\mathrm{e}^{-\pi\lambda t^2}$ monotone increasing for $\lambda \to 0$. On one hand, we use Theorem 1.3 ensured by boundedness and continuity of g, while on the other, employing Theorem 4.1 produces

$$\underbrace{\lim_{\lambda\to 0} g * h_\lambda(0)}_{=g(0)=\|f\|_2^2} = \int_{\mathbb{R}} \left|\widehat{f}(t)\right|^2\,\mathrm{d}t = \left\|\widehat{f}\right\|_2^2.$$

Furthermore, let $Y = \mathcal{F}[\mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})]$. Our goal will be to show that Y is dense in $\mathcal{L}^2(\mathbb{R})$. In particular, since $\mathcal{L}^2(\mathbb{R})$ is a Hilbert space [2] it suffices to show that the orthogonal complement is trivial, i.e., $Y^\perp = \{0\}$.

Consider a function $x \mapsto \mathrm{e}^{2\pi \mathrm{i}\alpha x}\mathrm{e}^{-\pi\lambda x^2} \in \mathcal{L}^1 \cap \mathcal{L}^2$, then applying $\mathcal{F}$ gives (by (1.3) and Example 1.1)

$$\underbrace{t \mapsto \int_{\mathbb{R}} \mathrm{e}^{2\pi \mathrm{i}tx}\mathrm{e}^{-\pi\lambda x^2}\mathrm{e}^{-2\pi \mathrm{i}tx}}_{h_\lambda(t-\alpha)} \in Y.$$

Now take $w \in Y^\perp$ and note

$$(h_\lambda * \overline{w})(\alpha) = \int_{\mathbb{R}} h_\lambda(\alpha - t)\overline{w}(t)\,\mathrm{d}t = \langle h_\lambda(\cdot - \alpha), w\rangle_{\mathcal{L}^2} = 0.$$

However, by Theorem 1.4 we also know

$$\lim_{\lambda\to 0} \|h_\lambda * \overline{w} - \overline{w}\|_2 = 0,$$

which, taking into account the previous computation, necessitates $w \overset{\text{a.e.}}{=} 0$, hence $Y^\perp = \{0\}$. In other words, Y is dense in $\mathcal{L}^2$, thus proving $\mathcal{F}$ is isometry.

Lastly, let $\chi_A := \chi_{[-A,A]}$ be an indicator function, i.e., we may write

$$\varphi_A(t) = \int_{-A}^{A} f(x)\mathrm{e}^{-2\pi \mathrm{i}tx}\,\mathrm{d}x = \int_{\mathbb{R}} (f \cdot \chi_A)\mathrm{e}^{-2\pi \mathrm{i}tx}\,\mathrm{d}x = \widehat{f \cdot \chi_A}(t).$$

Thus by $\mathcal{F}$ being isometry (see above), $\left\|\widehat{f} - \varphi_A\right\|_2^2 = \left\|\widehat{f} - \widehat{f \cdot \chi_A}\right\|_2^2 = \|f - f \cdot \chi_A\|_2^2$, which goes to 0 for $A \to \infty$. Analogously, we can also calculate (recall the definition of ψ_A)

$$\|f - \psi_A\|_2^2 = \left\|f(-x) - \widehat{\widehat{f} \cdot \chi_A}(-x)\right\|_2^2 = \left\|\widehat{f} - \widehat{f} \cdot \chi_A\right\|_2^2 \xrightarrow[A\to\infty]{} 0.$$

□

[2] Hilbert space is a real (or complex) inner product space which is also a complete metric space, i.e., a special case of Banach space.

Corollary 1.2. *Let $f \in \mathcal{L}^2, \widehat{f} \in \mathcal{L}^1$, then $f(x) \overset{a.e.}{=} \int_{\mathbb{R}} \widehat{f}(t) e^{2\pi i t x} \, dt$.*

Remark 1.5 (Heisenberg uncertainty principle). Consider $|f(x)|$ a density of a probability distribution, i.e.,

$$\|f\|_2^2 = \int_{\mathbb{R}} |f(x)|^2 \, dx = 1,$$

such that, for simplicity, its mean value is zero. Now $\int_{\mathbb{R}} x^2 |f(x)|^2 \, dx$ is the variance of said probability distribution. By isometry (see Theorem 1.6) we also know that $\int_{\mathbb{R}} \left|\widehat{f}(t)\right|^2 \, dt = 1$, i.e., $\left|\widehat{f}(t)\right|$ is also a probability distribution with its associated variance.

By computing

$$\int_{\mathbb{R}} x^2 |f(x)|^2 \, dx \cdot \int_{\mathbb{R}} t^2 \left|\widehat{f}(t)\right|^2 \, dt \geq C \|f\|_2^2 \cdot \left\|\widehat{f}\right\|_2^2 = C \|f\|_2^4 = C$$

it is possible to see that if one of the variances is small, the other has to be large in order to compensate. In other words, both the variance of $|f(x)|$ and $\left|\widehat{f}(t)\right|$ cannot be small at the same time.

Let us now determine the value of C. Recall from Remark 1.2 (6.) and the introductory motivation that $\widehat{f'}(t) = 2\pi i t \widehat{f}(t)$ if $f' \in \mathcal{L}^1$. Then by Theorem 1.6

$$\int_{\mathbb{R}} t^2 \left|\widehat{f}(t)\right|^2 \, dt = \frac{1}{4\pi^2} \int_{\mathbb{R}} \left|\widehat{f'}(t)\right|^2 \, dt = \frac{1}{4\pi^2} \int_{\mathbb{R}} |f'(x)|^2 \, dx,$$

and by Cauchy-Schwarz Theorem 4.4, it also follows that

$$\begin{aligned}\frac{1}{4\pi^2} \|x f(x)\|_2^2 \cdot \|f'\|_2^2 &\geq \frac{1}{4\pi^2} |\langle x f(x), f' \rangle|^2 = \frac{1}{4\pi^2} \left| \int_{\mathbb{R}} x f(x) \overline{f'(x)} \, dx \right| \\ &\geq \frac{1}{4\pi^2} \left| \mathfrak{Re} \int_{\mathbb{R}} x f(x) \overline{f'(x)} \, dx \right|^2 .\end{aligned}$$

Since

$$\mathfrak{Re}\left(f(x)\overline{f'(x)}\right) = \frac{1}{2}\left(f(x)\overline{f'(x)} + \overline{f(x)} f'(x)\right) = \frac{1}{2} (\langle f, f \rangle)' = \frac{1}{2}\left(|f(x)|^2\right)',$$

it yields (using integration by parts)

$$\|x f(x)\|_2^2 \cdot \left\|t \widehat{f}(t)\right\|_2^2 \geq \frac{1}{16\pi^2} \left| \int_{\mathbb{R}} x \left(|f(x)|^2\right)' \, dx \right|^2 = \frac{1}{16\pi^2} \left(\int_{\mathbb{R}} |f(x)|^2 \, dx \right)^2 = \frac{1}{16\pi^2} \|f\|_2^4 .$$

Lastly, the above attains equality (from Cauchy-Schwarz) when $f'(x) = C x f(x)$, i.e., $f(x) = e^{-\frac{C}{2} x^2}$ – f needs to be a restriction of an entire function to the real line (and we shall discuss this case in more depth later).

1.2 Fourier Transform as Complex-Valued Continuous Linear Functional

We have already seen that $(\mathcal{L}^1, +, *)$ is a Banach algebra. Now we would like to understand continuous linear function $\varphi : \mathcal{L}^1 \to \mathbb{C}$, such that

$$\varphi(f * g) = \varphi(f) \cdot \varphi(g),$$

e.g., φ could be the Fourier transform evaluated at a given point.

Proposition 1.1. *Using φ as defined above, it holds that*

$$\|\varphi\| = \sup_{\|x\| \leq 1} |\varphi(x)| \leq 1.$$

Proof. We shall prove this result by contradiction. Assume $\|\varphi\| > 1$, then there exists $x \in \mathcal{L}^1$ such that $\|x\| < 1$ and $\varphi(x) = 1$. Let $s_n = -\sum_{i=1}^n x^i$, where x^i is the i-fold convolution of x with itself. Since $\|x^n\| \leq \|x\|^n \xrightarrow[n\to\infty]{} 0$, (s_n) is a Cauchy sequence with $s := \lim_{n\to\infty} s_n$. Then

$$\begin{aligned} x * s_n &= s_n + x - x^n \\ &\Downarrow \text{ when } n \to \infty \\ x * s &= s + x \\ &\Downarrow \\ \varphi(x) \cdot \varphi(s) &= \varphi(s) + \varphi(x) \\ \varphi(s) &= \varphi(s) + 1, \end{aligned}$$

which is a contradiction. □

We have recalled before that $\mathcal{L}^2$ is a Hilbert space[3], and so is $\mathcal{L}^1$. Therefore by Riesz representation theorem 4.5, there exists $\beta \in \mathcal{L}^\infty$ such that $\varphi(f) = \int_{\mathbb{R}} f(x)\beta(x)\,\mathrm{d}x$. Then

$$\begin{aligned} \varphi(f * g) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x-y)g(y)\,\mathrm{d}y \right) \beta(x)\,\mathrm{d}x \\ &\overset{\text{Fubini}}{=} \int_{\mathbb{R}} g(y) \underbrace{\int_{\mathbb{R}} f(x-y)\beta(x)\,\mathrm{d}x}_{\varphi(f_y)}\,\mathrm{d}y \\ &= \int_{\mathbb{R}} g(y)\varphi(f_y)\,\mathrm{d}y, \end{aligned}$$

but also

$$\varphi(f * g) = \varphi(f) \cdot \varphi(g) = \varphi(f) \cdot \int_{\mathbb{R}} g(y)\beta(y)\,\mathrm{d}y.$$

[3]Hilbert space is a real (or complex) inner product space which is also a complete metric space, i.e., a special case of Banach space.

As these equalities hold for any $g \in \mathcal{L}^1$, it implies

$$\varphi(f)\beta(y) \overset{\text{a.e.}}{=} \varphi(f_y), \tag{1.6}$$

which is called the *Cauchy's exponential functional equation.* Even more,

$$\varphi(f_{x+y}) = \varphi(f)\beta(x+y) = \varphi((f_y)_x) = \varphi(f_y)\beta(x) = \varphi(f)\beta(y)\beta(x),$$

thus $\beta(x+y) = \beta(x)\beta(y)$. In addition, since $y \mapsto f_y$ is (uniformly) continuous, by Lemma 1.1, and φ is continuous by definition, it necessitates β is also continuous. Further, requirement $\beta \neq 0$ (at no point, otherwise $\beta \equiv 0$) together with $x = y = 0$ imply $\beta(0) = 1$.

Hence, there exists some $\delta > 0$ such that $\int_0^\delta \beta(x)\,\mathrm{d}x = c \neq 0$ with $c \in \mathbb{C}$. Moreover,

$$\beta(y) \overbrace{\int_0^\delta \beta(x)\,\mathrm{d}x}^{=c} = \int_0^\delta \beta(x+y)\,\mathrm{d}x = \int_y^{y+\delta} \beta(x)\,\mathrm{d}x,$$

so $\beta(y) = \frac{1}{c}\int_y^{y+\delta} \beta(x)\,\mathrm{d}x$. Thus, β is differential, as it is an integral of a continuous function. Lastly, applying $\frac{\mathrm{d}}{\mathrm{d}y}$ to both sides of (1.6) produces

$$\beta'(x+y) = \beta(x)\beta'(y),$$

and evaluating at $y = 0$ gives us

$$\beta'(x) = \beta'(0)\beta(x) \implies \beta(x) = \mathrm{e}^{\beta'(0)x}.$$

Using the fact that $\beta \in \mathcal{L}^\infty$, necessarily $\beta'(0) \in \mathrm{i}\mathbb{R}$, thus $\beta(x) = \mathrm{e}^{-2\pi\mathrm{i}tx}$. To conclude, there exists $t \in \mathbb{R}$ (dependent on φ) such that $\varphi(f) = \widehat{f}(t)$. In other words, the only linear continuous transform respecting convolution is precisely the Fourier transform.

1.3 Fourier Transform on $\mathbb{R}^n$

Remark 1.6. All the ideas used for defining and studying the Fourier transform on $\mathbb{R}$ can be transferred to $\mathbb{R}^n$. Indeed, for $f \in \mathcal{L}^1(\mathbb{R}^n)$ we set

$$\widehat{f}(\boldsymbol{t}) = \int_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\lambda(\boldsymbol{x})$$

where λ is the n-dimensional Lebesgue measure (also sometimes denoted λ^n). The inversion formula follows by similar arguments to the one-dimensional case, and reads

$$f(\boldsymbol{x}) = \int_{\mathbb{R}^n} \widehat{f}(\boldsymbol{t})\mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\lambda(\boldsymbol{t}).$$

Moreover, we can extend the Plancherel's theorem to $\mathcal{L}^2(\mathbb{R}^n)$, obtaining that $\mathcal{F}$ is an isometry on $\mathcal{L}^2(\mathbb{R}^n)$ and that $\mathcal{F}^4 = \mathrm{id}$. Hence, the eigenvalues of $\mathcal{F}$ are the fourth roots of unity, i.e., ± 1 and $\pm i$. In the exercises, we have shown that the Hermite functions

$$H_n(x) = \mathrm{e}^{\pi x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \mathrm{e}^{-2\pi x^2},$$

which are, in fact, of form $H_n(x) = \mathrm{e}^{-\pi x^2} \cdot P_n(x)$ (with P_n a polynomial of degree n), are eigenfunctions of $\mathcal{F}$, i.e.,

$$\mathcal{F}(H_n)(t) = (-i)^n H_n(t).$$

As such, these functions form a complete orthogonal system for $\mathcal{L}^2(\mathbb{R})$.

1.3.1 Fourier transform of Radial Functions

Consider a radial function F on $\mathbb{R}^n$, i.e., $F(\boldsymbol{x}) = f(\|\boldsymbol{x}\|) = f(r)$ for $r \geq 0$. Then by transformation to polar coordinates

$$\int_{\mathbb{R}^n} |F(\boldsymbol{x})| \,\mathrm{d}\lambda(\boldsymbol{x}) \overset{\boldsymbol{x}=r\boldsymbol{y},\, \|\boldsymbol{y}\|=1}{=} \int_0^\infty |f(r)|\, r^{n-1} \,\mathrm{d}r \underbrace{\int_{\mathbb{S}^{n-1}} \mathrm{d}\sigma}_{=:S(n-1)}, \tag{1.7}$$

where σ is the surface measure and $\mathbb{S}^{n-1}$ the unit ball in $\mathbb{R}^n$.

In particular, taking $F(\boldsymbol{x}) = \mathrm{e}^{-\pi\|\boldsymbol{x}\|^2}$ yields

$$\int_{\mathbb{R}^n} F(\boldsymbol{x}) \,\mathrm{d}\lambda(\boldsymbol{x}) = \left(\int_{\mathbb{R}} \mathrm{e}^{-\pi x^2} \,\mathrm{d}x \right)^n = 1,$$

as $e^{-\pi\|\boldsymbol{x}\|^2} = \mathrm{e}^{-\pi x_1^2} \cdot \ldots \cdot \mathrm{e}^{-\pi x_n^2}$. However, from (1.7) it follows that (see Definition 4.1)

$$\begin{aligned} 1 &= \int_0^\infty \mathrm{e}^{-\pi r^2} r^{n-1} \,\mathrm{d}r\, S(n-1) \\ &\overset{t=\pi r^2}{=} \int_0^\infty \mathrm{e}^{-t} \left(\frac{t}{\pi}\right)^{\frac{n-2}{n}} \frac{\mathrm{d}t}{2\pi} S(n-1) \\ &= \frac{S(n-1)}{2\pi \cdot \pi^{\frac{n-2}{2}}} \Gamma\left(\frac{n}{2}\right). \end{aligned}$$

thus

$$S(n-1) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}, \tag{1.8}$$

which we shall use throughout this section as a normalizing constant.

Let us now return back to the general case of $F(\boldsymbol{x})$, then by again applying the polar coordinates $\boldsymbol{x} = r\boldsymbol{y}$ with $\|\boldsymbol{y}\| = 1$,

$$\widehat{F}(\boldsymbol{t}) = \int_{\mathbb{R}^n} F(\boldsymbol{x}) \mathrm{e}^{-2\pi \mathrm{i}\langle \boldsymbol{t}, \boldsymbol{x}\rangle} \,\mathrm{d}\boldsymbol{\lambda}(\boldsymbol{x}) = \int_0^\infty f(r) r^{n-1} \int_{\mathbb{S}^{n-1}} \mathrm{e}^{-2\pi \mathrm{i} r\langle \boldsymbol{t}, \boldsymbol{y}\rangle} \,\mathrm{d}\sigma(\boldsymbol{y}) \,\mathrm{d}r.$$

Firstly, we expect $\widehat{f}(\boldsymbol{t})$ to be radial again, i.e., the inner integral should also depend only on $\|\boldsymbol{t}\|$. Specifically, we may write $\boldsymbol{t} = \|\boldsymbol{t}\|\,\boldsymbol{e}$, where $\boldsymbol{e}$ is any unit vector, such that its transformations correspond to rotations of $\boldsymbol{e}$. By radial symmetry, we fix $\boldsymbol{e}$ to our choice of $\boldsymbol{e}_n$, i.e., the n-th base unit vector of $\mathbb{R}^n$.

Hence

$$\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi \mathrm{i} r\langle \boldsymbol{t},\boldsymbol{y}\rangle}\,\mathrm{d}\sigma(\boldsymbol{y}) = \int_0^\pi \mathrm{e}^{-2\pi \mathrm{i} r\|\boldsymbol{t}\|\cos(\theta)} \sin(\theta)^{n-1}\,\mathrm{d}\theta \cdot C_n, \tag{1.9}$$

such that $C_n \int_0^\pi \sin(\theta)^{n-2}\,\mathrm{d}\theta = S(n-1)$, which follows from integration of all the variables that do not occur in $\boldsymbol{e}_n$. Denote by $\mathrm{B}(\alpha,\beta)$ the Beta function, which reads

$$\mathrm{B}(\alpha,\beta) := \int_0^{\frac{\pi}{2}} \sin(\theta)^{2\alpha-1}\cos^{2\beta-1}\,\mathrm{d}\theta = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

In total, we get by (1.7) and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,

$$C_n = \frac{S(n-1)}{2\mathrm{B}\left(\frac{n-1}{2},\frac{1}{2}\right)} = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n}{2}\right)}{2\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} = \frac{\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \tag{1.10}$$

Continuing with (1.9), notice that (by Taylor expansion)

$$\mathrm{e}^{-2\pi \mathrm{i} r\|\boldsymbol{t}\|\cos(\theta)} = \sum_{k=0}^\infty \frac{1}{k!}\left(-2\pi \mathrm{i} r\,\|\boldsymbol{t}\|\right)^k \cos(\theta)^k,$$

which converges uniformly (allowing us to commute $\sum_{k=0}^\infty$ and $\int_0^\pi \cdot\,\mathrm{d}\theta$). Thus,

$$\begin{aligned}
\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi \mathrm{i} r\langle \boldsymbol{t},\boldsymbol{y}\rangle}\,\mathrm{d}\sigma(\boldsymbol{y}) &= C_n \sum_{k=0}^\infty \frac{1}{k!}\left(-2\pi \mathrm{i} r\,\|\boldsymbol{t}\|\right)^k \underbrace{\int_0^\pi \cos(\theta)^k \sin(\theta)^{n-2}\,\mathrm{d}\theta}_{=0 \text{ for } k \text{ odd}} \\
&= C_n \sum_{k=0}^\infty \frac{1}{(2k)!}(-1)^k \left(2\pi r\,\|\boldsymbol{t}\|\right)^{2k} \cdot 2 \underbrace{\int_0^{\frac{\pi}{2}} \cos(\theta)^{2k} \sin(\theta)^{n-2}\,\mathrm{d}\theta}_{\mathrm{B}\left(k+\frac{1}{2},\frac{n-1}{2}\right) = \frac{\Gamma\left(k+\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{2k+n}{2}\right)}} \\
&= 2C_n\Gamma\left(\frac{n-1}{2}\right) \sum_{k=0}^\infty \frac{(-1)^k}{(2k)!}\left(2\pi r\,\|\boldsymbol{t}\|\right)^{2k} \frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(k+\frac{n}{2}\right)}.
\end{aligned}$$

Now because

$$\Gamma\left(k+\frac{1}{2}\right) = \Gamma\left(\frac{1}{2}\right)\prod_{i=0}^{k-1}\left(\frac{1}{2}+i\right) = \frac{\sqrt{\pi}}{2^k}\cdot 1\cdot 3\cdot 5\cdots(2k-1) = \frac{\sqrt{\pi}}{2^k}\frac{(2k)!}{2^k k!} = \frac{\sqrt{\pi}(2k)!}{4^k k!},$$

refer to the definition of the gamma function 4.1 for more information, then once again resuming on (1.9) with (1.10)

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-2\pi i r\langle \boldsymbol{t}, \boldsymbol{y}\rangle} \, d\sigma(\boldsymbol{y}) &= 2C_n \Gamma\left(\frac{n-1}{2}\right) \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} (2\pi r \|\boldsymbol{t}\|)^{2k} \frac{1}{\Gamma\left(k+\frac{n}{2}\right)} \frac{\sqrt{\pi}(2k)!}{4^k k!} \\
&= 2\sqrt{\pi} \frac{\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \Gamma\left(\frac{n-1}{2}\right) \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma\left(k+\frac{n}{2}\right)} (\pi r \|\boldsymbol{t}\|)^{2k} \\
&= 2\pi^{\frac{n}{2}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma\left(k+\frac{n}{2}\right)} (\pi r \|\boldsymbol{t}\|)^{2k},
\end{aligned}$$

which loosely corresponds to the *Bessel function of index* $\alpha \geq 0$, defined as

$$J_\alpha(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\alpha+1)} \left(\frac{x}{2}\right)^{2k+\alpha}.$$

Note that the Bessel function J_α solves the following *Bessel's differential equation* in $y(x)$,

$$x^2 y'' + 2xy' + (x^2 - \alpha^2) = 0.$$

In total, we get

$$\widehat{f}(\|\boldsymbol{t}\|) = \widehat{F}(\boldsymbol{t}) = 2\pi \frac{1}{\|\boldsymbol{t}\|^{\frac{n}{2}-1}} \int_0^\infty f(r) J_{\frac{n}{2}-1}(2\pi r \|\boldsymbol{t}\|) r^{\frac{n}{2}} \, dr,$$

and similarly,

$$F(\boldsymbol{x}) = \frac{2\pi}{\|\boldsymbol{x}\|^{\frac{n}{2}-1}} \int_0^\infty \widehat{f}(t) J_{\frac{n}{2}-1}(2\pi t \|\boldsymbol{x}\|) t^{\frac{n}{2}} \, dt.$$

Note that this is also called the *Bessel transform*.

1.3.2 Poisson Summation Formula

So far, we have studied various functions on $\mathbb{R}^n$ and how they behave under the Fourier transform. However, we have yet to discuss a very important class of functions, namely the periodic functions. More precisely, we shall be interested in what is a periodic analog to a given function on $\mathbb{R}^n$, i.e., what object corresponds to it on the n-torus $\mathbb{T}_n$ [1, pg. 250-253, 2, chpt. 3.2.2.].

To this end, let us have a function $f : \mathbb{R}^n \to \mathbb{R}$ (or $\mathbb{C}$) and a lattice $\Lambda \subseteq \mathbb{R}^n$, i.e., Λ is a **discrete subgroup**[4] such that $\operatorname{span}_{\mathbb{R}}(\Lambda) = \mathbb{R}^n$. For example,

$$\Lambda = \operatorname{span}_{\mathbb{Z}}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) = \left\{ \sum_{j=1}^{n} w_j \boldsymbol{v}_j \mid w_i, \dots, w_n \in \mathbb{Z} \right\} = \boldsymbol{A}\mathbb{Z}^n,$$

where $\boldsymbol{A} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ is a matrix.

[4] A topological group G is called a *discrete group* if there is not limit point (accumulation point) in it. In other words, for each element in G there exists a neighborhood which only contains that element.

In particular, we shall consider the *periodization of f*

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{\lambda}\in\Lambda} f(\boldsymbol{x}+\boldsymbol{\lambda}), \tag{1.11}$$

which, let us note, is a summation over a **countable set** (or *countable family*), thus it only makes sense if the series is **absolutely** convergent. Indeed, if F is in fact absolutely convergent, one can observe (as was desired) that F from (1.11) is Λ-periodic, i.e., $\forall \boldsymbol{\lambda} \in \Lambda$ it holds that $F(\boldsymbol{x}+\boldsymbol{\lambda}) = F(\boldsymbol{x})$ — shifting by $\boldsymbol{\lambda}$ simply re-arranges the sum.

The absolute convergence of (1.11) we may ensure by a pragmatic assumption

$$\exists C, \varepsilon > 0 \;:\; \forall \boldsymbol{x} \in \mathbb{R}^n \;:\; \left|f(\boldsymbol{x})\right| \le C(1+\|\boldsymbol{x}\|)^{-n-\varepsilon}. \tag{1.12}$$

Note that using only the exponent $-n$ instead of $-n-\varepsilon$ would not guarantee convergence.

Proposition 1.2. *The sum $\sum_{\substack{\boldsymbol{\lambda}\in\mathbb{Z}^n\\ \boldsymbol{\lambda}\neq\boldsymbol{0}}} \|\boldsymbol{\lambda}\|^{-\alpha}$ converges if, and only if $\alpha > n$.*

Proof. Surely,

$$\mathbb{Z}^n \setminus \{\boldsymbol{0}\} = \bigcup_{k=0}^{\infty} \underbrace{\left\{\boldsymbol{\lambda} \in \mathbb{Z}^n \mid 2^k \le \|\boldsymbol{\lambda}\| \le 2^{k+1}\right\}}_{B_k}.$$

Then $\#B_k \asymp 2^{nk}$, where $\asymp$ means that the two quantities are of the same order, i.e., they differ up to a constant multiple. Thus $\boldsymbol{\lambda} \in B_k$ implies $\|\boldsymbol{\lambda}\| \ge 2^k$ and

$$\sum_{\substack{\boldsymbol{\lambda}\in\mathbb{Z}^n\\ \boldsymbol{\lambda}\neq\boldsymbol{0}}} \|\boldsymbol{\lambda}\|^{-\alpha} \le \sum_{k=0}^{\infty} \frac{1}{2^{\alpha k}} \cdot \#B_k \approx \sum_{k=0}^{\infty} \frac{1}{2^{k(\alpha-n)}} < \infty$$

for $\alpha > n$. ☐

Notice that Proposition 1.2 holds also for general lattices and $\sum \|\boldsymbol{x}+\boldsymbol{\lambda}\|$ by offsetting the balls (in the case of a general lattice Λ these become scaled ellipsoids) used for computation, and that

$$\sum_{\boldsymbol{\lambda}\in\Lambda} \frac{C}{(1+\|\boldsymbol{x}+\boldsymbol{\lambda}\|)^{n+\varepsilon}} \le \overbrace{\frac{C}{(1+\|\boldsymbol{x}\|)^{n+\varepsilon}}}^{\boldsymbol{\lambda}=\boldsymbol{0}} + \sum_{\substack{\boldsymbol{\lambda}\in\Lambda\\ \boldsymbol{\lambda}\neq\boldsymbol{0}}} \frac{C}{\|\boldsymbol{x}+\boldsymbol{\lambda}\|^{n+\varepsilon}},$$

where the $\boldsymbol{\lambda} = \boldsymbol{0}$-part does not influence absolute convergence. In other words, Proposition 1.2 along with its proof justify the assumption (1.12).

We shall look for a Fourier(-like) transform for Λ-periodic functions – so far, the functions $\mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}$ formed a complete orthogonal system. Now, our goal will be to decompose $F(\boldsymbol{x})$ into a system of "simple" functions. Ideally, we could just use the "old" orthogonal basis $\mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}$ if we are able to choose $\boldsymbol{t}$ such that these functions are also Λ-periodic, i.e.,

$$\forall \boldsymbol{\lambda} \in \Lambda \;:\; \mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}+\boldsymbol{\lambda}\rangle} = \mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle} \iff \mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{\lambda}\rangle} = 1 \iff \langle \boldsymbol{t}, \boldsymbol{\lambda}\rangle \in \mathbb{Z}.$$

For this purpose we define the **dual lattice** Λ^* of Λ as

$$\Lambda^* := \{\boldsymbol{t} \in \mathbb{R}^n \mid \forall \boldsymbol{\lambda} \in \Lambda : \langle \boldsymbol{t}, \boldsymbol{\lambda} \rangle \in \mathbb{Z}\},$$

and it can be shown that $\Lambda^* = (A^\top)^{-1} \mathbb{Z}^n$.

Proposition 1.3. *The family of functions* $\left\{e^{2\pi i \langle \boldsymbol{t}, \boldsymbol{x} \rangle} \mid \boldsymbol{t} \in \Lambda^*\right\}$ *forms a* ***complete orthogonal*** *system of* Λ*-periodic functions.*

Proof. Firstly, orthogonality follows a similar argument as in the one-dimensional case, see for example [3, chpt. 8.4.2] (for simplicity, one might consider 1-torus-periodic functions). Secondly, notice that the system is closed under multiplication and complex conjugation. Furthermore, since Λ-periodic functions are functions on $\mathbb{R}^n/\Lambda$, which is compact, we may apply Stone-Weierstrass theorem 4.6 to get that

$$\operatorname{span}\left\{e^{2\pi i \langle \boldsymbol{t}, \boldsymbol{x} \rangle} \mid \boldsymbol{t} \in \Lambda^*\right\}$$

is dense in $C(\mathbb{R}^n/\Lambda)$ which, in turn, is dense in $\mathcal{L}^2(\mathbb{R}^n/\Lambda)$. □

We would like to obtain the "Fourier" series for F,

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{\mu} \in \Lambda^*} a_{\boldsymbol{\mu}} e^{2\pi i \langle \boldsymbol{\mu}, \boldsymbol{x} \rangle}, \tag{1.13}$$

which already holds in the $\mathcal{L}^2$ sense by Proposition 1.3. However, we want to obtain a point-wise version of this equality. Recall $\Lambda = \boldsymbol{A}\mathbb{Z}^n$ and that $\mathbb{R}^n/\Lambda$ inherits the Lebesgue measure of $\mathbb{R}^n$, therefore

$$\operatorname{vol}(\mathbb{R}^n/\Lambda) = \operatorname{vol}(\underbrace{\boldsymbol{A}[0,1]^n}_{U_\Lambda}) = |\det \boldsymbol{A}| =: |\Lambda|,$$

where $|\Lambda|$ is the **covolume** of Λ and $\operatorname{vol}(\mathbb{R}^n/\Lambda)$ gives the volume of a unit cell U_Λ. Remember that $F(\boldsymbol{x}) = \sum_{\boldsymbol{\lambda} \in \Lambda} f(\boldsymbol{x} + \boldsymbol{\lambda})$ is absolutely convergent, which allows us to compute $a_{\boldsymbol{\mu}}$ for a given $\boldsymbol{\mu} \in \Lambda^*$ as follows (note that $\int_{\mathbb{R}^n/\Lambda}$ again represents an integral over a unit cell)

$$\begin{aligned}
a_{\boldsymbol{\mu}} &= \frac{1}{|\Lambda|} \int_{\mathbb{R}^n/\Lambda} F(\boldsymbol{x}) e^{-2\pi i \langle \boldsymbol{\mu}, \boldsymbol{x} \rangle} \, d\lambda(\boldsymbol{x}) \overset{\text{abs. conv.}}{=} \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda} \in \Lambda} \int_{U_\Lambda} f(\boldsymbol{x} + \boldsymbol{\lambda}) e^{-2\pi i \overbrace{\langle \boldsymbol{\mu}, \boldsymbol{x} \rangle}^{=\langle \boldsymbol{\mu}, \boldsymbol{x} + \boldsymbol{\lambda} \rangle}} \, d\lambda(\boldsymbol{x}) \\
&= \frac{1}{|\Lambda|} \sum_{\boldsymbol{\lambda} \in \Lambda} \int_{\boldsymbol{\lambda} + U_\Lambda} f(\boldsymbol{x}) e^{-2\pi i \langle \boldsymbol{\mu}, \boldsymbol{x} \rangle} \, d\lambda(\boldsymbol{x}) = \frac{1}{|\Lambda|} \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-2\pi i \langle \boldsymbol{\mu}, \boldsymbol{x} \rangle} \, d\lambda(\boldsymbol{x}) \\
&= \frac{1}{|\Lambda|} \widehat{f}(\boldsymbol{\mu}).
\end{aligned}$$

Since we required absolute convergence of F, it must again be attained even by (1.13). In particular, placing the same assumption (1.12) on $\widehat{f}$ ensures the absolute convergence of (1.13)[5]. Then

$$F(\boldsymbol{x}) = \frac{1}{|\Lambda|} \sum_{\boldsymbol{\mu} \in \Lambda^*} \widehat{f}(\boldsymbol{\mu}) e^{2\pi i \langle \boldsymbol{\mu}, \boldsymbol{x} \rangle} = \sum_{\boldsymbol{\lambda} \in \Lambda} f(\boldsymbol{x} + \boldsymbol{\lambda}), \tag{1.14}$$

[5]While [1] uses the pragmatic assumption (1.12) for both f and $\widehat{f}$, [2] only restrict f this way and requires $\widehat{f}$ to be absolutely convergent on the lattice.

which is called the **Poisson summation formula**. Most frequently it is used for $\boldsymbol{x} = 0$ as

$$\frac{1}{|\Lambda|} \sum_{\boldsymbol{\mu} \in \Lambda^*} \widehat{f}(\boldsymbol{\mu}) = \sum_{\boldsymbol{\lambda} \in \Lambda} f(\boldsymbol{\lambda}).$$

Remark 1.7. In 2016, researchers employed the Poisson summation formula to solve optimal sphere packing problem in $n = 8$. They worked with a function f such that $\widehat{f}(\boldsymbol{t}) \geq 0$ and $f(\boldsymbol{\lambda}) \leq 0$ for $\|\boldsymbol{\lambda}\| \geq \min\{\|\boldsymbol{v}\| \mid \boldsymbol{v} \in \Lambda \,|\boldsymbol{0}|\}$, then $\frac{1}{|\Lambda|}\widehat{f}(\boldsymbol{0}) \leq f(\boldsymbol{0})$ and if f is a radial function, then f can "detect a ball" by being positive inside and negative outside of it. However, such function does not exists exist for every dimension. For $n = 8$, numerical experiments suggested there was such a function, namely the Gaussian-polynomial function.

1.4 Holomorphic Fourier Transform

In this section, we shall pay attention ot the growth of entire functions. Recall that restriction of a certain entire function to $\mathbb{R}$ happened to be lie in $\mathcal{L}^2(\mathbb{R})$ and thus existed their Fourier-Plancherel transform. In particular, we investigate the relationship between the growth of a function and the size of the support of its Fourier transform.

In the spirit of these findings, let $f \in \mathcal{H}(\mathbb{C})$, i.e., f is an entire function. If $\left|f(z)\right| \leq C(1 + |z|)^m$ then by results from complex analysis we know that f must be a polynomial — however a polynomial is unbounded in every direction, i.e., its restriction to $\mathbb{R}$ would not be in $\mathcal{L}^2$. Additionally, if f is bounded, then by theorem of Liouville 4.7 we get f is constant — again not in $\mathcal{L}^2$.

For this reason, consider $f \in \mathcal{H}(\mathbb{C})$ such that

$$\left|f(z)\right| \leq C\mathrm{e}^{A|z|^\alpha}$$

for $A, \alpha > 0$.

Definition 1.3 (Order & Type [4])**.** Let $f \in \mathcal{H}(\mathbb{C})$ and denote $M_f(r) = \max_{|z|=r} \left|f(z)\right|$. We say that f is a *function of finite order* if there exists $k > 0$ such that

$$M_f(r) > \mathrm{e}^{r^k}$$

holds for all sufficiently large r ($r > r_0(k)$). The infimum α among all such k is called the **(exact) order** of f, and can be equivalently characterized as

$$\mathrm{e}^{r^{\alpha-\varepsilon}} < M_f(r) < \mathrm{e}^{r^{\alpha+\varepsilon}}$$

for all sufficiently large r and any $\varepsilon > 0$.

Moreover, we define the **type** A of an entire function f *of order* α as the infimum of all $A' > 0$ which satisfy

$$M_f(r) < \mathrm{e}^{Ar^\alpha}$$

for sufficiently large r.

Proposition 1.4. *Let $f \in \mathcal{H}(\mathbb{C})$. Then f is an entire function of order α and type A if and only if the following equalities hold*

$$\alpha = \limsup_{|z|\to\infty} \frac{\log\log|f(z)|}{\log|z|} \quad \& \quad A = \limsup_{|z|\to\infty} \frac{\log|f(z)|}{|z|^{\alpha}}.$$

It can be shown that for $\alpha < 1$, such entire functions **cannot be bounded** on a line. Moreover, functions of order $\alpha > 1$ and type $2\pi A$, i.e., $|f(z)| \leq Ce^{2\pi A|z|}$ for all $z \in \mathbb{C}$, like sin, cos are bounded on $\mathbb{R}$. However they grow on the imaginary axis, and on $\mathbb{R}$ they are still not in $\mathcal{L}^2$. This finally leads us to the following theorem.

Theorem 1.7 (Paley-Wiener)**.** *Let $f \in \mathcal{H}(\mathbb{C})$ be an entire function such that $f|_{\mathbb{R}} \in \mathcal{L}^2(\mathbb{R})$. If f is of exponential type $2\pi A$ (and $\alpha = 1$), then there exists $F \in \mathcal{L}^2([-A, A])$ such that*

$$f(z) = \int_{-A}^{A} F(t)e^{2\pi i t z}\,dt.$$

Note that, in particular, we have now connected complex analysis, more precisely the growth of an entire function, with the (truncated) Fourier transform.

Proof. Let $f \in \mathcal{H}(\mathbb{C})$ be an entire function and consider its mollifier version $f_\varepsilon(x) = f(x)e^{-\varepsilon|x|}$ for $x \in \mathbb{R}$. Assume that

$$\lim_{\varepsilon\to 0} \int_{\mathbb{R}} f_\varepsilon(x)e^{-2\pi i t x}\,dx = 0 \tag{1.15}$$

for $t > |A|$. By the definition of f_ε we have $\lim_{\varepsilon\to 0} \|f - f_\varepsilon\|_2 = 0$, thus by Plancherel's theorem 1.6 we also get $\lim_{\varepsilon\to 0} \|\widehat{f} - \widehat{f_\varepsilon}\|_2 = 0$. Then from our assumption (1.15) follows that $\widehat{f}$ is supported only on $[-A, A]$, which proves the theorem. In other words, it suffices to show (1.15) holds.

