Integral Transforms and Computed Tomography

Tomas Sauer

Lehrstuhl für Mathematik mit Schwerpunkt Digitale Bildverarbeitung
FORWISS
University of Passau
Innstr. 43
94032 Passau

Version 0.0
Last modifications: 29.1.2023

Chaos is found in greatest abundance whereever order is being sought. It always defeats order, because it is better organized.

T. Pratchett, *Interesting times*

When the epoch of analogue (which was to say also the richness of language, of *analogy*) was giving way to the digital era, the final victory of the numerate over the literate.

S. Rushdie, *Fury*

The most incredible thing about miracles is that they happen.

G. K. Chesterton, *The Innocence of Father Brown*

And it didn't stop being magic just because you found out how it was done.

T. Pratchett, *Wee Free Men*

Tomas Sauer
Chair for Mathematical Image Processing
University of Passau
Innstr. 43
94032 Passau

Contents

What are Integral Transforms? 1

To understand one science perfectly means having a considerable knowledge of all other sciences

(P. J. Brebner, *Christopher Quarles College Professor and Master Detective*)

As the name may suggest, an **integral transform** is a transform that involves integrals. This, of course, is no help at all. So let us try to give an intuitive collection of what we may mean with it.

1.1 General considerations

A **transform** T takes a function f from a function space $\mathscr{X} : \{f : X \to \mathbb{C}\}$ to a function Tf from another function space $\mathscr{Y} = \{g : Y \to \mathbb{C}\}$. In some cases, we will have that $X = Y$, but in many cases it will happen that $Y \neq X$. This can be for various reasons:

1. The transform Tf may provide information of f that is not so easily visible from f directly.

2. We may be interested in a function f, but all we know is a measurement $g = Tf$ where T usually models the physics of the acquisition process of the measurements. This is what is usually called an inverse problem since in order to find f we have to invert T in some way or another[1] to obtain $f = T^{-1}g$.

Usually a transform, be it integral or not, is only defined for certain functions, so a careful definition of the space $\mathscr{X}$ and sometimes also of $\mathscr{Y}$ is usually important.

A transform is called an **integral transform** if it involves integrals of functions, i.e., is of the form

$$Tf(y) = \int_{X'} F(f, x, y)dx, \qquad X' \subset X.$$

Of course, this is still very general as the function $F : \mathscr{X} \times X \times Y \to \mathbb{C}$ can be very complicated as can be set X' over which integration is performed; and if all that is too easy for you, you may even replace dx by $d\mu(x)$ where μ is some measure on X. In reality, however, many integral operators are of a simpler structure. In

[1] In many practical applications T may not even be invertible or the inversion may numerically difficult or ill-conditioned.

particular, many of them are based of a **kernel** $K : X \times Y \to \mathbb{C}$ and the integral operator is of the form

$$Tf(y) = \int_X K(x,y) f(x) dx, \qquad f \in \mathscr{X}.$$

Such transforms are **linear**, i.e., $T(\alpha f) = \alpha Tf$ and $T(f+g) = Tf + Tg$ and many of their properties are usually derived from properties of the underlying kernel K.

1.2 Some examples

But before we get lost in more of the unnecessary general stuff, let us consider classical examples of integral transforms, in particular those that play a role in what is going to follow.

Example 1.2.1 (Fourier transform). The **Fourier transform** of a function $f : \mathbb{R}^d \to \mathbb{C}^d$ is defined as

$$Tf(\xi) := \hat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x}\, dx, \qquad \xi \in \mathbb{R}^d, \tag{1.2.1}$$

where $\xi \cdot x = \xi_1 x_1 + \cdots + \xi_d x_d$ denotes the inner product between ξ and x. This transform will be a fundamental tool in many of the things we consider later which will be reason why will dedicate Chapter 2 to it.

Example 1.2.2 (Fourier series). Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ denote the **torus** where we identify any two elements of $\mathbb{R}$ if they just differ by a multiple of 2π. As a set, $\mathbb{T} \simeq [-\pi, \pi]$, but any addition and multiplication is well-defined on $\mathbb{T}$ in contrast to the interval. For a function $f : \mathbb{T}^d \to \mathbb{C}$, which has a 2π-periodic extension to $\mathbb{R}^d$, we call

$$Tf(\alpha) = \hat{f}(\alpha) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}} f(x) e^{-i\alpha \cdot x}\, dx, \qquad \alpha \in \mathbb{Z}^d,$$

the coefficients of the **Fourier series** of f. Here we have a transform that maps functions defined on $\mathbb{T}^d$ to functions defined on $\mathbb{Z}^d$, so that the domain of the function and the domain of its transform are not the same any more.

Note that both the Fourier transform and the Fourier series are integral transforms with kernels, namely,

$$K(x,\xi) = e^{-ix\cdot\xi} \qquad \text{and} \qquad K(x,\alpha) = \frac{1}{(2\pi)^d} e^{-i\alpha\cdot x},$$

respectively. Also observe that in both cases the transform is only defined for functions that fulfill certain properties.

Example 1.2.3. A one-sided relative of the Fourier transform from Example 1.2.2 is the **Laplace transform**, defined as

$$Tf(s) := \int_0^\infty f(t) e^{-st}\, dt, \qquad s \in \mathbb{C}. \tag{1.2.2}$$

While the function f is only defined on $\mathbb{R}^d$, the Laplace transform, which plays a role in theory of systems and electric circuits, has a complex parameter that is usually interpreted as amplitude and frequency.

Example 1.2.4 (Wavelet transform). A function ψ with the property that[2]

$$\int_{\mathbb{R}^d} \frac{\left|\hat{\psi}(\xi)\right|^2}{|\xi|_1} d\xi < \infty$$

is called a (admissible) wavelet and the associated **wavelet transform** is defined as

$$Tf(x,u) = \int_{\mathbb{R}^d} \psi\left(\frac{x-t}{u}\right) f(t)\, dt, \qquad x \in \mathbb{R}^d,\ u \in \mathbb{R},$$

and has the property that the number of parameters of the transform is now even greater than the number of parameters of the function itself. The wavelet transform plays an important role in time-/frequency analysis.

Example 1.2.5 (Radon transform). A transform where function and transformed function have a completely difference structure is the **Radon transform** that we will consider in 2D here. For $f : \mathbb{R}^2 \to \mathbb{R}$ and a line

$$L = \{v_0 + tv : t \in \mathbb{R}\},$$

defined by the anchor point $v_0 \in \mathbb{R}^s$ and the direction $v \in \mathbb{R}^2 \setminus \{0\}$ one defines the **line integral**

$$Tf(L) = Rf(L) = \int_{\mathbb{R}} f(v_0 + tv)\, dt. \tag{1.2.3}$$

This is not a kernel integral and makes some stronger requirements on f since the integration is only over a line, which is a set of measure zero. What this means, we will see in a moment. The **Radon transform**, introduced by Johan Radon in 1917 has a physical interpretation that makes is the foundation of computed tomography as we will see later.

The list of examples that we have seen in this section is in no way exhaustive or complete, but it already shows some of the basic concepts and questions that we will encounter in the course of this lecture.

1.3 Integration and spaces

In this section we give a short and superficial introduction to what we me mean when we talk about integrals. Those who want to know details are for example referred to (Forster, 1984; Williamson, 1962). The most common integral in Analysis and Calculus[3] courses is the **Riemann integral** based on the integration of finite step functions on finite and mostly compact intervals. I works quite well for continuous functions on compact domains but fails for nasty functions like

$$f(x) = \begin{cases} 0 & x \in \mathbb{Q}, \\ 1 & x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases}$$

[2] This is called the *admissibility condition* and implies that $\int \psi = 0$.

[3] In the German system analysis is mostly calculus, the English and American system makes a distinction that is also visible in the title of books.

for which the lower integral is zero and the upper is the interval length. Also it notoriously troubled with *infinite* domains of integration, just try to compute

$$\int_0^\infty \cos(x)dx$$

by approximating it via $\int_0^R \cos(x)dx$ letting R tend to ∞. It simply does not work and actually the function $\cos(x)$ is not integrable on $\mathbb{R}$.