For $\alpha \in [-\pi, \pi]$ define Γ_α as the ray given by $z = e^{i\alpha}s$ with $s \in [0, \infty)$, and $\Pi_\alpha = \{z \in \mathbb{C} \mid \Re(ze^{i\alpha}) > A\}$ is a half-plane, see Figure 1.2.

Let $w \in \Pi_\alpha$ and set

$$\phi_\alpha(w) = \int_{\Gamma_\alpha} f(z)e^{-2\pi w z}\,dz \overset{z=e^{i\alpha}s}{=} e^{i\alpha}\int_0^\infty f(e^{i\alpha}s)e^{-2\pi i w e^{i\alpha}s}\,ds,$$

then surely (given f is of exponential type $2\pi A$)

$$|f(e^{i\alpha}s)| \leq Ce^{2\pi As} \implies |f(e^{i\alpha}s)| \cdot \left|e^{-2\pi i w e^{i\alpha}s}\right| \leq Ce^{2\pi As - 2\pi\Re(we^{i\alpha})s}.$$

In other words, ϕ_α converges for $w \in \Pi_\alpha$ (above we show it is bounded by an $\mathcal{L}^1$ function, then it follows from dominated convergence theorem 4.2). Continuity of ϕ_α follows from the fact that f is entire. Applying the theorems of Fubini and Morera 4.8 when considering integrals over closed smooth curves in Π_α yields holomorphy of ϕ_α. In particular, for $\alpha = 0$ and $\alpha = \pi$ we obtain that ϕ_0 is holomorphic for $\Re w > 0$ and ϕ_π for $\Re w < 0$, respectively, see Figure 1.3.

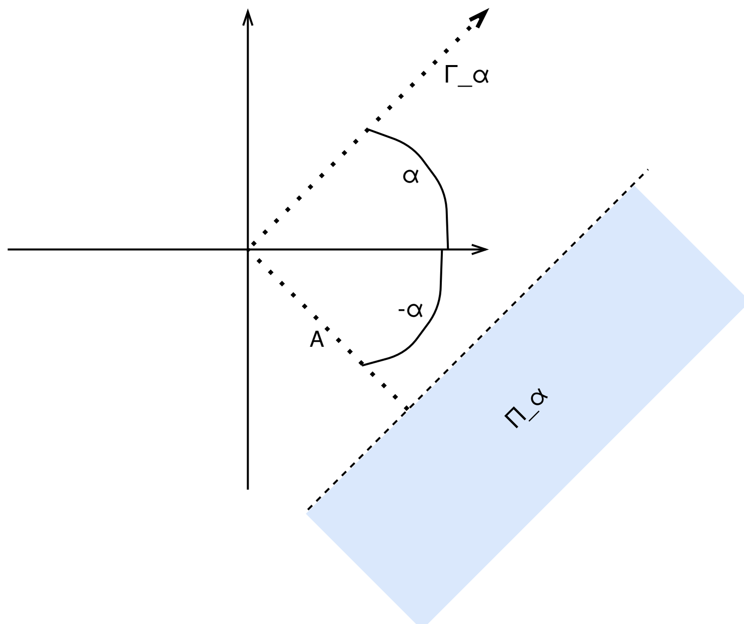

Figure 1.2: Illustration of Γ_α and Π_α

Finally, we can calculate

$$\int_{\mathbb{R}} f_\varepsilon(x)\mathrm{e}^{-2\pi\mathrm{i}tx}\,\mathrm{d}x = \overbrace{\int_0^\infty f(x)\mathrm{e}^{-(\varepsilon+2\pi\mathrm{i}t)x}\,\mathrm{d}x}^{\phi_0(\frac{\varepsilon}{2\pi}+\mathrm{i}t)} - \overbrace{\int_0^\infty f(x)\mathrm{e}^{-(-\varepsilon+2\pi\mathrm{i}t)(-x)}\,\mathrm{d}x}^{\phi_\pi(\frac{-\varepsilon}{2\pi}+\mathrm{i}t)}$$
$$= \phi_0\left(\frac{\varepsilon}{2\pi}+\mathrm{i}t\right) - \phi_\pi\left(\frac{-\varepsilon}{2\pi}+\mathrm{i}t\right),$$

i.e., the two values need to have the same limit as $\varepsilon \to 0$. We shall show this by demonstrating that any of two ϕ_α and ϕ_β agree on the intersection of their domains. Put differently, we shall prove that ϕ_α and ϕ_β are analytic continuations of each other. If we have that, we can replace ϕ_0 and ϕ_π by $\phi_{\frac{\pi}{2}}$ if $t < -A$ and by $\phi_{-\frac{\pi}{2}}$ if $t > A$, from which it follows rather obviously.

Take $0 < \beta - \alpha < \pi$ and $w \in \Pi_\alpha \cap \Pi_\beta \neq \varnothing$ and consider the R-radius curve $\mathcal{C}_R$, see Figure 1.4. By Cauchy's integral theorem 4.9 we know that $\oint_{\mathcal{C}_R} f(z)\mathrm{e}^{-2\pi wz}\,\mathrm{d}z = 0$. Certainly, for $z \in \mathcal{A}_R$ (the connecting arc of $\mathcal{C}_R$) holds that $|f(z)| \leq \mathrm{e}^{2\pi AR}$, and

$$\left|\mathrm{e}^{-2\pi wz}\right| \leq \mathrm{e}^{-2\pi R\cdot \overbrace{\min(\mathfrak{Re}\,\mathrm{e}^{\mathrm{i}\alpha}w,\mathfrak{Re}\,\mathrm{e}^{\mathrm{i}\beta}w)}^{>A}}.$$

Hence,

$$\left|\int_{\mathcal{A}_R} f(z)\mathrm{e}^{-2\pi wz}\,\mathrm{d}z\right| \leq \mathrm{e}^{2\pi RA}\cdot\mathrm{e}^{-2\pi R\min(\mathfrak{Re}\,\mathrm{e}^{\mathrm{i}\alpha}w,\mathfrak{Re}\,\mathrm{e}^{\mathrm{i}\beta}w)}\cdot R\cdot(\beta-\alpha) \xrightarrow[R\to\infty]{} 0.$$

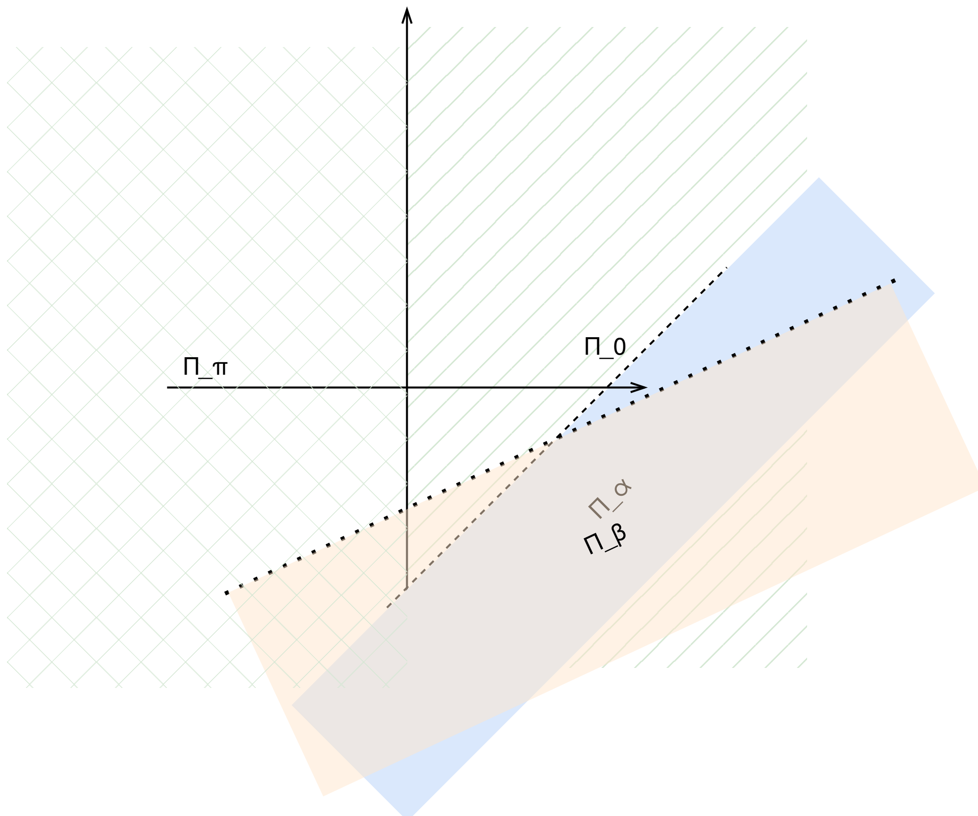

Figure 1.3: Illustration of various Π_α

Then

$$\begin{aligned}\phi_\alpha(w) - \phi_\beta(w) &= \int_{\Gamma_\alpha} f(z)\mathrm{e}^{-2\pi wz}\,\mathrm{d}z - \int_{\Gamma_\beta} f(z)\mathrm{e}^{-2\pi wz}\,\mathrm{d}z \\ &= \lim_{R\to\infty}\left(\oint_{C_R} f(z)\mathrm{e}^{-2\pi wz}\,\mathrm{d}z - \int_{A_R} \mathrm{e}^{-2\pi wz}\,\mathrm{d}z\right) = 0,\end{aligned}$$

i.e., ϕ_α and ϕ_β agree on the intersection of their domains. Thus, as we have suggested before, for $t > A$ we use

$$\phi_0\left(\frac{\varepsilon}{2\pi} + \mathrm{i}t\right) - \phi_\pi\left(\frac{-\varepsilon}{2\pi} + \mathrm{i}t\right) = \phi_{-\frac{\pi}{2}}\left(\frac{\varepsilon}{2\pi} + \mathrm{i}t\right) - \phi_{-\frac{\pi}{2}}\left(\frac{-\varepsilon}{2\pi} + \mathrm{i}t\right) \xrightarrow[\varepsilon\to 0]{} 0$$

and similarly for $t < -A$ with $\phi_{\frac{\pi}{2}}$. □

1.4.1 Phragmén-Lindelöf principle

Let us attempt to generalize the maximum principle to unbounded domains $U \subseteq \mathbb{C}$. In particular, taking $f \in \mathcal{H}(U) \cap C(\overline{U})$ we would like to get that f bounded on ∂U implies f is bounded on U.

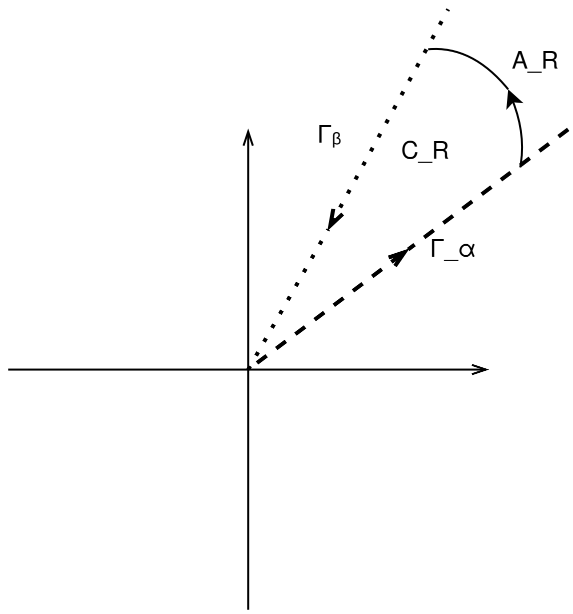

Figure 1.4: R-radius curve $\mathcal{C}_R$ between Γ_α and Γ_β with the connecting arc $\mathcal{A}_R$

Unfortunately for us, there exist counterexamples for $U = \mathbb{C}$ (or, in general, for unbounded U and any f as defined above). However, if we add a growth condition to f we can prove this statement (and for each type of unbounded U we get a different growth condition).

Theorem 1.8 (Phragmén-Lindelöf)**.** *Let $0 < \beta < \pi$, $\alpha < \frac{\pi}{2\beta}$ and set $S_\beta := \{z \in \mathbb{C} \mid |\operatorname{Arg} z| < \beta\}$. Consider f holomorphic on S_β and continuous on $\overline{S_\beta}$, i.e., $f \in \mathcal{H}(S_\beta) \cap \mathcal{C}(\overline{S_\beta})$. If there exists $A > 0$ such that $|f(z)| \leq e^{A|z|^\alpha}$ and f is bounded on ∂S_β, then f is bounded on S_β.*

Example 1.2. There are functions bounded on the boundary, yet unbounded on the interior. Take, for example, $\alpha = \frac{\pi}{2\beta}$ and $f = e^{Az^\alpha}$. Setting $\beta = \frac{\pi}{2}$ clearly gives that ∂S_β is the imaginary line where f is bounded, whereas it is clearly unbounded on the real line.

Proof. Take γ with $\alpha < \gamma < \frac{\pi}{2\beta}$ and $\varepsilon > 0$. Define the function

$$g_\varepsilon(z) = f(z)e^{-\varepsilon z^\gamma}$$

and notice g_ε is holomorphic on S_β and continuous on $\overline{S_\beta}$. Furthermore,

$$g_\varepsilon(e^{\pm i\beta}t) = f(e^{\pm i\beta}t)e^{-\varepsilon e^{\pm i\beta\gamma}t^\gamma} \quad \& \quad \mathfrak{Re}\left(e^{\pm i\beta\gamma}\right) > 0,$$

since $\beta\gamma < \frac{\pi}{2}$. In total, we have $|g_\varepsilon(e^{\pm i\beta}t)| \leq 1$ for $t \geq 0$. Also, it follows that $\left|e^{-\varepsilon e^{\pm i\beta\gamma}t^\gamma}\right| = e^{-\varepsilon \mathfrak{Re}\left(e^{\pm i\beta\gamma}\right)t^\gamma} \leq 1$, therefore $|g_\varepsilon(e^{\pm i\beta}t)| \leq |f(e^{\pm i\beta t}t)|$.

Thus, for any $R > 0$ and $\varphi \in [-\beta, \beta]$, we have

$$\left|g_\varepsilon(R\mathrm{e}^{\mathrm{i}\varphi})\right| \le \mathrm{e}^{AR^\alpha}\mathrm{e}^{-\varepsilon R^\gamma \mathrm{e}^{\mathrm{i}\varphi\gamma}} \le \mathrm{e}^{AR^\varphi - \varepsilon\cos(\beta\gamma)R^\gamma},$$

which holds since $\mathfrak{Re}\left(\mathrm{e}^{\mathrm{i}\varphi\gamma}\right) \ge \cos(\beta\gamma) > 0$. Thus, for $r \ge R$ large enough, we have $\left|g_\varepsilon(r\mathrm{e}^{\mathrm{i}\varphi})\right| \le 1$.

Consider the truncated sector

$$S_{\beta,R} := \left\{r\mathrm{e}^{\mathrm{i}\varphi} \mid 0 < r < R,\ \varphi \in (-\beta,\beta)\right\} \subset S_\beta.$$

Then $S_{\beta,R}$ is bounded and by maximum modulus principle 4.10 the maximum of $\left|g_\varepsilon\right|$ is attained on the boundary $\partial S_{\beta,R}$, i.e., on the $\pm\beta$-rays or the R-diameter arc. We assume that f is bounded on ∂S_β (which are the two $\pm\beta$-rays), hence $\left|g_\varepsilon\right| \le \left|f(z)\right| \le M$ for $z = r\mathrm{e}^{\pm\mathrm{i}\beta}$ and $0 < r < R$. By the above, $\left|g_\varepsilon(R\mathrm{e}^{\mathrm{i}\varphi})\right| \le 1$ for $\varphi \in [-\beta, \beta]$. In total, $\left|g_\varepsilon(z)\right| \le \max\{M, 1\}$ for $z \in S_{\beta,R}$.

Taking the limit $R \to \infty$ (where the boundedness on the arc still necessarily holds) gives that $\sup_{z\in S_\beta}\left|g_\varepsilon(z)\right| \le \max\{M,1\}$. Letting $\varepsilon \to 0$ produces $\mathrm{e}^{-\varepsilon z^\gamma} \to 1$ uniformly on compact subsets. Because the supremum bound does not depend on ε, we finally obtain $\sup_{z\in S_\beta}\left|f(z)\right| \le \max\{M,1\}$. □

Recall the uncertainty principle example 1.5 we have discussed above. There, we related the variance of f and $\hat{f}$ to each other, giving a lower bounded for product of the variances. Using the growth condition on f and $\hat{f}$ and the Phragmen-Lindelöf principle, we can strengthen this result to produce a single (small family of) distribution for f.

Theorem 1.9 (Hardy's uncertainty principle). *Let $f : \mathbb{R} \to \mathbb{C}$ be such that $\left|f(x)\right| \le \mathrm{e}^{-\pi x^2}$ and $\left|\hat{f}(t)\right| \le \mathrm{e}^{-\pi t^2}$. Then $f(x) = C\mathrm{e}^{-\pi x^2}$ for some $C > 0$.*

> **ℹ Note**
>
> I am not sure whether the proof was given in the lecture. In any case, I refer to [5] for more details and the proof.

Corollary 1.3. *Let $a, b > 0$ with $ab > 1$ and let $f : \mathbb{R} \to \mathbb{C}$ be such that $\left|f(x)\right| \le \mathrm{e}^{-\pi a x^2}$ and $\left|\hat{f}(t)\right| \le \mathrm{e}^{-\pi b t^2}$. Then $f \equiv 0$.*

Proof. Let, without loss of generality (we can make a change of variables, if necessary), $a = 1$. Then $\left|\hat{f}(t)\right| \le \mathrm{e}^{-\pi b t^2} \le \mathrm{e}^{-\pi t^2}$ as $b > 1$. By Theorem 1.9, $f(x) = c\mathrm{e}^{-\pi x^2}$ and thus (see Example 1.1) $\hat{f}(t) = C\mathrm{e}^{-\pi t^2}$. However, $C\mathrm{e}^{-\pi t^2} \le \mathrm{e}^{-\pi b t^2}$ for all $t > 0$, which only holds if $C = 0$. This, in turn, implies $c = 0$. □

Remark 1.8. Paley and Wiener have a small book [6][6], where they discuss a large collection of similar results. For example, one of the theorems concerns a generalization of Corollary 1.3 where we allow more growth on f. Then it can be shown, after a lot more work, that f must be polynomial.

[6] To the best of my knowledge, Prof. Grabner meant this book.

2 Distributions

2.1 Introduction

As a preparation, we shall study functionals on $C_c^\infty(\mathbb{R}^n)$, i.e., infinitely-differentiable compactly-supported functions. Before we show an example of a function from this class, consider $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 0, & x \le 0, \\ \mathrm{e}^{-\frac{1}{x}}, & x \ge 0, \end{cases}$$

which surely lies in $C^\infty(\mathbb{R})$. Now, set $\varphi : \mathbb{R}^n \to \mathbb{R}$ as $\varphi(\boldsymbol{x}) = cf(1 - \|\boldsymbol{x}\|)$ such that c is chosen to satisfy $\int_{\mathbb{R}^n} \varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$. Clearly, $\operatorname{supp}\varphi = \overline{B(\boldsymbol{0}, 1)}$, hence $\varphi \in C_c^\infty(\mathbb{R}^n)$.

Moreover, let $f \in \mathcal{L}^1(\mathbb{R}^n)$ and notice $\varphi_\varepsilon(\boldsymbol{x}) = \frac{1}{\varepsilon^n}\varphi\left(\frac{1}{\varepsilon}\boldsymbol{x}\right)$ is supported on $\overline{B(\boldsymbol{0}, \varepsilon)}$, whilst maintaining $\int_{\mathbb{R}^n} \varphi_\varepsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$. Then $f * \varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ by the definition of convolution 1.2.

Definition 2.1 (Urysohn function)**.** Let (X, τ) be a topological space and $V, U \subseteq X$ be open such that $V \subseteq \overline{V} \subseteq U$. We call a continuous map $f : X \to \mathbb{R}$ a *Urysohn function* for U and V, if $\mathbb{I}_{\overline{V}} \le f \le \mathbb{I}_U$.

Remark 2.1. In other words, the Urysohn function acts as *continuous* interpolant between $\mathbb{I}_{\overline{V}}$ and $\mathbb{I}_U$.

Can we find a "better" interpolation function f such that $f \in C^\infty(\mathbb{R}^n)$? Consider $V \subseteq \overline{V} \subseteq U \subseteq \mathbb{R}^n$ (assume there is positive distance between $\overline{V}$ and U) with V, U open and $\overline{V}$ now also *compact*, then $\mathbb{I}_V, \mathbb{I}_{\overline{V}} \in \mathcal{L}^1(\mathbb{R}^n)$. Setting $f = \mathbb{I}_{\overline{V}} * \varphi_\varepsilon$ notice it is supported on a subset of U for ε small enough, i.e., $\operatorname{supp} f \subseteq \overline{V} + \overline{B(\boldsymbol{0}, \varepsilon)}$, where $+$ represents the point-wise sum. Then, by the above reasoning, $f \in C_c^\infty(\mathbb{R}^n)$.

Definition 2.2 (Locally finite covering)**.** Let $\Omega \subseteq \mathbb{R}^n$ be an open set. The family of open sets $(U_i)_{i\in\mathbb{N}}$ is called the *locally finite covering* of Ω, if $\Omega \subseteq \bigcup_{i\in\mathbb{N}} U_i$ and every $x \in \Omega$ is contained in at most finitely many U_i, i.e., $\{i \in \mathbb{N} \mid x \in U_i\}$ is finite for all $x \in \Omega$.

Proposition 2.1. *Let $(U_i)_{i\in\mathbb{N}}$ be a locally finite covering of Ω. Then, every compact set $K \subseteq \Omega$ intersects only finitely many U_i.*

Proof. Exercise. □

Proposition 2.2 (Subordinate locally finite covering)**.** *For every locally finite cover* $(U_i)_{i\in\mathbb{N}}$ *of* Ω *there exists a* subordinate locally finite cover $(V_i)_{i\in\mathbb{N}}$, *i.e.,* $(V_i)_{i\in\mathbb{N}}$ *is a cover of* Ω*, such that for all* $i \in \mathbb{N}$ *holds* $\overline{V_i} \subseteq U_i$ *and* $\overline{V_i}$ *is compact*

Remark 2.2 (Resolution of one). Given a subordinate locally finite cover $(V_i)_{i\in\mathbb{N}}$ take $f_i = \mathbb{I}_{\overline{V_i}} * \varphi_{\varepsilon_i}$ with $\varepsilon_i > 0$ such that $\operatorname{supp} f_i \subseteq U_i$. Then $F(\boldsymbol{x}) := \sum_{i=1}^{\infty} f_i(\boldsymbol{x})$ is *finite* for every fixed $\boldsymbol{x}$, and $F(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Omega$. Taking $\psi_i(\boldsymbol{x}) = \frac{f_i(\boldsymbol{x})}{F(\boldsymbol{x})}$ we get functions ψ_i such that $\operatorname{supp} \psi_i \subseteq U_i$ and $\sum \psi_i = 1$. The family $(\psi_i)_{i\in\mathbb{N}}$ is called **resolution of one**. In other words, we have decomposed a constant function without compact support (either $\mathbb{I}_\Omega$ or $\mathbb{I}_{U_i}$, depending on the perspective) to compactly supported functions.

2.2 Topology of $C_c^\infty(\mathbb{R}^n)$ or $C_c^\infty(\Omega)$

Let us attempt to discern the topology of these function spaces. Firstly, we shall equip $C_c^\infty(\mathbb{R}^n)$ with a family of semi-norms

$$|||f|||_N = \sum_{|\boldsymbol{\alpha}|\le N} \left\| \left(1 + \|\boldsymbol{x}\|^2\right)^N \mathrm{D}^{\boldsymbol{\alpha}} f \right\|_\infty, \quad N \in \mathbb{N}_0,$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ is a multi-index ($|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_n$), i.e.,

$$\mathrm{D}^{\boldsymbol{\alpha}} = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \quad \& \quad \boldsymbol{x}^{\boldsymbol{\alpha}} = \prod_{i=1}^{n} x_i^{\alpha_i}. \tag{2.1}$$

As $f \in C_c^\infty(\mathbb{R}^n)$, $|||f|||_N$ is always finite, and if f is a polynomial of degree m then for $N > m$ we get $|||f|||_N = 0$, i.e., it is indeed a semi-norm.

We now equip $C_c^\infty(\mathbb{R}^n)$ with the topology induced by this family of semi-norms $\left(\|\cdot\|_N\right)_{N\in\mathbb{N}_0}$ (which is generated by all open balls w.r.t. all the semi-norms). Although this does not produce a normed space, we get that the topology is *metric* with

$$d(f,g) = \sum_{N=0}^{\infty} 2^{-N} \frac{|||f-g|||_N}{1 + |||f-g|||_N}.$$

Nonetheless, for practical purposes we would like to work with a complete space, hence we introduce the completion of $C_c^\infty(\mathbb{R}^n)$ with respect to d as the **Schwartz space**

$$\mathcal{S} := \left\{ f \in C^\infty(\mathbb{R}^n) \mid \forall N \in \mathbb{N}_0 : \; |||f|||_N < \infty \right\}. \tag{2.2}$$

For all functions $f \in \mathcal{S}$ we know that it and all its derivative vanish faster than any negative of $\|\boldsymbol{x}\|$. For example, it can be shown that $\mathrm{e}^{-\|\boldsymbol{x}\|^2} \in \mathcal{S}$.

If we take $f \in \mathcal{S}$ and calculate

$$\int_{\mathbb{R}^n} |f(\boldsymbol{x})|^p \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} \left|\left(1 + \|\boldsymbol{x}\|^2\right) f(\boldsymbol{x})\right|^p \frac{1}{\left(1 + \|\boldsymbol{x}\|^2\right)^p} \, \mathrm{d}x \le \int_{\mathbb{R}^n} \frac{|||f|||_1^p}{\left(1 + \|\boldsymbol{x}\|^2\right)^p} \, \mathrm{d}\boldsymbol{x}, \tag{2.3}$$

then as $\frac{1}{\left(1+\|x\|^2\right)^p} \leq \frac{1}{1+\|x\|^2} \in \mathcal{L}^1(\mathbb{R}^n)$[1], we obtain $\mathcal{S} \subseteq \mathcal{L}^p(\mathbb{R}^n)$. In addition, $\mathcal{S}$ is even dense in $\mathcal{L}^p(\mathbb{R}^n)$. However, this density is with respect to the $\mathcal{L}^p$-norm, i.e., we "lose" the topology of $\mathcal{S}$ defined above. Still, it follows from Plancherel's theorem 1.6 that we have Fourier transform defined on $\mathcal{S}$,

$$f \in \mathcal{S}: \quad \mathcal{F}f(\boldsymbol{t}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{x}, \boldsymbol{t}\rangle} \,\mathrm{d}\boldsymbol{x}. \tag{2.4}$$

Recalling that for $f \in \mathcal{S}$ we have $\mathrm{D}^{\boldsymbol{\alpha}} f \in \mathcal{S} \subseteq \mathcal{L}^2(\mathbb{R}^n)$ it yields

$$\mathcal{F}\left(\mathrm{D}^{\boldsymbol{\alpha}} f\right)(\boldsymbol{t}) = (-2\pi\mathrm{i})^{|\boldsymbol{\alpha}|} \boldsymbol{t}^{\boldsymbol{\alpha}} \cdot \mathcal{F}f(\boldsymbol{t}) \tag{2.5}$$

by (2.1) and computation similar as in (1.1). By the definition of fourier transform (2.4) we get that $\mathcal{F}f$ decays to 0 at infinity faster than any $\boldsymbol{t}^{\boldsymbol{\alpha}}$ grows, which combined with (2.5) leads to $\mathcal{F}f \in \mathcal{S}$. Applying similar reasoning to $\mathcal{F}^{-1}$ gives that $\mathcal{F}: \mathcal{S} \to \mathcal{S}$ is an **automorphism** of the Schwartz space.

Remark 2.3. We could have used the fact that $\mathcal{S}$ is dense in $\mathcal{L}^2$ (in the $\mathcal{L}^2$-norm) together with $\mathcal{F}$ being automorphism of $\mathcal{S}$ to prove the Plancherel's theorem 1.6.

2.3 Distributions

We turn our attention back to the space of infinitely-differentiable compactly-supported functions $C_c^\infty(\Omega)$ with $\Omega \subseteq \mathbb{R}^n$ open equipped with the family of norms

$$\|\varphi\|_N = \sum_{|\boldsymbol{\alpha}| \leq N} \left\|\mathrm{D}^{\boldsymbol{\alpha}} \varphi\right\|_\infty, \tag{2.6}$$

which induce a metric structure on $C_c^\infty(\Omega)$ (and thus also topology).

Definition 2.3 (Distribution)**.** A linear functional $\mathsf{T}: C_c^\infty(\Omega) \to \mathbb{R}$ is called a distribution, if for every compact $K \subseteq \Omega$ there exist constants $C > 0$ and $N \in \mathbb{N}_0$ such that for all $\varphi \in C_c^\infty(\Omega)$ with support contained in K, i.e., $\operatorname{supp} \varphi \subseteq K$, it holds that

$$\left|\mathsf{T}(\varphi)\right| \leq C \left\|\varphi\right\|_N. \tag{2.7}$$

Proposition 2.3. *Equivalently, we can characterize distributions as being* continuous linear functionals *with respect to the topology induced by (2.6), see [7, def. 6.7 and thm. 6.8].*

Let us note that C, N do depend on K, i.e., on the support of φ!

Definition 2.4 (Order of distribution)**.** A distribution is of order N, if $N \in \mathbb{N}_0$ is the smallest integer such that (2.7) holds for all compact $K \subseteq \Omega$.

[1]For completeness' sake, one can check $\arctan'(x) = \frac{1}{\tan'(\arctan(x))} = \frac{1}{1+x^2}$, where $\tan'(x) = \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = 1 + \tan^2(x)$.

Example 2.1. Firstly, let μ be the signed Borel measure on Ω. Then $\varphi \mapsto \int_\Omega \varphi \, \mathrm{d}\mu$ is a distribution of order 0. Indeed,

$$\mathsf{T}(\varphi) = \int_K \varphi \, \mathrm{d}\mu \le \mu(K) \sup_{\boldsymbol{x} \in K} \varphi = \overbrace{\mu(K) \left\| \varphi \right\|_\infty}^{N=0}.$$

Similarly, $\varphi \mapsto \mathrm{D}^{\boldsymbol{\alpha}} \varphi(\mathbf{0})$ is a distribution of order $|\boldsymbol{\alpha}|$.

Theorem 2.1. *Let* T *be a distribution of order 0. Then there exists a* signed measure μ *such that* $\mathsf{T}(\varphi) = \int_\Omega \varphi \, \mathrm{d}\mu$.

Before proving the theorem, let us introduce the following notation for the *space of test functions* $\mathcal{D}(\Omega) := C_c^\infty(\Omega)$ equipped with the topology considered above.

Proof. Let $(U_i)_{i\in\mathbb{N}}$ be a locally finite covering of Ω, see Definition 2.2. The distribution T extends to a continuous linear functional on $C(\overline{U_i})$ for $i \in \mathbb{N}$ by Hahn-Banach theorem. Additionally, by Riesz' representation theorem 4.5 there exists a signed measure μ_i on $\overline{U_i}$ such that

$$\forall \varphi \in C(\overline{U_i}) : \quad \mathsf{T}(\varphi) = \int_{\overline{U_i}} \varphi \, \mathrm{d}\mu_i.$$

Let $(\psi_i)_{i\in\mathbb{N}}$ be the *resolution of one*, see Remark 2.2, and $\varphi \in \mathcal{D}(\Omega)$, then $\varphi(\boldsymbol{x}) = \sum_{i\in\mathbb{N}} \varphi(\boldsymbol{x})\psi_i(\boldsymbol{x})$ as $\sum_{i\in\mathbb{N}} \psi(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \Omega$. Now, we can calculate

$$\mathsf{T}(\varphi) = \mathsf{T}\left(\sum_{i\in\mathbb{N}} \varphi \cdot \psi_i\right) \overset{\text{linear}}{=} \sum_{i\in\mathbb{N}} \mathsf{T}\overbrace{(\varphi \cdot \psi_i)}^{\operatorname{supp} \subseteq U_i} = \int_{\overline{U_i}} \varphi \cdot \psi_i \, \mathrm{d}\mu_i.$$

Taking $\mathrm{d}\mu = \sum_{i\in\mathbb{N}} \psi_i \, \mathrm{d}\mu_i$ gives $\mathsf{T}(\varphi) = \int_\Omega \varphi \, \mathrm{d}\mu$. As a closing note, let us remark that $\sum_{i\in\mathbb{N}} \varphi \cdot \psi_i$ is always finite (see Proposition 2.2) and thus linearity of T is applicable. □

Definition 2.5. A distribution T is called *positive*, if for every $\varphi \in \mathcal{D}(\Omega)$ such that $\varphi(\boldsymbol{x}) \ge 0$ for all $\boldsymbol{x} \in \Omega$ it holds that $\mathsf{T}(\varphi) \ge 0$.

Theorem 2.2. *Every positive distribution is of order 0.*

Proof. Let $K \subseteq \Omega$ be compact and $\chi \in \mathcal{D}(\Omega)$ such that $\chi(\boldsymbol{x}) = 1$ for all $x \in K$. For $\varphi \in \mathcal{D}(\Omega)$ with $\operatorname{supp} \varphi \subseteq K$ we have

$$-\left\| \varphi \right\|_\infty \chi(\boldsymbol{x}) \le \varphi(\boldsymbol{x}) \le \left\| \varphi \right\|_\infty \chi(\boldsymbol{x}).$$

Then continuity and positivity of T yields (recall (2.6))

$$-\left\| \varphi \right\|_\infty \mathsf{T}(\chi) \le \mathsf{T}(\varphi) \le \left\| \varphi \right\|_\infty \mathsf{T}(\chi) \implies \left| \mathsf{T}(\varphi) \right| \le \mathsf{T}(\chi) \left\| \varphi \right\|_\infty,$$

thus $N = 0$ in the sense of Definition 2.4. □

Corollary 2.1. *Every positive distribution "is" a positive measure.*

2.3.1 Calculus of Distributions

Definition 2.6 (Differentiation of distribution)**.** Let T be a distribution. Then we define for $\varphi \in \mathcal{D}(\Omega)$

$$\frac{\partial \mathsf{T}}{\partial x_j}(\varphi) := -\mathsf{T}\left(\frac{\partial \varphi}{\partial x_j}\right). \tag{2.8}$$

Intuitively, the minus sign in (2.8) comes from integration by parts.

Remark 2.4. Let T be a distribution of order N, i.e., $\left|\mathsf{T}(\varphi)\right| \leq C \left\|\varphi\right\|_N$. Then $\left|\frac{\partial \mathsf{T}}{\partial x_j}(\varphi)\right| \leq C \left\|\varphi\right\|_{N+1}$, i.e., the distribution $\frac{\partial \mathsf{T}}{\partial x_j}$ is of order $N+1$.

Also, it is important to note that the Schwarz' theorem holds for the differentiation of distributions, i.e., $\frac{\partial^2 \mathsf{T}}{\partial x_j \partial x_k} = \frac{\partial^2 \mathsf{T}}{\partial x_k \partial x_j}$.

Example 2.2. Firstly, let $\delta(\varphi) = \varphi(\mathbf{0})$ be the dirac distribution (centered at 0). Then

$$\mathrm{D}^{\boldsymbol{\alpha}}\, \delta(\varphi) = (-1)^{|\boldsymbol{\alpha}|} \delta\left(\mathrm{D}^{\boldsymbol{\alpha}}\, \varphi\right) = (-1)^{|\boldsymbol{\alpha}|}\, \mathrm{D}^{\boldsymbol{\alpha}}\, \varphi(\mathbf{0}).$$

Secondly, let $n = 1$ and $f(x) = |x|$. What is $f''(x)$? We can consider $\mathsf{T} : \varphi \mapsto \int_{\mathbb{R}} \varphi(x) \cdot |x| \, \mathrm{d}x$, then

$$\begin{aligned}
\mathsf{T}''(\varphi) &= \int_{\mathbb{R}} \varphi''(x)\, |x| \, \mathrm{d}x = \overbrace{\varphi'(x)\, |x| \Big|_{-\infty}^{\infty}}^{\text{supp}\, \varphi \text{ compact}} - \int_{\mathbb{R}} \varphi'(x) \, \mathrm{sign}(x) \, \mathrm{d}x \\
&= + \int_{-\infty}^{0} \varphi'(x) \, \mathrm{d}x - \int_{0}^{\infty} \varphi'(x) \, \mathrm{d}x = \left(\varphi(x)\big|_{-\infty}^{0}\right) - \left(\varphi(x)\big|_{0}^{\infty}\right) \\
&= \varphi(0) + \varphi(0) = 2\delta(\varphi) \\
&\Downarrow \\
f'' &= 2\delta.
\end{aligned}$$

Remark 2.5. Let $f \in \mathcal{C}(\Omega)$, then f corresponds to the distribution $\mathsf{T} : \varphi \mapsto \int_{\mathbb{R}^n} f(\boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}x$. Let us study the case $\Omega = \mathbb{R}^3$.