The **Lebesgue integral** on the other hand, is based on the notion of **measurable sets** in $\mathbb{R}^d$ and the defines for any countable decomposition of some measurable $E \subset \mathbb{R}^d$ into disjoint measurable sets E_j, $j \in \mathbb{N}$, the upper and lower sums

$$S^*(f) = \sum_{j=1}^{\infty} \left(\sup_{x \in E_j} f(x) \right) m(E_j), \qquad S_*(f) = \sum_{j=1}^{\infty} \left(\inf_{x \in E_j} f(x) \right) m(E_j),$$

and then the upper and lower integrals

$$\int_E^{\downarrow} f(x)dx := \inf_{E_1, E_2, \ldots} S^*(f), \qquad \int_E^{\uparrow} f(x)dx := \sup_{E_1, E_2, \ldots} S_*(f),$$

which is "the smallest from above" and "the largest from below"; the always satisfy $\int_E^{\downarrow} \geq \int_E^{\uparrow}$ and the function f is called **integrable** if the two coincide. When done properly this gives a proper notion of integration on $\mathbb{R}^d$ that agrees with the Riemann integral for $f \in C_C(\mathbb{R}^d)$, i.e., for all functions with compact support.

The integral has all nice properties that we expect of it, like linearity, positivity, monotonicity and additivity, at least as long as things are finite. For infinite operations, like limits or interchange of integrals or integrals and limits, one has to be careful. An example is the following result.

Theorem 1.3.1 (Lebesgue dominated convergence). *If f_n, $n \in \mathbb{N}$, is a sequence of integrable functions which is **dominated**, i.e, there exists an integrable function g such that[4] $|f_n| \leq g$, with $f_n \to f$, then f is integrable and*

$$\int f(x)\, dx = \lim_{n \to \infty} \int f_n(x)\, dx. \tag{1.3.1}$$

The point here is "dominated". In general the limit would neither be integrable nor would (1.3.1) hold without the sequence being dominated, cf. (Gelbaum and Olmstedt, 1964). And there exist quite a few of such subtleties for Lebesgue integration. We will not dwell on them in this lecture as we want to focus on other things and sometimes we may be a little bit "generous" in the arguments and refer to the theory. If you feel uncertain, take the effort to consult the standard literature, as in the end knowledge and even understanding is not a disadvantage.

Even the point evaluation of locally integrable functions is not a trivial issue. Of course, we can write $f(x)$, but since a locally integrable function can be modified

[4] This only has to hold almost everywhere, but we have not defined what this means nor do we intend to do so.

on a set of measure zero[5], the evaluation is not well defined and $f \mapsto f(x)$ is in no way a reasonable or continuous functional. To define a reasonable **point evaluation**, we use the **ball** of radius δ around x,

$$B_\delta(x) := \left\{ y \in \mathbb{R}^d \ : \ |y - x| \leq \delta \right\}, \qquad |B_\delta(x)| = \int_{B_\delta(x)} dt$$

to define the **average**

$$\mu_\delta f(x) := \frac{1}{V_\delta} \int_{B_\delta(x)} f(t)\, dt = \frac{1}{V_\delta} \int_{B_\delta(0)} f(t+x)\, dt \tag{1.3.2}$$

at x.

Theorem 1.3.2 (Lebesgue). *If f is locally integrable, one has for almost all $x \in \mathbb{R}^d$ that*

$$f(x) = \lim_{\delta \to 0} \mu_\delta f(x). \tag{1.3.3}$$

Any point x for which (1.3.3) holds is called a **Lebesgue point** of f and the Lebesgue's theorem, Theorem 1.3.2 states that the complement of the set of Lebesgue points has measure zero. Modifying the function there in an appropriate way, gives an integrable function where all points are Lebesgue points and we can assume that (1.3.3) holds everywhere.

In this lecture we will consider some standard spaces of functions that will be of use later.

Definition 1.3.3 (Continuous functions).

1. By $C(\mathbb{R}^d)$ we define the vector space of **continuous** functions of $\mathbb{R}^d$, which are functions such that for any $x \in \mathbb{R}^d$ and any $\varepsilon > 0$ there exists $\delta > 0$ such that[6]
$$|x - y| \leq \delta \qquad \Rightarrow \qquad |f(x) - f(y)| \leq \varepsilon.$$

2. Moreover, we write $C_u(\mathbb{R}^d)$ for the **uniformly continuous** function where for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any $x \in \mathbb{R}^d$
$$|x - y| \leq \delta \qquad \Rightarrow \qquad |f(x) - f(y)| \leq \varepsilon.$$

3. By $C_{00}(\mathbb{R}^d)$ we denote all functions f from $C(\mathbb{R}^d)$ that have **compact support**, that is, there exists a *compact* set $\Omega \subset \mathbb{R}^d$ such that $f(x) = 0$ for $x \notin \Omega$.

4. An important and particularly nice class of functions is $C^\infty_{00}(\mathbb{R}^d)$, the set of a **infinitely differentiable** functions with compact support.

Even if $C^\infty_{00}(\mathbb{R}^d)$ looks like a class of functions almost too good to be useful, they actually are quite frequent and useful: they are dense in the integrable functions.

[5] Even worse, the spaces of integrable functions that we consider are only equivalence classes modulo sets of measure zero.

[6] The norm $|\cdot|$ that measures the distance in $\mathbb{R}^d$ is not relevant here as all norms on finite dimensional spaces are equivalent. If this does not make sense to you, look it up.

Theorem 1.3.4. *For any integrable f and $\varepsilon > 0$ there exists $g \in C_{00}^{\infty}(\mathbb{R}^d)$ such that*

$$\int_{\mathbb{R}^d} |f(x) - g(x)| \, dx < \varepsilon. \tag{1.3.4}$$

A fundamental role will be played by functions whose powers are integrable which form the basic spaces for integral transforms.

Definition 1.3.5 (L_p spaces).

1. For $1 \le p < \infty$ the p-norm is defined as

$$\|f\|_p := \left(\int_{\mathbb{R}^d} |f(x)|^p \, dx \right)^{1/p}, \qquad f : \mathbb{R}^d \to \mathbb{C}, \tag{1.3.5}$$

   and the L_p-space consists of all functions with finite p-norm:

$$L_p(\mathbb{R}^d) := \left\{ f : \|f\|_p < \infty \right\} \tag{1.3.6}$$

2. For $p = \infty$ we consider

$$\|f\|_\infty = \text{ess-sup} \left\{ |f(x)| : x \in \mathbb{R}^d \right\}$$

   where the **essential supremum** allows to exclude the values of f on a set of measure zero. This is a somewhat ugly space.

3. The **dual norm** to $\|\cdot\|_p$ is $\|\cdot\|_q$ where $1/p + 1/q = 1$, $1 < p < \infty$ as well as the pair $1, \infty$.

Exercise 1.3.1 Show that

$$C_{00}(\mathbb{R}^d) \subset \bigcap_{1 \le p \le \infty} L_p(\mathbb{R}^d).$$

♢

A fundamental and frequently used property is that for $1/p + 1/q = 1$ and $f \in L_p(\mathbb{R}^d)$ as well as $g \in L_p(\mathbb{R}^d)$ one has that $fg \in L_1(\mathbb{R}^d)$ and the **Hölder inequality**

$$\|fg\|_1 \le \|f\|_p \, \|g\|_q \tag{1.3.7}$$

holds. The special case $p = q = \frac{1}{2}$ is known as the **Cauchy-Schwarz inequality**.

For $1 \le p \le \infty$, the space L_p is a **Banach space**, i.e., a complete normed space. "Normed" is clear since the spaces are defined by the norms $\|\cdot\|_p$, wile **complete** means that any **Cauchy sequence** of functions in $L_p(\mathbb{R}^d)$ has a limit, i.e., any squence f_n, $n \in \mathbb{N}$, of functions with the property that for any ε there exists $n_0 \in \mathbb{N}$ such that

$$\|f_n - f_m\|_p \le \varepsilon, \qquad m, n \ge n_0,$$

converges to a limit $f \in L_p(\mathbb{R}^n)$. The point is the existence of the limit and the proof is not completely trivial, cf. (Williamson, 1962).

Exercise 1.3.2 Give an example of a normed linear space that is *no* Banach space. What is the simplest example? ♢

The Mother of All Transforms

2

If God were omnipotent and omniscient in any literal sense, he wouldn't have bothered to make the universe at all. There is no success where there is no possibility of failure, no art without the resistance of the medium.

(R. Chandler, *Playback*)

The Fourier transform is *the* transform and it covers hardcore applications as well as very deep and absctract mathematics like working on locally compact Abelian groups (Gasquet and Witomski, 1998; Katznelson, 1976; Loomis, 1953).

2.1 Definition and Basic Properties

We begin with the Fourier transform of a function, defined on absolutely integrable functions.