If we consider $E(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{x}\|}$, what is ΔE? We know that $\Delta E(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \neq 0$, however we would like to understand it in the distributional sense. Recall the Gauss' divergence theorem, i.e.,

$$\oiint_{\partial B} \boldsymbol{v} \, \mathrm{d}\boldsymbol{A} = \iiint_B \mathrm{div}\, \boldsymbol{v} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

for a differentiable vector field $\boldsymbol{v}$, and take $\boldsymbol{v} = f \cdot \mathrm{grad}(\boldsymbol{g})$ where f, g are differentiable functions. Then

$$\oiint_{\partial B} f \, \mathrm{grad}(g) \, \mathrm{d}\boldsymbol{A} = \iiint_B \left(\langle \mathrm{grad}\, f, \mathrm{grad}\, g \rangle + f \Delta g\right) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z,$$

from which follows the *Gauss-Green formula*,

$$\oiint_{\partial B} (f \operatorname{grad}(g) - g \operatorname{grad}(f))\, \mathrm{d}\boldsymbol{A} = \iiint_B (f\Delta g - g\Delta f)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z. \tag{2.9}$$

Now let $g(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{x}\|}$ on $B = \{\boldsymbol{x} \in \mathbb{R}^3 \mid r \leq \|\boldsymbol{x}\| \leq \|R\|\}$ to exclude the singularity of g at 0. Lastly, take $f \in \mathcal{D}(\mathbb{R}^3)$ such that $\operatorname{supp} f \subseteq \{\boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x}\| < R\}$. Then $\operatorname{grad} g(\boldsymbol{x}) = \frac{-\boldsymbol{x}}{\|\boldsymbol{x}\|^3}$ and by (2.9)

$$\oiint_{\|\boldsymbol{x}\|=R} \Bigg(\overbrace{-f(\boldsymbol{x}) \frac{\boldsymbol{x}}{\|\boldsymbol{x}^3\|} - \frac{1}{\|\boldsymbol{x}\|} \operatorname{grad}(f)(\boldsymbol{x})}^{\text{outside of } \operatorname{supp} f \text{ with } f \text{ smooth } \implies 0} \Bigg) \mathrm{d}\boldsymbol{A} - \oiint_{\|\boldsymbol{x}\|=r} \left(-f(\boldsymbol{x}) \frac{\boldsymbol{x}}{\|\boldsymbol{x}^3\|} - \frac{1}{\|\boldsymbol{x}\|} \operatorname{grad}(f)(\boldsymbol{x}) \right) \overbrace{\mathrm{d}\boldsymbol{A}}^{\frac{\boldsymbol{x}}{r}\mathrm{d}A} = \iiint_B (\overbrace{f\Delta g}^{\Delta g = 0} - g\Delta f)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z,$$

thus

$$\begin{aligned} \iiint_B g\Delta f\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z &= -\oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x}) \overbrace{\frac{\boldsymbol{x}}{r^3} \cdot \frac{\boldsymbol{x}}{r}}^{\frac{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}{\|\boldsymbol{x}\|^4} = \frac{1}{r^2}} \mathrm{d}A - \frac{1}{r} \overbrace{\oiint_{\|\boldsymbol{x}\|=r} \operatorname{grad} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{A}}^{\text{divergence theorem}} \\ &= -\frac{1}{r^2} \oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x})\, \mathrm{d}A - \underbrace{\frac{1}{r} \iiint_{\|\boldsymbol{x}\|\leq r} \Delta f\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z}_{f \in \mathcal{D}(\mathbb{R}^3) \implies \frac{1}{r} O(r^3) \text{ so } \xrightarrow[r\to 0]{} 0}. \end{aligned}$$

Recall $f \in \mathcal{D}(\mathbb{R}^3)$ is (infinitely) differentiable, i.e., $f(\boldsymbol{x}) = f(\boldsymbol{0}) + O(\|\boldsymbol{x}\|)$, hence

$$\oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x})\, \mathrm{d}A = \oiint_{\|\boldsymbol{x}\|=r} (f(\boldsymbol{0}) + O(\|\boldsymbol{x}\|))\, \mathrm{d}A = (f(\boldsymbol{0}) + O(\|\boldsymbol{x}\|)) \underbrace{\oiint_{\|\boldsymbol{x}\|=r} \mathrm{d}A}_{\boldsymbol{x} \in \mathbb{R}^3 \implies 4\pi r^2} = 4\pi r^2 f(\boldsymbol{0}) + 4\pi O(r^3).$$

By letting $r \to 0$, we finally obtain

$$\iint_{\|\boldsymbol{x}\|\leq R} f\Delta g\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = \iint_{\|\boldsymbol{x}\|\leq R} \frac{\Delta f(\boldsymbol{x})}{\|\boldsymbol{x}\|}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = -4\pi f(\boldsymbol{0}) \implies \Delta g = -4\pi\delta_0 \tag{2.10}$$

as $\Delta g(f) = g(\Delta f)$ in the sense of the distributional Laplacian.

2.3.2 Supported Distributions

Definition 2.7 (Support of a distribution)**.** Let us define the support of a distribution T as

$$\begin{aligned} \operatorname{supp} \mathsf{T} &= \{\boldsymbol{x} \in \Omega \mid \forall \text{ open } U \ni \boldsymbol{x}\ \exists \varphi \in \mathcal{D}(\Omega) \text{ s.t. } \operatorname{supp}\varphi \subseteq U :\ \mathsf{T}(\varphi) \neq 0\} \\ &= \Omega \setminus \{\boldsymbol{x} \in \Omega \mid \exists U \ni \boldsymbol{x}\ \forall \varphi \in \mathcal{D}(\Omega) \text{ s.t. } \operatorname{supp}\varphi \subseteq U :\ \mathsf{T}(\varphi) = 0\}. \end{aligned}$$

Remark 2.6. Directly from the definition of differentiation of a distribution 2.6 follows that $\operatorname{supp} \mathrm{D}^\alpha \mathsf{T} \subseteq \operatorname{supp} \mathsf{T}$ as $\operatorname{supp} \mathrm{D}^\alpha \varphi \subseteq \operatorname{supp}\varphi$ for all $\varphi \in \mathcal{D}(\Omega)$ (recalling that $\operatorname{supp}\varphi := \overline{\{\boldsymbol{x} \in \Omega \mid \varphi(\boldsymbol{x}) \neq 0\}}$). Moreover, this set inclusion can even be strict, as illustrated by the previous Example 2.2.

Definition 2.8. Let T be a distribution on Ω and $a \in \mathcal{D}(\Omega)$. Then $a\mathsf{T}$ is defined by $(a\mathsf{T})(\varphi) = \mathsf{T}(a\varphi)$.

Remark 2.7. The product rule, and thus also the Leibniz rule, remains valid for $a\mathsf{T}$, i.e.,

$$\frac{\partial(a\mathsf{T})}{\partial x_j} = \frac{\partial a}{\partial x_j}\mathsf{T} + a\frac{\partial \mathsf{T}}{\partial x_j}.$$

Proposition 2.4. *Let* T *be a distribution of order N with compact support. Then there exists a constant $C > 0$, such that $\left|\mathsf{T}(\varphi)\right| \leq C \left\|\varphi\right\|_N$ (in the sense of (2.6)) independent on the support of φ.*

Proof. Let $\chi \in \mathcal{D}(\Omega)$ be a test function with $\chi(\boldsymbol{x}) = 1$ on the support of T and let $K \subseteq \Omega$ be a compact set such that $\operatorname{supp}\chi \subseteq K$. For a test function $\varphi \in \mathcal{D}(\Omega)$ it holds that $\operatorname{supp}(\chi\varphi) \subseteq K$, and $\mathsf{T}(\varphi\chi) = \mathsf{T}(\varphi)$ since $\varphi\chi - \varphi = 0$ on $\operatorname{supp}\mathsf{T}$. Now $\left|\mathsf{T}(\varphi)\right| \leq C \left\|\varphi\psi\right\|_N$ where C depends only on K (which includes $\operatorname{supp}(\varphi\chi)$), and because

$$\begin{aligned}
\left\|\varphi\chi\right\|_N &= \sum_{|\boldsymbol{\alpha}|\leq N} \left\|\mathrm{D}^{\boldsymbol{\alpha}}(\varphi\chi)\right\|_\infty = \sum_{|\boldsymbol{\alpha}|\leq N} \left\|\sum_{\boldsymbol{\beta}\leq\boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} \mathrm{D}^{\boldsymbol{\beta}}\varphi\, \mathrm{D}^{\boldsymbol{\alpha}-\boldsymbol{\beta}}\chi\right\|_\infty \\
&\leq \sum_{|\boldsymbol{\alpha}|\leq N} \sum_{\boldsymbol{\beta}\leq\boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} \left\|\mathrm{D}^{\boldsymbol{\beta}}\varphi\right\|_\infty \left\|\mathrm{D}^{\boldsymbol{\alpha}-\boldsymbol{\beta}}\chi\right\|_\infty \\
&\leq \tilde{C}\cdot\left(\sum_{|\boldsymbol{\alpha}|\leq N} \left\|\mathrm{D}^{\boldsymbol{\alpha}}\varphi\right\|_\infty\right)\left(\sum_{|\boldsymbol{\alpha}|\leq N} \left\|\mathrm{D}^{\boldsymbol{\alpha}}\chi\right\|_\infty\right) = \tilde{C}\left\|\varphi\right\|_N \left\|\chi\right\|_N,
\end{aligned}$$

we finally obtain $\left|\mathsf{T}(\varphi)\right| \leq C\cdot\tilde{C}\cdot\left\|\chi\right\|_N \left\|\psi_N\right\|$, where $C\cdot\tilde{C}\cdot\left\|\chi\right\|_N$ is independent of $\operatorname{supp}\varphi$. □

Remark 2.8. Let $\varphi, \chi \in \mathcal{D}(\Omega)$ be test functions such that $\chi(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in \operatorname{supp}\varphi$. Then for P the Taylor polynomial of φ of degree N at $\mathbf{0}$ we have

$$\varphi(\boldsymbol{x}) = \varphi(\boldsymbol{x})\chi(\boldsymbol{x}) = P(\boldsymbol{x})\chi(\boldsymbol{x}) + \Psi(\boldsymbol{x}) \quad \& \quad \left\|\Psi(\boldsymbol{x})\right\| \leq C\left\|\boldsymbol{x}\right\|^{N+1},$$

i.e., $\Psi(\boldsymbol{x})$ is also a test function. If, in addition, all derivatives of order $\leq N$ of φ vanish at $\mathbf{0}$ (so $P(\boldsymbol{x}) = 0$), then by differentiation

$$\left|\mathrm{D}^{\boldsymbol{\alpha}}\varphi\right|(\boldsymbol{x}) \leq C_{\boldsymbol{\alpha}}\left\|\boldsymbol{x}\right\|^{N+1-|\alpha|}$$

for $|\boldsymbol{\alpha}| \leq N$.

Theorem 2.3. *Let* T *be a distribution of order N with* $\operatorname{supp}\mathsf{T} \subseteq F$. *If φ is a test function such that φ and all its derivatives up to order N vanish on F, then $\mathsf{T}(\varphi) = 0$.*

Proof. Let $\varepsilon > 0$ and denote χ_ε the regularization of the indicator function of $F_{2\varepsilon} := \{\boldsymbol{x} \in \mathbb{R} \mid \operatorname{dist}(\boldsymbol{x}, F) \leq 2\varepsilon\}$, i.e.,

$$\chi_\varepsilon(\boldsymbol{x}) = \frac{1}{\varepsilon^n} \int_{F_{2\varepsilon}} \psi\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\varepsilon}\right) \mathrm{d}\boldsymbol{y},$$

where ψ is the *bump function* satisfying $\operatorname{supp} \psi \subseteq \overline{B(\boldsymbol{0}, 1)}$, $\psi \geq 0$, and $\int \psi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = 1$. Then $\operatorname{supp} \chi_\varepsilon \subseteq F_{4\varepsilon}$ and, by our assumption, $\mathsf{T}(\varphi) = \mathsf{T}(\varphi\chi_\varepsilon)$. As such, $\operatorname{supp} \varphi\chi_\varepsilon \subseteq F_{4\varepsilon}$ and also $\varphi\chi_\varepsilon = 0$ on F (as φ vanishes there). Thus

$$\left|\mathsf{T}(\varphi)\right| = \left|\mathsf{T}(\varphi\chi_\varepsilon)\right| \leq C \left\|\varphi\chi_\varepsilon\right\|_N.$$

Through differentiation we accumulate powers of ε obtaining $\left|\mathrm{D}^{\boldsymbol{\alpha}} \chi_\varepsilon(\boldsymbol{x})\right| \leq C_{\boldsymbol{\alpha}} \varepsilon^{-|\boldsymbol{\alpha}|}$. On the other hand, the last remark 2.8 dictates that $\left|\mathrm{D}^{\boldsymbol{\alpha}} \varphi\right| \leq C\varepsilon^{N+1-|\boldsymbol{\alpha}|}$ for $|\boldsymbol{\alpha}| \leq N$. Hence, on $F_{4\varepsilon}$ we have

$$\left|\mathrm{D}^{\boldsymbol{\alpha}}(\varphi\chi_\varepsilon)\right| \leq \sum_{\boldsymbol{\beta}+\boldsymbol{\gamma}=\boldsymbol{\alpha}} C_{\boldsymbol{\beta}} C_{\boldsymbol{\gamma}} \varepsilon^{N+1-|\boldsymbol{\beta}|} \varepsilon^{-|\boldsymbol{\gamma}|} \leq \tilde{C} \varepsilon^{N+1-|\boldsymbol{\alpha}|}.$$

Thus $\left\|\varphi\chi_\varepsilon\right\|_N \leq \tilde{C}\varepsilon$ in the sense of (2.6), which further implies

$$\left|\mathsf{T}(\varphi)\right| = \left|\mathsf{T}(\varphi\chi_\varepsilon)\right| \leq C \left\|\varphi\chi_\varepsilon\right\|_N \leq C \cdot \tilde{C} \cdot \varepsilon$$

and as $\varepsilon > 0$ was arbitrary it yields $\mathsf{T}(\varphi) = 0$. □

Corollary 2.2. *A distribution, whose support is a single point, is a linear combination of (evaluation of) φ and its derivatives.*

Proof. Without loss of generality let $\operatorname{supp} \mathsf{T} = \{\boldsymbol{0}\}$. Write, per Remark 2.8, $\varphi(\boldsymbol{x}) = P(\boldsymbol{x})\chi(\boldsymbol{x}) + \Psi(\boldsymbol{x})$ where P is the Taylor polynomial of order N and Ψ vanishes up to order N at $\boldsymbol{0}$. From Theorem 2.3 it follows $\mathsf{T}(\Psi) = 0$, therefore

$$\mathsf{T}(\varphi) = \mathsf{T}(P\chi) + \overbrace{\mathsf{T}(\Psi)}^{0} = \sum_{|\boldsymbol{\alpha}| \leq N} \frac{\mathrm{D}^{\boldsymbol{\alpha}} \varphi(\boldsymbol{0})}{\boldsymbol{\alpha}!} \underbrace{\mathsf{T}(\boldsymbol{x}^{\boldsymbol{\alpha}}\chi)}_{C_{\boldsymbol{\alpha}}} = \sum_{|\boldsymbol{\alpha}| \leq N} \frac{(-1)^{|\boldsymbol{\alpha}|} C_{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} (\mathrm{D}^{\boldsymbol{\alpha}} \delta)(\varphi),$$

as $P(\boldsymbol{x}) = \sum_{|\boldsymbol{\alpha}| \leq N} \frac{\mathrm{D}^{\boldsymbol{\alpha}} \varphi(\boldsymbol{0})}{\boldsymbol{\alpha}!} \boldsymbol{x}^{\boldsymbol{\alpha}}$ (recall (2.1)). □

2.3.3 Convolution of Distributions

Let $\varphi, \psi \in \mathcal{D}(\mathbb{R}^n)$ be test functions, then, as we already know, their convolution reads $\varphi * \psi(\boldsymbol{x}) = \int_{\mathbb{R}^n} \varphi(\boldsymbol{x}-\boldsymbol{y})\psi(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}$ (which is well-defined as φ, ψ are compactly supported). We shall now introduce the following notation:

- *reverse* of a test function φ,

$$\check{\varphi}(\boldsymbol{x}) := \varphi(-\boldsymbol{x}) \quad \& \quad \check{\mathsf{T}}(\varphi) := \mathsf{T}(\check{\varphi}); \tag{2.11}$$

- *translation* of φ by $\boldsymbol{h}$,

$$\mathcal{T}_{\boldsymbol{h}}\varphi(\boldsymbol{x}) := \varphi(\boldsymbol{x}-\boldsymbol{h}) \quad \& \quad \mathcal{T}_{\boldsymbol{h}}\mathsf{T}(\varphi) := \mathsf{T}\left(\mathcal{T}_{-\boldsymbol{h}}\varphi\right). \tag{2.12}$$

Caution

One should be careful not to forget the minus sign in the definition (2.12) of translation for distributions.

While this might seem arbitrary at first, an attentive reader might realize that

$$\varphi * \psi(\boldsymbol{x}) = \int_{\mathbb{R}^n} \varphi(\boldsymbol{y})(\mathcal{T}_{\boldsymbol{x}}\check{\psi})(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \varphi\left(\mathcal{T}_{\boldsymbol{x}}\check{\psi}\right), \tag{2.13}$$

which makes it rather natural to write $\varphi * \psi(\boldsymbol{x}) = \varphi\left(\mathcal{T}_{\boldsymbol{x}}\check{\psi}\right)$.

Remark 2.9. It can be shown that $\mathcal{T}_{\boldsymbol{h}}$ commutes with differentiation $\mathrm{D}^{\boldsymbol{\alpha}}$.

Definition 2.9 (Convolution of a distribution and test function)**.** Let T be a distribution and $\varphi \in \mathcal{D}(\mathbb{R}^n)$ be a test function. Then we define the *convolution of a distribution and test function* as $\mathsf{T} * \varphi(\boldsymbol{x}) := \mathsf{T}\left(\mathcal{T}_{\boldsymbol{x}}\check{\varphi}\right)$.

Remark 2.10. Firstly, let us note that, in general, $\mathsf{T} * \varphi \in C^\infty(\mathbb{R}^n)$, which is *not* compactly supported, i.e., we do not get back a test function.

However, if T is compactly supported, then so is $\mathsf{T} * \varphi$. Thus, in this case, $\mathsf{T} * \varphi \in \mathcal{D}(\mathbb{R}^n)$.

Theorem 2.4. *Let* $\mathsf{L} : \mathcal{D}(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ *be a* continuous linear map *which commutes with translation, then there exists exactly one distribution* T *such that* $\mathsf{L}\varphi = \mathsf{T} * \varphi$.

Proof. Since L commutes with translation, we can simply study $\mathsf{L}(\varphi)(\mathbf{0})$. Now this is a continuous linear functional on $\mathcal{D}(\mathbb{R}^n)$, thus by Proposition 2.3 a distribution, say $\check{\mathsf{T}}$. Hence, we may write $\mathsf{L}\varphi(\mathbf{0}) = \check{\mathsf{T}}(\varphi)$, and by commutation with $\mathcal{T}$, we have

$$\mathsf{L}\varphi(\boldsymbol{x}) = \mathsf{L}\left(\mathcal{T}_{-\boldsymbol{x}}\varphi\right)(\mathbf{0}) = \check{\mathsf{T}}\left(\mathcal{T}_{-\boldsymbol{x}}\varphi\right) = \mathsf{T}\left(\mathcal{T}_{\boldsymbol{x}}\check{\varphi}\right) = \mathsf{T} * \varphi(\boldsymbol{x}).$$

□

So far, we have discussed a convolution of a distribution with a test function. Now, we shall turn our attention to the convolution of two distributions!

Definition 2.10 (Convolution of distributions)**.** Let S be a *compactly supported* distribution and T a distribution. We define the convolution $\mathsf{T} * \mathsf{S}$ of a compactly supported distribution with a distribution as

$$(\mathsf{T} * \mathsf{S}) * \varphi := \mathsf{T} * (\mathsf{S} * \varphi)\,. \tag{2.14}$$

Remark 2.11. Consider the setting of Definition 2.10, then S gives by convolution in the sense of Definition 2.9 a map $\mathcal{D}(\mathbb{R}^n) \to \mathcal{D}(\mathbb{R}^n)$, while T gives a map $\mathcal{D}(\mathsf{T}) \to C^\infty(\mathbb{R}^n)$. Recall (2.14), and notice

$$(\mathsf{T} * \mathsf{S}) * \varphi = \underbrace{\mathsf{T} * \overbrace{(\mathsf{S} * \varphi)}^{\in \mathcal{D}(\mathbb{R}^n)}}_{\in C^\infty(\mathbb{R}^n)}$$

by Remark 2.10, i.e., $(\mathsf{T} * \mathsf{S}) * \cdot$ is a continuous linear map $\mathcal{D}(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$. From Theorem 2.4 it follows that this linear map uniquely corresponds to a distribution, which, in this case, is precisely $\mathsf{T} * \mathsf{S}$.

Proposition 2.5. *In the setting of Definition 2.10,* $\mathsf{T} * \mathsf{S} = \mathsf{S} * \mathsf{T}$ *holds.*

Proof. Let us show that $\mathsf{T} * \mathsf{S} * \varphi * \psi = \mathsf{S} * \mathsf{T} * \varphi * \psi$ for two test functions $\varphi, \psi \in \mathcal{D}(\mathbb{R}^n)$. If this holds, one may choose $\psi = \psi_\varepsilon$ (rescaled bump function) and let $\varepsilon \to 0$, which gives the result.

Hence, denote $a = \mathsf{T} * \psi \in C^\infty(\mathbb{R}^n)$ and calculate (using the fact that convolution between (test) functions is commutative)

$$\begin{aligned} \mathsf{T} * \mathsf{S} * \varphi * \psi \triangleq \mathsf{T} * \overbrace{(\mathsf{S} * \varphi)}^{\in \mathcal{D}(\mathbb{R}^n)} * \psi \overset{\text{com.}}{=} \underbrace{\mathsf{T} * \psi}_{a} * (\mathsf{S} * \varphi) = \mathsf{S} * \varphi * a \\ = \mathsf{S} * \underbrace{(\mathsf{T} * \psi)}_{a} * \varphi = \mathsf{S} * \mathsf{T} * \varphi * \psi. \end{aligned}$$

□

2.3.4 Harmonic Distributions

Definition 2.11 (Harmonic distribution)**.** A solution to the equation $\Delta f = 0$ is called a *harmonic* distribution (or function, if applicable).

Theorem 2.5 (Weyl [8 (pp. 127)])**.** *Let* T *be a distribution satisfying* $\Delta \mathsf{T} = 0$*. Then* T *is a function.*

Proof. We must show that T is (at least) a C^2-function, because then the notions of distributional and usual Laplacian coincide, and T appears as an ordinary harmonic function. To this end let $\boldsymbol{z} \in \mathbb{R}^n$ be fixed and χ be a test function, which satisfies $\chi(\boldsymbol{x}) = 1$ for $\|\boldsymbol{x} - \boldsymbol{z}\| \leq r$. We then consider $\mathsf{S} = \chi \mathsf{T}$ and show that S is a C^∞-function on $\|\boldsymbol{x} - \boldsymbol{z}\| < r$ (which, in turn, implies T is a smooth function on a neighborhood of arbitrarily chosen $\boldsymbol{z}$).

From the definition of S and T being harmonic, we see right away that $\Delta \mathsf{S}$ has a compact support and vanishes on $\|\boldsymbol{x} - \boldsymbol{z}\| < r$ (as $\Delta \mathsf{S}(\varphi) = 0$ for $\operatorname{supp} \varphi \subseteq \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{z}\| < r\}$).

Let $E(\boldsymbol{x}) = -\frac{1}{(n-2)\omega_n}\frac{1}{\|\boldsymbol{x}\|^{n-2}}$ with ω_n denoting the area of a unit sphere in $\mathbb{R}^n$ and recall Remark 2.5, where we have shown $\Delta E = \delta_{\mathbf{0}}$ (for $n = 3$, but this holds in general). Take a test function φ such that $\varphi(\boldsymbol{x}) = 1$ for $\|\boldsymbol{x}\| < \varepsilon$ and $\varphi(\boldsymbol{x}) = 0$ for $\|\boldsymbol{x}\| > 2\varepsilon$. Moreover,

$$E(\boldsymbol{x}) = \underbrace{\varphi(\boldsymbol{x})E(\boldsymbol{x})}_{E_1(\boldsymbol{x})} + \underbrace{(1-\varphi(\boldsymbol{x}))E(\boldsymbol{x})}_{E_2(\boldsymbol{x})},$$

and notice that $E_2 \in C^\infty$ (the singularity is hidden by $1-\varphi$) and that E_1 is distribution with a compact support (by Definition 2.8, as we can view E as a distribution[2]). By previous computation we get $E(\Delta\varphi) = \Delta E(\varphi) = \delta_{\mathbf{0}}\varphi = \varphi(\mathbf{0})$, hence,

$$\mathsf{S} = \delta * \mathsf{S} = \Delta E * \mathsf{S} = E * \Delta\mathsf{S} = E_1 * \Delta\mathsf{S} + E_2 * \Delta\mathsf{S},$$

where $E_1 * \Delta S$ is a convolution of 2 compactly supported distributions, and $E_2 * \Delta S$ is a C^∞ function (by being a convolution of a C^∞-function with compactly supported distribution, see Remark 2.10). Using the fact that $\operatorname{supp}\Delta\mathsf{S} \subseteq \{\|\boldsymbol{x}-\boldsymbol{z}\| > r\}$ we get $\operatorname{supp}(E_1 * \Delta\mathsf{S}) \subseteq \{\|\boldsymbol{x}-\boldsymbol{z}\| > r-2\varepsilon\}$. Thus on $\{\|\boldsymbol{x}-\boldsymbol{z}\| < r-2\varepsilon\}$ the distribution S is in fact a C^∞-function, and therefore by the argumentation above T is a C^∞-function. □

Remark 2.12. A solution to $\Delta\mathsf{T} = 0$ is not only C^2 but even C^∞, i.e., the Laplacian forces functions in its kernel to be C^∞.

Let us again restrict ourselves, for simplicity, to $n = 3$ and consider a harmonic function f. In Remark 2.5, we have shown

$$\oiint_{\partial B}(f \operatorname{grad} g - g \operatorname{grad} f)\, \mathrm{d}\boldsymbol{A} = \iiint_B (f\Delta g - g\Delta f)\, \mathrm{d}V.$$

Moreover, take $g(\boldsymbol{x}) = -\frac{1}{4\pi}\frac{1}{\|\boldsymbol{x}\|}$ (as in the proof of Theorem 2.5), i.e., $\Delta g = 0$ for $\boldsymbol{x} \neq \mathbf{0}$ and $\operatorname{grad} g(\boldsymbol{x}) = \frac{1}{4\pi}\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|^3}$, with $B = \{\boldsymbol{x} \in \mathbb{R}^n \mid r \leq \|\boldsymbol{x}\| \leq R\}$. Then

$$\oiint_{\|\boldsymbol{x}\|=R}(f \operatorname{grad} g - g \operatorname{grad} f)\, \mathrm{d}\boldsymbol{A} - \oiint_{\|\boldsymbol{x}\|=r}(f \operatorname{grad} g - g \operatorname{grad} f)\, \mathrm{d}\boldsymbol{A} = 0$$
$$\Downarrow$$
$$\oiint_{\|\boldsymbol{x}\|=r}(f \operatorname{grad} g - g \operatorname{grad} f)\, \mathrm{d}\boldsymbol{A} = V,$$

where V does not depend on r. Hence,

$$V = \frac{1}{4\pi}\oiint_{\|\boldsymbol{x}\|=r}\left(f(\boldsymbol{x})\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|^3} - \frac{1}{r}\operatorname{grad} f\right)\overbrace{\frac{\boldsymbol{x}}{r}\,\mathrm{d}A}^{\mathrm{d}\boldsymbol{A}} = \frac{1}{4\pi r^2}\oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x})\,\mathrm{d}A - \frac{1}{4\pi r}\underbrace{\oiint_{\|\boldsymbol{x}\|=r}\operatorname{grad} f\,\mathrm{d}\boldsymbol{A}}_{\iiint_{\|\boldsymbol{x}\|<r}\Delta f\,\mathrm{d}V = 0}.$$

[2]Canonically, to a function f we assign the distribution $\mathsf{T}_f\varphi = \int_{\mathbb{R}^n} f(\boldsymbol{x})\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ for $\varphi \in \mathcal{D}(\mathbb{R}^n)$. Therefore, we do not need the function f to be defined (for evaluation) everywhere, but (local) integrability against test functions $\mathcal{D}(\mathbb{R}^n)$ is sufficient. In this sense, while $E(\boldsymbol{x})$ is not defined at 0 it still defines a distribution.

In other words, we have obtained that $\frac{1}{4\pi r^2}\oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x})\,\mathrm{d}A$ does not depend on r! Furthermore, if f is continuous, then $\lim_{r\to 0}\frac{1}{4\pi r^2}\oiint_{\|\boldsymbol{x}\|=r} f(\boldsymbol{x})\,\mathrm{d}A = f(\mathbf{0})$. Combined with the result above, we have shown the *mean value property* for harmonic functions

$$f(\boldsymbol{x}_0) = \frac{1}{4\pi r^2}\oiint_{\|\boldsymbol{x}-\boldsymbol{x}_0\|=r} f(\boldsymbol{x}-\boldsymbol{x}_0)\,\mathrm{d}A = \frac{1}{n\omega_n r^{n-1}}\oint_{\|\boldsymbol{x}-\boldsymbol{x}_0\|=r} f(\boldsymbol{x}-\boldsymbol{x}_0)\,\mathrm{d}A(\boldsymbol{x}).$$

In particular, for $n = 2$ we get the Cauchy integral theorem 4.9 as in $\mathbb{R}^2$ harmonic functions form the real part of holomorphic functions (in simply connected domains, to be precise).

Conversely, let f be a continuous function satisfying the mean-value property. Take ψ_ε a regularization function, which, you might recall, is radial and supported on $[0,\varepsilon]$, then

$$\begin{aligned}\overbrace{f * \psi_\varepsilon}^{\in C^\infty}(\boldsymbol{x}) &= \int_{\mathbb{R}^3} f(\boldsymbol{x}-\boldsymbol{y})\psi_\varepsilon(\boldsymbol{y})\,\mathrm{d}V(\boldsymbol{y}) = \int_0^\infty \overbrace{\int_{\|\boldsymbol{y}\|=r} f(\boldsymbol{x}-\boldsymbol{y})\,\mathrm{d}A(\boldsymbol{y})}^{4\pi r^2 f(\boldsymbol{x})}\,\psi_\varepsilon(r)\,\mathrm{d}r \\ &= 4\pi f(\boldsymbol{x}) \underbrace{\int_0^\varepsilon r^2\psi_\varepsilon(r)\,\mathrm{d}r}_{\int_{\mathbb{R}^3}\psi_\varepsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}=1 \implies \frac{1}{4\pi}} = f(\boldsymbol{x}),\end{aligned}$$

i.e., f is a C^∞-function!

2.4 Tempered Distributions

Let us now take test functions rapidly decaying in infinity (whereas before we had compactly supported test functions). Thus, we also need to consider distribution which do not grow fast at infinity — in fact, the name "tempered distributions" is a short for "distribution of temperate growth", where "temperate growth" means *polynomial growth*. Unfortunately, this introduces a bit of a chaos to the naming where both variants ("*tempered distributions*" [3, 9, 10] and "*temperate distributions*" [8]) are common.

Nevertheless, we shall recall the Schwartz space (2.2) of functions such that they and all their derivatives decay at infinity faster than any negative power of $\|\boldsymbol{x}\|$. To this end, we defined a family of semi-norms

$$|||f|||_k := \sum_{|\boldsymbol{\alpha}|\le k} \left\| \left(1+\|\boldsymbol{x}\|^2\right)^k \mathrm{D}^{\boldsymbol{\alpha}} f \right\|_\infty \tag{2.15}$$

such that the Schwartz space (2.2) is given

$$\mathcal{S} = \left\{ f \in C^\infty(\mathbb{R}^n) \mid \forall k \in \mathbb{N} : \ |||f|||_k < \infty \right\}.$$

Moreover, remember that we equip $\mathcal{S}$ with the topology induced by these semi-norms, and that $\mathcal{S}$ is a complete metric space (however not a normed one).

Theorem 2.6. *The space of test functions $\mathcal{D}(\mathbb{R}^n)$ is dense in $\mathcal{S}$.*

This theorem should come as rather natural, because surely compactly supported functions decay faster than any negative power of $\|\boldsymbol{x}\|$ at infinity.

Proof. Let φ be a test function such that $\varphi(\boldsymbol{x}) = 1$ for $\|\boldsymbol{x}\| \leq 1$ and $\varphi(\boldsymbol{x}) = 0$ for $\|\boldsymbol{x}\| \geq 2$. Take $f \in \mathcal{S}$ and $f_n(\boldsymbol{x}) = \varphi\left(\frac{\boldsymbol{x}}{n}\right) f(\boldsymbol{x}) \in \mathcal{D}(\mathbb{R}^n)$. To prove the density of $\mathcal{D}(\mathbb{R}^n)$ in $\mathcal{S}$ it suffices to show that $|||f_n - f|||_k \to 0$ for all k.

Surely,

$$\mathrm{D}^{\boldsymbol{\alpha}}(f_n - f) = \sum_{\boldsymbol{\beta}\leq\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}!}{\boldsymbol{\beta}!(\boldsymbol{\alpha}-\boldsymbol{\beta})!} \mathrm{D}^{\boldsymbol{\beta}}\left(1 - \varphi\left(\frac{\boldsymbol{x}}{n}\right)\right) \mathrm{D}^{\boldsymbol{\alpha}-\boldsymbol{\beta}} f(\boldsymbol{x})$$

by Leibniz' product rule. Now as $\varphi \in \mathcal{D}(\mathbb{R}^n)$, we have[3]

$$\left|\mathrm{D}^{\boldsymbol{\beta}}\left(1 - \varphi\left(\frac{\boldsymbol{x}}{n}\right)\right)\right| \leq \frac{C_{\varphi,\boldsymbol{\beta}}}{n^{|\boldsymbol{\beta}|}} \leq \frac{C_{\varphi,\boldsymbol{\beta}}}{n}$$

for $\boldsymbol{\beta} \neq \boldsymbol{0}$. Then

$$\left\|\left(1 + \|\boldsymbol{x}\|^2\right)^k \mathrm{D}^{\boldsymbol{\alpha}}(f_n - f)\right\|_\infty \leq \overbrace{\sum_{\boldsymbol{0}<\boldsymbol{\beta}\leq\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}!}{\boldsymbol{\beta}!(\boldsymbol{\alpha}-\boldsymbol{\beta})!} \frac{C_{\varphi,\boldsymbol{\beta}}}{n} \left\|\left(1 + \|\boldsymbol{x}\|^2\right)^k \mathrm{D}^{\boldsymbol{\alpha}-\boldsymbol{\beta}} f\right\|_\infty}^{\boldsymbol{\beta}\neq\boldsymbol{0}} + \|1 - \varphi\|_\infty \cdot \overbrace{\sup_{\|\boldsymbol{x}\|\geq n} \left|\left(1 + \|\boldsymbol{x}\|^2\right)^k \mathrm{D}^{\boldsymbol{\alpha}} f(\boldsymbol{x})\right|}^{f\in\mathcal{S} \implies \text{tends to 0 for } n\to\infty},$$

which goes to 0 for $n \to \infty$, thus $f_n \to f$ in the semi-norms (note that $|\boldsymbol{\alpha}|$ depends on k, not n). ☐

Remark 2.13. Let us remark that $\mathcal{S}$ is closed under differentiation (as its elements are infinitely differentiable), and that every continuous linear functional on $\mathcal{S}$ is a distribution (as $\mathcal{D}(\mathbb{R}^n) \subseteq \mathcal{S}$) by Proposition 2.3.

Definition 2.12 (Tempered distribution)**.** A continuous linear functional on $\mathcal{S}$ is called a *tempered*[4] distribution.

Remark 2.14. Clearly, a distribution with compact support is tempered. However, there are distributions that are *not* tempered, e.g., for $n = 1$ the distribution induced by $f(x) = \mathrm{e}^x$ diverges for $\varphi(x) = \mathrm{e}^{-(1+x^2)^{\frac{1}{4}}} \in \mathcal{S}$ (for compactly supported $\psi \in \mathcal{D}(\mathbb{R})$, it would not be an issue).

Proposition 2.6. *Let μ be a signed measure such that $|\mu|(B(\boldsymbol{0}, R)) \leq CR^N$ for some $C > 0$ and some $N > 0$. Then μ canonically induces a tempered distribution $\mathsf{T}_\mu(\varphi) = \int_{\mathbb{R}^n} \varphi(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$.*

Proof. It holds that

$$\left|\mathsf{T}_\mu(\varphi)\right| \leq \int_{\mathbb{R}^n} |\varphi(\boldsymbol{x})| \,\mathrm{d}\,|\mu|\,(\boldsymbol{x}),$$

[3]A test function $\varphi \in \mathcal{D}(\mathbb{R}^n)$ has all its derivatives continuous and compactly supported, thus bounded.

[4]As was mentioned in the beginning of this section, alternatively we may call it *temperate* distribution.

and as $\left|\varphi(\boldsymbol{x})\right| \leq \frac{|||\varphi|||_k}{\left(1+\|\boldsymbol{x}\|^2\right)^k}$ (from (2.15)), we further obtain

$$\int_{\mathbb{R}^n} \frac{1}{\left(1+\|\boldsymbol{x}\|^2\right)^k} \,\mathrm{d}\left|\mu\right|(\boldsymbol{x}) \leq \underbrace{\int_{\|\boldsymbol{x}\|\leq 1} \frac{1}{\left(1+\|\boldsymbol{x}\|^2\right)^k} \,\mathrm{d}\left|\mu\right|(\boldsymbol{x})}_{I} + \sum_{l=0}^{\infty} \underbrace{\int_{2^l\leq\|\boldsymbol{x}\|\leq 2^{l+1}} \frac{1}{\left(1+\|\boldsymbol{x}\|^2\right)^k} \,\mathrm{d}\left|\mu\right|(\boldsymbol{x})}_{\leq \frac{1}{4^{kl}} C 2^{(l+1)N}}$$

$$\leq I + C2^N \sum_{l=0}^{\infty} 2^{-(2k-N)l},$$

which converges for $2k > N$. In other words, we have shown that T_μ is a bounded[5], thus also continuous, linear operator. □

2.4.1 Fourier Transform of Tempered Distributions

> **i** Fourier transform on the Schwartz spaced
>
> Recall that we have already discussed relationship of Schwartz space $\mathcal{S}$ and the Fourier transform in Section 2.2. In particular, for $f \in \mathcal{S}$ we have
>
> $$\widehat{f}(\boldsymbol{t}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{x},\boldsymbol{t}\rangle} \,\mathrm{d}\boldsymbol{x},$$
>
> which converges and is well-defined (by (2.3)). Moreover, from (2.5) we deduced that $\widehat{f} \in \mathcal{S}$ and that $\mathcal{F}$ is an *automorphism* on the Schwartz space — indeed, the identity $\widehat{\widehat{f}} = \check{f}$, where $\check{f}$ is the "reverse" of the function, see (2.11), further implies that the Fourier transform is a *continuous bijection* on $\mathcal{S}$.

Definition 2.13 (Fourier transform of tempered distributions)**.** Let T be a tempered distribution. We define its Fourier transform as $\widehat{\mathsf{T}}(\varphi) := \mathsf{T}(\widehat{\varphi})$.

One can notice right away that since $\varphi, \widehat{\varphi} \in \mathcal{S}$ both T, and $\widehat{\mathsf{T}}$ a tempered.