Definition 2.1.1 (Fourier transform). The **Fourier transform** of a function $f \in L_1(\mathbb{R}^d)$ is defined as

$$\mathscr{F} f(\xi) := \widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x)\, e^{-i\,\xi^T x}\, dx, \qquad \xi \in \mathbb{R}^d. \tag{2.1.1}$$

The Fourier transform maps functions defined on $\mathbb{R}^d$ to functions defined on $\mathbb{R}^d$, however, the two instances of $\mathbb{R}^d$ play a somewhat different role. The keyword is **dual group**, cf. (Katznelson, 1976), but accidentially $\mathbb{R}^d$ is its own dual. Here

$$x^T \xi = \sum_{j=1}^{d} x_j \xi_j$$

is the standard **inner product** on $\mathbb{R}^d$ yielding that

$$e^{-i\,\xi^T x} = e^{-i(x_1\xi_1 + \cdots + x_d\xi_d)} = \prod_{j=1}^{d} e^{-ix_j\xi}$$

Since[1] $\left|e^{-ix}\right| = 1$ for $x \in \mathbb{R}$, we also have that $\left|e^{-ix^T\xi}\right| = 1$ and therefore

$$\left| \int_{\mathbb{R}^d} f(x)\, e^{-i\,\xi^T x}\, dx \right| \le \int_{\mathbb{R}^d} \left| f(x)\, e^{-i\,\xi^T x} \right| dx = \int_{\mathbb{R}^d} |f(x)|\, dx,$$

[1] Just to make it clear: we are talking about the *complex* modulus here.

hence the Fourier transform is *well-defined* for $f \in L_1(\mathbb{R}^d)$. Historically, the *Fourier transform* is a relatively modern concept, Fourier[2] himself "only" introduced *Fourier series* as a tool to solve the heat equation. Of course, Fourier series can do much more, for example they form the foundation of musical acoustics (Benson, 2007; Helmholtz, 1885).

Remark 2.1.2. 1. The physical and technical interpretation of the univariate Fourier transform of a function, interprested as signal over time, is the contribution of a certain frequency to the signal. To do so, the *complex* value $\widehat{f}(\xi) = \left|\widehat{f}(\xi)\right| e^{i\,\theta(\xi)}$ is interpreted as **amplitude** $\left|\widehat{f}(\xi)\right|$ and **phase** $\theta(\xi)$ associated to the frequency.

2. The condition $f \in L_1$ ensures that

$$\left|\widehat{f}(\xi)\right| \leq \|f\|_1 . \tag{2.1.2}$$

Keep in mind, however, that the condition is only *sufficient* but not *necessary* for the existence of the Fourier transform.

3. Sometimes the Fourier transform is equipped with a factor $(2\pi)^{d/2}$, and we will see soon why this is the case. However, when using various sources of literature or libraries, it is important to check which factor is used there. Otherwise it can lead to quite bad errors.

4. The Fourier transfrom not only exists on $\mathbb{R}^d$ but on locally compact abelian groups using the associated Haar measure, cf. (Loomis, 1953). This is a nice theory intertwining mesure theory, algebra, topology and analysis, but not needed in our applications here.

5. An important operation on $\mathbb{R}^d$ is the **translation** operator τ_y, $y \in \mathbb{R}^d$, defined as

$$\tau_y f = f(\cdot + y) . \tag{2.1.3}$$

Moreover, any *nonsingular* matrix $A \in \mathbb{R}^{d\times d}$ defines a **dilation**

$$\sigma_A f = f(A\cdot) . \tag{2.1.4}$$

Proposition 2.1.3. *For $f \in L_1$ we have that $\hat{f} \in C_u(\mathbb{R}^d)$.*

Proof: For $\xi, \delta \in \mathbb{R}^d$, we estimate the difference

$$\begin{aligned}\left|\widehat{f}(\xi+\delta) - \widehat{f}(\xi)\right| &= \left|\int_{\mathbb{R}^d} f(x)\left(e^{-i(\xi+\delta)^T x} - e^{-i\xi^T x}\right) dx\right| \\ &\leq \int_{\mathbb{R}^d} |f(x)| \underbrace{\left|e^{-i\xi^T x}\right|}_{=1} \left|e^{-i\delta^T x} - 1\right| dx;\end{aligned}$$

[2] Jean Baptiste Fourier, 1768–1830, , French mathematician and politician, was not only member of the "Académie des Sciences", but also participated int the expedition of Napoleon Bonaparte to Egypt as a scientific advisor. Later he became governor of the Department Isère with its capital Grenoble where he supported Champollion and thus contributed to decyphering the hieroglyphs.

since $f \in L_1$ there exists for any $\varepsilon > 0$ a number $M > 0$ such that[3]

$$\int_{\mathbb{R}^d \setminus [-M,M]^d} |f(x)| \, dx < \varepsilon$$

ist. Then, using Exercise 2.1.1,

$$\begin{aligned}
&\left|\widehat{f}\,(\xi+\delta) - \widehat{f}(\xi)\right| \\
&\quad\leq \int_{[-M,M]^d} |f\,(x)| \left|e^{-i\delta^T x} - 1\right| \, dx + \int_{\mathbb{R}^d \setminus [-M,M]^d} |f\,(x)| \underbrace{\left|e^{-i\delta^T x} - 1\right|}_{\leq 2} \, dx \\
&\quad\leq \int_{[-M,M]^d} |f\,(x)| \left|e^{-i\delta^T x} - 1\right| \, dx + 2\varepsilon \\
&\quad\leq M \, \|\delta\|_1 \int_{\mathbb{R}^d} |f\,(x)| + 2\varepsilon = M \, \|\delta\|_1 \, \|f\|_1 + 2\varepsilon.
\end{aligned}$$

This expression can be made arbitrarily small *indepently of* ξ since any given ε implies an M and according to this number we can choose $\|\delta\|_1 \leq \epsilon/(M\|f\|_1)$ to obtain

$$\left|\widehat{f}\,(\xi+\delta) - \widehat{f}(\xi)\right| \leq 3\varepsilon$$

independently of ξ. □

Exercise 2.1.1 Show: For $a \in \mathbb{R}^d$ and $1 \leq p, q \leq \infty$ such that $1/p + 1/q = 1$, one has

$$\left|e^{-ia^T x} - 1\right| \leq \|a\|_p \, \|x\|_q \, .$$

♢

The next is an important operation that is not only closely related to the Fourier transform and of great value in signal processing, but can also be seen as an alternative multiplication of functions.

Definition 2.1.4 (Convolution). For $f, g \in L(\mathbb{R}^d)$ we define the **convolution**

$$f * g := \int_{\mathbb{R}^d} f(\cdot - t) \, g(t) \, dt \qquad * : L(\mathbb{R}) \times L(\mathbb{R}) \to L(\mathbb{R}), \tag{2.1.5}$$

whenever the integral on the right hand side exists.

The convolution is symmetric since a simple change of variables yields

$$f * g = \int_{\mathbb{R}^d} f(\cdot - t) \, g(t) \, dt = \int_{\mathbb{R}^d} f(t) \, g(\cdot - t) \, dt = g * f.$$

We also have that[4]

$$\begin{aligned}
\|f * g\|_1 &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-t) \, g(t) \, dt \right| dx \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-t) \, g(t)| \, dt \, dx \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-t) \, g(t)| \, dx \, dt = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)| \, |\, g(t)| \, dx \, dt = \|f\|_1 \|g\|_1,
\end{aligned}$$

[3] This is one of the "magic" properties of the Lebesgue integral

[4] The interchanging of the integrals due to Fubini's theorem is yet another Lebesgue subtlety that we are not working out in detail.

so that the convolution maps $L_1 \times L_1$ to L_1; this is the "multiplication" property mentioned before.

Exercise 2.1.2 Show that $\left(L_1(\mathbb{R}^d), *\right)$ is a **Banach algebra**, i.e., a Banach space with a multiplication operation "$*$" that satisfies the usual laws of commutativity and distributivity and is compatible with the norm of the Banach space. ◊

Next, we collect fundamental properties of the Fourier transform that we will use frequently in what follows.