Proposition 2.7. *The following properties follow directly from Definition 2.13:*

1. $\mathcal{F}\left(\mathrm{D}^{\boldsymbol{\alpha}}\,\mathsf{T}\right) = (2\pi\mathrm{i}\boldsymbol{t})^{\boldsymbol{\alpha}}\widehat{\mathsf{T}}$;
2. $\mathrm{D}^{\boldsymbol{\alpha}}\,\widehat{\mathsf{T}} = \mathcal{F}((-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}\mathsf{T})$;
3. $\mathcal{F}\left(\mathcal{T}_h\mathsf{T}\right) = \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{h},\boldsymbol{t}\rangle}\widehat{\mathsf{T}}$;
4. $\mathcal{T}_h\widehat{\mathsf{T}} = \mathcal{F}\left(\mathrm{e}^{2\pi\mathrm{i}\langle \boldsymbol{h},\boldsymbol{t}\rangle}\mathsf{T}\right)$.

Proof. Recall Remark 1.2, otherwise obvious. □

[5] In the sense that $\mathsf{T} : X \to Y$ is a linear operator which said to be *bounded* if there exists a constant $C > 0$ such that $\|\mathsf{T}\boldsymbol{x}\|_Y \leq C\,\|\boldsymbol{x}\|_X$ for all $x \in \mathcal{D}\,(\mathsf{T})$. This, in turn, implies continuity of T (defined as usual).

Theorem 2.7. *Let μ be a signed measure with $|\mu|(\mathbb{R}^n) < \infty$, i.e., with* finite total mass. *Then its Fourier transform is the continuous and bounded function $\widehat{\mu}(\boldsymbol{t})$ defined by*

$$\widehat{\mu}(\boldsymbol{t}) = \int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{x},\boldsymbol{t}\rangle} \,\mathrm{d}\mu(\boldsymbol{x}).$$

Proof. Firstly, the defining integral surely exists as μ has finite total mass $\|\mu\| = \int_{\mathbb{R}^n} \mathrm{d}\,|\mu|(\boldsymbol{x})$. The exponential has always modulus 1, thus boundedness easily follows from $|\widehat{\mu}(\boldsymbol{t})| \leq \left|\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle} \,\mathrm{d}\mu(\boldsymbol{x})\right| = \|\mu\|$. The continuity stems of the function $\widehat{\mu}$ from the fact that as $\boldsymbol{t}_k$ converge to $\boldsymbol{t}_0$, the integrands $\mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t}_k,\boldsymbol{x}\rangle}$ converge point-wise to $\mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t}_0,\boldsymbol{x}\rangle}$ and are bounded by the integrable function $+1$ (μ has finite total mass), which hence allows the use dominated convergence theorem 4.2.

Lastly, it suffices to show that $\widehat{\mu}$ is indeed the Fourier transform. To that end, let $\varphi \in \mathcal{S}$, then

$$\begin{aligned}
\mathsf{T}_\mu \widehat{\varphi} = \int_{\mathbb{R}^n} \widehat{\varphi}(\boldsymbol{t}) \,\mathrm{d}\mu(\boldsymbol{t}) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \varphi(\boldsymbol{x}) \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{x},\boldsymbol{t}\rangle} \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\mu(\boldsymbol{t}) \\
&= \int_{\mathbb{R}^n} \varphi(\boldsymbol{x}) \underbrace{\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle} \,\mathrm{d}\mu(\boldsymbol{t})}_{\widehat{\mu}(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \varphi(\boldsymbol{x}) \widehat{\mu}(\boldsymbol{x}) \,\mathrm{d}x = \widehat{\mathsf{T}}_{\boldsymbol{\mu}} \varphi.
\end{aligned}$$

□

Example 2.3. Recall the Poisson summation formula (1.14) from Section 1.3.2,

$$\sum_{\boldsymbol{\lambda}\in\Lambda} f(\boldsymbol{\lambda}) = \sum_{\boldsymbol{\mu}\in\Lambda^*} \widehat{f}(\boldsymbol{\mu})$$

for a lattice Λ in $\mathbb{R}^n$. For the identity to hold, we required f and $\widehat{f}$ to be absolutely convergent, which we ensured by the assumption (1.12) on both functions. Clearly, if $f \in \mathcal{S}$ it is trivially satisfied, and by the properties of the Fourier transform on the Schwartz space it is also satisfied for $\widehat{f}$. In other words, this formula holds for all $f \in \mathcal{S}$.

However, we can also translate this into the distributional setting by defining the *Dirac comb* distribution

$$\mathsf{T} = \sum_{\boldsymbol{\lambda}\in\Lambda} \delta_{\boldsymbol{\lambda}} \implies \widehat{\mathsf{T}} = \sum_{\boldsymbol{\mu}\in\Lambda^*} \delta_{\boldsymbol{\mu}}.$$

In particular, it holds that $\widehat{\delta}_{\boldsymbol{0}}\varphi = \delta_{\boldsymbol{0}}\widehat{\varphi} = \int_{\mathbb{R}^n} \varphi(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ and as such $\widehat{\delta}_{\boldsymbol{0}} = 1$ in the *functional* sense (by the canonical induction of distributions from functions).

Definition 2.14 (Convolution measure)**.** Let μ and ν be finite measures on $\mathbb{R}^n$. Then we define the convolution measure as

$$\mu * \nu(A) = \int_{\mathbb{R}^n} \mu(\boldsymbol{x} + A) \,\mathrm{d}\nu(\boldsymbol{x}),$$

where $\boldsymbol{x} + A$ denotes the set A translated by $\boldsymbol{x}$.

Surely, $\mu * \nu$ is also a finite measure such that $\|\mu * \nu\| \leq \|\mu\| \cdot \|\nu\|$. Also, requirements of Proposition 2.6 are satisfied, thus the convolution measure "is" a tempered distribution. In other words, we can calculate

$$\widehat{\mu * \nu}(\boldsymbol{t}) = \int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle} \,\mathrm{d}(\mu * \nu)(\boldsymbol{x}) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{y}+\boldsymbol{x}\rangle} \,\mathrm{d}\mu(\boldsymbol{y})\,\mathrm{d}\nu(\boldsymbol{x}) = \widehat{\mu}(\boldsymbol{t})\widehat{\nu}(\boldsymbol{t}).$$

Recall now the Paley-Wiener theorem 1.7 which hinted at entire functions of restricted growth being Fourier transform of $\mathcal{L}^2$-integrable truncated functions on the real line.

Theorem 2.8. *Let* T *be a distribution with* $\operatorname{supp}\mathsf{T} \subseteq B(\mathbf{0}, A)$*. Then* $\widehat{\mathsf{T}}$ *is a function on* $\mathbb{R}^n$ *that extends to entire function on* $\mathbb{C}^n$*. This function satisfies*

$$\left|\widehat{\mathsf{T}}(\boldsymbol{z})\right| \leq C\left(1 + \|\boldsymbol{z}\|^2\right)^{\frac{N}{2}} \mathrm{e}^{2\pi A\|\operatorname{Im}\boldsymbol{z}\|}$$

such that $\widehat{\mathsf{T}}(\boldsymbol{z}) = \mathsf{T}\left(\mathrm{e}^{-2\pi\mathrm{i}\langle\cdot,\boldsymbol{z}\rangle}\right)$.

Proof. We shall start by writing down the Taylor expansion in $\boldsymbol{z}$ of $\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{z}\rangle} = \sum_{\boldsymbol{\alpha}} \frac{(-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}\boldsymbol{z}^{\boldsymbol{\alpha}}$ and defining

$$\mathsf{T}\Big(\overbrace{(-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}}^{\notin\mathcal{D}(\mathbb{R}^n)}\Big) = \mathsf{T}\Big(\overbrace{(-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}\chi(\boldsymbol{x})}^{\in\mathcal{D}(\mathbb{R}^n)}\Big)$$

for $\chi \in \mathcal{D}(\mathbb{R}^n)$ such that $\chi(\boldsymbol{x}) = 1$ when $\|\boldsymbol{x}\| \leq A$. Then $\left|\mathsf{T}\left((-2\pi\mathrm{i}\boldsymbol{x})^{\alpha}\right)\right| \leq C\left\|\left|(-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}\chi(\boldsymbol{x})\right\|\right|_N$ for some $C > 0$ and some $N > 0$, where the semi-norm is polynomial in $\boldsymbol{\alpha}$ (by the internal differentiation). Thus $\mathsf{T}\left(\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{z}\rangle}\right) = \sum_{\boldsymbol{\alpha}} \frac{\mathsf{T}((-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}})\boldsymbol{z}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}$ converges uniformly on compact subsets of $\mathbb{C}^n$ due to the polynomial behavior of $\mathsf{T}\left((-2\pi\mathrm{i}\boldsymbol{x})^{\boldsymbol{\alpha}}\right)$. Now[6], by Definition 2.3,

$$\left|\mathsf{T}\left(\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{z}\rangle}\right)\right| \leq C' \sum_{|\boldsymbol{\alpha}|\leq N} \sup_{\|\boldsymbol{x}\|\leq 2\pi A} \underbrace{\left\|\mathrm{D}^{\boldsymbol{\alpha}}\,\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{z}\rangle}\right\|}_{\left\|(-2\pi\mathrm{i}\boldsymbol{z})^{\boldsymbol{\alpha}}\mathrm{e}^{-2\pi\langle\boldsymbol{x},\boldsymbol{z}\rangle}\right\|},$$

where $\|-2\pi\mathrm{i}\boldsymbol{z}\|^{\boldsymbol{\alpha}} \leq C\left(1 + \|\boldsymbol{z}\|\right)^{\frac{N}{2}}$ and $\left|\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{z}\rangle}\right| \leq \mathrm{e}^{2\pi A\|\operatorname{Im}(\boldsymbol{z})\|}$.

Lastly, we shall verify that we have indeed found the Fourier transform. Let $\varphi \in \mathcal{S}$ be a test function (such that $\widehat{\varphi}$ is a function of the variable $\boldsymbol{t}$), then

$$\mathsf{T}(\widehat{\varphi}) = \mathsf{T}\left(\int_{\mathbb{R}^n} \varphi(\boldsymbol{x})\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\boldsymbol{x}\right) = \int_{\mathbb{R}^n} \varphi(\boldsymbol{x})\mathsf{T}\left(\mathrm{e}^{-2\pi\mathrm{i}\langle\boldsymbol{x},\boldsymbol{t}\rangle}\right)\,\mathrm{d}\boldsymbol{x} = \widehat{\mathsf{T}}(\varphi),$$

where the second operation is justified by the fact that we can approximate an integral by Riemann sums and the fact that T is continuous and linear. ∎

[6]For clarity, let us mention the supremum is equivalent with the $\mathcal{L}^\infty$-norm.

2.4.2 Convolutions of Tempered Distributions

Remark 2.15. Let $f, g \in \mathcal{S}$ be Schwartz functions, then $f * g$ is defined (as $f, g \in \mathcal{L}^1$, see (2.3)) and $\widehat{f * g} = \hat{f} \cdot \hat{g} \in \mathcal{S}$ (both $\hat{f}, \hat{g} \in \mathcal{S}$), thus also $f * g \in \mathcal{S}$.

We shall now consider convolution of a tempered distribution and a Schwartz test function analogously to the "general distribution" case, see Definition 2.9. In other words, for T a tempered distribution and $f \in \mathcal{S}$, we have

$$\mathsf{T} * f(\boldsymbol{x}) := \mathsf{T}(\underbrace{\mathcal{T}_{\boldsymbol{x}} \check{f}}_{\in \mathcal{S}}), \tag{2.16}$$

which is surely well-defined.

Theorem 2.9. *Let* T *be a tempered distribution and* $f \in \mathcal{S}$*. Then* $\mathsf{T} * f$ *is a* C^∞*-function and satisfies*

$$\left|\mathsf{T} * f(\boldsymbol{x})\right| \leq C\left(1 + \|\boldsymbol{x}\|^2\right)^N$$

for some C, N*.*

In other words, Theorem 2.9 states that $\mathsf{T} * f$ is a function growing at most polynomially (whilst being in C^∞).

Proof. Right away, we have for some C, N (see Definition 2.3)

$$\begin{aligned}
\left|\mathsf{T} * f(\boldsymbol{h})\right| \triangleq \left|\mathsf{T}\left(\mathcal{T}_{\boldsymbol{h}} \check{f}\right)\right| \leq C \left\|\left|\mathcal{T}_{\boldsymbol{h}} \check{f}\right|\right\|_N \\
&= C \sum_{|\boldsymbol{\alpha}| \leq N} \left\| \left(1 + \|\boldsymbol{x}\|^2\right)^N \mathrm{D}^{\boldsymbol{\alpha}} f(\boldsymbol{x} - \boldsymbol{h}) \right\|_\infty \\
&\overset{\text{subs.}}{=} C \sum_{|\boldsymbol{\alpha}| \leq N} \left\| \left(1 + \|\boldsymbol{x} + \boldsymbol{h}\|^2\right)^N \mathrm{D}^{\boldsymbol{\alpha}} f(\boldsymbol{x}) \right\|_\infty \\
&\leq 2C \left\|\left| f \right|\right\|_N \left(1 + \|\boldsymbol{h}\|^2\right)^N,
\end{aligned}$$

where we used $(1 + \|\boldsymbol{x} + \boldsymbol{h}\|^2) \leq 2(1 + \|\boldsymbol{x}\|^2)(1 + \|\boldsymbol{h}\|^2)$. Let us remark the $\mathcal{L}^\infty$-norm is understood w.r.t. $\boldsymbol{x}$. The fact that $\mathsf{T} * f \in C^\infty$ stems from the relationship of distributions and differential operators, see Definition 2.6, and f being Schwartz. □

Let us now state an analog of Theorem 2.4 for tempered distributions.

Theorem 2.10. *Let* L *be a* continuous *linear map* $\mathcal{S} \to C^\infty$*, which commutes with all translations. Then* L *is a convolution with a tempered distribution.*

Proof. Analogous to the proof of Theorem 2.4. □

Theorem 2.11. *Let* T *be a tempered distribution and* $f \in \mathcal{S}$*. Then* $\widehat{\mathsf{T} * f} = \hat{f} \cdot \hat{\mathsf{T}}$*.*

Proof. Since T is tempered, also $\widehat{\mathsf{T}}$ is tempered and thus $\widehat{f}\cdot\widehat{\mathsf{T}}$ is tempered. Furthermore, even $\mathsf{T}*f$ is a tempered distribution, since it is a $\mathcal{C}^\infty$-function of at most polynomial growth (by Theorem 2.9) and as such we can integrate it against any $g\in\mathcal{S}$ and get convergence. To complete the proof we calculate for an arbitrary $g\in\mathcal{S}$

$$\begin{aligned}
\widehat{\mathsf{T}*f}(g) &\triangleq (\mathsf{T}*f)(\widehat{g}) = \mathsf{T}*f*\check{\widehat{g}}(\mathbf{0}) = \mathsf{T}*\left(\check{f}*\widehat{g}\right)^{\vee}(\mathbf{0})\\
&= \mathsf{T}\left(\check{f}*\widehat{g}\right) = \mathsf{T}\left(\left(f*\check{\widehat{g}}\right)^{\vee}\right) = \widehat{\mathsf{T}}\left(\widehat{f*\check{\widehat{g}}}\right)\\
&= \widehat{\mathsf{T}}\left(\widehat{f}\cdot\widehat{\check{\widehat{g}}}\right) = \widehat{\mathsf{T}}\left(\widehat{f}\cdot\check{\check{g}}\right) = \widehat{\mathsf{T}}\left(\widehat{f}\cdot g\right)\\
&= \widehat{f}\cdot\widehat{\mathsf{T}}(g).
\end{aligned}$$

using Theorem 2.10 and (2.13) for the second step (with $\widehat{g} = \mathcal{T}_{\mathbf{0}}\widehat{g}$), and $\widehat{\widehat{f}} = \check{f}$ for the eighth step, as well as the fact that

$$\begin{aligned}
\left(f*g\right)^{\vee}(\boldsymbol{x}) = (f*g)(-\boldsymbol{x}) &= \int_{\mathbb{R}^n} f(-\boldsymbol{x}-\boldsymbol{y})g(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{\mathbb{R}^n} \check{f}(\boldsymbol{x}+\boldsymbol{y})\check{g}(-\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\\
&\overset{\boldsymbol{s}=-\boldsymbol{y}}{=} \int_{\mathbb{R}^n} \check{f}(\boldsymbol{x}-\boldsymbol{s})\check{g}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = (\check{f}*\check{g})(\boldsymbol{x}).
\end{aligned}$$

□

Theorem 2.12. *Let* R *be a compactly supported distribution and* T *a tempered one. Then* $\mathsf{T}*\mathsf{R}$ *is tempered and* $\widehat{\mathsf{T}*\mathsf{R}} = \widehat{\mathsf{R}}\cdot\widehat{\mathsf{T}}$.

Proof. Firstly, observe that convolution with R continuously takes $\mathcal{S}$ into itself. And since $\widehat{\mathsf{R}}$ is a $\mathcal{C}^\infty$-function of polynomial growth by Theorem 2.8, multiplication (as in Definition 2.8) with $\widehat{\mathsf{R}}$ also takes $\mathcal{S}$ into itself continuously, i.e., the problem is well-defined.

Now, $\mathsf{T}*\mathsf{R}$ is clearly a continuous linear map in the sense of Remark 2.11, thus a tempered distribution. Then

$$\begin{aligned}
\widehat{\mathsf{T}*\mathsf{R}}(f) &\triangleq (\mathsf{T}*\mathsf{R})(\widehat{f}) = \mathsf{T}*\mathsf{R}*\check{\widehat{f}}(\mathbf{0})\\
&= \mathsf{T}\left(\widehat{\widehat{\mathsf{R}}\cdot f}\right) = \widehat{\mathsf{T}}\left(\widehat{\mathsf{R}}f\right) = \widehat{\mathsf{R}}\cdot\widehat{\mathsf{T}}(f),
\end{aligned}$$

where we used the same “tricks” as in the proof of Theorem 2.11. In particular, by the introductory remarks to this proof, $\widehat{\mathsf{R}*\check{\widehat{f}}} = \widehat{\mathsf{R}}\cdot f$ is a Schwartz function, and $\left(\mathsf{R}*\check{\widehat{f}}\right)^{\vee} = \widehat{\widehat{\mathsf{R}}\cdot f}$. □

2.5 Analytic Continuation

Remark 2.16 (Homogeneous distribution [8, chpt. 23]). Let f be a locally integrable function on $\mathbb{R}^n$ (which thus canonically prescribes a distribution) and $\boldsymbol{L}$ an invertible linear transformation of

$\mathbb{R}^n$ into itself. Such mapping $\boldsymbol{L}$ surely induces a new function $(f \circ \boldsymbol{L})(\boldsymbol{x}) = f(\boldsymbol{L}\boldsymbol{x})$, which is again invertible. Using the distribution perspective, we have

$$(f \circ \boldsymbol{L})(\varphi) = \int_{\mathbb{R}^n} f(\boldsymbol{L}\boldsymbol{x})\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \overset{\boldsymbol{y}=\boldsymbol{L}\boldsymbol{x}}{=} \int_{\mathbb{R}^n} f(\boldsymbol{y})\varphi(\boldsymbol{L}^{-1}\boldsymbol{y})\left|\det \boldsymbol{L}^{-1}\right|\mathrm{d}\boldsymbol{y} = f\left(\frac{\varphi \circ \boldsymbol{L}^{-1}}{|\det \boldsymbol{L}|}\right),$$

where the last equality holds in the distributional sense. Following this principle, we define $(\mathsf{T} \circ \boldsymbol{L})(\varphi) := \frac{1}{|\det \boldsymbol{L}|}\mathsf{T}(\varphi \circ \boldsymbol{L}^{-1})$, and as such the composition with $\boldsymbol{L}$ defines a linear mapping of the space of distributions onto itself (for distributions with $\boldsymbol{L}$-invariant domains).

An example of particular importance for us is $\boldsymbol{L}_\varepsilon = \varepsilon \boldsymbol{I}_n$ for some positive number ε. Here $|\det \boldsymbol{L}_\varepsilon| = \varepsilon^n$ and

$$(\mathsf{T} \circ \boldsymbol{L}_\varepsilon)(\varphi) = \frac{1}{\varepsilon^n}\mathsf{T}\left(\varphi\left(\frac{\cdot}{\varepsilon}\right)\right). \tag{2.17}$$

Using this equality, a distribution T is **homogeneous of degree** k, if for all positive ε, it holds that $\mathsf{T} \circ \boldsymbol{L}_\varepsilon = \varepsilon^k \mathsf{T}$. In particular, for the Dirac distribution at origin we have

$$(\delta_{\boldsymbol{0}} \circ \boldsymbol{L}_\varepsilon)(\varphi) = \frac{1}{\varepsilon^n}\delta_{\boldsymbol{0}}\left(\varphi\left(\frac{\cdot}{\varepsilon}\right)\right) = \frac{1}{\varepsilon^n}\varphi\left(\frac{0}{\varepsilon}\right) = \varepsilon^{-n}\delta_{\boldsymbol{0}}(\varphi), \tag{2.18}$$

i.e., the Dirac distribution at origin $\delta_{\boldsymbol{0}}$ is homogeneous of degree $-n$.

Caution

The following example has been adapted from Prof. Grabner's lecture and [8] (page 156 and onward).

Example 2.4. Consider the function $f(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{x}\|^{n-\lambda}}$ for $\mathfrak{Re}\,\lambda > 0$. Then f is locally integrable, i.e., for every ball we can evaluate the integral[7], and thus canonically corresponds to a tempered distribution. We shall discuss the canonical distribution T_λ is meromorphic[8] with respect to λ.

For $0 < \mathfrak{Re}\,\lambda < \frac{n}{2}$ we can write (so that the "$\mathcal{L}^2$-part" is still square-integrable)

$$f(\boldsymbol{x}) = \underbrace{\chi_{\|\cdot\|\leq 1}(\boldsymbol{x})\frac{1}{\|\boldsymbol{x}\|^{n-\lambda}}}_{\in\mathcal{L}^1} + \underbrace{\chi_{\|\cdot\|>1}(\boldsymbol{x})\frac{1}{\|\boldsymbol{x}\|^{n-\lambda}}}_{\in\mathcal{L}^2}, \tag{2.19}$$

and because we can compute the Fourier transform of both parts (using Definition 1.1 and Theorem 1.6), by linearity of $\mathcal{F}$ we have that the Fourier transform of f is again a function. Since f is radial, $\widehat{f}$ is also radial. Furthermore, because f is homogeneous of degree $\lambda - n$, i.e., $f(t\boldsymbol{x}) = t^{\lambda-n}f(\boldsymbol{x})$ for $t > 0$, $\widehat{f}$ has to also be homogeneous of degree $-n-(\lambda-n) = -\lambda$. In total, we obtain $\widehat{f}(\boldsymbol{t}) = C(n,\lambda)\frac{1}{\|\boldsymbol{t}\|^\lambda}$.

[7]The reason for the local integrability may be illustrated with $\int_{\|\boldsymbol{x}\|\leq 1} \frac{1}{\|\boldsymbol{x}\|^{n-2}}\,\mathrm{d}\boldsymbol{x} = \int_0^1 \frac{1}{r^{n-2}} r^{n-1}\,\mathrm{d}r \cdot \omega_n$, which is clearly a convergent integral.

[8]Meromorphic function is holomorphic on $\mathbb{C}$ expect for at most countably many isolated poles.

Now, we can use just one well-chosen test function to discern the full form of $C(n,\lambda)$. Indeed, let $\psi(\boldsymbol{x}) = \mathrm{e}^{-\pi\|\boldsymbol{x}\|^2}$ (which equals its own Fourier transform), then using the equality $\mathsf{T}_{\widehat{f}}\psi = \widehat{\mathsf{T}}_f\psi = \mathsf{T}_f\widehat{\psi}$, we have

$$\begin{aligned}
\mathsf{T}_f\widehat{\psi} &= \int_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{e}^{-\pi\|\boldsymbol{x}\|^2}\,\mathrm{d}\boldsymbol{x} = \omega_n\int_0^\infty r^{\lambda-n}\mathrm{e}^{-\pi r^2}r^{n-1}\,\mathrm{d}r = \omega_n\int_0^\infty r^{\lambda-1}\mathrm{e}^{-\pi r^2}\,\mathrm{d}r\\
&\overset{v=\pi r^2}{=} \frac{\omega}{2\pi}\int_0^\infty\left(\frac{v}{\pi}\right)^{\frac{\lambda-2}{2}}\mathrm{e}^{-v}\,\mathrm{d}v = \frac{\omega_n}{2}\pi^{-\frac{\lambda}{2}}\int_0^\infty v^{\frac{\lambda}{2}-1}\mathrm{e}^{-v}\,\mathrm{d}v\\
&= \frac{\omega_n}{2}\pi^{-\frac{\lambda}{2}}\Gamma\left(\frac{\lambda}{2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{T}_{\widehat{f}}\psi &= \int_{\mathbb{R}^n}\widehat{f}(\boldsymbol{t})\mathrm{e}^{-\pi\|\boldsymbol{t}\|^2}\,\mathrm{d}\boldsymbol{t} = \omega_n C(n,\lambda)\int_0^\infty r^{-\lambda}\mathrm{e}^{-\pi r^2}r^{n-1}\,\mathrm{d}r\\
&= \omega_n C(n,\lambda)\int_0^\infty r^{n-\lambda-1}\mathrm{e}^{-\pi r^2}\,\mathrm{d}r\\
&\overset{v=\pi r^2}{=} \frac{\omega_n}{2\pi}C(n,\lambda)\int_0^\infty\left(\frac{v}{\pi}\right)^{\frac{n-\lambda-2}{2}}\mathrm{e}^{-v}\,\mathrm{d}v = \frac{\omega_n}{2}\pi^{\frac{\lambda-n}{2}}C(n,\lambda)\int_0^\infty v^{\frac{n-\lambda}{2}-1}\mathrm{e}^{-v}\,\mathrm{d}v\\
&= \frac{\omega_n}{2}\pi^{\frac{\lambda-n}{2}}C(n,\lambda)\Gamma\left(\frac{n-\lambda}{2}\right).
\end{aligned}$$

Comparing both sides yields

$$C(n,\lambda) = \pi^{\frac{n}{2}-\lambda}\frac{\Gamma\left(\frac{\lambda}{2}\right)}{\Gamma\left(\frac{n-\lambda}{2}\right)} \implies \widehat{f}(\boldsymbol{t}) = \pi^{\frac{n}{2}-\lambda}\frac{\Gamma\left(\frac{\lambda}{2}\right)}{\Gamma\left(\frac{n-\lambda}{2}\right)}\|\boldsymbol{t}\|^{-\lambda}, \tag{2.20}$$

which even holds for $0 < \mathfrak{Re}\,\lambda < n$ (by analytic continuation). Note that (a multiple of) T_f is typically called the *Riesz potential* w.r.t λ.

Indeed, recall that λ is a complex parameter, and the dependence of $\int_{\mathbb{R}^n}\frac{1}{\|\boldsymbol{x}\|^{n-\lambda}}\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is holomorphic in λ (for a fixed φ and $\mathfrak{Re}\,\lambda > 0$). Such mapping is even meromorphic for $\lambda\in\mathbb{C}$ (save for even non-positive integers, as can be seen in (2.20)). So far, we relied on the fact that f was integrable around 0 (see (2.19)), which simply does not hold for $\mathfrak{Re}\,\lambda > n$; we shall come back to this issue later.

Example 2.5. For simplicity, let us make the situation easier by taking $n = 1$. For $f(x) = |x|$ we have already computed in Example 2.2 that $\mathrm{D}^2 f = 2\delta$ (in the distribution sense). Also, by Example 2.3 $\widehat{\mathrm{D}^2 f} = 2 \implies$, so

$$-4\pi^2 t^2\widehat{f}(t) = 2 \implies \widehat{f}(t) = -\frac{1}{2\pi^2 t^2}$$

for $t\neq 0$. This is even suggested by homogenity argument used above. Consider now the following distribution

$$\mathsf{S}(\varphi) := \frac{1}{\pi^2}\int_0^\infty\left(\varphi(0) - \frac{\varphi(x)+\varphi(-x)}{2}\right)\frac{\mathrm{d}x}{x^2},\quad \varphi\in\mathcal{S}(\mathbb{R}),$$

where we can interpret $\frac{\varphi(x)+\varphi(-x)}{2}$ as an average over sphere of all radii (with $\int \mathrm{d}x$), which vanishes for $x \to \infty$ by $\varphi \in \mathcal{S}(\mathbb{R})$. In other words, the integrand behaves primarily like $\frac{1}{x^2}$ for large x, and is bounded near origin (by Taylor, for example). Therefore, $\mathsf{S}(\varphi)$ is convergent for every $\varphi \in \mathcal{S}(\mathbb{R})$. However, the distribution prescribed by $\widehat{f}$ is $-\frac{1}{2\pi^2}\int_{-\infty}^{\infty}\varphi(t)\frac{\mathrm{d}t}{t^2}$ (notice the different lower bound), which is *not* convergent[9] for every $\varphi \in \mathcal{S}(\mathbb{R})$. If φ vanishes at origin, we indeed get $\mathsf{S} = \mathsf{T}_{\widehat{f}}(= \widehat{f})$, thus it must hold $\mathsf{S} - \widehat{f} = c\delta_0$ for an appropriate constant c. By noticing

$$\begin{aligned}(\mathsf{S}\circ \boldsymbol{L}_\varepsilon)(\varphi) &= \frac{1}{\varepsilon}\mathsf{S}(\varphi\circ \boldsymbol{L}_\varepsilon^{-1}) = \frac{1}{\pi^2\varepsilon}\int_0^\infty \left(\varphi(0) - \frac{\varphi\left(\frac{x}{\varepsilon}\right)+\varphi\left(-\frac{x}{\varepsilon}\right)}{2}\right)\frac{\mathrm{d}x}{x^2} \\ &\overset{y=\frac{x}{\varepsilon}}{=} \frac{1}{\pi^2\varepsilon}\int_0^\infty \left(\varphi(0) - \frac{\varphi(y)+\varphi(-y)}{2}\right)\frac{\frac{\mathrm{d}x}{\varepsilon}}{\frac{x^2}{\varepsilon^2}} = \varepsilon^{-2}\mathsf{S}(\varphi),\end{aligned} \tag{2.21}$$

we get that both S and $\widehat{f}$ are homogeneous of degree -2, whereas δ_0 is only homogeneous of degree -1 (see (2.18)), thus $c = 0$ and $\mathsf{S} = \widehat{f}$. To put it differently, S is an analytic continuation (or canonical extension) of $\widehat{f}$.

Similarly for general n and $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ we have that $\mathsf{T}_f(\varphi) = \int_{\mathbb{R}^n}\|\boldsymbol{x}\|\,\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is a tempered distribution ($\varphi \in \mathcal{S}(\mathbb{R}^n)$ decays faster than any negative power of $\|\boldsymbol{x}\|$ and is in $\mathcal{L}^p$-space). Moreover, we again expect the Fourier transform to be $\widehat{f}(\boldsymbol{t}) = -\frac{c}{\|\boldsymbol{t}\|^{n+1}}$ by homogenity argument (for $\boldsymbol{t} \neq \boldsymbol{0}$). Now take

$$\mathsf{S}(\varphi) := c\int_0^\infty \left(\varphi(\boldsymbol{0}) - \frac{1}{\omega_n}\int_{\|\boldsymbol{x}\|=1}\varphi(r\boldsymbol{x})\,\mathrm{d}A(\boldsymbol{x})\right)\frac{\mathrm{d}r}{r^2},$$

and again $\mathsf{S}(\varphi) = \widehat{f}(\varphi)$ for $\varphi(\boldsymbol{0}) = \boldsymbol{0}$, hence $\mathsf{S} - \widehat{f} = \mu\delta_{\boldsymbol{0}}$. The terms on the LHS are both homogeneous of degree $-n-1$, whereas $\delta_{\boldsymbol{0}}$ is homogeneous of degree $-n$ (see (2.18)), which once again leads to $\mu = 0$.

Lastly, to determine c we calculate $\operatorname{grad} f = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ and

$$\begin{gathered}\frac{\partial}{\partial x_1}\frac{x_1}{\|\boldsymbol{x}\|} = \frac{1}{\|\boldsymbol{x}\|} + x_1\cdot\frac{2x_1}{\|\boldsymbol{x}\|^3}\cdot\left(-\frac{1}{2}\right) = \frac{1}{\|\boldsymbol{x}\|} - \frac{x_1^2}{\|\boldsymbol{x}\|^3} \\ \Downarrow \\ \Delta f = \frac{n}{\|\boldsymbol{x}\|} - \frac{\|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^3} = \frac{n-1}{\|\boldsymbol{x}\|}.\end{gathered} \tag{2.22}$$

Then using (1.1) for each term of Δf and (2.20) (with $\lambda = n-1$) yields

$$\left.\begin{aligned}\widehat{\Delta f} &= -4\pi\|\boldsymbol{t}\|^2\,\widehat{f}(\boldsymbol{t}) = -4\pi\|\boldsymbol{t}\|^2\,\mathsf{S} \\ \widehat{\left(\frac{n-1}{\|\boldsymbol{x}\|}\right)} &= 2\frac{n-1}{2}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\pi^{1-\frac{n}{2}}\frac{1}{\|\boldsymbol{t}\|^{n-1}}\end{aligned}\right\} \overset{\Gamma(\frac{1}{2})=\sqrt{\pi}}{\Longrightarrow} c = 2\Gamma\left(\frac{n+1}{2}\right)\pi^{-\frac{n-1}{2}}.$$

[9]Indeed, the averaging over a sphere (here in $\mathbb{R}$) and the deliberate form of the integrand of S allow for integrability around origin.

Remark 2.17. In practice, we can also compute the Fourier transform of higher powers of $\|\boldsymbol{x}\|$, i.e., $\|\boldsymbol{x}\|^{\lambda}$, however we would need more correction terms in S (apart from the sphere-averaging) for derivatives in φ.

Remark 2.18. Similarly to above, we can try to construct an analytic continuation of the Gamma function Γ. Indeed, recalling Definition 4.1 we know that Γ is defined for $\mathfrak{Re}\, z > 0$, then

$$\begin{aligned}
\Gamma(z) &= \int_0^\infty t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t \\
&= \int_0^1 t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t + \overbrace{\int_1^\infty t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t}^{\text{convergent for any } z} \\
&= \int_0^1 t^{z-1}(\mathrm{e}^{-t}-1)\,\mathrm{d}t + \underbrace{\int_0^1 t^{s-1}\,\mathrm{d}t}_{\frac{1}{z}} + \int_1^\infty t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t,
\end{aligned}$$

and $\int_0^1 t^{z-1}(\mathrm{e}^{-t}-1)\,\mathrm{d}t$ converges for $\mathfrak{Re}\, z > -1$ (by, for example, Taylor expansion of e^{-t}). Notice that this is the same idea as we did above (subtracting the problematic term), which gives us the analytic continuation. Specifically for $-1 < \mathfrak{Re}\, z < 0$ we have $\int_0^1 t^{z-1}\,\mathrm{d}t = \frac{1}{z} = -\int_1^\infty t^{z-1}\,\mathrm{d}t$, hence

$$\Gamma(z) = \int_0^1 t^{z-1}(\mathrm{e}^{-t}-1)\,\mathrm{d}t - \int_1^\infty t^{z-1}\,\mathrm{d}t + \int_1^\infty t^{z-1}\mathrm{e}^{-t}\,\mathrm{d}t = \int_0^\infty (\mathrm{e}^{-t}-1)t^{z-1}\,\mathrm{d}t.$$

We shall continue with Example 2.5 in the spirit of Remark 2.17. Firstly, notice that for $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ we determined $\widehat{f}$ as an explicit tempered distribution such that $\operatorname{supp}\widehat{f} = \mathbb{R}^n$. In more generality, consider $f_\lambda(\boldsymbol{x}) = \|\boldsymbol{x}\|^{\lambda}$, and, for simplicity choose $\lambda = 2$, then by (1.1)

$$\begin{aligned}
(-4\pi^2)\mathsf{T}_{\widehat{f_2}}\varphi = (-4\pi^2)\mathsf{T}_{f_2}\widehat{\varphi} &= \int_{\mathbb{R}^n} (-4\pi^2)\|\boldsymbol{t}\|^2\,\widehat{\varphi}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \Delta_{\boldsymbol{x}}\varphi(\boldsymbol{x})\mathrm{e}^{-2\pi\langle \boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{t} \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{R}^n} \Delta_{\boldsymbol{x}}\varphi(\boldsymbol{x})\int_{\mathbb{R}^n} \mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\boldsymbol{t}\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \Delta\varphi(\boldsymbol{x})\delta_{\boldsymbol{0}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \Delta\varphi(\boldsymbol{0}) \\
&\implies \widehat{f_2} = -\frac{1}{4\pi^2}\Delta\delta_{\boldsymbol{0}},
\end{aligned}$$

because the following distributional identity holds for the Dirac delta at origin,

$$\int_{R^n}\varphi(\boldsymbol{x})\int_{\mathbb{R}^n}\mathrm{e}^{-2\pi\mathrm{i}\langle \boldsymbol{t},\boldsymbol{x}\rangle}\,\mathrm{d}\boldsymbol{t}\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n}\widehat{\varphi}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \int_{\mathbb{R}^n}\underbrace{\widehat{\delta_{\boldsymbol{0}}}(\boldsymbol{x})}_{1}\,\widehat{\varphi}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \delta_{\boldsymbol{0}}\widehat{\widehat{\varphi}} = \delta_{\boldsymbol{0}}\check{\varphi} = \delta_{\boldsymbol{0}}\varphi.$$

This is rather remarkable, as it is now a local operator with $\operatorname{supp}\widehat{f_2} = \{\boldsymbol{0}\}$ (whilst before it acted globally). Moreover, observe that for a monomial $P(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{\alpha}}$ it further holds by the same argument that

$$\int_{\mathbb{R}^n} (2\pi \mathrm{i})^{|\boldsymbol{\alpha}|} \boldsymbol{t}^{\boldsymbol{\alpha}} \widehat{\varphi}(\boldsymbol{t}) \, \mathrm{d}\boldsymbol{t} = \mathrm{D}^{\boldsymbol{\alpha}} \varphi(\mathbf{0}) = \mathrm{D}^{\boldsymbol{\alpha}} \delta_{\mathbf{0}} \varphi,$$

i.e., $\widehat{P} = \frac{1}{(2\pi \mathrm{i})^{|\boldsymbol{\alpha}|}} \mathrm{D}^{\boldsymbol{\alpha}} \delta_{\mathbf{0}}$, which is again local.

Let us phrase this approach (of extending/continuing distributions) more generally. To this end, assume T_λ is a distributions depending holomorphically on $\lambda \in U \subseteq \mathbb{C}$, i.e., for every test function φ the mapping $\lambda \mapsto \mathsf{T}_\lambda(\varphi)$ is holomorphic on U. Now for any $\lambda_0 \in U$ we can express $\mathsf{T}_\lambda(\varphi)$ as a power series around λ_0,

$$\mathsf{T}_\lambda(\varphi) = \sum_{k=0}^{\infty} \mathsf{A}_k(\varphi)(\lambda - \lambda_0)^k$$

with $\mathsf{A}_k(\varphi) = \frac{1}{k!} \frac{\mathrm{d}^k \mathsf{T}_\lambda(\varphi)}{\mathrm{d}\lambda^k}\Big|_{\lambda=\lambda_0}$, which follows from holomorphy. Surely, we can exploit this!

Example 2.6 (Adapted from the lecture and [8, chpt. 24])**.** To begin in a simpler setting, let $n = 1$ and choose $f_\lambda(x) = \begin{cases} 0 & x < 0 \\ x^{\lambda-1} & x > 0 \end{cases}$ for $\mathfrak{Re}\,\lambda > 0$, hence f is *locally* integrable (and thus a distribution). Now we shall closely mirror what we did for the Gamma function Γ in Remark 2.18, i.e., take the Taylor polynomial of order $N-1$ (chosen arbitrarily) of $\varphi(x) = P_{N-1}(x) + x^N g(x)$, then

$$\begin{aligned}
\mathsf{T}_{f_\lambda}(\varphi) &= \int_{\mathbb{R}} f_\lambda(x)\varphi(x)\,\mathrm{d}x = \int_0^\infty x^{\lambda-1}\varphi(x)\,\mathrm{d}x \\
&= \int_0^1 P_{N-1}(x)x^{\lambda-1}\,\mathrm{d}x + \int_0^1 x^{N+\lambda-1}g(x)\,\mathrm{d}x + \int_1^\infty x^{\lambda-1}\varphi(x)\,\mathrm{d}x \\
&= \int_0^1 \overbrace{\sum_{k=0}^{N-1} \frac{1}{k!} \mathrm{D}^k \varphi(0) x^k}^{P_{N-1}(x)} x^{\lambda-1}\,\mathrm{d}x + \text{remaining terms} \\
&= \sum_{k=0}^{N-1} \frac{1}{k!} \frac{\mathrm{D}^k \varphi(0)}{k+\lambda} \cdot (1-0) + \text{remaining terms} \\
&= \sum_{k=0}^{N-1} \frac{1}{k!(k+\lambda)} \mathrm{D}^k \varphi(0) + \underbrace{\int_0^1 x^{N+\lambda-1}g(x)\,\mathrm{d}x}_{\text{holomorphic for } \mathfrak{Re}\,\lambda > -N} + \underbrace{\int_1^\infty x^{\lambda-1}\varphi(x)\,\mathrm{d}x}_{\text{holomorphic by Morera}},
\end{aligned}$$

where the first term is just a rational function of λ, the second is clearly holomorphic for $\mathfrak{Re}\,\lambda > -N$, and last term is even entire by Morera's theorem 4.8[10]. Since φ and N were chosen arbitrarily, it is evident we can continue/extend the distribution for $\lambda \in \mathbb{C} \setminus (-\mathbb{N}_0)$ and that non-positive integers form poles of the mapping.

However, one might recall that the Gamma function has the same poles, thus it might render advantageous to consider the distribution $\mathsf{T}_\lambda = \frac{f_\lambda}{\Gamma(\lambda)}$. As $\frac{1}{\Gamma(\lambda)}$ only rescales the distribution, then $\mathrm{D}\,\mathsf{T}_\lambda = \frac{\lambda-1}{\Gamma(\lambda)} f_{\lambda-1} = \mathsf{T}_{\lambda-1}$. Noticing $\mathrm{D}\,\mathsf{T}_1 =: \mathsf{T}_0 = \delta_0$ (as T_1 is the Heaviside function, which has the

[10] Indeed $\oint_C \int_1^\infty x^{\lambda-1}\varphi(x)\,\mathrm{d}x\,\mathrm{d}\lambda = \int_1^\infty \varphi(x) \oint_C \mathrm{e}^{\log(x)(\lambda-1)}\,\mathrm{d}\lambda\,\mathrm{d}x = \int_1^\infty \varphi(x)\cdot 0\,\mathrm{d}x = 0$.

Dirac distribution as its derivative, see Example 2.2), we finally obtain $\mathsf{T}_{-k} = \mathrm{D}^k\,\delta_0$. The LHS is surely a tempered distribution, thus also $\widehat{\mathsf{T}}_{-k} = (-2\pi i t)^k$.

Example 2.7 (Continuation of Example 2.4, [8, chpt. 24])**.** Last, but not least, we shall consider the continuation of Example 2.4, where we shall address the issues discussed at the end of the example. To this end, let $f_\lambda(\boldsymbol{x}) = \|\boldsymbol{x}\|^{\lambda-n}$ which is surely locally integrable for $\mathfrak{Re}\,\lambda > 0$, and therefore a distribution on the same λ-domain. Now by an analogous construction as in Example 2.6,

$$\begin{aligned}\mathsf{T}_{f_\lambda}\varphi &= \int_{\mathbb{R}^n} f_\lambda(\boldsymbol{x})\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &= \underbrace{\int_{\|\boldsymbol{x}\|\le 1} P_{N-1}(\boldsymbol{x})\,\|\boldsymbol{x}\|^{\lambda-n}\,\mathrm{d}\boldsymbol{x}}_{R_\lambda(\boldsymbol{x})} + \int_{\|\boldsymbol{x}\|\le 1} \underbrace{(\varphi - P_{N-1})\,(\boldsymbol{x})}_{O\left(\|\boldsymbol{x}\|^N\right)}\,\|\boldsymbol{x}\|^{\lambda-n}\,\mathrm{d}\boldsymbol{x} + \int_{\|\boldsymbol{x}\|>1} \|\boldsymbol{x}\|^{\lambda-n}\,\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},\end{aligned} \tag{2.23}$$

where again the second term is holomorphic for $\mathfrak{Re}\,\lambda > -N$ and last term is an entire function of λ. The function $R_\lambda(\boldsymbol{x})$ has terms $\int_{\|\boldsymbol{x}\|\le 1} \boldsymbol{x}^{\boldsymbol{\alpha}}\,\|\boldsymbol{x}\|^{\lambda-n}\,\mathrm{d}\boldsymbol{x}$ which vanish if at least one of the components of $\boldsymbol{\alpha}$ is odd (by symmetry), thus only terms with $\boldsymbol{\alpha} = 2\boldsymbol{\beta}$ contribute. As such $R_\lambda(\boldsymbol{x})$, and by extension T_λ, has poles at *non-positive even* integers. This hints at the possible usefulness of $\mathsf{T}_\lambda = \frac{1}{\Gamma\left(\frac{\lambda}{2}\right)} f_\lambda$, where we take $\Gamma\left(\frac{\lambda}{2}\right)$ because it too has poles at non-positive even integers.

As a side note, let $f : \mathbb{R}^n \to \mathbb{R}$ (or $\mathbb{C}$) be a radial function, i.e., $f(\boldsymbol{x}) = g(\|\boldsymbol{x}\|)$, then

$$\frac{1}{r^{n-1}}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^{n-1}g'\right) = \frac{1}{r^{n-1}}\left((n-1)r^{n-2}g' + r^{n-1}g''\right) = \frac{n-1}{r}g' + g''$$

and $\operatorname{grad} f = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}g'$, thus the laplacian is (using (2.22))

$$\Delta f = \frac{n-1}{\|\boldsymbol{x}\|}g' + \left\langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}g''\right\rangle \implies \Delta g := \Delta f = \frac{1}{r^{n-1}}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^{n-1}g'\right). \tag{2.24}$$

Using the above argumentation and the side note with (2.24) for $g(r) = r^{\lambda-n}$, we may observe

$$\Delta\mathsf{T}_\lambda = \frac{(\lambda-n)(\lambda-2)}{\Gamma\left(\frac{\lambda}{2}\right)} f_{\lambda-2} = 2(\lambda-n)\mathsf{T}_{\lambda-2}$$

by $\Gamma\left(\frac{\lambda}{2}\right) = \left(\frac{\lambda}{2}-1\right)\Gamma\left(\frac{\lambda-2}{2}\right)$. Specifically, let $\mathsf{T}_2 = \frac{1}{\|\boldsymbol{x}\|^{n-2}}$ and recall $\Delta\mathsf{T}_2 = -(n-2)\omega_n\delta_{\boldsymbol{0}}$ (from Remark 2.5, where we did the calculation for $n = 3$) which also equals $2(2-n)\mathsf{T}_0$, it follows that $\mathsf{T}_0 = -\frac{\omega_n}{2}\delta_{\boldsymbol{0}}$ and iteration yields $\mathsf{T}_{-2k} = c_k\Delta\delta_{\boldsymbol{0}}$. Still, for non-integer λ with $\mathfrak{Re}\,\lambda < 0$, we can use the original expression (2.23), where the term $R_\lambda(\boldsymbol{x})$ only has poles at non-positive integers, thus it turns a sum of constants $C_{2\boldsymbol{\beta}}(\lambda)$.

To conclude, let us note that in Example 2.4 we computed (2.20), which one can re-write as $\widehat{\mathsf{T}}_\lambda = \mathsf{T}_{n-\lambda}\pi^{\frac{n}{2}-\lambda}$ after some re-arranging. This used to hold for $\mathfrak{Re}\,\lambda \in \left(0, \frac{n}{2}\right)$, and we suggested it also holds for $\mathfrak{Re}\,\lambda \in (0, n)$ due to the form of (2.20). However, now we have extended it to hold for all $\lambda \in \mathbb{C}$ by the means of analytic continuation.

3 Differential Geometry

3.1 Introduction to Multilinear Algebra

Definition 3.1 (Multilinear map)**.** Let $V_1, \ldots, V_k, V$ be vector spaces over a field $\mathbb{K}$. A map $m : V_1 \times \cdots \times V_k \to V$ is called *multilinear* if it is linear in each coordinate, i.e.,

$$m(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \lambda \boldsymbol{v}_i + \boldsymbol{w}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_k) = \lambda m(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) + m(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{w}, \boldsymbol{v}_{i+1}, \ldots, \boldsymbol{v}_k)$$

for all $i \in [k]$ and every $\lambda \in \mathbb{K}$, and $\boldsymbol{w} \in V_i$.

It turns out we can characterize, or more precisely decompose, multilinear functions rather nicely. To this end, we would like to construct a vector space W and a multilinear map $\iota : V_1 \times \cdots \times V_k \to W$ such that **every** multilinear map $m : V_1 \times \cdots \times V_k \to V$ can be decomposed into ι and a linear map φ with $m = \varphi \circ \iota$, where ι does **not** depend on m, but φ does – we call this (ability to always construct a) decomposition a **universal property**.

$$\begin{tikzcd}
V_1 \times \cdots \times V_k \arrow[rr, "m"] \arrow[dr, "\iota"'] & & V \\
& W \arrow[ur, "\varphi"'] &
\end{tikzcd}$$

Figure 3.1: Decomposition of the multilinear map m into the *m-independent* part ι and *dependent* part φ.

Crucially, we can make the following observation.

Proposition 3.1. *If such a pair (W, ι), as described above conforming to the universal property, exists, then it is unique up to an isomorphism.*

Proof. Indeed, as ι does *not* depend on m it must map the space of all *multilinear* functions $V_1 \times \cdots \times V_k \to V$ to space of *linear* functions $W \to V$ (if it was a subspace, we could take that as W). Moreover, ι is also injective, as if there was $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ such that $\boldsymbol{v} \neq \boldsymbol{0}$ but $\iota(\boldsymbol{v}) = 0_W$, then for multilinear m such that $m(\boldsymbol{v}) \neq 0_V$ we have

$$m(\boldsymbol{v}) = (\varphi_m \circ \iota)(\boldsymbol{v}) = \varphi_m(\iota(\boldsymbol{v})) = \varphi_m(0_w) \overset{\text{linear}}{=} 0_V,$$

hence we have a contradiction. Thus ι is a linear bijection between $V_1 \times \cdots \times V_K$ and W, i.e., “various” W’s differ only up-to isomorphism. $\square$

Now we define that the tensor product $\otimes$ for a k-linear function $f(\boldsymbol{x}_{1,\dots,k})$ and a l-linear function $g(\boldsymbol{x}_{k+1,\dots,k+l})$ as

$$(f \otimes g)(\boldsymbol{x}_{1,\dots,k+l}) = f(\boldsymbol{x}_{1,\dots,k}) \cdot g(\boldsymbol{x}_{k+1,\dots,k+l}). \tag{3.1}$$

In particular, we can take f, g be covectors of given vector spaces! Note that the tensor product is not necessarily associative, however it is distributive, i.e., in general

$$f \otimes g \neq g \otimes f \quad \& \quad (f \otimes g) \otimes h = f \otimes (g \otimes h).$$

3.1.1 Abstract Tensor Product of Vector Spaces

Firstly, let us note this section was added on top of the material of the lecture is mainly based on [11, chpt. 12.1.1].

So far, we have basically shown that multilinear functions may be viewed as the span of $\boldsymbol{v}^1 \otimes \ldots \otimes \boldsymbol{v}^k$, where crucially $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k$ are **covectors** (i.e., linear functionals on V_i). However this construction can also be extended to general finite-dimensional vector spaces.

Let us have vector spaces $V_1, \dots, V_k$ over the field $\mathbb{K}$. To define the vector product of these spaces, we first need to understand *formal linear combinations*. Intuitively, formal linear combinations of $V(\mathbb{K})$ are expressions of form $\sum_{i=1}^m a_i \boldsymbol{v}_i$ for $a_i \in \mathbb{K}$ and $\boldsymbol{v}_i \in V$ — however we may also define them for a general set S without the necessary underlying algebraic structure that would allow for multiplication by scalars and/or addition of these elements.

Definition 3.2 (Formal linear combination)**.** For any set S, a *formal linear combination* of elements of S is a function $f : S \to \mathbb{K}$ such that $f(s) = 0$ for all but finitely many $s \in S$. Consequently, the **free vector space** on S, denoted $\mathscr{F}(S)$, is the set of all formal linear combinations of elements of S.

Intuitively, we can think of every $f \in \mathscr{F}(S)$ as "$\sum_{i=1}^n a_i s_i$" (note that we cannot, a priori, add or multiply elements from S) such that f would return the scalar a_i for each s_i in the sum (and zero for all other elements of S not present).

Evidently, under point-wise addition (for $f, g \in \mathscr{F}(S)$ we have $(f+g)(s) = f(s) + g(s)$) and scalar multiplication (for $\lambda \in K$ and $f \in \mathscr{F}(S)$ we have $f(\lambda s) = \lambda f(s)$), $\mathscr{F}(S)$ becomes a vector over $\mathbb{K}$. Furthermore, for each element $x \in S$, there is a function $\delta_x \in \mathscr{F}(S)$ that takes the value 1 on x and zero on all other elements of S, and we *identify* this function with x itself, thus S can be viewed as a subset of $\mathscr{F}(S)$. Every element $f \in \mathscr{F}(S)$ can then be uniquely written as $f = \sum_{i=1}^m a_i x_i$ where $x_1, \dots, x_m$ are the elements of S for which $f(x_i) \neq 0$, and $a_i = f(x_i)$. In other words, S is a basis of $\mathscr{F}(S)$, which is thus finite-dimensional if and only if S is a finite set.

Proposition 3.2. *For any set S and any vector space W, every map $\boldsymbol{A} : S \to W$ has a unique extension to a **linear** map $\overline{\mathbf{A}} : \mathscr{F}(S) \to W$.*

Now we focus on $V_1, \dots, V_k$ vector spaces over $\mathbb{K}$. We form the *free vector space* $\mathscr{F}(V_1 \times \cdots \times V_k)$ on their Cartesian product, which is the set of all formal linear combinations of k-tuples. Moreover, let $\mathscr{R} \subset \mathscr{F}(V_1 \times \cdots \times V_k)$ spanned by all elements of the following forms:

$$\begin{gathered}(\boldsymbol{v}_1, \dots, a\boldsymbol{v}_i, \dots, \boldsymbol{v}_k) - a(\boldsymbol{v}_1, \dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_k) \\ (\boldsymbol{v}_1, \dots, \boldsymbol{v}_i + \boldsymbol{v}_i', \dots, \boldsymbol{v}_k) - (\boldsymbol{v}_1, \dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_k) - (\boldsymbol{v}_1, \dots, \boldsymbol{v}_i', \dots, \boldsymbol{v}_k),\end{gathered} \tag{3.2}$$

with $\boldsymbol{v}_i, \boldsymbol{v}_i' \in V_i$ for all $i \in [k]$ and $a \in \mathbb{K}$. We finally define the **(abstract) tensor product of the spaces** $V_1, \dots, V_k$, denoted $V_1 \otimes \dots \otimes V_k$ to be the following quotient space:

$$V_1 \otimes \dots \otimes V_k := {}^{\mathscr{F}(V_1 \times \cdots \times V_k)}\Big/_{\mathscr{R}}, \tag{3.3}$$

and let $\Pi : \mathscr{F}(V_1 \times \cdots \times V_k) \to V_1 \otimes \dots \otimes V_k$ be the natural projection. Moreover, the equivalence class of an element $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) \in V_1 \otimes \dots \otimes V_k$ is denoted

$$\boldsymbol{v}_1 \otimes \dots \otimes \boldsymbol{v}_k = \Pi(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k), \tag{3.4}$$

and is called the **(abstract) tensor** of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$. It then follows from the definition of the tensor product $V_1 \otimes \dots \otimes V_k$ that $\boldsymbol{v}_1 \otimes \dots \otimes \boldsymbol{v}_k$ is multilinear in the sense of Definition 3.1. Intuitively, elements $(\boldsymbol{v}_1 + \boldsymbol{v}_1', \dots, \boldsymbol{v}_k)$ and $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) + (\boldsymbol{v}_1', \dots, \boldsymbol{v}_k)$ are different symbols/elements $\mathscr{F}(V_1 \times \cdots \times V_k)$ (both are valid as $V_1, \dots, V_k$ are vector spaces, as well as $\mathscr{F}(S)$ for any set S), i.e., their "coefficients functions" might not return the same results. We "fix" this discrepancy by identifying them into one equivalence by $\mathscr{R}$, thus enforcing the multilinearity.

Using the observation Proposition 3.1 and the ideas from Section 3.1.1, we define

$$\begin{aligned} W &:= \operatorname{span}\{\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \dots \otimes \boldsymbol{v}_k \mid \boldsymbol{v}_1 \in V_1, \dots, \boldsymbol{v}_k \in V_k\} \\ &= \operatorname{span}\left\{\boldsymbol{b}_{1j_1} \otimes \boldsymbol{b}_{2j_2} \otimes \dots \otimes \boldsymbol{b}_{kj_k} \mid \boldsymbol{b}_{i1}, \dots, \boldsymbol{b}_{i\dim V_i} \text{ the basis of } V_i\right\}, \end{aligned} \tag{3.5}$$

where $W = V_1 \otimes V_2 \otimes \dots \otimes V_k$ in setting of Proposition 3.1. Note that elements $\boldsymbol{b}_{1j_1} \otimes \boldsymbol{b}_{2j_2} \otimes \dots \otimes \boldsymbol{b}_{kj_k}$ form the basis of W.

Remark 3.1 (Basis-free canonical objects). Let us explore turning a "supposedly" basis-dependent object into a canonical one. Let V be a vector space, then $H(V,V)$ is the vector space *endomorphism* of V,

$$\begin{gathered} H(V,V) := \{\varphi : V \to V \mid \varphi \text{ is linear}\} \\ \Downarrow \\ H(V,V) \cong V \otimes V^*, \end{gathered}$$

i.e., $H(V,V)$ is isomorphic to vector space of elements $v \otimes l : V \to V$ (where $l \in V^*$ is a covector, i.e., linear functional), hence $w \mapsto l(w) \cdot v$. For illustration, let $V = \mathbb{R}^n$ with basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, then to each linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ there uniquely corresponds a matrix $\boldsymbol{A} = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$, for which thus holds (per (3.5))

$$\boldsymbol{A}\boldsymbol{x} = \sum_{i,j=1}^n a_{ij}(\overbrace{\boldsymbol{e}_i}^{\in V} \otimes \overbrace{\boldsymbol{e}^j}^{\in V^*})(\boldsymbol{x}) = \sum_{i,j=1}^n a_{ij}\boldsymbol{e}^j(\boldsymbol{x})\boldsymbol{e}_i = \sum_{i,j=1}^n a_{ij}\left\langle \boldsymbol{e}_j, \boldsymbol{x}\right\rangle \boldsymbol{e}_i.$$

Now consider the trace $\operatorname{tr}: V \otimes V^* \to \mathbb{K}$ with $v \otimes l \mapsto l(v)$. Returning back to the special case $V = \mathbb{R}^n$, we see

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i,j=1}^{n} a_{ij} \boldsymbol{e}^j(\boldsymbol{e}_i) = \sum_{i,j=1}^{n} a_{ij} \underbrace{\langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle}_{\delta_{i,j}} = \sum_{i=1}^{n} a_{ii}.$$

In other words, we have prescribed the trace without the employing any base, which was to be done.

3.1.2 Exterior Product of Vector Spaces

Definition 3.3 (Alternating map)**.** Let V, Y be a vector spaces. A multilinear map $m: V^k \to Y$ is called *alternating* if[1]

$$m(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) = \operatorname{sign}(\pi) m\left(\boldsymbol{v}_{\pi(1)}, \dots, \boldsymbol{v}_{\pi(k)}\right) \quad \forall \pi \in S_k.$$

Remark 3.2. Such maps can only be non-trivial if $k \leq \dim V$. In other case, we can use [11, lemma 14.1.] and calculate several linear combinations to obtain $m = -m$ (by the pigeon hole principle).

Remark 3.3. If $k = \dim V$, m is uniquely defined by its value at $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ (basis). Then, $m(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) = \det(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) \cdot \boldsymbol{y}$ for some $\boldsymbol{y} \in Y$

Definition 3.4 (Exterior (wedge) product)**.** For a vector space V we define

$$\Lambda^k V := \operatorname{span}\{\boldsymbol{v}_1 \wedge \boldsymbol{v}_2 \wedge \dots \wedge \boldsymbol{v}_k \mid \boldsymbol{v}_i \in V\}$$

the *k-th exterior power* of V as the alternating analogue of the tensor product. Note the convention $\Lambda^0 V = \mathbb{K}$ for a vector space V over the field $\mathbb{K}$. To this end, we set

$$\operatorname{Alt}: \boldsymbol{v}_1 \otimes \dots \otimes \boldsymbol{v}_k \mapsto \frac{1}{k!} \sum_{\pi \in S_k} \operatorname{sign}(\pi)(\boldsymbol{v}_{\pi(1)} \otimes \dots \otimes \boldsymbol{v}_{\pi(k)}),$$

then $\Lambda^k V = \operatorname{Alt}(\otimes^k V)$. Alternatively, we take the perspective

$$\Lambda^k V \cong {}^{\otimes^k V}\Big/_{\operatorname{span}\left\{\boldsymbol{v}_1 \otimes \dots \otimes \boldsymbol{v}_k \mid \boldsymbol{v}_i = \boldsymbol{v}_j \text{ for some } i \neq j\right\}}. \tag{3.6}$$

Intuitively, Alt is the alternating variant of ι from the introduction, and in (3.6) we simply "mod it out".

Remark 3.4. Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in V$, then $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are linear independent if and only if $\boldsymbol{v}_1 \wedge \dots \wedge \boldsymbol{v}_k \neq 0$. One can think of this as an analogue for determinant, which also works for $k < \dim V$.

[1]As a reminder, S_k denotes the *symmetric group* of permutations of $[k]$.

Remark 3.5. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in V$ are linearly independent, then

$$\{\boldsymbol{x} \in V \mid \boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_k \wedge \boldsymbol{x} = 0\} = \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\},$$

as per Remark 3.4 it is the set of all linearly *dependent* vectors in V.

Example 3.1. Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^3$ be linearly independent, then we can define the plane spanned by $\boldsymbol{v}_1, \boldsymbol{v}_2$ in the following ways:

$$\begin{aligned} \boldsymbol{v}_1 \wedge \boldsymbol{v}_2 \wedge \boldsymbol{x} &= 0, \\ \det(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{x}) &= 0, \\ \langle \boldsymbol{v}_1 \times \boldsymbol{v}_2, x\rangle &= 0. \end{aligned}$$

Similarly, $\boldsymbol{v} \wedge \boldsymbol{x} = 0$ describes a line, and for basis-representations $\boldsymbol{v} = (a, b, c)$, $\boldsymbol{x} = (x, y, z)$ we have that

$$\begin{aligned} \boldsymbol{v} \wedge \boldsymbol{x} &= (a\boldsymbol{e}_1 + b\boldsymbol{e}_2 + c\boldsymbol{e}_3) \wedge (x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e}_3) \\ &= (ay - bx)(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) + (bz - cy)(\boldsymbol{e}_2 \wedge \boldsymbol{e}_3) + (cx - az)(\boldsymbol{e}_3 \wedge \boldsymbol{e}_1) \\ &\Downarrow \\ \boldsymbol{v} \wedge \boldsymbol{x} = 0 &\implies \begin{cases} ay - bx = 0, \\ bz - cy = 0, \\ cz - az = 0, \end{cases} \end{aligned}$$

i.e., we get a system of linear equations.

Remark 3.6. Not every element of $\Lambda^k V$ is a (wedge-)product!

Proposition 3.3. *Let $\boldsymbol{v} \in \Lambda^k V \neq 0_{\Lambda^k V}$. Then $\boldsymbol{v}$ can be expressed as (a scalar multiple of) $\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_k$ if and only if the space $L = \{\boldsymbol{x} \in V \mid \boldsymbol{v} \wedge \boldsymbol{x} = 0\}$ is k-dimensional.*

For the proof of this proposition, we will need the following observation.

Lemma 3.1. *If for $\boldsymbol{z} \in \Lambda^k V$ holds $\boldsymbol{z} \wedge \boldsymbol{w} = 0$ for any $\boldsymbol{w} \in \Lambda^m V$, where $k + m \leq n$, then $\boldsymbol{z} = \boldsymbol{0}$.*

Proof. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ be its basis, then recall $\boldsymbol{z} = \sum_{j_1 < \ldots j_k} a_{j_1,\ldots,j_k} \boldsymbol{v}_{j_1} \wedge \ldots \wedge \boldsymbol{v}_{j_k} \in \Lambda^k V$. If $k + m \leq n$ and we pick an arbitrary basis vector $\boldsymbol{w}$ of $\Lambda^m V$, i.e., $\boldsymbol{w} = \boldsymbol{v}_{l_1} \wedge \ldots \wedge \boldsymbol{v}_{l_m}$ with $l_1 < \cdots < l_m$, then

$$\boldsymbol{z} \wedge \boldsymbol{w} = \sum_{j_1 < \cdots < j_k} a_{j_1,\ldots,j_k} \boldsymbol{v}_{j_1} \wedge \ldots \wedge \boldsymbol{v}_{j_k} \wedge \boldsymbol{v}_{l_1} \wedge \ldots \wedge \boldsymbol{v}_{l_m}.$$

Every term in this sum is either zero or a basis element of $\Lambda^{k+m} V$, which only occurs once (for *only one* choice of $\boldsymbol{w}$ among the basis vectors of $\Lambda^k V$). However, as it has to be identically zero for *all*, this means that $a_{j_1,\ldots,j_k}$ must be 0 if the given term is a basis vector of $\Lambda^{k+m} V$. Repeating over all choices of $\boldsymbol{w}$ yields $\boldsymbol{z} = \boldsymbol{0}$. $\square$

Proof. Let us now prove Proposition 3.3.

" $\Longrightarrow$ ": From $\mathbf{0} \neq \boldsymbol{v} = \boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_k$ we get linear independence of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ by Remark 3.4, thus their span (by Remark 3.5) is k-dimensional.

" $\Longleftarrow$ ": If L is k-dimensional, it may be written as a span of linearly independent vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, i.e., $L = \operatorname{span}\{\boldsymbol{v}_1, \ldots \boldsymbol{v}_k\}$, such that they can be extended to form the basis V, namely $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n$. Now choose c such that

$$(\boldsymbol{v} - c\boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_k) \wedge \boldsymbol{v}_{k+1} \wedge \ldots \wedge \boldsymbol{v}_n = \mathbf{0},$$

i.e., $\boldsymbol{v} \wedge \boldsymbol{v}_{k+1} \wedge \ldots \wedge \boldsymbol{v}_n = c\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_n$. This is possible since both sides of the equality belong in $\Lambda^n V$, which is 1-dimensional, thus the quantities are proportional. Now take any $\boldsymbol{w} \in \Lambda^{n-k} V$ and consider $\boldsymbol{u} = (\boldsymbol{v} - c\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_k) \wedge \boldsymbol{w}$. Without loss of generality, we can consider $\boldsymbol{w}$ to be a basis element of

$$\Lambda^{n-k} V = \operatorname{span}\left\{\boldsymbol{v}_{j_1} \wedge \ldots \wedge \boldsymbol{v}_{j_{n-k}} \mid \boldsymbol{v}_i \in V \text{ and } j_1, \ldots, j_{n-k} \text{ pairwise distinct}\right\}.$$

Now there are two possible cases for the form of $\boldsymbol{w}$:

1. there exists some $j_l \leq k$, and without loss of generality, $l = 1 \implies \boldsymbol{u} = \mathbf{0}$ by the definition of L;
2. all $j_l > k$, then all j_l are exactly $k+1, \ldots, n$ and, once again without loss of generality, we can also assume in that order $\implies$ by the definition of c, it yields $\boldsymbol{u} = \mathbf{0}$.

In total, we obtained $(\boldsymbol{v} - c\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_n) \wedge \boldsymbol{w} = \mathbf{0}$ for all $\boldsymbol{w} \in \Lambda^{n-k} V$, and by Lemma 3.1, it finally yields $\boldsymbol{v} = c\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_n$. $\square$

Example 3.2. For illustration take $V = \mathbb{R}^4$ and $\boldsymbol{v} = \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 + \boldsymbol{e}_3 \wedge \boldsymbol{e}_4$, which is notably *not a product*. Our goal will be to find $\boldsymbol{x} = \sum_{i=1}^4 x_i \boldsymbol{e}_i$ such that $\boldsymbol{v} \wedge \boldsymbol{x} = \mathbf{0}$. Then

$$\boldsymbol{v} \wedge \boldsymbol{x} = x_3 \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 + x_4 \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_4 + x_1 \boldsymbol{e}_1 \wedge \boldsymbol{e}_3 \wedge \boldsymbol{e}_4 + x_2 \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 \wedge \boldsymbol{e}_4 = 0,$$

and notice that each of the terms corresponds to a different basis vector in $\Lambda^3 \mathbb{R}^4$, i.e., they cannot subtract and thus $\boldsymbol{x} = \mathbf{0}$ (whereas if $\boldsymbol{v}$ *was* a product, we would not have this restriction).

Let $\boldsymbol{A} : V \to W$ be linear. Our goal will now be to extend this map to k-th exterior powers $\Lambda^k V$ and $\Lambda^k W$. To this end, we naturally define

$$\boldsymbol{v}_1 \wedge \ldots \wedge \boldsymbol{v}_k \mapsto (\boldsymbol{A}\boldsymbol{v}_1) \wedge \ldots \wedge (\boldsymbol{A}\boldsymbol{v}_k). \tag{3.7}$$

Let now V be equipped with a scalar product, i.e., a non-degenerate bilinear form. This can also be extended to the exterior power $\Lambda^k V$. Indeed, take $\boldsymbol{v} = \boldsymbol{v}_1 \wedge \ldots \boldsymbol{v}_k \in \Lambda^k V$ and $\boldsymbol{w} = \boldsymbol{w}_1 \wedge \ldots \wedge \boldsymbol{w}_k \in \Lambda^k V$ and set

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle := \det\left(\left\langle \boldsymbol{v}_i, \boldsymbol{w}_j \right\rangle_{i,j}\right). \tag{3.8}$$

Since the Gram matrix is regular for $\boldsymbol{v} \neq \mathbf{0}$, this scalar product is non-degenerate on $\Lambda^k V$, as we wanted.

Lastly, let V be an n-dimensional vector space with a scalar product and an orientation — By [11, prp. 15.3] orientation of a vector space is given by an *ordered* basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Now every non-zero element $\boldsymbol{w} \in \Lambda^n(V^*)$ determines an orientation $O_{\boldsymbol{w}}$ of V as follows:

- if $n \geq 1$, then O_w is the set of ordered bases $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$ such that $O_w(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) > 0$;
- if $n = 0$, then O_w is $+1$ if $w > 0$, and -1 if $w < 0$.

Two nonzero n-covectors determine the same orientation if and only if each is positive multiple of each other! This, in turn, also implies that for 2 bases $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ and $\boldsymbol{w}_1, \dots, \boldsymbol{w}_n$ of V we can write $\boldsymbol{v}_1 \wedge \dots \wedge \boldsymbol{v}_n = c\boldsymbol{w}_1 \wedge \dots \wedge \boldsymbol{w}_n$ by Proposition 3.3 with $c > 0$ if the bases are oriented the same.

Definition 3.5 (Hodge star functional)**.** Every $\boldsymbol{\lambda} \in \Lambda^k V$ defines a linear map $L_{\boldsymbol{\lambda}} : \Lambda^{n-k} V \to \mathbb{R}$, where

$$\boldsymbol{\mu} \mapsto \boldsymbol{\lambda} \wedge \boldsymbol{\mu} =: L_{\boldsymbol{\lambda}}(\boldsymbol{u}) \underbrace{\boldsymbol{e}_1 \wedge \dots \wedge \boldsymbol{e}_n}_{\|\cdot\|=1}, \tag{3.9}$$

where $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ is an oriented orthonormal basis of V.

Since V has orthonormal basis, it also has a scalar product, and thus by (3.8) also Λ^{n-k} is equipped with a non-degenerate scalar product[2]. Hence, there exists $*\boldsymbol{\lambda} \in \Lambda^{n-k} V$ such that $L_{\boldsymbol{\lambda}}(\boldsymbol{\mu}) = \langle *\boldsymbol{\lambda}, \boldsymbol{\mu} \rangle$ (by Theorem 4.5). This further induces an **isomorphic** linear map $* : \Lambda^k V \to \Lambda^{n-k} V$ called the **Hodge star**, i.e.,

$$\boldsymbol{\lambda} \wedge \boldsymbol{\mu} = \langle *\boldsymbol{\lambda}, \mu \rangle \, \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_n. \tag{3.10}$$

! Important 1: Notation

Let us now introduce the following notation: For $1 \leq j_1 < \cdots < j_m \leq n$, we write $H = \{j_1, \dots, j_m\}$ and $\boldsymbol{e}_H = \boldsymbol{e}_{j_1} \wedge \dots \wedge \boldsymbol{e}_{j_m} \in \Lambda^m V$, where $\boldsymbol{e}_i$ denotes a basis vector of V.

Let $\boldsymbol{e}_i$ be an orthonormal basis (and we still require orientation). For K an index set, $\#K = n - k$, disjoint to H and $N = [n]$ we have $\boldsymbol{e}_H \wedge \boldsymbol{e}_K = \langle *\boldsymbol{e}_H, \boldsymbol{e}_K \rangle \, \boldsymbol{e}_N$ by Definition 3.5. Of course, if H and K are only disjoint if $K = N \setminus H = H^c$, otherwise we would get $\boldsymbol{e}_H \wedge \boldsymbol{e}_K = 0$. In particular, as $\boldsymbol{e}_i$ is an orthonormal basis, it yields (for $K = H^c$)

$$\boldsymbol{e}_H \wedge \boldsymbol{e}_K = \operatorname{sign}(H, K)\boldsymbol{e}_N \implies *\boldsymbol{e}_H = \operatorname{sign}(H, K)\boldsymbol{e}_K,$$

where $\operatorname{sign}(H, K)$ is the orientation of the basis with indices following the concatenation of H and K.

Example 3.3 (Hodge star and cross product)**.** Let us now restrict ourselves to $\mathbb{R}^3$ for simplicity. Then $*\boldsymbol{e}_1 = \pm\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$, which further implies

$$\boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 = \langle \pm\boldsymbol{e}_2 \wedge \boldsymbol{e}_3, \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 \rangle \, \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3,$$

hence $*\boldsymbol{e}_1 = +\boldsymbol{e}_2 \wedge \boldsymbol{e}_3$. Similarly, we can compute that $*\boldsymbol{e}_2 = -\boldsymbol{e}_1 \wedge \boldsymbol{e}_3$ and $*\boldsymbol{e}_3 = +\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$. Moreover, analogous calculation immediately yields the following

$$*(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2) = +\boldsymbol{e}_3, \quad *(\boldsymbol{e}_1 \wedge \boldsymbol{e}_3) = -\boldsymbol{e}_2, \quad *(\boldsymbol{e}_2 \wedge \boldsymbol{e}_3) = +\boldsymbol{e}_1.$$

The similarity with cross-product is *not coincidental.* In fact, $\boldsymbol{x} \times \boldsymbol{y} = *(\boldsymbol{x} \wedge \boldsymbol{y})$ holds in $\mathbb{R}^3$!

[2] This is also the reason why we may write $\|\boldsymbol{e}_1 \wedge \dots \wedge \boldsymbol{e}_n\| = 1$.

In general, we have that $**\boldsymbol{e}_H = (-1)^{k(n-k)}\boldsymbol{e}_h$, and thus $*** * * = *^4 = \mathrm{id}$. Also, later[3] we shall need the exterior power of the dual space V^*. For $l_1, \dots, l_k \in V^*$ is the exterior product $l_1 \wedge \dots \wedge l_k$ a multilinear map $V^k \to \mathbb{R}$ set by

$$l_1 \wedge \dots \wedge l_k(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) := \det\left((l_i(\boldsymbol{v}_j))_{i,j}\right).$$

This map is alternating, see Definition 3.3, in the input $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$, which justifies interpreting it as a linear functional on $\Lambda^k V$. In other words, any $l_1 \wedge \dots \wedge l_k \in \Lambda^k(V^*)$ is an element from the dual $(\Lambda^k V)^*$, i.e., $\Lambda^k V^* \cong \Lambda^k(V^*)$.

! Important 2: Convention: Einstein summation

If one index occurs twice in an expression, once as an upper index and once as a lower index, this index is summed over the according index space. Lower indices are called **covariant**, whereas upper indices **contravariant**.