Theorem 2.1.5 (Properties of the Fourier transform). *For $f \in L_1$, the following holds true:*

1. *for any $y \in \mathbb{R}^d$,*
$$\left(\tau_y f\right)^\wedge(\xi) = e^{iy^T\xi}\,\widehat{f}(\xi), \qquad \xi \in \mathbb{R}^d. \tag{2.1.6}$$
*That is, translations are turned into **phase shifts**.*

2. *for any nonsingular $A \in \mathbb{R}^{d\times d}$ one has*
$$(\sigma_A f)^\wedge(\xi) = \frac{\widehat{f}\left(A^{-T}\xi\right)}{|\det A|}, \qquad \xi \in \mathbb{R}^d. \tag{2.1.7}$$

3. *for $g \in L_1$,*
$$(f * g)^\wedge(\xi) = \widehat{f}(\xi)\,\widehat{g}(\xi), \qquad \xi \in \mathbb{R}^d. \tag{2.1.8}$$

4. *if $\frac{\partial}{\partial x_j} f \in L_1$, $j = 1, \ldots, d$, then*
$$\left(\frac{\partial}{\partial x_j} f\right)^\wedge(\xi) = i\xi_j\,\widehat{f}(\xi), \qquad j = 1, \ldots, d, \quad \xi \in \mathbb{R}^d. \tag{2.1.9}$$

5. *if $(\cdot)_j\, f \in L_1$, $j = 1, \ldots, d$, then $\widehat{f}$ is differentiable and*
$$\frac{\partial}{\partial \xi_j}\widehat{f}(\xi) = \left(-i(\cdot)_j\, f\right)^\wedge(\xi), \qquad j = 1, \ldots, d, \quad \xi \in \mathbb{R}^d. \tag{2.1.10}$$

6. *if $\widehat{f} \in L_1(\mathbb{R})$, then*
$$f(x) = \left(\widehat{f}\right)^\vee(x) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi \tag{2.1.11}$$

Definition 2.1.6. The operation
$$f \mapsto f^\vee := \frac{1}{(2\pi)^d} f^\wedge(-\cdot)$$
is called the **inverse Fourier transform**.

Proof: For 1), we compute

$$\begin{aligned}\left(\tau_y f\right)^\wedge(\xi) &= \int_{\mathbb{R}^d} f(x+y)\, e^{-i\xi^T x}\, dx = \int_{\mathbb{R}^d} f(x)\, e^{-i\xi(x-y)}\, dx = e^{iy^T\xi} \int_{\mathbb{R}^d} f(x)\, e^{-i\xi^T x}\, dx \\ &= e^{iy^T\xi}\, \widehat{f}(\xi),\end{aligned}$$

and 2) we obtain because of

$$(\sigma_A f)^\wedge(\xi) = \int_{\mathbb{R}^d} f(Ax)\, e^{-i\xi^T x}\, dx = \left|\det A^{-1}\right| \int_{\mathbb{R}^d} f(x)\, e^{-i\xi^T A^{-1}x}\, dx = \frac{\widehat{f}\left(A^{-T}\xi\right)}{|\det A|};$$

in both cases it is helpful that the transformations map $\mathbb{R}^d$ *bijectively* to itself.

The statement 3) is due to

$$\begin{aligned}(f*g)^\wedge(\xi) &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(s) g(x-s)\, ds \right) e^{-i\xi^T x}\, dx \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(s)\, e^{i\xi^T s}\, g(x-s)\, e^{i\xi^T(x-s)}\, ds dx = \widehat{f}(\xi)\, \widehat{g}(\xi),\end{aligned}$$

while for 4) we use integration by parts which is justified since $C_{00}(\mathbb{R}^d)$ is dense in $L_1(\mathbb{R}^d)$ and yields for $j = 1, \ldots, d$

$$\begin{aligned}\left(\frac{\partial}{\partial x_j} f\right)^\wedge(\xi) &= \int_{\mathbb{R}^d} \frac{\partial f}{\partial x_j}(x) e^{-i\xi^T x}\, dx = -\int_{\mathbb{R}^d} f(x) \frac{\partial}{\partial x_j} e^{-i\xi^T x}\, dx \\ &= i\xi \int_{\mathbb{R}^d} f(x) e^{-i\xi^T x}\, dx = i\xi\, \widehat{f}(\xi).\end{aligned}$$

5) is obtained by computing for $h > 0$ and $\eta \in \mathbb{R}^d$ the **difference quotient**

$$\begin{aligned}\frac{\widehat{f}(\xi + h\eta) - \widehat{f}(\xi)}{h} &= \int_{\mathbb{R}^d} f(x)\, \frac{e^{-i(\xi+h\eta)^T x} - e^{-i\xi^T x}}{h}\, dx \\ &= \int_{\mathbb{R}^d} f(x)\, e^{-i\xi^T x}\, \frac{e^{-ih\eta^T x} - 1}{h}\, dx.\end{aligned}$$

Since

$$\lim_{h\to 0} \frac{e^{-ih\eta^T x} - 1}{h} = \lim_{h\to 0} (-i\eta^T x)\, e^{-ihx} = -i\eta^T x$$

and since

$$\left(\eta^T x\right) f = \sum_{j=1}^d \eta_j\, x_j f(x)$$

the Fourier integral exists and is linear in η due to our assumption that $(\cdot)_j f \in L_1(\mathbb{R}^d)$. Therefore, $\hat{f}$ is differentiable and (2.1.10) follows by choosing $\eta = e_j$, $j = 1, \ldots, d$.

The proof of 6) requires a bit more effort and uses the **Féjer kernel**

$$F_\lambda := \lambda^d F(\lambda\cdot), \quad \lambda > 0, \qquad F(x) := \frac{1}{(2\pi)^d} \int_{[-1,1]^d} \prod_{j=1}^d \left(1 - \left|t_j\right|\right) e^{ixt}\, dt, \quad x \in \mathbb{R}^d,$$

which is the inverse Fourier transform of a hat function. These kernels have the property[5] that for any $f \in L_1\left(\mathbb{R}^d\right)$

$$\lim_{\lambda\to\infty} \|f - f * F_\lambda\|_1 = 0 \tag{2.1.12}$$

holds, see (Katznelson, 1976, S. 124–126) for the case $d = 1$ which can easily be extended by a straightforward tensor product argument. Therefore[6] some subsequence of $f * F_\lambda \to f$ converges almost everywhere for $\lambda \to \infty$, cf. (Forster, 1984, S. 96), and we get for almost all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
f * F_\lambda(x) &= \int_{\mathbb{R}^d} f(t)\, F_\lambda(x-t)\, dt \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(t) \left(\lambda \int\limits_{[-1,1]^d} \prod_{j=1}^{d} \left(1 - |\xi_j|\right)\, e^{i(x-t)^T \lambda\xi}\, d\xi \right) dt \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}} f(t) \int\limits_{[-\lambda,\lambda]^d} \prod_{j=1}^{d} \left(1 - \frac{|\xi_j|}{\lambda}\right) e^{i(x-t)^T\xi}\, d\xi\, dt \\
&= \frac{1}{(2\pi)^d} \int\limits_{[-\lambda,\lambda]^d} \prod_{j=1}^{d} \left(1 - \frac{|\xi_j|}{\lambda}\right) \underbrace{\int_{\mathbb{R}^d} f(t)\, e^{-i\xi^T t}\, dt}_{=\widehat{f}(\xi)}\, e^{ix^T\xi}\, d\xi \\
&= \underbrace{\frac{1}{(2\pi)^d} \int\limits_{[-\sqrt{\lambda},\sqrt{\lambda}]} \overbrace{\prod_{j=1}^{d} \left(1 - \frac{|\xi_j|}{\lambda}\right)}^{(1-1/\sqrt{\lambda})^d \le \cdot \le 1} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi}_{\to (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi} \\
&\quad + \underbrace{\frac{1}{(2\pi)^d} \int\limits_{\sqrt{\lambda} \le |\xi_j| \le \lambda} \prod_{j=1}^{d} \left(1 - \frac{|\xi_j|}{\lambda}\right) \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi}_{\to 0}, \\
&\to \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi,
\end{aligned}$$

[5] That we do not prove here!

[6] One more of these statements that need some effort to be verified in detail.

since

$$\left| \int\limits_{\sqrt{\lambda}\le|\xi_j|\le\lambda} \prod_{j=1}^{d} \left(1 - \frac{|\xi_j|}{\lambda}\right) \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi \right|$$

$$\le \int\limits_{\sqrt{\lambda}\le|\xi_j|\le\lambda} \underbrace{\prod_{j=1}^{d} \left|1 - \frac{|\xi_j|}{\lambda}\right|}_{\le 1} \left|\widehat{f}(\xi)\right| \underbrace{\left|e^{ix^T\xi}\right|}_{=1} d\xi \le \int\limits_{\xi\in\mathbb{R}^d\setminus\left[-\sqrt{\lambda},\sqrt{\lambda}\right]^d} \left|\widehat{f}(\xi)\right| d\xi \to 0$$

da $\widehat{f} \in L_1(\mathbb{R})$. Hence, the inverse Fourier transform $\widehat{f}^\vee$ is a uniformly continuous function that coincides with f almost everwhere, so that we may even assume f is continuous. □

Exercise 2.1.3 Prove the following statement without using (2.1.9): if $f, f' \in L_1(\mathbb{R})$, then $(f')^\wedge(0) = 0$.
Hint: integration by parts. ◊

The next classic gives us information on the behavior of the Fourier transform of a function "far out".