Example 3.4. Concretely, we may write $\boldsymbol{x} = x^i \boldsymbol{e}_i = \sum_i x^i \boldsymbol{e}_i$. A bit more involved example may be for $\#H = k$:

$$x^H \boldsymbol{e}_H = \sum_{\#H=k} x^H \boldsymbol{e}_H.$$

Similarly, if $\boldsymbol{a} = y_i \boldsymbol{e}^i = \sum_{i=1}^n y_i \boldsymbol{e}^i$, where $\boldsymbol{e}^i$ is the dual basis of V^*, then

$$\boldsymbol{a}(\boldsymbol{x}) = x^i y_j \boldsymbol{e}^j(\boldsymbol{e}_i) = x^i y_j \delta_i^j = x^j y_j.$$

Remark 3.7. Let V be a vector space with bases $(\boldsymbol{e}_i)_i$ and consider a vector $\boldsymbol{x} \in V$, i.e., $\boldsymbol{x} = x^i \boldsymbol{e}_i$. Now we shall define a new basis $(\boldsymbol{e}'_j)_j$ by a linear change of coordinates $\boldsymbol{A} = (a^i_j)_{i,j}$, thus $\boldsymbol{e}'_j = a^i_j \boldsymbol{e}_i$. Notice that by invertibility of $\boldsymbol{A}$ we get that $\boldsymbol{A}^{-1} = (a^{-1})^j_i$, i.e.,

$$a^i_j \, (a^{-1})^j_k = \delta^i_k.$$

We shall now write the vector $\boldsymbol{x}$ in the basis $(\boldsymbol{e}'_j)_j$ by its representation in $(\boldsymbol{e}_i)_i$. To this end, calculate

$$x^i \boldsymbol{e}_i = \boldsymbol{x} = x'^j \boldsymbol{e}'_j = x'^j a^i_j \boldsymbol{e}_i \implies x'^j = (a^{-1})^j_i \, x^i,$$

hence the coordinates x^i transform **with inverse** of change-of-basis (which we call *contravariant* behavior). On the other hand, it can be analogously shown that coordinates of covectors **transform the same** as change-of-basis in the original vector space, i.e., the "co-vary" as the original basis making them *covariant*.

[3] In fact, we have already used it when taking about orientation!

3.2 Differentiable Manifolds

Let $\mathcal{M}$ be a topological space, which we shall equip with a certain structure. Consider $(U_i)_{i\in I}$ open sets in $\mathcal{M}$ with the property $\bigcup_{i\in I} U_i = \mathcal{M}$. Furthermore, we consider **charts** (φ_i, U_i), where $\varphi_i : U_i \to V_i \subseteq \mathbb{R}^n$ are homeomorphism[4], such that they are **compatible**. Compatibility states that for $U_i \cap U_j \neq \emptyset$ the mapping

$$\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j) \in \mathcal{C}^\infty, \tag{3.11}$$

i.e., it is infinitely differentiable (for simplicity). Note that in theory, it is sufficient to require that $\varphi_i \circ \varphi_j^{-1}$ is only a diffeomorphism[5], however then we have to always "count" the number of allowed derivatives such that all the operations are well-defined. For illustration, see Figure 3.2.

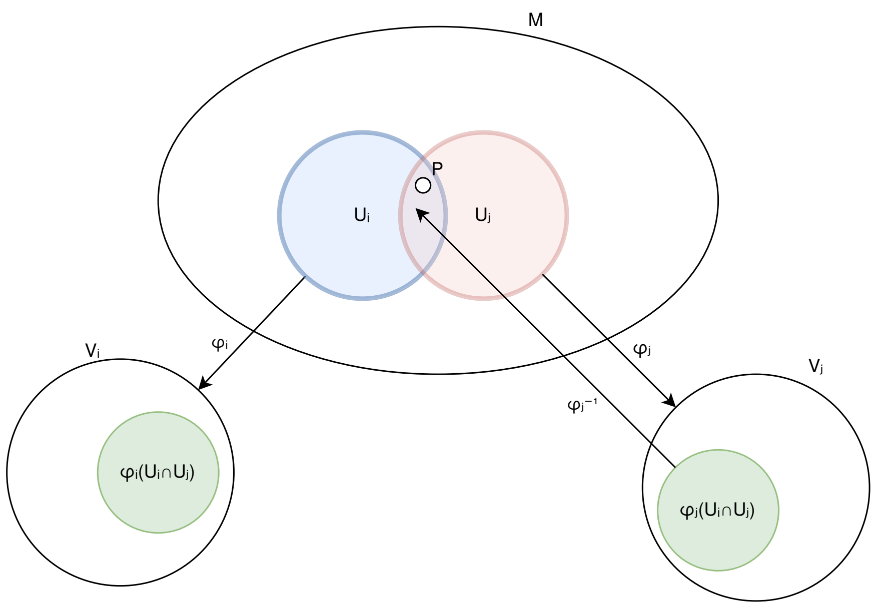

Figure 3.2: Illustration of a differentiable manifold with an example of (a part of) an atlas

Let us call **atlas** the collection of all charts $(\varphi_i, U_i)_{i\in I}$. Having an atlas allows us to define differentiability of functions $f : \mathcal{M} \to \mathbb{R}$ by identifying it with the differentiability of $f \circ \varphi_i^{-1} : V_i \to \mathbb{R}$. Importantly, due to the compatibility of charts (3.11), the differentiability of f is well-defined on the intersections. Hence, an atlas defines a differentiable structure on $\mathcal{M}$. We remark that the *maximal atlas* contains all possible chart maps and domains, and that it surely exists by Zorn's lemma 4.2.

[4]A function $f : X \to Y$ between two topological spaces is a **homeomorphism** if f is a *continuous bijection* and the inverse is also *continuous*, i.e., f is continuously invertible.

[5]A function $f : X \to Y$ between two topological spaces is a **diffeomorphism** if f and its inverse f^{-1} are both differentiable.

3.2.1 Tangent and Cotangent Spaces

So far, we have seen that an atlas provides a differentiable structure to $\mathcal{M}$. However, it is not a priori trivially clear how to define tangents, as $\mathcal{M}$ might not be a subset of $\mathbb{R}^n$. Intuitively speaking, in this section we shall define the notion of said tangents by identifying directions with directional derivatives of functions on $\mathcal{M}$.

Let us fix a point $P \in \mathcal{M}$, for which have (multiple) $(u_0^1, \dots, u_0^n) \mapsto \varphi_i^{-1}(u_0^1, \dots, u_0^n) = P$ with $\varphi_i^{-1} : V_i \to U_i$. Now consider $f : \mathcal{M} \to \mathbb{R}$ and take $f \circ \varphi_I^{-1} : V_i \to \mathbb{R}$. This allows us to abuse the notation to write $f(u_0^1, \dots, u_0^n)$ and we define the **tangent space** $\mathrm{T}_P\mathcal{M}$ at the point P as the vector space of all directional derivatives in $(u_0^1, \dots, u_0^n)$. In other words, we transform the problem of tangents of functions on the manifold $\mathcal{M}$ to the tangents of functions on the chart domains $V_i \subseteq \mathbb{R}^n$ (where we know how to do it).

Remark 3.8. An obvious basis for such space are $\frac{\partial}{\partial u^i} = \partial_i$ **at the point** P — the span of which produces a vector space of operators. In particular, this gives us tangent vectors that can be applied functions. Note that while this basis depends on the chart we are using (by the u^i), the change of a chart corresponds only to a change of basis (by compatibility (3.11) are the charts isomorphic, even diffeomorphic, on the intersections). As such, fixing P and then choosing a corresponding chart is indeed a viable well-defined strategy.

Indeed, let $P \in \mathcal{M}$ be such that $P \in U_i \cap U_j$ with corresponding charts (φ_i, U_i) and (φ_j, U_j), recall Figure 3.2. Let $(u^1, \dots, u^n)$ be a basis of V_i and $(v^1, \dots, v^n)$ the basis of V_j, then

$$\mathrm{T}_P\mathcal{M} = \operatorname{span}\left\{\frac{\partial}{\partial u^1}, \dots, \frac{\partial}{\partial u^n}\right\} = \operatorname{span}\left\{\frac{\partial}{\partial v^1}, \dots, \frac{\partial}{\partial v^n}\right\} \quad \& \quad \frac{\partial}{\partial v^l} = \frac{\partial u^k}{\partial v^l}\frac{\partial}{\partial u^k} \tag{3.12}$$

at the point P (where we use the Einstein notation, see Important 2). This is obviously only a change of basis.

By collecting tangent spaces at all points of the manifold $\mathcal{M}$ we construct the **tangent bundle** as

$$\mathrm{T}\mathcal{M} = \biguplus_{P \in \mathcal{M}} \mathrm{T}_P\mathcal{M},$$

where $\uplus$ is the disjoint union, as the individual tangents spaces are not simply subsets of $\mathbb{R}^n$, but are bound to their respective points P and they also carry this information. Finally, there exists a map $\Pi : \mathrm{T}\mathcal{M} \to \mathcal{M}$ such that $(P, \boldsymbol{v}) \mapsto P$ with $\boldsymbol{v} \in \mathrm{T}_P\mathcal{M}$.

Definition 3.6 (Vector field)**.** A vector field $\boldsymbol{X}$ is a section from the tangent bundle $\mathrm{T}\mathcal{M}$, i.e., for each $P \in \mathcal{M}$ we take a vector from $\mathrm{T}_P\mathcal{M}$. Thus $\boldsymbol{X} = x^k \frac{\partial}{\partial u^k}$ in the corresponding chart (φ_i, U_i).

Clearly, the vector field $\boldsymbol{X}$ is differentiable if coordinate functions $x^k : V_i \to \mathbb{R}$ are differentiable in every chart (there we know what "differentiability" even means). Of course, the coordinate function changes by considering different charts (even for a given point, though keep in mind Remark 3.8).

For every function $f : \mathcal{M} \to \mathbb{R}$ we can compute $\boldsymbol{X} f$, i.e., every vector field **acts on functions**. In particular, we define the **differential** of f as

$$\mathrm{d} f(\boldsymbol{X}) = \boldsymbol{X} f, \tag{3.13}$$

which offers two different perspectives on the same object. Namely, the differential $\mathrm{d} f$ can be viewed as a linear functional on $\mathrm{T}_P\mathcal{M}$ for every $P \in \mathcal{M}$ (recall that a vector space is only a section of the tangent bundle $\mathrm{T}\mathcal{M}$). In other words, this locally defines $\mathrm{d} f$ as an element of the dual space $\mathrm{T}^*_P\mathcal{M}$ by

$$\mathrm{d} f = \frac{\partial f}{\partial u^k} \mathrm{d} u^k,$$

where $\mathrm{d} u^k$ form a covector base of $\mathrm{T}^*_P\mathcal{M}$ (dual to the basis $\frac{\partial}{\partial u^1}, \dots, \frac{\partial}{\partial u^n}$ of $\mathrm{T}_P\mathcal{M}$). Moreover, for a basis covector $\mathrm{d} u^k$ holds

$$\mathrm{d} u^k \left(\frac{\partial}{\partial u^l} \right) = \frac{\partial u^k}{\partial u^l} = \delta^k_l.$$

As $\mathrm{d} u^k$ are covariant, they transform with the change-of-basis matrix (unlike the $\frac{\partial}{\partial u^l}$ basis of contravariant vectors, see (3.12)), i.e., $\mathrm{d} u^k = \frac{\partial u^k}{\partial v^l} \mathrm{d} v^l$. Finally, we denote by $\mathrm{T}^*\mathcal{M} = \biguplus_{P \in \mathcal{M}} \mathrm{T}^*_P\mathcal{M}$ the **cotangent bundle** (as an analogue to the *tangent bundle*).

Tip

Recall that in Analysis 2, we had

$$f(u^1, \dots, u^n) = f(u^1_0, \dots, u^n_0) + \operatorname{grad} f(u^1_0, \dots, u^n_0) \begin{pmatrix} u^1 - u^1_0 \\ \vdots \\ u^n - u^n_0 \end{pmatrix} + \text{ error }.$$

Here, we typically transport $\mathrm{d} f(u^1_0, \dots, u^n_0)$ via the scalar product to the tangent space (as the gradient $\operatorname{grad} f(u^1_0, \dots, u^n_0)$).

3.2.2 Differential Forms

First and foremost, one should note that expressions of form $\omega = a_k \mathrm{d} u^k$ (as such, also the basis covectors of $\mathrm{T}^*_P\mathcal{M}$ themselves) are **differential forms** of order 1 with $\omega(P) \in \mathrm{T}^*_P\mathcal{M}$. Moreover, every function f induces a differential form $\omega = \mathrm{d} f$, and such forms are called **exact forms** — more precisely, a form ω is called **exact** if it may be written as $\omega = \mathrm{d}\alpha$, where α is another differential form. However, the converse does not hold even with C^∞! It is easy to see that when an exact form $\omega = \omega_i \mathrm{d} u^i$ corresponds to at least twice differentiable potential function f[6], we obtain a nice

[6]In case of the differential forms, the coordinate functions ω_i are by assumption smooth everywhere on $\mathcal{M}$, hence this is always satisfied.

necessary condition stemming from the symmetry of second derivatives of f,

$$\frac{\partial \omega_i}{\partial u^j} = \frac{\partial}{\partial u^j}\frac{\partial f}{\partial u^i} = \frac{\partial^2 f}{\partial u^i \partial u^j} = \frac{\partial}{\partial u^i}\frac{\partial f}{\partial u^j} = \frac{\partial \omega_j}{\partial u^i}$$

$$\Updownarrow$$

$$\frac{\partial \omega_i}{\partial u^j} - \frac{\partial \omega_j}{\partial u^i} = 0. \tag{3.14}$$

A form ω which follows this necessary condition (3.14) is called **closed**.

Remark 3.9 (Lie bracket). Let $\boldsymbol{X}, \boldsymbol{Y}$ be vector fields and f be a function $\mathcal{M} \to \mathbb{R}$, then $\boldsymbol{X}f$, $\boldsymbol{Y}f$ are also functions $\mathcal{M} \to \mathbb{R}$. Thus $\boldsymbol{XY}f$ and $\boldsymbol{YX}f$ are again functions, but are $\boldsymbol{XY}$ and/or $\boldsymbol{YX}$ vector fields?

Take $\boldsymbol{X} = x^k \frac{\partial}{\partial u^k}$ and $\boldsymbol{Y} = y^k \frac{\partial}{\partial u^k}$, then

$$\boldsymbol{XY}f = x^k \frac{\partial}{\partial u^k}\left(y^l \frac{\partial}{\partial u^l} f\right) = x^k \Bigg(\underbrace{\frac{\partial y^l}{\partial u^k}\frac{\partial f}{\partial u^l}}_{\in \mathrm{T}_P\mathcal{M}} + \underbrace{y^l \frac{\partial^2 f}{\partial u^l \partial u^k}}_{\notin \mathrm{T}_P\mathcal{M}} \Bigg),$$

hence $\boldsymbol{XY}$ is not a vector field (due to the inclusion of second order terms). However, one might consider the so called **Lie bracket** $[\boldsymbol{X}, \boldsymbol{Y}] = (\boldsymbol{XY} - \boldsymbol{YX})$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$, for which holds

$$[\boldsymbol{X}, \boldsymbol{Y}] = \underbrace{\left(x^k \frac{\partial y^l}{\partial u^k}\frac{\partial}{\partial u^l} - y^l \frac{\partial x^k}{\partial u^l}\frac{\partial}{\partial u^k} \right)}_{\in \mathrm{T}_P\mathcal{M}} f.$$

Specifically, it is easy to see that the Lie bracket is symmetric.

Definition 3.7 (Differential form)**.** A differential form of degree k is a section in $\Lambda^k \mathrm{T}^*\mathcal{M} = \biguplus_{P \in \mathcal{M}} \Lambda^k \mathrm{T}^*_P\mathcal{M}$.

A differential form may be expressed as $\omega = a_H \, \mathrm{d}u^H$ in every chart, where we use the notation introduced in Important 1. For illustration. for a degree k differential form the index set $H = \{h_1, \ldots, h_k\}$ goes through all k-element subsets of $[n]$, i.e., $\#H = k$, and $\mathrm{d}u^H = \mathrm{d}u^{h_1} \wedge \ldots \wedge \mathrm{d}u^{h_k}$. Concretely, $\omega(P) \in \Lambda^k \mathrm{T}^*_P\mathcal{M}$ is a multilinear form on the tangent space, such that we can evaluate it as $\omega(P)(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ for $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathrm{T}_P\mathcal{M}$.

Proposition 3.4 (Exterior derivative)**.** *There is exactly one* ***linear*** *map* $\mathrm{d} : \Lambda^k \mathrm{T}^*\mathcal{M} \to \Lambda^{k+1}\mathrm{T}^*\mathcal{M}$ *with the following properties:*

1. $\mathrm{d}(\lambda \wedge \mu) = \mathrm{d}\lambda \wedge \mu + (-1)^{\deg \lambda} \lambda \wedge \mathrm{d}\mu$ *(graded product/Leibniz rule);*
2. $\mathrm{d}(\mathrm{d}\omega) = 0$ *(Poincaré lemma);*

3. *for* $f \in \Lambda^0 \mathrm{T}^* \mathcal{M}$ *(thus* f *is a function)* $\mathrm{d}(f) = \mathrm{d}f$ *(coordinate-free by (3.13)*[7]*).*

Let us note that the second property considered with functions f, g (i.e., 0-forms) reads as (recalling tensor product for functions)

$$\mathrm{d}(fg) = \mathrm{d}(f \wedge g) = g\mathrm{d}f + f\mathrm{d}g, \tag{3.15}$$

hence we recover the standard product rule for differentiation.

Proof. We shall split the proof into 6 steps [12, thm. 30.4., 11, prp. 14.23. & thm 14.24.].

Step 1.: Let us verify uniqueness of d. Firstly, we show that properties 1. and 2. imply that for any forms $\omega_1, \dots, \omega_k$, we have $\mathrm{d}(\omega_1 \wedge \cdots \wedge \omega_k) = 0$. Let us proceed by induction over k. The base case of $k = 1$ is a consequence of 3. Now suppose it holds for $k - 1$ and set $\eta = (\mathrm{d}\omega_2 \wedge \cdots \wedge \mathrm{d}\omega_k)$, which by 1. can be re-written as

$$\mathrm{d}(\mathrm{d}\omega_1 \wedge \eta) = \mathrm{d}(\mathrm{d}\omega_1) \wedge \eta \pm \mathrm{d}\omega_1 \wedge \mathrm{d}\eta.$$

The first term vanishes by 2., and the $\mathrm{d}\eta = 0$ by the induction hypothesis. Furthermore, we show that for any k-form ω, the form $\mathrm{d}\omega$ is determined entirely by the value of d on 0-forms, which are, in turn, specified by 3. Since d is linear, it suffices to consider the case $\omega = f\mathrm{d}u^H$ (here f is a 0-form). Compute

$$\begin{aligned}\mathrm{d}\omega = \mathrm{d}(f\mathrm{d}u^H) &= \mathrm{d}(f \wedge \mathrm{d}u^H) \\ &\overset{1.}{=} \mathrm{d}(f) \wedge \mathrm{d}u^H + f \wedge \mathrm{d}(\mathrm{d}u^H) \\ &\overset{3.}{=} \mathrm{d}f \wedge \mathrm{d}u^H\end{aligned}$$

and the the result we have just proved (here the necessary form of 2.). This gives that $\mathrm{d}\omega$ is determined solely by the value of d on 0-form f. This is a coordinate-free object, which we shall use to lift the restriction to a single chart (hence we can only study the object on a single chart domain).

Step 2.: We now define d. It's value is specified by 3. Not only it is in accordance with the uniqueness results above, but it also gives us a recipe to define it for forms of positive degree: If V is open set in $\mathbb{R}^n$ and if ω is a k-form on V (i.e., on a given chart domain), we write ω uniquely in the form $\omega = f_H \mathrm{d}u^H$, and define

$$\mathrm{d}\omega = \mathrm{d}\left(f_H \mathrm{d}u^H\right) := \mathrm{d}f_H \wedge \mathrm{d}u^H. \tag{3.16}$$

Let us note that here $\mathrm{d}u^H$ come from the basis representation, and thus all H's are ascending sets of indices. We check that $\mathrm{d}\omega$ is indeed infinitely differentiable. For this purpose, we first compute

$$\mathrm{d}\omega = \left(\frac{\partial f_H}{\partial u^i}\mathrm{d}u^i\right) \wedge \mathrm{d}u^H,$$

where terms with $i \in H$ will vanish. The $(k+1)$-forms of remaining terms can be rearranged so that the indices are in ascending order (producing only ± 1). In total, we see that it is a linear

[7] While we do need coordinates to write down $\mathrm{d}f$ from (3.13), it is defined coordinate-free by a $\mathrm{T}\mathcal{M}$-section, see Definition 3.6.

combination of $(k+1)$-forms with coefficients $\frac{\partial f_H}{\partial u^i}$. Then smoothness of f implies $\mathrm{d}\omega$ is itself infinitely differentiable.

Now let us show d is linear on k-forms with $k > 0$. Let

$$\omega = f_H \mathrm{d}u^H \quad \& \quad \eta = g_H \mathrm{d}u^H$$

be k-forms, then by (3.16) and the fact that such d is linear on 0-forms

$$\begin{aligned}
\mathrm{d}(a\omega + b\eta) &= \mathrm{d}\left((af_H + bg_H)\,\mathrm{d}u^H\right) \\
&= \mathrm{d}\,(af_H + bg_H) \wedge \mathrm{d}u^H \\
&\overset{\text{linear}}{=} (a\,\mathrm{d}f_H + b\,\mathrm{d}g_H) \wedge \mathrm{d}u^H \\
&= a\,\mathrm{d}\omega + b\,\mathrm{d}\eta.
\end{aligned}$$

Step 3.: We now show that if J is an *arbitrary* k-tuple of indices from $[n]$, then $\mathrm{d}(f \wedge \mathrm{d}u^J) = \mathrm{d}f \wedge \mathrm{d}u^J$. Certainly, this holds if two of the indices in J are the same, since $\mathrm{d}u^J = 0$ in such case. Thus suppose the indices in J are distinct and let H be the index set obtained by rearranging indices of J in ascending order (by permutation $\pi \in S_k$). As the exterior product $\wedge$ is alternating, see Definition 3.3 and Definition 3.4, it implies $\mathrm{d}u^H = \mathrm{sign}(\pi)\mathrm{d}u^J$. The definition (3.16) of d in *Step 2.* together with its linearity now yield

$$\mathrm{sign}(\pi)\mathrm{d}(f \wedge \mathrm{d}u^J) = \mathrm{sign}(\pi)\mathrm{d}f \wedge \mathrm{d}u^J.$$

Step 4.: Let us focus on the *graded product rule* property 1. In the case both μ, λ are 0-forms, the graded product rule simplifies to the standard product rule, which we have shown as a small remark in (3.15). Consider μ, λ be k-forms of positive order (i.e., both are non-zero-forms) and by linearity of d, it suffices to consider the case $\lambda = f\mathrm{d}u^I$ and $\mu = g\mathrm{d}u^J$. Then

$$\begin{aligned}
\mathrm{d}(\lambda \wedge \mu) &= \mathrm{d}\left(fg\mathrm{d}u^I \wedge \mathrm{d}u^J\right) \\
&\overset{\text{Step 3.}}{=} \mathrm{d}(fg) \wedge \mathrm{d}u^I \wedge \mathrm{d}u^J \\
&= (g\,\mathrm{d}f + f\,\mathrm{d}g) \wedge \mathrm{d}u^I \wedge \mathrm{d}u^J \\
&= (\mathrm{d}f \wedge \mathrm{d}x^I) \wedge (g \wedge \mathrm{d}x^J) + (-1)^k (f \wedge \mathrm{d}u^I) \wedge (\mathrm{d}g \wedge \mathrm{d}u^J) \\
&= \mathrm{d}\lambda \wedge \mu + (-1)^k \lambda \wedge \mathrm{d}\mu,
\end{aligned}$$

where the $(-1)^k$ factor comes from the fact that we need to exchange the k-form $\mathrm{d}u^I$ and the 1-form $\mathrm{d}g$. Finally, the case where one of the forms is actually a function (i.e., a 0-form) proceeds analogously to the case we have just shown.

Step 5.: Let us finally focus on the Poincaré's lemma property 2. Assume, for now, that f is a differential 0-form, then

$$\begin{aligned}
\mathrm{d}(\mathrm{d}f) &\triangleq \mathrm{d}\left(\frac{\partial f}{\partial u^i}\mathrm{d}u^i\right) \\
&\overset{\text{Step 2.}}{=} \mathrm{d}\left(\frac{\partial f}{\partial u^i}\right)\mathrm{d}u^i \\
&= \left(\frac{\partial}{\partial u^j}\frac{\partial}{\partial u^i} f\right)\mathrm{d}u^j \wedge \mathrm{d}u^i,
\end{aligned}$$

and since by Schwarz' theorem the second derivatives of a smooth function f are *symmetric*, but the exterior product is *alternating* (or *anti-symmetric*), we get the desired result. Note that we can also rewrite this as

$$\mathrm{d}(\mathrm{d}f) = \sum_{i<j} \left(\frac{\partial^2}{\partial u^i \partial u^j} - \frac{\partial^2}{\partial u^j \partial u^i} \right) f \mathrm{d}u^i \wedge \mathrm{d}u^j.$$

Step 6.: Last, but not least, we show that if ω is a differential k-form with $k > 0$, we still get $\mathrm{d}(\mathrm{d}\omega) = 0$. Since d is linear, it suffices to focus on $\omega = f\mathrm{d}u^H$. Then (3.16) implies

$$\mathrm{d}(\mathrm{d}\omega) = \mathrm{d}\left(\mathrm{d}f \wedge \mathrm{d}u^H\right) \overset{1.}{=} \mathrm{d}(\mathrm{d}f) \wedge \mathrm{d}u^H - \mathrm{d}f \wedge \mathrm{d}(\mathrm{d}u^H),$$

where $\mathrm{d}(\mathrm{d}f) = 0$ by *Step 5.*, and

$$\mathrm{d}(\mathrm{d}u^H) = \mathrm{d}(1\mathrm{d}u^H) = \mathrm{d}(1) \wedge \mathrm{d}u^H = 0$$

by definition. Hence $\mathrm{d}(\mathrm{d}\omega) = 0$ and this concludes the proof. □

Example 3.5. Let us turn our attention to the problem of integrating a vector field (e.g., a line integral or a surface integral). Let (P, Q, R) be the coordinates of a given vector field in $\mathbb{R}^3$, and thus $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ be the *coordinates of motion*, we then have

$$\int P\mathrm{d}x + Q\mathrm{d}y + R\mathrm{d}z,$$

where the integrand takes form a scalar product, and is obviously a differential form ω. then

$$\begin{aligned}
\mathrm{d}\omega &= \left(\frac{\partial P}{\partial x}\mathrm{d}x + \frac{\partial P}{\partial y}\mathrm{d}y + \frac{\partial P}{\partial z}\mathrm{d}z \right) \wedge \mathrm{d}x + \left(\frac{\partial Q}{\partial x}\mathrm{d}x + \frac{\partial Q}{\partial y}\mathrm{d}y + \frac{\partial Q}{\partial z}\mathrm{d}z \right) \wedge \mathrm{d}y + \left(\frac{\partial R}{\partial x}\mathrm{d}x + \frac{\partial R}{\partial y}\mathrm{d}y + \frac{\partial R}{\partial z}\mathrm{d}z \right) \wedge \mathrm{d}z \\
&= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \mathrm{d}y \wedge \mathrm{d}z + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \mathrm{d}z \wedge \mathrm{d}x + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathrm{d}x \wedge \mathrm{d}y,
\end{aligned}$$

where we get the coordinates of the **curl** of the given vector field.

Similarly for a 2-form,

$$\begin{aligned}
\mu &= P\mathrm{d}y \wedge \mathrm{d}z + Q\mathrm{d}z \wedge \mathrm{d}x + R\mathrm{d}x \wedge \mathrm{d}y \\
\mathrm{d}\mu &= \frac{\partial P}{\partial x}\mathrm{d}x \wedge \mathrm{d}y \wedge dz + \frac{\partial Q}{\partial y}\mathrm{d}y \wedge \mathrm{d}z \wedge \mathrm{d}x + \frac{\partial R}{\partial z}\mathrm{d}z \wedge \mathrm{d}x \wedge \mathrm{d}y \\
&= \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z,
\end{aligned}$$

we get the **divergence** of the vector field instead!

We shall now study the relationship between the exact and closed forms, in particular we shall show that locally every closed form is exact!

Let $\mathscr{F}^k(X)$ denote the space of *smooth* differential forms of degree k on X. Moreover, we shall define a "pullback", firstly for smooth functions, and then also for norms.

Definition 3.8 (Pullback of functions)**.** Given a smooth map $f : N \to M$ between differentiable manifolds M, N and a smooth function $\varphi : M \to \mathbb{R}$, the *pullback* $f\varphi$ is the function defined by $(f\varphi)(\boldsymbol{x}) = \varphi(f(\boldsymbol{x}))$ for $x \in N$.

Definition 3.9 (Pullback of differential forms)**.** Let $f : N \to M$ be a smooth map between differentiable manifolds M, N and let $\omega \in \mathscr{F}^k(M)$ be a smooth k-form on M. The pullback $f\omega \in \mathscr{F}^k(N)$ is the unique differential k-form on N such that (for a given point $P \in N$)

$$(f\omega)_P(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k) = \omega_{f(P)}(\mathrm{d}f_P(\boldsymbol{v}_1), \dots, \mathrm{d}f_P(\boldsymbol{v}_k)),$$

where $\boldsymbol{v}_i \in \mathrm{T}_P M$. Here, $\mathrm{d}f_P : \mathrm{T}_P N \to \mathrm{T}_{f(P)} M$ is the *differential pushforward* of f at P.

From linearity of the k-form follows also linearity of the pullback. Lastly, let $\mathsf{K} : \mathscr{F}^{p+1}(I \times U) \to \mathscr{F}^p(U)$, where $I = [0, 1]$, be the **homotopy** operator, which maps

$$\begin{aligned} A\,\mathrm{d}u^H &\mapsto 0 \\ B(t, \boldsymbol{x})\,\mathrm{d}t \wedge \mathrm{d}u^J &\mapsto \int_0^1 B(t, \boldsymbol{u})\,\mathrm{d}t\,\mathrm{d}u^J, \end{aligned} \tag{3.17}$$

where $B \in C^\infty$.

Proposition 3.5. *If one recalls $j_0 : U \to I \times U$ with $\boldsymbol{u} \mapsto (0, \boldsymbol{u})$ and analogously $\boldsymbol{u} \overset{j_1}{\mapsto} (1, \boldsymbol{u})$, it then follows* $\mathsf{K}(\mathrm{d}\omega) + \mathrm{d}\mathsf{K}(\omega) = j_1\omega - j_0\omega$.

Proof. Firstly, let us note that by a simple calculation

$$j_1(\mathrm{d}u^i) = \mathrm{d}u^i, \quad j_1(\mathrm{d}t) = 0, \quad j_0(\mathrm{d}u^i) = \mathrm{d}u^i, \quad j_0(\mathrm{d}t) = 0.$$

Consider now the simpler case of $\omega = A\,\mathrm{d}u^H$, where, by definition of K, we have that ω maps to $\mathsf{K}\omega = 0$ and

$$\mathrm{d}\omega \triangleq \mathrm{d}A \wedge \mathrm{d}u^H = \frac{\partial A}{\partial t}\mathrm{d}t \wedge \mathrm{d}u^H + \text{ other terms without } \mathrm{d}t,$$

thus $\mathsf{K}(\mathrm{d}\omega) = \int_0^1 \frac{\partial A}{\partial t}\,\mathrm{d}t\,\mathrm{d}u^H = (A(1, \boldsymbol{u}) - A(0, \boldsymbol{u}))\,\mathrm{d}u^H$ as the rest of the terms without $\mathrm{d}t$ is annihilated by K. In total, $\mathsf{K}(\mathrm{d}\omega) + \mathrm{d}\mathsf{K}(\omega) = (A(1, \boldsymbol{u}) - A(0, \boldsymbol{u}))\,\mathrm{d}u^H = j_1\omega - j_0\omega$.

Turning our attention to $\omega = B\,\mathrm{d}t \wedge \mathrm{d}u^J$ firstly yields (notice that the dependence of B on t always vanishes by $\wedge\mathrm{d}t$)

$$\mathrm{d}\omega = \mathrm{d}B \wedge \mathrm{d}t \wedge \mathrm{d}u^J = \frac{\partial B}{\partial u^i}\mathrm{d}u^i \wedge \mathrm{d}t \wedge \mathrm{d}u^J = -\frac{\partial B}{\partial u^i}\mathrm{d}t \wedge \mathrm{d}u^i \wedge \mathrm{d}u^J,$$

hence $\mathsf{K}(\mathrm{d}\omega) = -\int_0^1 \frac{\partial B}{\partial u^i}(t, \boldsymbol{u})\,\mathrm{d}t\,\mathrm{d}u^i \wedge \mathrm{d}u^J$. Additionally,

$$\mathsf{K}\omega \triangleq \int_0^1 B(t, \boldsymbol{u})\,\mathrm{d}t\,\mathrm{d}u^J \implies \mathrm{d}(\mathsf{K}\omega) = \int_0^1 \frac{\partial B}{\partial u^i}(t, \boldsymbol{u})\,\mathrm{d}t\,\mathrm{d}u^i \wedge \mathrm{d}u^J,$$

which, combined together with the last result, produces $\mathsf{K}(\mathrm{d}\omega) + \mathrm{d}(\mathsf{K}\omega) = 0$. This is indeed consistent, as for $\omega = B\mathrm{d}t \wedge \mathrm{d}u^J$ also

$$j_i\omega \triangleq \underbrace{j_i(\mathrm{d}t)}_{0} \wedge\ j_i(B\mathrm{d}u^J) = 0.$$

$\square$

Definition 3.10 (Contractible set)**.** A subset $U \subseteq \mathbb{R}^n$ is called *contractible* if there exists a *continuous* map, called **null-homotopy**, $\phi : I \times U \to U$ with properties

1. $\phi(1, \boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in U$;
2. $\phi(0, \boldsymbol{x}) = \boldsymbol{x}_0$ for all $\boldsymbol{x} \in U$ and some fixed $\boldsymbol{x}_0 \in U$.

If U is contractible, we can make the following computation:

$$(\phi \circ j_1)(\boldsymbol{x}) = \boldsymbol{x} \quad \& \quad (\phi \circ j_0)(\boldsymbol{x}) = \boldsymbol{x}_0.$$

Let now ω be a k-form on U, then

$$\underbrace{(\ j_1\circ\ \ \phi)}_{(\phi\circ j_1)}\omega = \omega \quad \& \quad (\ j_0\circ\ \ \phi)\omega = 0.$$

We can apply Proposition 3.5 to $\ \phi\omega$, where we now consider only **closed** forms (i.e., $\mathrm{d}\omega = 0$), which tells us

$$\mathsf{K}(\underbrace{\mathrm{d}\ \ \phi\omega}_{\triangleq\phi\mathrm{d}\omega}) + \mathrm{d}\mathsf{K}(\ \phi\omega) = j_1\ \ \phi\omega - \ j_0\ \ \phi\omega = \omega,$$

and by $\mathrm{d}\omega = 0$, it finally yields $\omega = \mathrm{d}(\mathsf{K}\ \ \phi\omega)$, i.e., ω is **exact**! In other words, for contractible domains the notions of **exact** and **closed** differential forms **coincide**! Note that we have shown that every exact form is also closed earlier.

Example 3.6. When solving differential equations, e.g., in Analysis 2, one may sometimes use the mapping $t \mapsto t(\xi, \eta)$ to compute the line integral of $\omega = P\mathrm{d}\xi + Q\mathrm{d}\eta$,

$$\int_{(0,0)}^{(\xi,\eta)} P\mathrm{d}\xi + Q\mathrm{d}\eta = \int_0^1 \left(P(t\xi, t\eta)\xi + Q(t\xi, t\eta)\eta\right)\,\mathrm{d}t,$$

which was often much easier to compute. Note that this works precisely because a line (through origin) is contractible to the origin!

3.3 Integration of Differential Forms

So far, we introduced the probleme where one has a vector field, and wants to compute its integral over a surface (there we need a normal of the surface) or over a curve (there we need a tangent to the curve) — notice, however, that in both times orientation is needed!

Definition 3.11 (Oriented manifold)**.** An n-dimensional manifold $\mathcal{M}$ is oriented, if there exists an n-form ω on $\mathcal{M}$, such that $\omega(P) \neq 0$ for every $P \in \mathcal{M}$.

Remark 3.10. At every $P \in \mathcal{M}$ we can use ω to check orientation of the basis $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ of $\mathrm{T}_p\mathcal{M}$ by checking the sign of $\omega(P)(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$.

Remark 3.11 (Oriented atlas). Let $\mathcal{A} = (U_i, \varphi_i)_{i\in I}$ be an atlas. If for every i, j with $U_i \cap U_j \neq \emptyset$ it holds that[8] $\det\left(\frac{\partial(\varphi_i \circ \varphi_j^{-1})}{\partial(u^1, \dots, u^n)}\right) > 0$, we call $\mathcal{A}$ **oriented**. Moreover, manifold $\mathcal{M}$ is oriented[9] if it admits an oriented atlas.

Let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ be an *oriented basis* of $\mathbb{R}^n$ and consider the **oriented n-simplex** S, defined as[10]

$$S = \left\{ t^1\boldsymbol{v}_1 + \cdots + t^n\boldsymbol{v}_n \mid t^i \geq 0, \ \sum_{i=1}^{n} t^i = 1 \right\}. \tag{3.18}$$

Note that the orientation of the simplex S comes from the orientation of the basis $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$. Then for an n-form $\omega = A(\boldsymbol{x})\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ we set

$$\int_S \omega := \int \!\! \cdots \!\! \int_S A(\boldsymbol{x})\, \mathrm{d}x^1 \dots \mathrm{d}x^n, \tag{3.19}$$

where the right-hand side is a Riemann integral! Importantly, notice that using this definition, we can integrate n-forms over n-simplices.

Now we shall study how integrals of differential forms on $\mathcal{M}$ over (general) simplices behave, and how we could generalize this. To this end, consider m-simplex S in $\mathbb{R}^m$ with $m \leq n$, which can embedded into $\mathbb{R}^n$ by embedding $\iota : (t^1, \dots, t^m) \mapsto (t^1, \dots, t^m, 0, \dots, 0) \in \mathbb{R}^n$. Further, denote by T the image of S under $\varphi_i^{-1} \circ \iota$ (i.e., the image of m-dimensional simplex on n-dimensional manifold $\mathcal{M}$ by φ_i^{-1}). Lastly, take ω and m-form on $\mathcal{M}$ given by $\omega = a_H \mathrm{d}u^H$ in U_i. By pullback of ι, i.e., $\iota\omega$ we get an m-form on $\mathbb{R}^m$, hence analogously to (3.19) we write $\int_T \omega := \int_S \iota\omega$. Figure 3.3 illustrates this setting visually.

Remark 3.12. Using the construction above, we can integrate over diffeomorphic images of simplices.

Definition 3.12 (Chain)**.** A chain C is a formal sum of (images of) simplices, $C = \sum_i \alpha_i S_i$, where S_i are simplices and $\alpha_i \in \mathbb{R}$ (or $\mathbb{Z}$).

[8]These determinants can, in fact, be used to construct a form prescribed by them.
[9]Hence, for example, the Möbius strip is *not orientable*.
[10]Beware that I found that simplex S is sometimes defined with $\sum_{i=1}^{n} t^i \leq 1$, whereas other times with $\sum_{i=1}^{n} t^i = 1$.

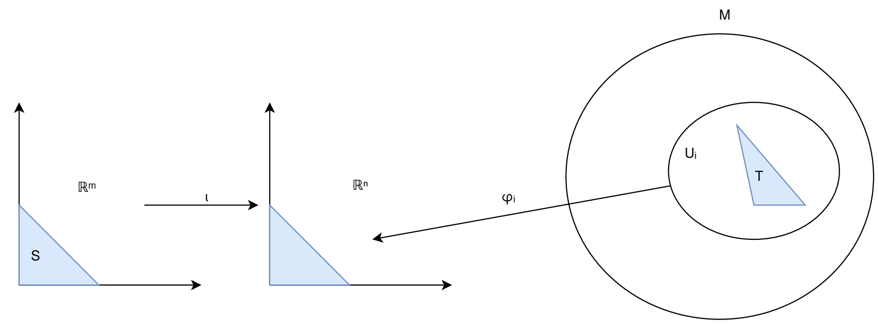

Figure 3.3: Illustration of an m-simplex embedded into $\mathbb{R}^n$ and further projected to T on the manifold $\mathcal{M}$

Denoting $\phi = \varphi_i^{-1} \circ \iota$, we can extend integration over simplices to integration over chains as follows,

$$\phi(C) = \sum_i \alpha_i \phi(S_i) \implies \int_c \phi(\omega) = \sum_i \alpha_i \int_{S_i} \phi(\omega) \triangleq \sum_i \alpha_i \int_{\phi(S_i)} \omega =: \int_{\phi(C)} \omega.$$

Say we have a simplex $S = \left\{\sum_{i=0}^n t^i P_i \mid t^i \geq 0,\ \sum_{i=0}^n t^i = 1\right\}$ in $\mathbb{R}^n$, which has $n+1$ indices $(P_0, \dots, P_n)$ and their order determines the orientation of the simplex. In other words, for $\pi \in S_{n+1}$ we have $(P_{\pi(0)}, \dots, P_{\pi(n)}) = \mathrm{sign}(\pi)(P_0, \dots, P_n)$. Before rigorously defining the boundary ∂S of a simplex S, we first show it on simple examples, see Figure 3.4:

1. $n = 1$ (Figure 3.4 a)): $\partial S : (P_1) - (P_0)$, where $(P_1), (P_0)$ are 0-dimensional simplices;
2. $n = 2$ (Figure 3.4 b)):

$$\partial S : (P_0, P_1) + (P_1, P_2) + (P_2, P_0) = (P_1, P_2) - (P_0, P_2) + (P_0, P_1),$$

   where after rearranging the k-th term is missing exactly the vertex P_k;
3. $n = 3$ (Figure 3.4 c)):

$$\begin{aligned}\partial S :&(P_1, P_2, P_3) + (P_0, P_1, P_3) + (P_2, P_0, P_3) + (P_0, P_2, P_1)\\ &= (P_1, P_2, P_3) - (P_0, P_2, P_3) + (P_0, P_1, P_3) - (P_0, P_1, P_2).\end{aligned}$$

Hence, in general for a simplex $S = (P_0, \dots, P_n)$ we have

$$\partial S := \sum_{i=0}^n (-1)^i (P_0, \dots, P_{i-1}, P_{i+1}, \dots, P_n) = \sum_{i=0}^n (-1)^i (P_0, \dots, \cancel{P_i}, \dots, P_n). \tag{3.20}$$

Furthermore, let us observe that $\partial\partial S = 0$; indeed, compute

$$\begin{aligned}\partial\partial S &\triangleq \sum_{i=0}^n (-1)^i \partial (P_0, \dots, \cancel{P_i}, \dots, P_n)\\ &= \sum_{i=0}^n (-1)^i \left(\sum_{j=0}^{i-1} (-1)^j (P_0, \dots, \cancel{P_j}, \dots, \cancel{P_i}, \dots, P_n) - \sum_{j=i+1}^n (-1)^j (P_0, \dots, \cancel{P_i}, \dots, \cancel{P_j}, \dots, P_n) \right)\\ &= 0,\end{aligned} \tag{3.21}$$

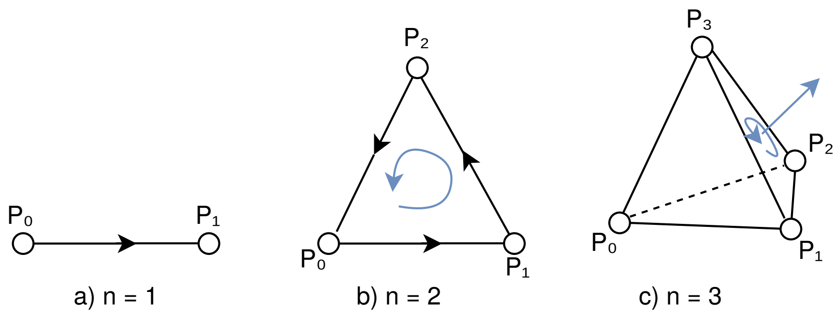

Figure 3.4: Visualizations of basic simplices for $n \in [3]$

as we iterate through all combinations of i, j, which subtract to result in 0. Moreover, this extends to chains right away by linearity, i.e., $\partial\partial C = 0$.

Definition 3.13 (Cycle & boundary)**.** Let C be a p-chain[11], then

- C is called a **cycle**, if $\partial C = 0$;
- C is called a **boundary**, if there is a $(p+1)$-chain D such that $\partial D = C$.

Lemma 3.2. *Every* boundary *is a cycle (but not vice versa).*

Proof. Follows trivially from the chain version of (3.21). □

Theorem 3.1 (Stokes')**.** *Let C be a p-chain and $\omega \in \mathscr{F}^{p-1}(\mathcal{M})$, then $\int_C \mathrm{d}\omega = \int_{\partial C} \omega$.*

Example 3.7. Before proving the Stokes' theorem 3.1 in full generality, we shall present several special cases, where it reduces to another theorem from Analysis 2 or 3.

Right away, for $n = 2$ and $p = 2$ we get (here, C is surface in $\mathbb{R}^2$ and ∂C is a line in $\mathbb{R}^2$, thus a curve in $\mathcal{M}$)

$$\int_{\partial C} P\,\mathrm{d}x + Q\,\mathrm{d}y = \int_C \left(\frac{\partial P}{\partial x}\mathrm{d}x + \frac{\partial P}{\partial y}\mathrm{d}y\right) \wedge \mathrm{d}x + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}x \wedge \mathrm{d}y,$$

which, one might recall, is the Gauss' theorem.

Furthermore, for $n = 3$ and $p = 2$ it yields

$$\begin{aligned}\int_{\partial C} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z &= \int_C \left(\frac{\partial P}{\partial x}\mathrm{d}x + \frac{\partial P}{\partial y}\mathrm{d}y + \frac{\partial P}{\partial z}\mathrm{d}z\right) \wedge \mathrm{d}x + \ldots \\ &= \int_C \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathrm{d}y \wedge \mathrm{d}z + \left(\frac{\partial P}{\partial y} - \frac{\partial R}{\partial x}\right)\mathrm{d}z \wedge \mathrm{d}x + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x \wedge \mathrm{d}x,\end{aligned}$$

[11]A p-chain is a chain of p simplices.

which the Stokes' theorem as we know it from introductory analysis lectures.

If we take $n = 3$ and $n = 3$

$$\int_{\partial C} P\,\mathrm{d}y \wedge \mathrm{d}z + Q\,\mathrm{d}z \wedge \mathrm{d}x + R\,\mathrm{d}x \wedge \mathrm{d}y = \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z,$$

we get the so called divergence theorem.

Lastly, for $n = 1$, thus necessarily $p = 1$, the form ω now becomes of degree 0, i.e., a function f, hence for $C = (P_0, P_1)$ we get

$$\int_C \mathrm{d}f = f(P_1) - f(P_0).$$

In other words, we have generalized the fundamental theorem of calculus.

Proof. By linearity, it suffices to prove the theorem for a p-form $\omega = A\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^p$, in $\mathbb{R}^{p+1}$, and $C = (P_0, \dots, P_{p+1})$ (hence boundary is p-dimensional) with P_0 and $P_j = \big(0, \dots, \underset{j}{\underset{\sim}{1}}, \dots, 0\big)^\top$. Then $\mathrm{d}\omega = (-1)^p \frac{\partial A}{\partial x^{p+1}} \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^{p+1}$ and

$$\partial C = \sum_{i=0}^{p+1} (-1)^i (P_0, \dots, \cancel{P_i}, \dots, P_{p+1}) = (P_1, \dots, P_{p+1}) + (-1)^{p+1}(P_0, \dots, P_p) + \text{ other summands }. \tag{3.22}$$

Firstly notice that $\int_S \omega = 0$ if S if one of the "other summands", as one of the variables, say x^j, is constant 0 for some $j = 1, \dots, p$, and as such $\mathrm{d}x^j = 0$. Recall the definition of S (3.18), where the conditions $x^j \geq 0 \wedge \sum_{j=1}^{p+1} x^j = 1$ are equivalent to $0 \leq x^{p+1} \leq 1 - \sum_{j=1}^{p} x^j$ (such that the sum is at most 1). Hence, by (3.19)

$$\begin{aligned}
\int_C \mathrm{d}\omega &= (-1)^p \int_C \frac{\partial A}{\partial x^{p+1}} \mathrm{d}x^1 \dots \mathrm{d}x^{p+1} \\
&= (-1)^p \int_{\sum_{j=1}^p x^j \leq 1} \int_{x^{p+1}=0}^{1-\sum_{j=1}^p x^j} \frac{\partial A}{\partial x^{p+1}} \mathrm{d}x^{p+1}\,\mathrm{d}x^1 \dots \mathrm{d}x^p \\
&\overset{\text{Fubini}}{=} \int_{\sum_{j=1}^p x^j \leq 1} \left(A\left(x^1, \dots, x^p, 1 - \sum_{j=1}^{p} x^j\right) - A\,(x^1, \dots, x^p, 0)\right) \mathrm{d}x^1 \dots \mathrm{d}x^p.
\end{aligned} \tag{3.23}$$

Moreover, simplex definition (3.18) yields

$$\begin{aligned}
(P_1, \dots, P_{p+1}) &= \left\{(x^1, \dots, x^{p+1}) \mid x^1 + \cdots + x^{p+1} = 1,\ x^j \geq 0\right\}, \\
(P_0, \dots, P_p) &= \left\{(x^1, \dots, x^p, 0) \mid x^1 + \cdots + x^p \leq 1,\ x^j \geq 0\right\},
\end{aligned}$$

as for $(P_0, \dots, P_p)$ we have $\sum_{j=0}^{p} x^j = 1$. Thus, by (3.19) we also get

$$
\begin{aligned}
(-1)^{p+1} \int_{(P_0,\dots,P_p)} \omega &= (-1)^{p+1} \int_{x^1+\dots+x^p \le 1} A(x^1, \dots, x^p, 0)\, \mathrm{d}x^1 \dots x^p \\
&\& \\
\int_{(P_1,\dots,P_{p+1})} \omega &= \int_{(P_1,\dots,P_0)} A\left(x^1, \dots, x^p, 1 - \sum_{j=1}^{p} x^j\right) \mathrm{d}x^1 \dots \mathrm{d}x^p \\
&= (-1)^p \int_{(P_0,\dots,P_p)} \mathrm{d}x^1 \dots \mathrm{d}x^p,
\end{aligned} \tag{3.24}
$$

where for the second integral we first fixed x^{p+1} by $x^1, \dots, x^p$ in A, which then allowed us to project the simplex $(P_1, \dots, P_{p+1})$ to $(P_1, \dots, P_p, P_0)$.

Finally, comparing (3.23) with (3.24) plugged into (3.22) produces the desired equality. □

Let us remark that for each step of the proof of Stokes' theorem 3.1 we need the fundamental theorem of calculus.

Remark 3.13. In particular, the integral gives a pairing between p-chain C and p-form ω, namely $(C, \omega) \mapsto \int_C \omega$. Moreover, taking $C = \tilde{C} + \partial D$ and $\omega = \tilde{\omega} + \mathrm{d}\eta$ where $\tilde{C}$ is a cycle 3.13 and $\tilde{\omega}$ is closed, yields

$$
\int_C \omega = \int_{\tilde{C}+\partial D} \tilde{\omega} + \mathrm{d}\eta = \int_{\tilde{C}} \tilde{\omega} + \int_{\partial D} \tilde{\omega} + \int_{\tilde{C}} \mathrm{d}\eta + \int_{\partial D} \mathrm{d}\eta \implies \int_C \omega = \int_{\tilde{C}} \tilde{\omega},
$$

as by Theorem 3.1 we have

- $\tilde{C}$ is a cycle, i.e., $\partial\tilde{C} = 0 \implies \int_{\tilde{C}} \mathrm{d}\eta = \int_{\partial\tilde{C}} \eta = 0$;
- $\tilde{\omega}$ is closed, then $\int_{\partial D} \tilde{\omega} = \int_D \mathrm{d}\tilde{\omega} = 0$;
- $\int_{\partial D} \mathrm{d}\eta = \int_{\partial\partial D} \eta = 0$ by (3.21).

In other words, this pairing of p-cycles and p-forms is isomorphic to an analogous pairing of quotient groups of cycles modulo boundaries and forms modulo closed forms.

Remark 3.14. So far, we have seen that while every exact form is closed, the converse is not true in general. However, *locally* every closed form is indeed exact (*Poincaré's lemma*, see Proposition 3.5 and Definition 3.10), which we have shown for contractible neighborhoods of points. However, a stronger statement may be formulated using cycles!

Definition 3.14 (Period of a form)**.** Let ω be a k-form on $\mathcal{M}$ and C be a k-cycle on $\mathcal{M}$. Then the integral $\int_C \omega$ is called a *period* (of ω).

Theorem 3.2 (de Rham)**.** *A closed form is exact if and only if all its periods vanish.*

As a special case follows the path independence $\int_C \boldsymbol{F}\, \mathrm{d}\boldsymbol{r}$ of a conservative vector field $\boldsymbol{F}$ — indeed, for us this translates into integrals over 1-forms, and by Theorem 3.2 this property follows.

3.4 Riemannian Metric on $\mathcal{M}$

Let us now equip every tangent space $\mathrm{T}_P\mathcal{M}$ with a norm (i.e., a non-degenerate bilinear form), which might be achieved by prescribing scalar products $\left\langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \right\rangle =: g_{ij} = g\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right)$ of basis vectors of $\mathrm{T}_P\mathcal{M}$. Here g_{ij} are local to the point P of $\mathrm{T}_P\mathcal{M}$ and we assume $(g_{ij})_{ij}$ is strictly positive definite. Note that g_{ij} are functions of the point P (given a suitable coordinate chart), and we require them to be infinitely differentiable.

A scalar product on the tangent space $\mathrm{T}_P\mathcal{M}$ canonically induces a corresponding scalar product on the cotangent space $\mathrm{T}_P^*\mathcal{M}$, i.e., for $g_{ij} = g\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right)$ we have in the dual basis $g(\mathrm{d}u^i, \mathrm{d}u^j) =: g^{ij}$ with $g^{il}g_{lj} = \delta^i_j$. In other words, the matrices $(g^{ij})_{ij}$ and $(g_{ij})_{ij}$ are inverse. Additionally, this scalar product also induces a scalar product on $\Lambda^k\mathrm{T}_P^*\mathcal{M}$ by *ground determinants*

$$g(\mathrm{d}u^H, \mathrm{d}u^K) = \det\left((g^{ij})_{i\in H, j\in K}\right),$$

where H, K are subsets of $[n]$ of cardinality k.

By equipping $\Lambda^k\mathrm{T}_P^*\mathcal{M}$ with a scalar product, thus a norm and an orientation, we can define the Hodge-star operator 3.5, see (3.10). Recall that for a vector space V and $\boldsymbol{\alpha} \in \Lambda^k V$, $\boldsymbol{\beta} \in \Lambda^{n-k}V$, the Hodge-star operator $* : \Lambda^k V \to \Lambda^{n-k}V$ it holds

$$(*\boldsymbol{\alpha}) \wedge (*\boldsymbol{\beta}) \triangleq \langle * * \boldsymbol{\alpha}, *\boldsymbol{\beta}\rangle\, \boldsymbol{\omega} = (-1)^{k(n-k)} \langle \boldsymbol{\alpha}, *\boldsymbol{\beta}\rangle\, \boldsymbol{\omega} = (-1)^{k(n-k)} \boldsymbol{\beta} \wedge \boldsymbol{\alpha} = \boldsymbol{\alpha} \wedge \boldsymbol{\beta},$$

where the last equality holds as $\wedge$ is alternating. Combining $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \triangleq \langle *\boldsymbol{\alpha}, \boldsymbol{\beta}\rangle\, \boldsymbol{\omega}$ with an intermediate result of the above computation yields that for $* : \Lambda^k V \to \Lambda^{n-k} V$ the “complementary” Hodge-star operator $* : \Lambda^{n-k} V \to \Lambda^k V$ is also the **adjoint**.

Let us use these results to understand the surface of the manifold. For this purpose, we define a surface measure on an oriented manifold $\mathcal{M}$, i.e., we have $\tilde{\omega} \in \mathscr{F}^n(\mathcal{M})$ such that $\tilde{\omega}(P) \neq 0$ for all $P \in \mathcal{M}$. We define the volume form[12] by $\mathrm{d}\,\mathrm{vol} = \omega = \frac{\tilde{\omega}}{\|\tilde{\omega}\|}$ where $\|\tilde{\omega}\| = \sqrt{g(\tilde{\omega}, \tilde{\omega})}$. Specifically, the form ω assigns volume 1 to a unit cube. Furthermore, we would like to write as multiple of $\mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n$. To this end, notice $\left\|\mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n\right\|^2 = \det((g^{ij})_{ij}) = \det((g_{ij})_{ij})^{-1}$. Thus indeed $\omega = \sqrt{\det((g_{ij})_{ij})}\mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n$.

We use the volume form to define a “pre-Hilbert” space structure on $\mathscr{F}^k(\mathcal{M})$; for $\alpha, \beta \in \mathscr{F}^k(\mathcal{M})$ we set

$$\langle \alpha, \beta\rangle := \int_{\mathcal{M}} g(\alpha, \beta)\omega = \int_{\mathcal{M}} (*\alpha) \wedge \beta = \int_{\mathcal{M}} \alpha \wedge (*\beta).$$

Our goal will now be to find the adjoint of the exterior derivative $\mathrm{d} : \mathscr{F}^k(\mathcal{M}) \to \mathscr{F}^{k+1}(\mathcal{M})$ with respect to the above-defined scalar product. To this end, given $\alpha \in \mathscr{F}^k(\mathcal{M})$, $\beta \in \mathscr{F}^{k+1}(\mathcal{M})$ we

[12] Let us note $\mathscr{F}^n(\mathcal{M})$ for n-dimensional manifold $\mathcal{M}$ is a one-dimensional space, hence we have effectively two choices for the volume form.

compute

$$\begin{aligned}
\langle \mathrm{d}\alpha, \beta \rangle &= \int_{\mathcal{M}} \mathrm{d}\alpha \wedge \beta \\
&= \int_{\mathcal{M}} \mathrm{d}(\alpha \wedge (*\beta)) - (-1)^k \int_{\mathcal{M}} \alpha \wedge \mathrm{d}(*\beta) \\
&\overset{\text{Stokes}}{=} \int_{\partial\mathcal{M}} \alpha \wedge (*\beta) - (-1)^k \int_{\mathcal{M}} \alpha \wedge \mathrm{d}(*\beta)
\end{aligned}$$

where we used the fact $\mathrm{d}(\alpha \wedge (*\beta)) \triangleq \mathrm{d}\alpha \wedge (*\beta) + (-1)^k \alpha \wedge \mathrm{d}(*\beta)$ and a generalization of Stokes' theorem 3.1 to manifolds. The first term vanishes if $\mathcal{M}$ does not have a boundary, hence by $** = (-1)^{n(k-k)}\mathrm{id}$

$$\begin{aligned}
\langle \mathrm{d}\alpha, \beta \rangle &= (-1)^{k+1} \int_{\mathcal{M}} \alpha \wedge \mathrm{d}(*\beta) \\
&= (-1)^{k+1}(-1)^{k(n-k)} \int_{\mathcal{M}} \alpha \wedge *(*\mathrm{d} * \beta) \\
&= (-1)^{k+1+nk-k^2} \int_{\mathcal{M}} (*\alpha) \wedge (*\mathrm{d} * \beta) \\
&\triangleq (-1)^{nk+1} \langle \alpha, *\mathrm{d} * \beta \rangle ,
\end{aligned}$$

where the last equality holds as $k - k^2 \equiv 0 \mod 2$. Thus, $\delta := (-1)^{nk+1} * \mathrm{d}*$ is the adjoint of d with respect to the scalar product on $\mathscr{F}^k(\mathcal{M})$.

Henceforth, we have can define the **self-adjoint** operator $\Delta := \delta\mathrm{d} + \mathrm{d}\delta$ on $\mathscr{F}^k(\mathcal{M})$, which is typically the **Laplace** operator or **Laplacian**. In the case of $k = 0$, it simplifies to

$$\Delta f = \delta \mathrm{d} f + \mathrm{d}\underbrace{\delta f}_{0} = - * \mathrm{d} * \mathrm{d} f,$$

where $\delta f = 0$ because $\delta : \mathscr{F}^k(\mathcal{M}) \to \mathscr{F}^{k-1}(\mathcal{M})$ with $\mathscr{F}^{-1}(\mathcal{M}) = \{0\}$. The important takeaway here is that the laplacian exists **only if we have a metric**. Recall that typically the laplacian is defined as the divergence of gradient $\mathrm{d}f(\boldsymbol{v}) = \langle \operatorname{grad} f, \boldsymbol{v} \rangle$, which, in turn, depends on the scalar product. In other words, this reliance on a scalar product was part of the definition even in $\mathbb{R}^3$.

Let us now focus on computing the Hodge-star of a 1-form $\lambda = a_i \mathrm{d}u^i$. In fact, it suffices to compute only $*\mathrm{d}u^i = b_i^j \mathrm{d}\hat{u}^j$ with $\mathrm{d}\hat{u}^j = \mathrm{d}u^1 \wedge \cdots \wedge \cancel{\mathrm{d}u^j} \wedge \cdots \wedge \mathrm{d}u^n$. Clearly,

$$*\mathrm{d}u^i \wedge \mathrm{d}u^l \triangleq \left\langle (-1)^{(k+1)(n-(k+1))} \mathrm{d}u^i, \mathrm{d}u^j \right\rangle \omega \overset{k=1}{=} \left\langle \mathrm{d}u^i, \mathrm{d}u^j \right\rangle \omega,$$

where the left-hand side expands to sum over $b_i^j \mathrm{d}\hat{u}^j \wedge \mathrm{d}u^l$ with the only non-zero term for $j = l$. Moving l to the correct position produces

$$(-1)^{n-l} b_i^l \mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n = *\mathrm{d}u^i \wedge \mathrm{d}u^l = g^{il} \underbrace{\sqrt{g}\,\mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n}_{\omega}$$

$$\Downarrow$$

$$b_i^l = (-1)^{n-l}\sqrt{g} \implies *\mathrm{d}u^i = \sum_l (-1)^{n-l} g^{il} \sqrt{g}\, \mathrm{d}\hat{u}^l.$$

3.4.1 Green's Formula

Consider now a sub-manifold $\mathcal{N} \subseteq \mathcal{M}$ with $\dim \mathcal{N} = \dim \mathcal{M}$ a two functions (0-forms) φ, ψ on $\mathcal{M}$. A natural definition of the $\mathcal{N}$-restricted scalar product of functions on $\mathcal{N}$, i.e., a bilinear form on φ, ψ, is (by inclusion into $\mathcal{M}$ and) employing the scalar product of k-forms,

$$\int_{\mathcal{N}} \mathrm{d}\varphi \wedge * \mathrm{d}\psi = \int_{\mathcal{N}} * \mathrm{d}\varphi \wedge \mathrm{d}\psi = \int_{\mathcal{N}} \langle \mathrm{d}\varphi, \mathrm{d}\psi \rangle \, \omega.$$

Then as $\mathrm{d}\varphi = \frac{\partial \varphi}{\partial u^j} \mathrm{d}u^j$ and $* \mathrm{d}\psi = * \left(\frac{\partial \psi}{\partial u^i} \mathrm{d}u^i \right) = \sum_l (-1)^{n-l} g^{il} \sqrt{g} \frac{\partial \psi}{\partial u^i} \mathrm{d}\hat{u}^l$, we get

$$\begin{aligned}
\mathrm{d}\varphi \wedge * \mathrm{d}\psi &= \sum_l (-1)^{n-l} g^{il} \sqrt{g} \frac{\partial \psi}{\partial u^i} \frac{\partial \varphi}{\partial u^j} \underbrace{\mathrm{d}u^j \wedge \mathrm{d}\hat{u}^l}_{0 \iff j \neq} \\
&= (-1)^{n-1} g^{il} \sqrt{g} \frac{\partial \psi}{\partial u^i} \frac{\partial \varphi}{\partial u^l} \, \mathrm{d}u^1 \wedge \cdots \wedge \mathrm{d}u^n \\
&= (-1)^{n-1} g^{il} \frac{\partial \psi}{\partial u^i} \frac{\partial \varphi}{\partial u^l} \omega,
\end{aligned}$$

which clearly is a quadratic form applied to the derivatives. Before continuing with the main idea of this section, notice that by (3.10)

$$* \overbrace{\mathrm{d} * \mathrm{d}\psi}^{\text{degree } n} = \Delta\psi \implies \mathrm{d} * \mathrm{d}\psi = \Delta\psi\omega.$$

Furthermore, by the manifold-version of Stokes' theorem 3.1 we get (keep in mind $\partial\mathcal{N}$ is $n-1$ dimensional space)

$$\begin{aligned}
\int_{\partial\mathcal{N}} \varphi \underbrace{* \mathrm{d}\psi}_{\text{degree } n-1} &= \int_{\mathcal{N}} \mathrm{d}(\varphi * \mathrm{d}\psi) \triangleq \int_{\mathcal{N}} \mathrm{d}\varphi \wedge * \mathrm{d}\psi + \int_{\mathcal{N}} \varphi \underbrace{\overbrace{\mathrm{d} * \mathrm{d}\psi}^{\text{degree } n}}_{\Delta\psi \cdot \omega} \\
&= \int_{\mathcal{N}} \mathrm{d}\varphi \wedge * \mathrm{d}\psi + \int_{\mathcal{N}} \varphi \Delta\psi\omega.
\end{aligned}$$

By exchanging the order of φ and ψ in the result above and comparing them, we finally get the **Green's formula**,

$$\int_{\partial\mathcal{N}} \varphi * \mathrm{d}\psi - \psi * \mathrm{d}\varphi = \int_{\mathcal{N}} (\varphi\Delta\psi - \psi\Delta\varphi) \, \omega.$$

The main take-home message here is that integration of forms over (oriented) differentiable manifolds generalizes the identities and theorems from classical analysis.

3.4.2 Harmonic Forms and Hodge Decomposition

Definition 3.15 (Harmonic form)**.** A form $\omega \in \mathscr{F}^k(\mathcal{M})$ is called *harmonic*, if $\Delta\omega = 0$.

Immediately, we can see that (recall δ is adjoint of d)

$$0 = \langle \Delta\omega, \omega \rangle = \langle (\delta \mathrm{d} + \mathrm{d}\delta)\omega, \omega \rangle = \langle \mathrm{d}\omega, \mathrm{d}\omega \rangle + \langle \delta\omega, \delta\omega \rangle \implies \mathrm{d}\omega = \delta\omega = 0,$$

as the scalar product is, by definition, non-degenerate. Also $\langle \mathrm{d}\alpha, \delta\beta \rangle = \langle \mathrm{dd}\alpha, \beta \rangle = 0$ using Proposition 3.4, and for $\Delta\omega = 0$

$$\langle \mathrm{d}\alpha, \omega \rangle = \langle \alpha, \delta\omega \rangle = 0, \langle \delta\beta, \omega \rangle = \langle \beta, \mathrm{d}\omega \rangle \qquad\qquad = 0,$$

hence the images of d, δ and the space of harmonic forms are **mutually orthogonal**. Therefore we can phrase the following statement

Theorem 3.3 (Hodge decomposition)**.** *Let $\eta \in \mathscr{F}^k(\mathcal{M})$. Then there exist $\alpha \in \mathscr{F}^{k-1}(\mathcal{M})$, $\beta \in \mathscr{F}^{k+1}(\mathcal{M})$, and $\omega \in \mathscr{F}^k(\mathcal{M})$ with $\Delta\omega = 0$ such that $\eta = \mathrm{d}\alpha + \delta\beta + \omega$.*

Remark 3.15. Importantly, this composition is unique in the sense that $\mathrm{d}\alpha$ is unique (however not α, as we could clearly perturb it by any closed form and get the same result). To see that, consider $0 = \mathrm{d}\alpha + \delta\beta + \omega$. Applying d, and δ to it, respectively, gives

$$\left.\begin{aligned} 0 = \mathrm{dd}\alpha + \mathrm{d}\delta\beta + \mathrm{d}\omega &\implies \mathrm{d}\delta\beta = 0 \\ 0 = \delta\mathrm{d}\alpha + \delta\delta\beta + \delta\omega &\implies \delta\mathrm{d}\alpha = 0 \end{aligned}\right\} \implies \begin{cases} 0 = \langle \mathrm{d}\delta\beta, \beta \rangle = \langle \delta\beta, \delta\beta \rangle & \implies \delta\beta = 0, \\ 0 = \langle \delta\mathrm{d}\alpha, \alpha \rangle = \langle \mathrm{d}\alpha, \mathrm{d}\alpha \rangle & \implies \mathrm{d}\alpha = 0, \end{cases}$$

and thus also $\omega = 0$. In other words, this decomposition maps to 0 if and only if all the terms are 0.

The map $h : \ker(\Delta) \to H^p(\mathcal{M})$, where $\ker(\Delta)$ is the space of harmonic forms and the ***p*-th cohomology space** $H^p(\mathcal{M})$ is the quotient space of closed p-forms modulo exact p-forms, with $\omega \mapsto [\omega]$ is an isomorphism.

As a side-note, also recall that we have an isomorphism between $H^p(\mathcal{M})$ and p-chains modulo p-boundaries.

Proof. Indeed, one can verify that

- h is injective: if $[\omega] = 0$ then ω is an exact form, i.e., $\omega = \mathrm{d}\alpha$. As ω is also harmonic, we get $\delta\omega = 0$, hence
$$0 = \langle \delta\omega, \alpha \rangle = \langle \omega, \mathrm{d}\alpha \rangle = \langle \omega, \omega \rangle \implies \omega = 0;$$
- h is surjective: take a representative $\eta \in [\eta]$, then, by definition, η is closed. Using Theorem 3.3, we get $\eta = \mathrm{d}\alpha + \delta\beta + \gamma$, and
$$0 = \mathrm{d}\eta = 0 + \mathrm{d}\delta\beta + 0 \implies \mathrm{d}\delta\beta = 0 \implies \langle \mathrm{d}\delta\beta, \beta \rangle = \langle \delta\beta, \delta\beta \rangle = 0 \implies \delta\beta = 0.$$
Thus $\eta = \mathrm{d}\alpha + \gamma$, which, in turn, implies $\gamma \in [\eta]$ (as they differ up to a closed form $\omega = \mathrm{d}\alpha$).

In particular, this shows that p-th cohomology space $H^p(\mathcal{M})$ is finite dimensional (although both “original” spaces are infinite dimensional), as it is isomorphic to a finite dimensional space. □

3.5 A (Little) Bit on Riemannian Geometry

Consider again a differentiable manifold $\mathcal{M}$ with the tangent spaces $\mathrm{T}_P\mathcal{M}$ forming the tangent bundle $\mathrm{T}\mathcal{M}$. Moreover, let $\boldsymbol{X}$ a vector field, and recall that vector fields act on functions by $\boldsymbol{X}f$ (as directional derivatives), see (3.13).

Let now $\boldsymbol{X}, \boldsymbol{Y}$ be vector fields, then as we have discussed in Remark 3.9, $\boldsymbol{Y}\boldsymbol{X}$ is *not* a vector field, i.e., it does not coincide with directional derivatives. To remedy this, we have introduced the **Lie bracket** $[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}$, which was again a vector field. Nonetheless, we would like to also have directional derivatives of vector fields, as the naive construction above clearly does not work/suffice.

Definition 3.16 (Connection)**.** A map $\nabla : \mathrm{T}\mathcal{M} \times$ vector fields $\to \mathrm{T}\mathcal{M}$ is called a *connection*, if it satisfies

1. (*linearity in direction*): $\nabla_{\alpha\xi+\beta\eta}\boldsymbol{X} = \alpha\nabla_\xi\boldsymbol{X} + \beta\nabla_\eta\boldsymbol{X}$ for every point $P \in \mathcal{M}$ and $\xi, \eta \in \mathrm{T}_P\mathcal{M}$ and
2. (*function-vector field product rule*): $\nabla_\xi f\boldsymbol{X} = \xi f \cdot \boldsymbol{X} + f\nabla_{\boldsymbol{\xi}}\boldsymbol{X}$ at every point P such that[13] $\xi \in \mathrm{T}_P\mathcal{M}$

.

Remark 3.16. In coordinates we have $\nabla_{\frac{\partial}{\partial u^i}}\frac{\partial}{\partial u^j} = \Gamma^l_{ij}\frac{\partial}{\partial u^l}$, where Γ^l_{ij} are called the **Christoffel** symbols (or the *connection coefficients*).

Example 3.8. Take for simplicity $\mathbb{R}^3$ and manifold $\mathcal{M}$ prescribed by $\boldsymbol{F} : U \to \mathbb{R}^3$ such that $U \subseteq \mathbb{R}^2$ and assume $\frac{\partial \boldsymbol{F}}{\partial u^1}, \frac{\partial \boldsymbol{F}}{\partial u^2}$ are *linearly independent* for all $\boldsymbol{u} = (u^1, u^2) \in U$ (see Figure 3.5 for illustration). Then

$$\underbrace{\frac{\partial^2 \boldsymbol{F}}{\partial u^1 \partial u^2}}_{\in\mathbb{R}^3} = \underbrace{\Gamma^l_{ij}\frac{\partial \boldsymbol{F}}{\partial u^l}}_{\in\mathrm{T}_P\mathcal{M}} + h_{ij}\boldsymbol{n},$$

where $\boldsymbol{n}$ is a normal vector such that $\|\boldsymbol{n}\| = 1$. Clearly, h_{ij} holds the curvature information.

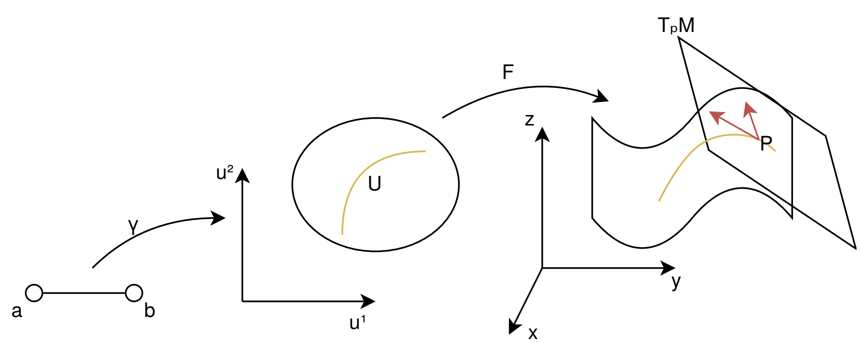

Figure 3.5: Differential manifold and a parametrized curve in $\mathbb{R}^3$

[13] Recall that ξf is the directional derivative of f in the direction of ξ at P.

Now consider a parametrized curve $\mathcal{C}(t)$ in $\mathcal{M}$, which reads $\mathcal{C}(t) = \boldsymbol{F} \circ \gamma(t)$. For $\gamma_i(t) = u^i(t)$ we have $\dot{\mathcal{C}}(t) = \frac{\partial \boldsymbol{F}}{\partial u^i}\dot{u}^i(t)$. To get the curvature, we need second derivatives, i.e.,

$$\ddot{\mathcal{C}}(t) = \frac{\partial^2 \boldsymbol{F}}{\partial u^i \partial u^j}\dot{u}^i(t)\dot{u}^j(t) + \frac{\partial \boldsymbol{F}}{\partial u^i}\ddot{u}^i(t),$$

where we assumed that $\mathcal{C}(t)$ was parametrized by arc-length (more on this later). The curvature is then $\langle \ddot{\mathcal{C}}, \boldsymbol{n} \rangle = h_{ij}\dot{u}^i\dot{u}^j$.

In the general case, consider a curve $\mathcal{C}$ on $\mathcal{M}$ given by parametrization γ in a chart, i.e., $\gamma : u^i = u^i(t)$ for $t \in [a, b]$. Then $\boldsymbol{X} \circ \mathcal{C}(t)$ is the vector field along the curve $\mathcal{C}$. Moreover, let $T(t)$ be the tangent vector of $\mathcal{C}$, i.e., $T(t) = \frac{\partial u^i}{\partial t}\frac{\partial}{\partial u^i} \in \mathrm{T}_{\mathcal{C}(t)}\mathcal{M}$. Then for $\boldsymbol{X} = x^i \frac{\partial}{\partial u^i}$ we have (by Remark 3.16)

$$\nabla_T \boldsymbol{X} \overset{\text{linear}}{=} T x^i \frac{\partial}{\partial u^i} + x^i \nabla_T \frac{\partial}{\partial u^i} = \frac{\partial x^i}{\partial u^j}\frac{\partial u^j}{\partial t}\frac{\partial}{\partial u^i} + x^i \underbrace{\Gamma^k_{ij}\frac{\partial u^j}{\partial t}\frac{\partial}{\partial u^i}}_{\text{in coordinates}}. \tag{3.25}$$

Definition 3.17 (Parallel vector field)**.** A vector field $\boldsymbol{X}$ is called *parallel* along the curve $\mathcal{C}$, if $\nabla_T \boldsymbol{X} = 0$ for all $t \in [a, b]$.