Proposition 2.1.7 (**Riemann-Lebesgue lemma**). *For $f \in L_1(\mathbb{R}^d)$, we have that*

$$\lim_{|\xi|\to\infty} \hat{f}(\xi) = 0. \tag{2.1.13}$$

Proof: If, in addition, $\frac{\partial}{\partial x_j} f \in L_1$, $j = 1, \ldots, n$, then (2.1.13) is a direct consequence of (2.1.9) and (2.1.2):

$$\left\| \frac{\partial f}{\partial x_j} \right\|_1 \ge \left| \left(\frac{\partial f}{\partial x_j} \right)^\wedge (\xi) \right| = |\xi_j| \left| \widehat{f}(\xi) \right| \qquad \xi \in \mathbb{R},\ j = 1, \ldots, d.$$

Summing over j, we then find that

$$\sum_{j=1}^{d} \left\| \frac{\partial f}{\partial x_j} \right\|_1 \ge \left| \widehat{f}(\xi) \right| \sum_{j=1}^{d} |\xi_j| = |\xi|_1 \left| \widehat{f}(\xi) \right|,$$

hence,

$$\left| \widehat{f}(\xi) \right| \le \frac{1}{|\xi|_1} \sum_{j=1}^{d} \left\| \frac{\partial f}{\partial x_j} \right\|_1 \to 0, \qquad |\xi| \to \infty.$$

For arbitrary $f \in L_1$ and $g \in C_{00}^\infty(\mathbb{R}^d)$ such that[7] $\|f - g\|_1 \le \varepsilon$, we get, on the other hand, that

$$\varepsilon \ge \|f - g\|_1 \ge \left| \widehat{f}(\xi) - \widehat{g}(\xi) \right| \ge \left| \widehat{f}(\xi) \right| - |\widehat{g}(\xi)|,$$

hence

$$\lim_{\|\xi\|\to\infty} \left| \widehat{f}(\xi) \right| \le \lim_{\|\xi\|\to\infty} |\widehat{g}(\xi)| + \|f - g\|_1 \le \varepsilon,$$

[7] Once more it is the density of the smooth functions that is relevant here.

and since ε can be chosen arbitrarily small, the claim follows. □

The third classic is a result that tells us that properly normalized the Fourier transform can be extended to an *isometry* on $L_2(\mathbb{R}^d)$.

Theorem 2.1.8 (Parseval[8]/Plancherel). *For $f, g \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ we have[9]*

$$\int_{\mathbb{R}^d} f(x)\, g(x)\, dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, \overline{\widehat{g}(\xi)}\, d\xi, \tag{2.1.14}$$

and especially with $f = g$,

$$\|f\|_2 = \frac{1}{(2\pi)^{d/2}} \|\widehat{f}\|_2. \tag{2.1.15}$$

Theorem 2.1.8 allows us to extend the Fourier transform to $L_2(\mathbb{R}^d)$ by considering, for $f \in L_2$ a sequence

$$f_n := \chi_{[-n,n]^d} \cdot f \qquad \in L_1 \cap L_2, \qquad n \in \mathbb{N},$$

cf. Exercise 2.1.4, of "cut off" function that converge to f for $n \to \infty$ in the norm $\|\cdot\|_2$. Since, according to Theorem 2.1.8,

$$\left\| \widehat{f_{n+k}} - \widehat{f_n} \right\|_2 = \left\| (f_{n+k} - f_n)^\wedge \right\|_2 = (2\pi)^{d/2} \, \|f_{n+k} - f_n\|_2 \,, \qquad k, n \in \mathbb{N},$$

$\widehat{f_n}$ is a Cauchy sequence and due to completeness it has a limit in L_2, which we *define to be* $\widehat{f}$. This gives a formally correct and consistent definition of a Fourier transform on $L_2(\mathbb{R}^d)$.

This construction and (2.1.15) explain why the Fourier transform is often defined as

$$\widehat{f}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x)\, e^{-i\xi^T x}\, dx$$

since then it is an **isometry** on $L_2(\mathbb{R}^d)$, i.e.,

$$\|\hat{f}\|_2 = \|f\|_2, \qquad f \in L_2(\mathbb{R}^d). \tag{2.1.16}$$

Exercise 2.1.4 If $f \in L_2$ is compactly supported, the $f \in L_1$. ◊

Proof of Theorem 2.1.8: We define

$$h(x) = \int_{\mathbb{R}^d} f(t)\, g\,(t - x)\; dt = (f * g(-\cdot))\,(x), \qquad x \in \mathbb{R}^d,$$

so that $h(0) = \int fg$. Moreover,

$$\widehat{h}(\xi) = \widehat{f}(\xi) \underbrace{\left(g(-\cdot)\right)^\wedge(\xi)}_{=\overline{\widehat{g}(\xi)}} = \widehat{f}(\xi)\, \overline{\widehat{g}(\xi)}, \qquad \xi \in \mathbb{R}^d.$$

[8] Marc–Antoine Parseval des Chênes, 1755–1836, contemporary of Fourier, was quite involved in the troubles of the French revolution and published 5 *(five)* papers in his lifetime, but all of them were presented to the *Académie des Sciences* which was not ordinary

[9] Keep in mind that the complex inner product involves complex conjugation in order to be definite.

If f and g are so "nice" that $\widehat{f}, \widehat{g} \in L_2$ ist, for example when they are differentiable, then (2.1.11) implies

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, \overline{\widehat{g}(\xi)}\, d\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{h}(\xi)\, e^{i0\xi}\, d\xi = h(0) = \int_{\mathbb{R}^d} f(x)\, g(x)\, dx,$$

which gives (2.1.14). And the **Plancherel identity** (2.1.15) is finally a direct consequence of the **Parseval formula** (2.1.14). □

2.2 Fourier Series and Periodization

We obtain a different form of a Fourier transform if we consider functions on the **torus**

$$\mathbb{T}^d := \mathbb{R}^d / \left(2\pi\mathbb{Z}^d\right) \simeq [-\pi, \pi]^d .$$

Keep in mind that the torus is more than just the set $[-\pi, \pi]^d$ since all additions and multiplications are well defined on the torus, just modulo 2π in each component. Because of that, all functions from $L_p\left(\mathbb{T}^d\right)$ are 2π **periodic** which means that

$$f\left(\cdot + 2\pi\alpha\right) = f, \qquad \alpha \in \mathbb{Z}^d \tag{2.2.1}$$

and gives a natural extension of these functions to $\mathbb{R}^d$. The converse can also be done.

Definition 2.2.1. The **periodization** of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$f_* = \sum_{\alpha\in\mathbb{Z}^d} f\left(\cdot + 2\pi\alpha\right), \tag{2.2.2}$$

provided that the sum on the right hand side exists.

That the sum on the right hand side exists relies on conditions on f, for example, the constant function $f = 1$ cannot be periodized. Moreover, it depends on the sense in which the function should exist, pointwise or as an element of a function space like L_p.