Note that the definition of a parallel vector field to a given curve *does not* depend on parametrization of the curve. Moreover, from (3.25) follows that for a given curve $\mathcal{C}$ (with parametrization by $u^i(t)$) a parallel vector field solves (as $\frac{\partial x^i}{\partial u^j}\frac{\partial u^j}{\partial t} = \frac{\mathrm{d}x^i}{\mathrm{d}t}$)

$$\nabla_T \boldsymbol{X} = \left(\frac{\mathrm{d}x^k}{\mathrm{d}t} + x^i \Gamma^k_{ij}\frac{\partial u^j}{\partial t}\right)\frac{\partial}{\partial u^k} = 0, \tag{3.26}$$

which is a linear system of ODEs for the coordinate functions $x^k = x^k(t)$ of the vector field $\boldsymbol{X}$ (along the curve $\mathcal{C}$). Classical results from the theory of linear ODE systems imply that for a fixed choice of $\boldsymbol{X}(a) \in \mathrm{T}_{\gamma(a)}\mathcal{M}$ we have a unique solution, and that linearly independent choice of $\boldsymbol{X}(a)$ yield linearly independent solutions for all t. Hence, $\boldsymbol{X}(a) \in \mathrm{T}_{\gamma(a)}\mathcal{M} \mapsto \boldsymbol{X}(b) \in \mathrm{T}_{\gamma(b)}\mathcal{M}$ is an **isomorphism** of tangent spaces!

In other words, as soon as have defined ∇_ξ (which *depends on the curve*), we can connect tangent spaces by these isomorphism. In fact, this is where the name "connection" comes from.

Remark 3.17. Let us repeat that the connection isomorphism depends on the curve $\mathcal{C}$, however not on its parametrization.

A special, but very important, case of a parallel vector field to a given curve arises when the curve is parallel to itself! These curves are called geodesics, and serve as generalizations of straight lines to a "curved" manifold $\mathcal{M}$. We say that geodesic curves *parallely transport their own tangents*.

Definition 3.18 (Geodesic curve)**.** A curve $\mathcal{C}$ is called *geodesic* if its tangent vector $T(t)$ satisfies $\nabla_T T = 0$.

A curve $\mathcal{C}$ induces a vector field (defined only along $\mathcal{C}$) via its tangent $T(t)$, i.e., $x^k = \frac{\partial u^k}{\partial t} = \dot{u}^k$. Then the linear ODE system then give(3.26) becomes

$$\ddot{u}^k + \Gamma^k_{ij}\dot{u}^i\dot{u}^j = 0, \tag{3.27}$$

which is, however, non-linear 2nd order system of ODEs.

Let, for simplicity, $a = 0$ and $b = 1$. Choosing initial conditions $P = \gamma(0)$ (i.e., also $u^i(0)$ give P) and $T(0) = \boldsymbol{\xi} \in \mathrm{T}_P\mathcal{M}$ suffices to provide a unique solution $of\, u^i(t)$, thus to unique prescribe $\gamma_{\boldsymbol{\xi}}(t)$.

In total, this gives an **exponential map** $\exp_p : \mathrm{T}_P\mathcal{M} \to \mathcal{M}$ with $\boldsymbol{\xi} \mapsto \gamma_{\boldsymbol{\xi}}(1)$. Intuitively, the exponential map $\exp_p$ gives us points connected to P by straight lines in $\mathcal{M}M$. Note that as $\mathrm{T}_P\mathcal{M}$ is a vector space, straight lines take the usual notion.

3.5.1 Riemannian Manifolds

Assume from now on that we have a Riemannian metric on $\mathcal{M}$, i.e., a positive definite bilinear form $g : \mathrm{T}_P\mathcal{M} \times \mathrm{T}_P\mathcal{M} \to \mathbb{R}$ on all tangent spaces, and set $g(\boldsymbol{X}, \boldsymbol{Y}) =: \langle \boldsymbol{X}, \boldsymbol{Y} \rangle$ as the scalar product.

Definition 3.19 (Levi-Civita connection)**.** A connection ∇ is called *Levi-Civita connection*, if it satisfies for all vector fields $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ the following properties:

1. $\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} = [\boldsymbol{X}, \boldsymbol{Y}]$;
2. $\boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \nabla_{\boldsymbol{Z}}\boldsymbol{X}, \boldsymbol{Y} \rangle + \langle \boldsymbol{X}, \nabla_{\boldsymbol{Z}}\boldsymbol{Y} \rangle$.

Proposition 3.6. *A Levi-Civita connection ∇ is given uniquely by the required properties.*

Proof. Firstly, notice that giving $\langle \nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z} \rangle$ for any $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ uniquely determines $\nabla_{\boldsymbol{X}}\boldsymbol{Y}$. Thus

$$\begin{aligned}
\langle \nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z} \rangle &\overset{2.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \underbrace{\langle \boldsymbol{Y}, \nabla_{\boldsymbol{X}}\boldsymbol{Z} \rangle} \overset{1.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \underbrace{\langle \boldsymbol{Y}, \nabla_{\boldsymbol{Z}}\boldsymbol{X} \rangle} - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle \\
&\overset{2.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle + \underbrace{\langle \boldsymbol{X}, \nabla_{\boldsymbol{Z}}\boldsymbol{Y} \rangle} - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle \\
&\overset{1.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle + \underbrace{\langle \boldsymbol{X}, \nabla_{\boldsymbol{Y}}\boldsymbol{Z} \rangle} - \langle \boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}] \rangle - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle \\
&\overset{2.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle + \boldsymbol{Y}\langle \boldsymbol{X}, \boldsymbol{Z} \rangle - \underbrace{\langle \nabla_{\boldsymbol{Y}}\boldsymbol{X}, \boldsymbol{Z} \rangle} - \langle \boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}] \rangle - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle \\
&\overset{1.}{=} \boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle - \boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle + \boldsymbol{Y}\langle \boldsymbol{X}, \boldsymbol{Z} \rangle - \langle \nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z} \rangle + \langle \boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}] \rangle - \langle \boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}] \rangle - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle,
\end{aligned}$$

hence

$$\langle \nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z} \rangle = \frac{1}{2}\left(\boldsymbol{X}\langle \boldsymbol{Y}, \boldsymbol{Z} \rangle + \boldsymbol{Y}\langle \boldsymbol{X}, \boldsymbol{Z} \rangle - \boldsymbol{Z}\langle \boldsymbol{X}, \boldsymbol{Y} \rangle - \langle \boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}] \rangle - \langle \boldsymbol{Y}, [\boldsymbol{X}, \boldsymbol{Z}] \rangle + \langle \boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}] \rangle\right). \tag{3.28}$$

As the right-hand side of said identity does *not* depend on ∇, we have successfully shown that a Levi-Civita connection is given uniquely. In other words, if we want the connection ∇ to be Levi-Civita, we must necessarily use (3.28). $\square$

Let us now use (3.28) to compute the Christoffel symbols Γ^k_{ij}. Consider the following (chosen) setting

$$\boldsymbol{X} = \frac{\partial}{\partial u^i},\ \boldsymbol{Y} = \frac{\partial}{\partial u^j},\ \boldsymbol{Z} = \frac{\partial}{\partial u^l} \implies \nabla_{\boldsymbol{X}}\boldsymbol{Y} \triangleq \Gamma^k_{ij}\frac{\partial}{\partial u^k}.$$

Then by Remark 3.9 we have $[\boldsymbol{X},\boldsymbol{Y}] = [\boldsymbol{Y},\boldsymbol{Z}] = [\boldsymbol{Z},\boldsymbol{X}] = 0$ and

$$\langle \nabla_{\boldsymbol{X}}\boldsymbol{Y}, \boldsymbol{Z}\rangle = \Gamma^k_{ij}g_{kl} \overset{\text{by above}}{=} \frac{1}{2}\left(\frac{\partial}{\partial u^i}g_{jl} + \frac{\partial}{\partial u^j}g_{il} - \frac{\partial}{\partial u^l}g_{ij}\right),$$

thus

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial}{\partial u^i}g_{jl} + \frac{\partial}{\partial u^j}g_{il} - \frac{\partial}{\partial u^l}g_{ij}\right). \tag{3.29}$$

Hence also $\Gamma^k_{ij} = \Gamma^k_{ji}$ as a consequence of property 1. of Definition 3.19. Unless noted otherwise, we will assume we are working with the Levi-Civita connection (induced by the chosen scalar product).

Previously, we have shown that from connections we get the (notion of) parallel transport 3.17 of vector fields along curves. Moreover, we have discussed how vector fields parallel to curves provide **isomorphism** of tangent spaces at points along this curve. As we are now in the Riemannian manifold setting (i.e., we have a scalar product), we strengthen this result to **isometries**.

Proposition 3.7. *Parallel transport along curves induces* isometries *on tangent spaces.*

Proof. Let C be a curve given by a parametrization γ and $T = \dot{\gamma}$ is tangent vector. Let $\boldsymbol{X}$ be a vector field parallel to C, i.e., $\nabla_T\boldsymbol{X} = 0$. By the 2nd property of Definition 3.19 if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are parallel along C, then

$$T\langle \boldsymbol{X}, \boldsymbol{Y}\rangle = \langle \nabla_T\boldsymbol{X}, \boldsymbol{Y}\rangle = \langle \boldsymbol{X}, \nabla_T\boldsymbol{Y}\rangle = 0$$

implies $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle$ is constant along C (as it has zero tangent-directional derivative at all points of the curve). Hence the isomorphism also preserves the Riemannian metric, i.e., it is actually an *isometry*. □

Remark 3.18. Recall Definition 3.18, which says that a curve C given by γ is called a *geodesic*, if $\nabla_T T = 0$, where T is its tangent vector. This, by Proposition 3.7, implies that $\|T\|$ is constant along C. Without a significant restriction we can require $\|T\| = 1$.

So far, we have only mentioned geodesics in the sense that they parallely transport its own tangent. However, it turns out they also provide, in a certain sense, shortest paths.

Definition 3.20 (Arclength)**.** Let C be a curve given by a parametrization $\gamma : u^i = u^i(t)$ for $t \in [a, b]$. We define the *length of the tangent* as[14]

$$\mathrm{d}s^2 = g_{ij}\mathrm{d}u^i\mathrm{d}u^j = g_{ij}\dot{u}^i\dot{u}^j\,\mathrm{d}t^2,$$

which, in turn, prescribes the *arclength* of C as

$$\ell(C) = \int_a^b \mathrm{d}s = \int_a^b \sqrt{g_{ij}\dot{u}^i\dot{u}^j}\,\mathrm{d}t.$$

[14]Keep in mind that $\mathrm{d}s^2$ is **not** a form!

Our goal will now be to find curves C that minimize $\ell(C)$ for given end-point of the curve, see Figure 3.6.

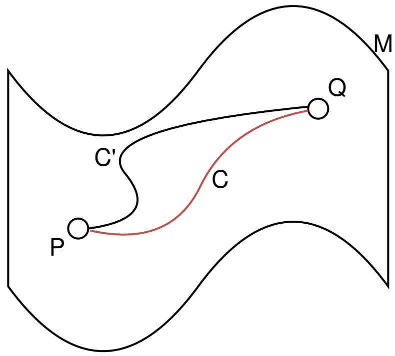

Figure 3.6: Illustration of a geodesic curve C and a non-geodesic C' between points P and Q

Let us define a displaced curve C_ε with parametrization $\gamma_\varepsilon(t)$ by $u^i = u^i(t) + \varepsilon h^i(t)$ with $h^i(a) = h^i(b) = 0$ for $i \in [n]$. We will seek curves such that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\ell(C_\varepsilon)\bigg|_{\varepsilon=0} = 0,$$

i.e., they are stationary points with respect to displacement. Now by Definition 3.20

$$\ell(C_\varepsilon) = \int_a^b \sqrt{g_{ij}(\dot{u}^i + \varepsilon\dot{h}^i)(\dot{u}^j + \varepsilon\dot{h}^j)}\,\mathrm{d}t,$$

where importantly g_{ij} also depend on ε and h^i (as we move u^i by the displacements, and thus change the tangent space, which carries g_{ij}), and we shall assume $g_{ij}\dot{u}^i\dot{u}^j = 1$ (which is just a change of variables). By viewing the derivative as a linearization, we get (by our assumption)

$$\begin{aligned}\sqrt{g_{ij}(\dot{u}^i + \varepsilon\dot{h}^i)(\dot{u}^j + \varepsilon\dot{h}^j)} &= \sqrt{1 + \varepsilon\left(2g_{ij}\dot{u}^i\dot{h}^j + \frac{\partial}{\partial u^l}g_{ij}h^l\dot{u}^i\dot{u}^j\right) + \text{h.o.t.}} \\ &= 1 + \frac{\varepsilon}{2}\left(2g_{ij}\dot{u}^i\dot{h}^j + \frac{\partial}{\partial u^l}g_{ij}h^l\dot{u}^i\dot{u}^j\right) + \text{h.o.t.}\end{aligned}$$

where the second equality comes from linearization of $\sqrt{x}$ to $\frac{x}{2}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\ell(C_\varepsilon)\bigg|_{\varepsilon=0} = \frac{1}{2}\int_a^b \left(2g_{ij}\dot{u}^i\dot{h}^j + \frac{\partial}{\partial u^l}g_{ij}h^l\dot{u}^i\dot{u}^j\right)\mathrm{d}t \overset{!}{=} 0$$

for all choices of h^i such that the boundary conditions are satisfied. Notice that the first term of the integrand depends on $\dot{h}^j(t)$, whereas the second term only on $h^l(t)$, which suggest integration by parts. Indeed

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\ell(C_\varepsilon) &= \frac{1}{2}\Bigg(\underbrace{2g_{il}\dot{u}^i h^l\Big|_{t=a}^b}_{h^l(a)=h^l(b)=0} - \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}(2g_{il}\dot{u}^i)h^l\,\mathrm{d}t + \int_a^b \frac{\partial}{\partial u^l}g_{ij}h^l\dot{u}^i\dot{u}^j\,\mathrm{d}t\Bigg) \\ &= \frac{1}{2}\int_a^b\left(-2\frac{\partial u^j}{\partial g}{}_{il}\,\dot{u}^i\dot{u}^j - 2g_{il}\ddot{u}^i + \frac{\partial}{\partial u^l}g_{ij}\dot{u}^i\dot{u}^j\right)h^l\,\mathrm{d}t \overset{!}{=} 0.\end{aligned}$$

Now for this to vanish for all choices of h^l, the term in the bracket must vanish identically. Interestingly, this is where the theory of distributions 2 comes from, as the h^l plays a role of the test functions and want the terms in brackets to be an identically zero distributions. In total, we now have

$$g_{il}\ddot{u}^i + \frac{1}{2}\frac{\partial}{\partial u^j}g_{il}\dot{u}^i\dot{u}^j + \frac{1}{2}\frac{\partial}{\partial u^i}g_{jl}\dot{u}^i\dot{u}^i - \frac{1}{2}\frac{\partial}{\partial u^l}g_{ij}\dot{u}^i\dot{u}^l = 0,$$

which we can multiply by the inverse matrix g^{lk} to isolate the first term, yielding

$$\ddot{u}^k + \underbrace{\frac{1}{2}g^{lk}\left(\frac{\partial}{\partial u^j}g_{il} + \frac{\partial}{\partial u^i}g_{jl} - \frac{\partial}{\partial u^l}g_{ij}\right)}_{\triangleq\Gamma^k_{ij}}\dot{u}^i\dot{u}^j = 0.$$

This finally results in the differential equation

$$\ddot{u}^k + \Gamma^k_{ij}\dot{u}^i\dot{u}^j = 0, \tag{3.30}$$

which by (3.27) is equivalent with $\nabla_T T = 0$. Under our assumptions on g_{ij} (and, by extension, Γ^k_{ij} via the Levi-Civita connection) for any initial data[15] $u^i(0) = P^i$ and $\dot{u}^i = \xi^i$ there exists and interval $(\underbrace{t_{\text{min}}}_{<0}, \underbrace{t_{\text{max}}}_{>0})$ such that the solution $u^i(t)$ of (3.30) exists on this interval. For $\boldsymbol{\xi} \in \mathrm{T}_P\mathcal{M}$ define the function $\gamma_{\boldsymbol{\xi}}(t)$ as the unique solution satisfying $\gamma_{\boldsymbol{\xi}}(0) = P$ and $\dot{\gamma}_{\boldsymbol{\xi}}(0) = \boldsymbol{\xi}$, then $\gamma_{t\boldsymbol{\xi}}(s) = \gamma_{\boldsymbol{\xi}}(ts)$ (when everything exists).

Moreover, these results can be used to define an **exponential map** $\exp_P : U \to \mathcal{M}$, where U is an open neighborhood of $\mathbf{0} \in \mathrm{T}_P\mathcal{M}$, by $\boldsymbol{\xi} \mapsto \gamma_{\boldsymbol{\xi}}(1)$, see Figure 3.7 a).

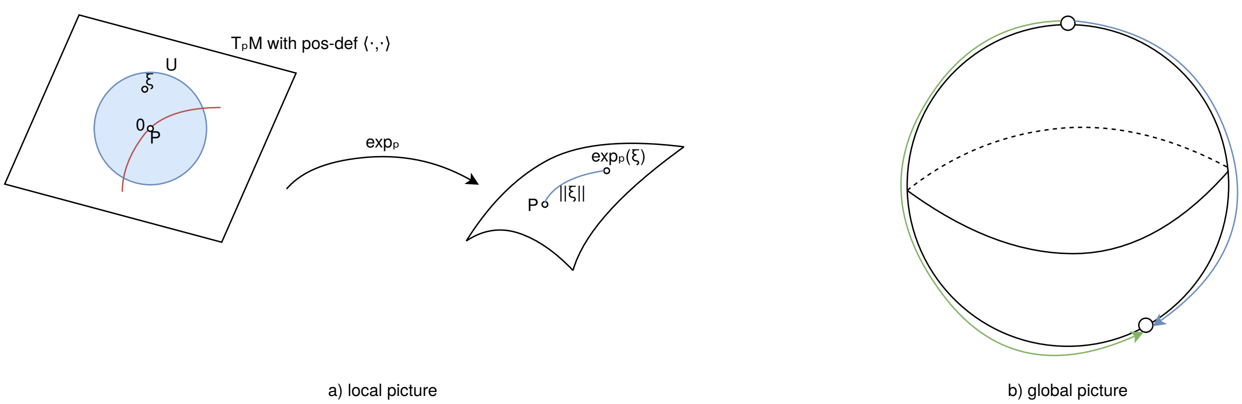

Figure 3.7: Illustration of the exponential map $\exp_P$ in both local and global perspectives

Importantly, while geodesics themselves define straight lines in a global manner, interpreting them as shortest path is necessarily possible only via a local viewpoint (as should be hinted by the fact that we only required them to be *stationary points* with respect to displacements, not global minimizers). If, for example, we take Earth and move along a meridian from the north pole to the south pole and even a bit further (i.e., we "overshoot" the south pole), then such geodesic is surely

[15] Note that we can move from $[a, b]$ to a time interval starting at 0 simply by scaling the time.

longer than if we went from the north pole the other way around. For illustration see Figure 3.7 b), where both curves plotted are geodesics. Furthermore, we can use geodesics to define a metric space structure on $\mathcal{M}$.

Proposition 3.8. *Take a curve $\omega : [0,b] \to \mathcal{M}$ such that there exist $0 = t_0 < t_1 < \cdots < t_n = b$ with $\omega|_{(t_i,t_{i+1})} \in C^\infty$ satisfying $\nabla_{\dot\omega}\dot\omega = 0$. We call such an ω a* ***piecewise geodesic****. Using this notion, we define a metric on $\mathcal{M}$ as*

$$d(P,Q) := \inf\{\ell(\omega) \mid \omega(0) = P \wedge \omega(b) = Q\}. \tag{3.31}$$

Proof. The function in (3.31) is surely a metric, as the following properties hold:

- $d(P,Q) = d(Q,P)$, which follows from parametrizing the minimizing piecewise geodesic in the opposite dimension;
- $d(P,Q) \geq 0$;
- $d(P,R) \leq d(P,Q) + d(Q,R)$, where as we fix Q we effectively take 2 infima over smaller sets (thus both of them can only be larger).

Lastly, we need to show that $d(P,Q) = 0 \implies P = Q$. To this end, recall that g (the bilinear form underlying our scalar product on $\mathrm{T}_P\mathcal{M}$) is positive definite, thus for a compact subset K of U we have that $\|\boldsymbol{\xi}\|_{\mathrm{T}_P\mathcal{M}} \geq \lambda \left\|\mathrm{d}\varphi_P(\boldsymbol{\xi})\right\|_{\mathbb{R}^n}$ for the corresponding chart map φ. Let now $\omega : [0,b] \to \mathcal{M}$ be any piecewise C^1 curve (namely any piecewise geodesic) staying in U. Its arclength is then by Definition 3.20

$$\ell(\omega) = \int_0^b \sqrt{g_{ij}\dot u^i \dot u^j}\,\mathrm{d}t = \int_0^b \left\|\dot\omega(t)\right\|_{\mathrm{T}_{\omega(t)}\mathcal{M}} \mathrm{d}t \geq \lambda \int_0^b \left\|\mathrm{d}\varphi_{\omega(t)}(\dot\omega(t))\right\|_{\mathbb{R}^n} \mathrm{d}t = \lambda \int_0^b \left\|\frac{\mathrm{d}}{\mathrm{d}t}(\varphi\circ\omega)(t)\right\|_{\mathbb{R}^n} \mathrm{d}t,$$

and hence we're in the usual Euclidean setting. Here, we directly get

$$\ell(\omega) \geq \lambda \int_0^b \left\|\frac{\mathrm{d}}{\mathrm{d}t}(\varphi\circ\omega)(t)\right\|_{\mathbb{R}^n} \mathrm{d}t \geq \lambda \left\|\varphi(\omega(b)) - \varphi(\omega(0))\right\|_{\mathbb{R}^n} = \lambda \left\|\varphi(Q) - \varphi(P)\right\|_{\mathbb{R}^n},$$

hence the desired implication. □

Definition 3.21 (Geodesically complete manifold)**.** We call a manifold $\mathcal{M}$ *geodesically complete*, if $\exp_P$ exists for all $\boldsymbol{\xi} \in \mathrm{T}_P\mathcal{M}$ and all $P \in \mathcal{M}$.

Remark 3.19. In a geodesically complete manifold, we can view the exponential map exp as a mapping $\mathcal{M} \times \mathrm{T}\mathcal{M} \to \mathcal{M}$ such that $(P, \boldsymbol{\xi}) \mapsto \exp_P(\boldsymbol{\xi})$.

Before we introduce (and prove) the two Rinov-Hopf theorems concluding this course, we shall state a supplementary result, which provides a local description of the metric using geodesics.

Proposition 3.9. *Let $\mathcal{M}$ be a Riemannian manifold. Then for every $P \in \mathcal{M}$ there exists $\varepsilon(P) > 0$ and a neighborhood U of P such that:*

1. *any two points in U are connected by a geodesic of length $< \varepsilon(P)$, and*

2. *for all* $Q \in U$ *we have that* $\exp_Q\big|_{B(\mathbf{0},\varepsilon(P))}$ *is a diffeomorphism.*

Theorem 3.4 (Rinov-Hopf on minimal geodesics)**.** *Let* $\mathcal{M}$ *be a* connected *and* geodesically complete *manifold. Then any pair of points on* $\mathcal{M}$ *can be connected by a minimal geodesic*[16].

Let us note that this minimal geodesic is not necessarily unique (for example, there is no unique minimal geodesic connection the north and the south pole of Earth).

Proof. Let $P, Q \in \mathcal{M}$, $P \neq Q$ and $\delta = d(P,Q) = \inf_{\omega \in \mathrm{Paths}(P,Q)} \ell(\omega)$, where $\mathrm{Paths}(P,Q)$ denotes the set all paths connecting P and Q. By Proposition 3.9 take $\varepsilon(P)$ and $0 < \delta_0 < \min(\varepsilon, \delta)$. Further consider $S(P,r) = \{S \in \mathcal{M} \mid d(S,P) = r\}$ a sphere of radius r (in the geometry of $\mathcal{M}$). For illustration see Figure 3.8.

Take now $P_0 \in S(P,\delta_0)$ such that $d(P_0,Q) = d(S(P,\delta_0),Q)$, i.e., the element of the sphere $S(P,\delta_0)$ that is closest to Q. By the definition of ε (Proposition 3.9 2.) we can take $\boldsymbol{\xi} = \frac{1}{\delta_0}\exp_P^{-1}(P_0) \in \mathrm{T}_P\mathcal{M}$, which we can do as $\exp_P$ is diffeomorphic ε-close to P, then $\|\boldsymbol{\xi}\| = 1$. We will show that $d(\gamma_{\boldsymbol{\xi}}(t), Q) = \delta - t$ for all $0 \le t \le \delta$, which would prove the theorem. Note that we already defined $\gamma_{\boldsymbol{\xi}}(t)$ to be distance t from P.

We shall first prove the statement for $t = \delta_0$. Then by Proposition 3.8

$$\delta = d(P,Q) \le \underbrace{d(P,P_0)}_{\delta_0} + d(P_0,Q) = \delta_0 + d(P_0,Q).$$

Let $\omega : [0,1] \to \mathcal{M}$ be any path connecting P and Q with $\omega(0) = P$ and $\omega(1) = Q$. Then there exists α such that $\omega(\alpha) \in S(P,\delta_0)$ and

$$\ell(\omega) = \ell(\omega[0,\alpha]) + \ell(\omega[\alpha,1]) \ge \delta_0 + d(\omega(\alpha),Q) \ge \delta_0 + d(P_0,Q) \ge \delta,$$

as P_0 was defined as the closest element of $S(P_0,\delta_0)$ to Q. Now, recalling $P_0 = \gamma_{\boldsymbol{\xi}}(\delta_0)$, we get

$$\delta \triangleq \inf_{\omega} \ell(\omega) \ge \delta_0 + d(P_0,Q) \ge \delta \implies d(\gamma_{\boldsymbol{\xi}}(\delta_0),Q) = \delta - \delta_0.$$

Moreover, we can repeat this argument for $t < \delta_0$ by cutting the path again, i.e., it yields that $d(\gamma_{\boldsymbol{\xi}}(t),Q) = \delta - t$ for $t \in [0,\delta_0]$ (such that it holds on the closed interval by continuity).

Assume now there is a maximal $\delta_1 = \max\left\{t \in [0,\delta] \mid d(\gamma_{\boldsymbol{\xi}}(t),Q) = \delta - t\right\} < \delta$ such that the relation holds. We shall show this leads to a contradiction. If $\delta_1 < \delta$ then there exist $P_1 = \gamma_{\boldsymbol{\xi}}(\delta_1)$ and take $0 < \delta_2 < \min(\delta - \delta_1, \varepsilon(P_1))$. Analogously to P_0, take $P_2 \in S(P_1,\delta_2)$ such that $d(P_2,Q) = d(S(P_1,\delta_2),Q)$. We can now repeat the same idea as above for P_1, P_2 and Q to produce

$$\delta_2 + d(P_2,Q) = \delta - \delta_1 \implies d(P_2,Q) = \delta - (\delta_1 + \delta_2), \tag{3.32}$$

thus we only need to make sure nothing broke at P_1 (i.e., that we are not using a piecewise geodesic). As $\delta_2 < \varepsilon(P_1)$, where $\exp_{P_1}$ is diffeomorphic, P_1 and P_2 are connected by a unique geodesic, which we can view, by geodesic completeness of $\mathcal{M}$, as the extension of $\gamma_{\boldsymbol{\xi}}$ (this extension together with $\gamma_{\boldsymbol{\xi}}[0,\delta_1]$ is still a geodesic). This in combination with (3.32) contradicts the maximality of δ_1. $\square$

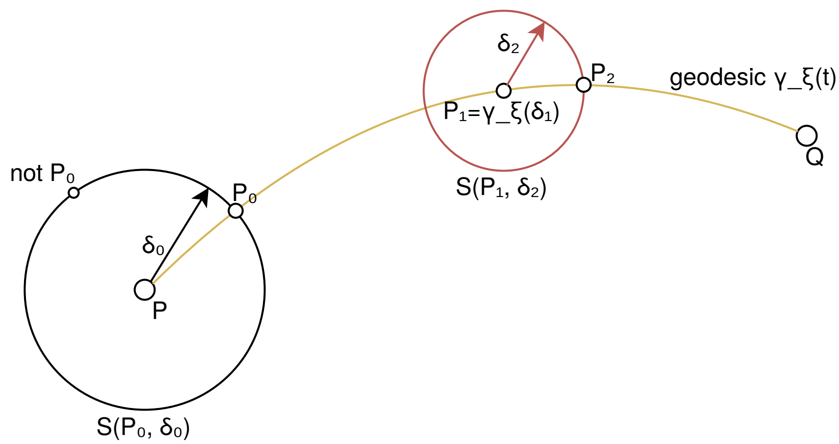

Figure 3.8: Illustration of a minimal geodesic γ_ξ and the construction of the proof of Theorem 3.4

Theorem 3.5 (Rinov-Hopf on geodesical completeness)**.** *If $\mathcal{M}$ is a* complete *metric space, then $\mathcal{M}$ is* geodesically complete.

This theorem can be re-phrased such that the only way for geodesics to end is to "fall off flat earth", i.e., to run outside their domain.

Proof. Let $\mathcal{M}$ be a complete metric space. Assume there exists $P \in \mathcal{M}$, $\boldsymbol{\xi} \in \mathrm{T}_P\mathcal{M}$ such that $\gamma_\xi(t)$ exists for all $\alpha < t < \beta$, where $\alpha < 0$ and $\beta > 0$. Without loss of generality, assume $\alpha > -\infty$. Take t_n a decreasing sequence to α, then for $m > n$

$$d(\gamma_\xi(t_n), \gamma_\xi(t_m)) \leq \ell\big(\gamma_\xi([t_n, t_M])\big) = \|\boldsymbol{\xi}\| \cdot |t_m - t_n|$$

by Definition 3.20 (and that $u^i = \xi^i$). Hence $\gamma_\xi(t_n)$ is a Cauchy sequence, which, by completeness of $\mathcal{M}$, attains a limit $Q = \lim_{n\to\infty} \gamma_\xi(t_n) \in \mathcal{M}$. As $\dot{\gamma}_\xi(t_n) \in \mathrm{T}_{\gamma_\xi(t_n)}$ then by Proposition 3.7 $\left\|\dot{\gamma}_\xi(t_n)\right\| = \|\boldsymbol{\xi}\|$ (though one should be careful here, as both norms are with respect to different tangent spaces). Along with continuity of the norm and this isometry, it implies that in the tangent bundle $\mathrm{T}\mathcal{M}$ we are in a compact subset. Hence there is a subsequence τ_n such that $\dot{\gamma}_\xi(t_n) \to \boldsymbol{\eta}$. Now consider the geodesic γ_η emanating from Q. Then these geodesics agree, i.e., $\gamma_\eta(t-\alpha) = \gamma_\xi(t)$, and as γ_η can be continued, it contradicts the choice of α. $\square$

[16]In other words, we do not need the piecewise geodesics to attain the minimum in the metric distance between two points.

4 Relevant Results

Theorem 4.1 (Monotone convergence [13]). *Let (S, Σ, μ) be a measure space and $X \in \Sigma$. If (f_n) is a sequence of non-negative measurable functions on X such that*

$$0 \leq f_1(x) \leq f_2(x) \leq \dots$$

*for all $x \in X$, i.e., (f_n) is **monotone increasing**, then the point-wise supremum $f := \sup_n f_n$ is measurable and*

$$\int_X f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \sup_n \int_X f_n \, \mathrm{d}\mu.$$

Theorem 4.2 (Dominated convergence [14]). *Let (f_n) be a sequence of complex-valued* measurable *functions on a given measure space (S, Σ, μ). Suppose that $f_n \to f$, i.e., f_n converges point-wise to f,*

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for every $x \in S$. Assume that f_n is dominated by some integrable *function g (called* majorant*) in the sense*

$$|f_n(x)| \leq g(x)$$

for all points $x \in S$ and indices n. Then f_n, f are integrable *and*

$$\lim_{n\to\infty} \int_S f_n \, \mathrm{d}\mu = \int_S \lim_{n\to\infty} f_n \, \mathrm{d}\mu = \int_S f \, \mathrm{d}\mu.$$

Proposition 4.1 (Jensen's inequality [15]). *Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. Let $f : \Omega \to \mathbb{R}$ be a μ-measurable function and $\varphi : \mathbb{R} \to \mathbb{R}$ **convex**. Then*

$$\varphi\left(\int_\Omega f \, \mathrm{d}\mu\right) \leq \int_\Omega \varphi \circ f \, \mathrm{d}\mu,$$

and, in particular,

$$\int_\Omega f(x) \, \mathrm{d}\mu(x) \leq \left(\int_\Omega |f(x)|^p \, \mathrm{d}\mu(x)\right)^{\frac{1}{p}}.$$

Lemma 4.1 (Fatou's [16]). *Given a measure space $(\Omega, \mathcal{A}, \mu)$ and a set $X \in \mathcal{A}$, let (f_n) be a sequence of μ-measurable **non-negative** functions on X. Define $f(x) = \liminf_{n\to\infty} f_n(x)$ for every $x \in X$. Then f is μ-measurable, and*

$$\int_X f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu.$$

Theorem 4.3 (Riesz-Fischer [17]). *For* $1 \leq p < \infty$, $\mathcal{L}^p(E)$ *is a Banach space*[1]. *Furthermore, if* $f_n \xrightarrow[\mathcal{L}^p]{} f$, *then* $(f_n)_n$ *has a subsequence that converges to* f *point-wise almost everywhere on* E.

Theorem 4.4 (Cauchy-Schwarz inequality [18]). *For all vectors* $\boldsymbol{u}, \boldsymbol{v}$ *of an* inner product space, *the following holds*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|u\| \, \|v\| \, .$$

Theorem 4.5 (Riesz representation [19]). *Let* H *be a Hilbert space*[2] *whose inner product* $\langle x, y \rangle$ *is linear in its first argument and* anti-linear *in its second argument. For* ***every continuous linear functional*** $\varphi \in H^*$, *there exists a unique vector* $f_\varphi \in H$, *called the* Riesz representation of φ, *such that*

$$\varphi(x) = \left\langle x, f_\varphi \right\rangle \quad \forall x \in H.$$

Importantly, for complex *HIlbert spaces,* f_φ *is always located in the* anti-linear *coordinate of the inner product.*

Definition 4.1 (Gamma function). We define a factorial extension to complex numbers Γ as

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t} \, \mathrm{d}t,$$

for $\mathfrak{Re}\, z > 0$. Moreover, it holds that

$$\Gamma(z+1) = z\Gamma(z)$$

Theorem 4.6 (Stone-Weierstrass). *Suppose* X *is a compact Hausdorff space*[3] *and* A *is subalgebra of* $C(X, \mathbb{R})$ *which contains a non-zero constant function. Then* A *is dense in* $C(X, \mathbb{R})$ *if and only if it separates points.*

Theorem 4.7 (Liouville's). *Bounded entire function is constant.*

Theorem 4.8 (Morera's). *Let* f *be continuous on an open subset* $D \subseteq \mathbb{C}$ *such that* $\oint_\gamma f(z) \, \mathrm{d}z = 0$ *for every closed piecewise-*C^1 *curve* γ *in* D. *Then* f *is holomorphic on* D.

Theorem 4.9 (Cauchy's Integral Theorem). *Let* $D \subseteq \mathbb{C}$ *be simply connected open subset and* f *is holomorphic in* D. *Then for every smoothed closed curve* γ *in* D *(i.e.,* $[\gamma] \subseteq D$*) holds* $\oint_\gamma f(z) \, \mathrm{d}z = 0$.

Theorem 4.10 (Maximum principle). *Let* f *be holomorphic on bounded connected open subset* $D \subset C$ *and continuous on* $\overline{D}$, *i.e.,* $f \in \mathcal{H}(D) \cap C(\overline{D})$. *If* $\left|f(z)\right| \leq M$ *for all* $z \in \partial D$ *and some* $M > 0$, *then* $\left|f(z)\right| \leq M$ *for all* $z \in D$ *such that* $\left|f(z)\right| = M$ *occurs for* $z \in D$ *only if* f *is constant on* D.

---

[1] Banach space is a complete normed space.

[2] Hilbert space is a real (or complex) inner product space which is also a complete metric space, i.e., a special case of Banach space.

[3] Hausdorff space is a topological space where distinct points have disjoint neighborhoods

Lemma 4.2 (Zorn's)**.** *Let P be a partially ordered set that satisfies the following two properties:*

1. *P is nonempty;*
2. *Every chain in P has an upper bound in P.*

Then P has at least one maximal element.

References

[1] STEIN, E. M. (2016). *Introduction to fourier analysis on euclidean spaces (PMS-32)*. Princeton University Press, s.l.

[2] GRAFAKOS, L. (2014). *Classical fourier analysis*. Springer New York.

[3] OSGOOD, B. (2007). *The fourier transform and its applications*.

[4] LEVIN, B. (1980). *Distribution of zeros of entire functions*. American Mathematical Society, Providence, R.I.

[5] TAO, T. (2009). Hardy's uncertainty principle.

[6] PALEY, R. E. A. C. and WIENER, N. (2012). *Fourier transforms in the complex domain*. American Mathematical Society, New York.

[7] RUDIN, W. (2013). *Real and complex analysis*. McGraw-Hill, New York, NY [u.a.].

[8] DONOGHUE, W. F. (1969). *Distributions and fourier transforms*. Academic Press, New York.

[9] RUDIN, W. (2007). *Functional analysis*. McGraw-Hill, Boston, Mass. [u.a.].

[10] BREDIES, K. and LORENZ, D. (2018). *Mathematical image processing*. Springer International Publishing.

[11] LEE, J. M. (2012). *Introduction to smooth manifolds*. Springer New York.

[12] MUNKRES, J. R. (1994). *Analysis on manifolds*. Addison-Wesley, Reading, Mass. [u.a.].

[13] CONTRIBUTORS, W. (2026). Monotone convergence theorem.

[14] CONTRIBUTORS, W. (2025). Dominated convergence theorem.

[15] CONTRIBUTORS, W. (2026). Jensen's inequality.

[16] CONTRIBUTORS, W. (2025). Fatou's lemma.

[17] Meziani, H. (2022). Real analysis MAA 6616 lecture 25 convergence in lp spaces and completeness - Lecture25.

[18] Contributors, W. (2026). Cauchy–Schwarz inequality.

[19] Contributors, W. (2025). Riesz representation theorem.