Lemma 2.2.2 (Periodization). *If $f \in L_1\left(\mathbb{R}^d\right)$ then $f_* \in L_1\left(\mathbb{T}^d\right)$ with $\|f_*\|_1 \le \|f\|_1$.*

Proof: Since

$$\begin{aligned}
\|f_*\|_1 &= \int_{\mathbb{T}^d} |f_*(x)|\, dx = \int_{\mathbb{T}^d} \left| \sum_{\alpha\in\mathbb{Z}^d} f\left(x + 2\pi\alpha\right) \right| dx \le \int_{\mathbb{T}^d} \sum_{\alpha\in\mathbb{Z}^d} \left| f\left(x + 2\pi\alpha\right) \right| dx \\
&= \sum_{\alpha\in\mathbb{Z}^d} \int_{2\pi\alpha+[-\pi,\pi]^d} \left| f\left(x + 2\pi\alpha\right) \right| dx = \int_{\mathbb{R}^d} |f(x)|\, dx = \|f\|_1 ,
\end{aligned}$$

we $f_* \in L_1\left(\mathbb{T}^d\right)$ and 2π periodicity of the function is obtained by shifting the summation index in

$$f_*\left(\cdot + 2\pi\alpha\right) = \sum_{\beta\in\mathbb{Z}^d} f\left(\cdot + 2\pi(\alpha+\beta)\right) = \sum_{\beta\in\mathbb{Z}^d} f\left(\cdot + 2\pi\beta\right) = f_*$$

as $\beta \to \beta - \alpha$. □

Exercise 2.2.1 Give an example where $\|f_*\|_1 < \|f\|_1$. ◊

Definition 2.2.3. For $f \in L_1(\mathbb{R}^d)$, the αth **Fourier coefficient** is defined as

$$\widehat{f_*}(\alpha) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x)\, e^{-i\alpha^T x}\, dx, \qquad \alpha \in \mathbb{Z}^d, \tag{2.2.3}$$

and the associated **Fourier series** is

$$f_*(x) \simeq \sum_{\alpha\in\mathbb{Z}^d} \widehat{f_*}(\alpha)\, e^{i\alpha^T x}, \qquad x \in \mathbb{T}^d, \tag{2.2.4}$$

again provided the series converges.

Convergence of Fourier series is a nontrivial issue, even in one variable. There exists even the example, due to Du Bois–Reymond in 1873, of a *continuous* function whose Fourier series was *divergent* at some point. This observation even triggered a new field of mathematics, Approximation Theory, cf. (Sauer, 2017).

Remark 2.2.4. In Fourier series, the function belongs to $L_1(\mathbb{T}^d)$ while the Fourier coefficients belong to $\mathbb{Z}^d$ and the Fourier series takes them back to a function defined on $\mathbb{T}^d$. Therefore, $\mathbb{T}^d$ and $\mathbb{Z}^d$ are **dual groups** in the vague way that we mentioned before. In contrast to $\mathbb{R}^d$, there is no more self-duality.

That the Fourier coefficients of the periodization f_* should be connected to the Fourier series of f, is shown by the following computation:

$$\begin{aligned}
\widehat{f_*}(\alpha) &= \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} \sum_{\beta\in\mathbb{Z}^d} f(x+2\pi\beta)\; e^{-i\alpha^T x}\, dx \\
&= \frac{1}{(2\pi)^d} \sum_{\beta\in\mathbb{Z}^d} \int_{\mathbb{T}^d+2\pi\beta} f(x)\, e^{-i\alpha^T(x-2\pi\beta)}\, dx \\
&= \frac{1}{(2\pi)^d} \sum_{\beta\in\mathbb{Z}^d} \int_{\mathbb{T}^d+2\pi\beta} f(x)\, e^{-i\alpha^T x} \underbrace{e^{i2\pi\alpha^T\beta}}_{=1}\, dx \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(x)\, e^{-i\alpha^T x}\, dx = \frac{1}{(2\pi)^d}\, \widehat{f}(\alpha),
\end{aligned}$$

which brings us to our next result.

Theorem 2.2.5 (Poisson[10] formula). *If $f, \widehat{f} \in L_1$ we have the **Poisson summation formula***

$$\sum_{\alpha\in\mathbb{Z}^d} f(2\pi\alpha) = \frac{1}{(2\pi)^d} \sum_{\alpha\in\mathbb{Z}^d} \widehat{f}(\alpha) \qquad \textit{and} \qquad \sum_{\alpha\in\mathbb{Z}^d} f(\alpha) = \sum_{\alpha\in\mathbb{Z}^d} \widehat{f}(2\pi\alpha)\,. \tag{2.2.5}$$

[10] Siméon Denis Poisson, 1781–1840, student of Laplace and Legendre, between 300 and 400 publications on Mathematics and Physics.

Proof: If the partial sums of f_* converge[11], then $\widehat{f}(\alpha) = \widehat{f_*}(\alpha)$ f"ur $\alpha \in \mathbb{Z}^d$, implies that

$$\begin{aligned}\frac{1}{(2\pi)^d}\sum_{\alpha\in\mathbb{Z}}\widehat{f}(\alpha) &= \sum_{\alpha\in\mathbb{Z}^d}\widehat{f_*}(\alpha)\underbrace{e^{i\alpha^T 0}}_{=1} = \left(\sum_{\alpha\in\mathbb{Z}^d}\widehat{f_*}(\alpha)\,e^{i\alpha^T\cdot}\right)(0) = f_*(0)\\ &= \sum_{\alpha\in\mathbb{Z}^d} f\left(0+2\pi\alpha\right) = \sum_{\alpha\in\mathbb{Z}^d} f\left(2\pi\alpha\right),\end{aligned}$$

and yields the first identity. With this and (2.1.7) a "change of variables" gives

$$\begin{aligned}\sum_{\alpha\in\mathbb{Z}^d} f(\alpha) &= \sum_{\alpha\in\mathbb{Z}^d}\left(\sigma_{(2\pi)^{-1}I}f\right)(2\pi\alpha) = \frac{1}{(2\pi)^d}\sum_{\alpha\in\mathbb{Z}^d}\left(\sigma_{(2\pi)^{-1}I}f\right)^{\wedge}(\alpha)\\ &= \sum_{\alpha\in\mathbb{Z}^d}\widehat{f}\left(2\pi\alpha\right).\end{aligned}$$

□

Remark 2.2.6. The Poisson formula (2.2.5) is a pointwise statement and uses evaluations of f and $\hat{f}$ only at the discrete points $\mathbb{Z}^d$ and $2\pi\mathbb{Z}^d$, respectively. These points are a set of measure zero and any L_p function is defined only up to such a set. However, since we assume that both $f, \hat{f}$ belong to L_1, they also belong to $C_u(\mathbb{R}^d)$, since they are related by (inverse) Fourier transforms and therefore the point evaluation is a continuous linear functional and all sums in (2.2.5) are well-defined.

2.3 Distributions and a Strange Derivative

If we want a function for which the Fourier transform and the inverse Fourier are defined, it has to be of the form $f, \widehat{f} \in L_1(\mathbb{R}^d)$, at least so far. A way to bypass this is to use objects that go beyond functions, but include L_1 functions in a natural way. This will be the concept of tempered distributions. For details on various types of distributions see (Yosida, 1965) which is not easy to read, but a very compact source of valuable information.

Definition 2.3.1 (Test functions and distributions).

1. A **test function** $f \in \mathscr{T}\left(\mathbb{R}^d\right)$ is a element of the linear space of $C^\infty_{00}(\mathbb{R}^d)$ with uniformly bounded derivatives,

$$\mathscr{T}\left(\mathbb{R}^d\right) = \left\{\phi \in C^\infty_{00}\left(\mathbb{R}^d\right) \;:\; \sup_{\alpha\in\mathbb{N}_0^d}\sup_{x\in\mathbb{R}^d}\left|\frac{\partial^{|\alpha|}}{\partial x^\alpha}\phi(x)\right| < \infty\right\} \tag{2.3.1}$$

   topologized by the family of seminorms

$$|\phi|_\alpha = \max_{x\in\mathbb{R}^d}\left|\frac{\partial^{|\alpha|}}{\partial x^\alpha}\phi(x)\right|, \qquad \alpha \in \mathbb{N}_0. \tag{2.3.2}$$

[11] Otherwise we would have to apply a proper summation method, like Féjer kernels, cf. (Sauer, 2017).

2. A sequence $(\phi_n : n \in \mathbb{N})$ of test functions is said to converge to some ϕ if there is a compact set $\Omega \subset \mathbb{R}^d$ such that $\phi_n(x) = 0$, $x \notin \Omega$ and

$$\lim_{n\to\infty} |\phi_n - \phi|_\alpha = 0, \qquad \alpha \in \mathbb{N}_0^d. \tag{2.3.3}$$

3. A **distribution** T is a continuous linear functional[12] $T : \mathscr{T}\left(\mathbb{R}^d\right) \to \mathbb{C}$.

4. A distribution is called a **regular distribution** if there exists a function $f \in L_1\left(\mathbb{R}^d\right)$ such that

$$T(\phi) = T_f(\phi) = \int_{\mathbb{R}^d} f(x)\phi(x)\,dx, \qquad \phi \in \mathscr{T}(\mathbb{R}^d). \tag{2.3.4}$$

Remark 2.3.2. 1. The topology of test functions is needed to ensure that the limit of test functions is a test function again and in particular compactly supported. Even the limit of uniformly convergent compactly supported functions need not be compactly suppoted any more as the example

$$f_n = f\chi_{[-n,n]}, \qquad f|_{[k-1,k]} = f|_{[-k,1-k]} = \frac{1}{k}, \quad k = 1, 2, \ldots \tag{2.3.5}$$

shows.

2. Any expression of the form (2.3.4) is a continuous linear functional since for $\phi, \phi' \in \mathscr{T}(\mathbb{R}^d)$ we also have that $\phi - \phi' \in \mathscr{T}(\mathbb{R}^d)$ and

$$\begin{aligned} |T(\phi) - T(\phi')| &\leq \int_{\mathbb{R}^d} |f(x)|\,|\phi(x) - \phi'(x)|\,dx \leq \max_{x\in\mathbb{R}^d} |\phi(x) - \phi'(x)|\;\|f\|_1 \\ &\leq \|\phi - \phi'\|\,\|f\|_1, \end{aligned}$$

3. For $x \in \mathbb{R}^d$ the **point evaluation** $\phi \mapsto \phi(x)$ is a continuous linear functional $\mathscr{T}(\mathbb{R}^d)$ since convergence in $\mathscr{T}(\mathbb{R}^d)$ is uniform convergence of *all* derivatives, hence especially uniform convergence and therefore pointwise convergence. Hence $\phi_n \to \phi$ implies $\phi_n(x) \to \phi(x)$.

4. While the point evaluation is a distribution, it is not a *regular* distribution - there exists no L_1 function that represents it. Nevertheless it is the limit of the sequence

$$T_n(\phi) = \int_{\mathbb{R}^d} f_n(t)\,\phi(t)\,dt, \qquad f_n = \frac{n}{2}\,\chi_{[x-1/n,x+1/n]}, \tag{2.3.6}$$

of regular distributions. In other words, the regular distributions are not closed.

[12] A **functional** is a mapping from a (function) space to the underlying field, in our case $\mathbb{C}$. A *linear* functional is continuous iff it is bounded, see (Kreyszig, 1978; Taylor and Lay, 1980; Yosida, 1965).

5. While we will use distributions as some replacements for functions, they have the obvious problem that, in contrast to $f(x)$ for a function, the "value" of a distribution somewhere is not defined at all. They are just linear functionals.

6. Nevertheless, distributions are sometimes even called **generalized functions**.

Exercise 2.3.1 Turn (2.3.5) into an example where f and all the f_n are continuous. ◊

The example of the point evaluation shows us where the advantage of distributions lies: while the sequence f_n from (2.3.6) does not converge[13] in L_1, it still converges in the distributional sense, distributions are much more "forgiving". Moreover, regular distributions for differentiable f motivate the definition of the *derivative* of a distribution by simple partial integration:

$$\begin{aligned}\left(\frac{\partial^{|\alpha|}}{\partial x^\alpha} T\right)(\phi) &= \int_{\mathbb{R}^d} \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(x)\phi(x)\,dx \\ &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} f(x) \frac{\partial^{|\alpha|}}{\partial x^\alpha}\phi(x)\,dx = (-1)^{|\alpha|}\, T\left(\frac{\partial^{|\alpha|}}{\partial x^\alpha}\phi\right). \quad (2.3.7)\end{aligned}$$

The right hand side of (2.3.7) does not depend on f any more and is a continuous linear functional on $\mathscr{T}(\mathbb{R}^d)$ due to the way how we normed the space, we have the following observation.

Proposition 2.3.3. *Any distribution is infinitely differentiable with*

$$\left(\frac{\partial^{|\alpha|}}{\partial x^\alpha} T\right)(\phi) = (-1)^{|\alpha|}\, T\left(\frac{\partial^{|\alpha|}}{\partial x^\alpha}\phi\right).$$

Moreover, in contrast to functions, differentiation and limit can be exchanged and a lot of things that were forbidden in Analysis of functions are possible for distributions. One *disadvantage*, on the other hand, is that it is very difficult to check for a given distribution whether it is positive. We not dwell on details here, the basics of the theory are, for example, nicely summarized in (Forster, 1984).

To define a Fourier transform for distributions, we need a slightly different class of distributions.

Definition 2.3.4 (Tempered distributions).

1. The **Schwartz class**[14] $\mathscr{S}\left(\mathbb{R}^d\right)$ consists of all **rapidly decaying functions**

$$\mathscr{S}\left(\mathbb{R}^d\right) = \left\{ f \in C^\infty\left(\mathbb{R}^d\right) \;:\; \sup_{x\in\mathbb{R}^d} \left| p(x)\, \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(x) \right| < \infty,\; p \in \Pi,\; \alpha \in \mathbb{N}_0^d \right\}, \quad (2.3.8)$$

[13] It also is not a Cauchy sequence!

[14] Named after *Laurent Schwartz*, one of the fathers of distribution theory, and note after *Hermann Amandus Schwarz*, the Cauchy-Schwarz Schwarz.

with the topology based on the family

$$\|f\|_m := \sup_{x\in\mathbb{R}^d} \max_{|\alpha|\le m} \max_{p\in\Pi_m} \left| p(x) \frac{\partial^{|\alpha|}}{\partial x^\alpha} f(x) \right|, \qquad m \in \mathbb{N}_0, \tag{2.3.9}$$

of seminorms. This means that a sequence $f_n \in \mathscr{S}(\mathbb{R}^d)$ converges to $f \in \mathscr{S}(\mathbb{R}^d)$ if

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} \sup_{\alpha\in\mathbb{N}_0^d} \sup_{p\in\Pi} \left| p(x) \frac{\partial^{|\alpha|}}{\partial x^\alpha} (f_n - f)(x) \right| = 0. \tag{2.3.10}$$

2. A **tempered distribution** T is a continuous linear functional on $\mathscr{S}(\mathbb{R}^d)$.

Note $\mathscr{T}\left(\mathbb{R}^d\right) \subset \mathscr{S}\left(\mathbb{R}^d\right)$ not only in the sense of an inclusion of sets, but also with respect to the respective topology: the topology induced by (2.3.9) is *stronger* than the one induced by (2.3.2) which means that any sequence that converges with respect to the stronger distribution also converges with respect to the weaker one, but maybe not conversely. Since the continous functionals for a superset are a subset of the continuous functionals for the set[15], any tempered distribution is also a distribution so that the wording makes sense.

Proposition 2.3.5. *$f \in \mathscr{S}(\mathbb{R}^d)$ if and only if $\hat{f} \in \mathscr{S}(\mathbb{R}^d)$.*

Proof: Since $\mathscr{S}(\mathbb{R}^d) \subset L_1(\mathbb{R}^d)$ because of the fast decay, $\hat{f}$ exists. By Theorem (2.1.10),

$$\frac{\partial^{|\alpha|}}{\partial \xi^\alpha} \hat{f}(\xi) = \left((-i(\cdot))^\alpha f \right)(\xi), \qquad \alpha \in \mathbb{N}_0^s \tag{2.3.11}$$

and since $f \in \mathscr{S}(\mathbb{R}^d)$ also implies that $pf \in \mathscr{S}(\mathbb{R}^d)$ for any $p \in \Pi$, the right hand side of (2.3.11) belongs to $L_1(\mathbb{R}^d)$ and therefore the derivative on the left hand side is a continuous function. Moreover, for any $p \in \Pi$,

$$p(\xi) \frac{\partial^{|\alpha|}}{\partial \xi^\alpha} \hat{f}(\xi) = \left(p(-iD) \, (-i(\cdot))^\alpha f \right), \tag{2.3.12}$$

where

$$p(-iD) = \sum_\alpha p_\alpha i^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial x^\alpha}, \qquad p = \sum_\alpha p_\alpha (\cdot)^\alpha,$$

and the right hand side of (2.3.12) is still in $\mathscr{S}(\mathbb{R}^d)$, so that the Riemann-Lebesque Lemmma, Proposition 2.1.7, yields the fast decay. The converse is clear since the inverse Fourier transform has the same properties. □

Exercise 2.3.2 Prove that $\mathscr{S}(\mathbb{R}^d) \subset C_u(\mathbb{R}^d)$ (*strict* inclusion). ◊

Now we are in business: due to Proposition 2.3.5 we can apply the same idea used for derivatives to define the Fourier transform $\widehat{T}$ of a tempered distribution as

$$\widehat{T}(\phi) = T\left(\widehat{\phi}\right), \qquad \phi \in \mathscr{S}\left(\mathbb{R}^d\right); \tag{2.3.13}$$

[15] Being continuous and linear on a *larger* set means more constraints.

and the inverse Fourier transform is almost trivial:

$$T^\vee(\phi) = T\left(\phi^\vee\right), \qquad \phi \in \mathscr{S}\left(\mathbb{R}^d\right), \tag{2.3.14}$$

which immediately yields that

$$\widehat{T}^\vee(\phi) = T^\vee\left(\widehat{\phi}\right) = T\left(\widehat{\phi}^\vee\right) = T(\phi), \qquad \phi \in \mathscr{S}\left(\mathbb{R}^d\right),$$

hence $\widehat{T}^\vee = T$. For the Fourier transform of tempered distributions we have similar properties as for the Fourier transform of functions, for example,

$$\begin{aligned}\left(\frac{\partial}{\partial x_j}T\right)^\wedge(\phi) &= -\widehat{T}\left(\frac{\partial}{\partial x_j}\phi\right) = -T\left(\left(\frac{\partial}{\partial x_j}\phi\right)^\wedge\right) = -T\left(-i\xi_j\widehat{\phi}\right)\\ &= \left(ix_jT\right)\left(\widehat{\phi}\right) = \left(ix_j\widehat{T}\right)(\phi).\end{aligned}$$

More on the Fourier transform for tempered distributions can be found in (Yosida, 1965).

What does all that have to do with derivative. Well, introducing the **Laplace operator**

$$\Delta = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$$

we note that due to (2.1.10)

$$(\Delta f)^\wedge(\xi) = \sum_{j=1}^d \left(-i\xi_j\right)^2 \hat{f}(\xi) = -\hat{f}(\xi)\sum_{j=1}^d \left(\xi_j\right)^2 = -|\xi|_2^2\,\hat{f}(\xi), \qquad \xi \in \mathbb{R}^d,$$

and we can easily solve the differential equation[16]

$$\Delta f = g \qquad \Leftrightarrow \qquad \widehat{f} = -\left(\frac{1}{|\cdot|_2^2}\right)^\wedge \widehat{g} = -\left(|\cdot|_2^{-2}\right)^\wedge \widehat{g}$$

using the Fourier transform. The Laplace operator is a second order differential operation which fits well with the exponent 2 in the norm. We generalize this idea a little bit to obtain a fairly fundamental concept.

Definition 2.3.6 (Riesz potential). For $r \in \mathbb{R}$ we define the *Riesz potential*[17] of order r is defined via its Fourier transform as

$$(I^r f)^\wedge(\xi) = |\xi|_2^{-r}\,\widehat{f}(\xi), \qquad \xi \in \mathbb{R}^d. \tag{2.3.15}$$

In this context, $r < 0$ corresponds to a differential operator of order r, whiel for $r > 0$ it corresponds to an "inverse" diffential operator which is some sort of integral operator and has a smoothing effect.

[16] This is called the **heat equation** and solving it was Fourier's original motivation to introduce Fourier series.

[17] Marcel and Frederic Riesz were two brothers of equal mathematical success. In particular, both contributed to Functional Analysis.

The Riesz potential can be understood as a differentiation of order r where now we can even define **fractional derivatives** of arbitrary order, which we will write as $\Delta^{r/2}$. For $r = 2$ we recover the Laplace operator. For distributions, these "differential operators" are of course well-defined and we will call a function differentiable of order $r > 0$ in the Fourier sense if

$$\lim_{\xi \to \infty} \left(1 + |\xi|_2^2\right)^{r/2} \widehat{f}(\xi) = 0. \tag{2.3.16}$$

Replacing the decay rate $|\xi|_2^2$ by $1 + |\xi|_2^2$ does not make a difference for $|\xi|_2^2$, it only makes our lives easier for $|\xi| \to 0$.

Definition 2.3.7. The **Sobolev space** $H^r\left(\mathbb{R}^d\right)$ consists of all functions $f \in L_2\left(\mathbb{R}^d\right)$ for which the norm

$$\|f\|_{H^r}^2 := \int_{\mathbb{R}^d} \left(1 + |\xi|_2^2\right)^r \left|\widehat{f}(\xi)\right|^2 \, d\xi \tag{2.3.17}$$

is finite.

And in fact, functions from the Sobolev space are differentiable, at least under some conditions.

Theorem 2.3.8 (Sobolev embedding theorem). *If $r > \frac{d}{2} + k$, then $H^r\left(\mathbb{R}^d\right) \subset C^k\left(\mathbb{R}^d\right)$.*

Proof: Choose α with $|\alpha| = k$ and consider, for $f \in H^r(\mathbb{R})$ the function $g(\xi) = (-i\xi)^\alpha \widehat{f}(\xi)$, $\xi \in \mathbb{R}^d$. To show that $g \in L_1\left(\mathbb{R}^d\right)$, i.e., that

$$\int_{\mathbb{R}^d} |g(\xi)| \, d\xi < \infty,$$

we choose $C > 0$, split the integral and use the Cauchy-Schwarz inequalitly to obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} |g(\xi)| \, d\xi &= \int\limits_{\|\xi\|_2 \le C} |g(\xi)| \, d\xi + \int\limits_{\|\xi\|_2 \ge C} |g(\xi)| \, d\xi \\
&= \int\limits_{\|\xi\|_2 \le C} \left|\xi^\alpha \widehat{f}(\xi)\right| \, d\xi + \int\limits_{\|\xi\|_2 \ge C} \left|\xi^\alpha \widehat{f}(\xi)\right| \, d\xi \\
&\le \underbrace{\sqrt{\int\limits_{\|\xi\|_2 \le C} |\xi^\alpha|^2}}_{\le C^k} \underbrace{\sqrt{\int\limits_{\|\xi\|_2 \le C} \left|\widehat{f}(\xi)\right|^2 \, d\xi}}_{\le \left\|\widehat{f}\right\|_2 = (2\pi)^{d/2} \|f\|_2} + \underbrace{\sqrt{\int\limits_{\|\xi\|_2 > C} \left|\left(1 + \|\xi\|_2^2\right)^{r/2} \widehat{f}(\xi)\right|^2 \, d\xi}}_{= \|f\|_{H^r}} \times \\
&\quad \times \sqrt{\int\limits_{\|\xi\|_2 > C} \left|\left(1 + \|\xi\|_2^2\right)^{-r/2} \xi^\alpha\right|^2 \, d\xi}.
\end{aligned}$$

The first three quantities are finite, and for the third integral we use the fact that $r > \frac{d}{2} + k$ and the polar coordinates $\xi = sv$, $s \in \mathbb{R}_+$, $\|v\|_2 = 1$ with $d\xi = s^{d-1} ds \, dv$ to

conclude that

$$\begin{aligned}
&\int\limits_{\|\xi\|_2>C} \left|\left(1+\|\xi\|_2^2\right)^{-r/2}\xi^\alpha\right|^2 d\xi \\
&\quad= \int_C^\infty \int_{\|v\|_2=1} \left(1+s^2\right)^{-r} s^{2k} \underbrace{(v^\alpha)^2}_{\leq 1} s^{d-1}\, dv\, ds \\
&\quad\leq \left(\int_{\|v\|=1} dv\right) \int_C^\infty \left(1+C^{-2}\right) s^{2k-2r+d-1}\, ds;
\end{aligned}$$

the integral exists if the exponent is < -1 which is equivalent to $2k - 2r + d < 0$ or $r > k + \frac{d}{2}$.

If $g \in L_1(\mathbb{R}^d)$, there exists the inverse Fourier transform

$$F(x) := g^\vee(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} g(\xi)\, e^{ix^T\xi} d\xi$$

as a uniformly continuous function and since

$$\left(\frac{\partial^k}{\partial x^\alpha} F\right)^\wedge (\xi) = g(\xi) = \left(\frac{\partial^k}{\partial x^\alpha} f\right)^\wedge (\xi)$$

the functions f and F have to coincide almost everywhere. □

The Radon Transform and Its Relatives

3

If your wish is to become really a man of science and not merely a petty experimentalist, I should advise you to apply to every branch of natural philosophy, including mathematics.

(M. Shelley, *Frankenstein*))

Now we will focus on the main integral transform of this lecture, namely the *Radon transform* and its relatives. The main task of this chapter will be to define and invert it *formally* and to consider questions of uniqueness of the transform. Numerical methods will come later.

3.1 The Physical Motivation

The main importance of the Radon transform comes from its interpretation in the context of **Computed Tomography**. If an **X-ray** beam is sent through inhomogeneous material, a part of its energy is absorbed by the material it passes, the rest of the energy continues its way through the object, see Fig. 3.1.1. This model ignores a lot of physical effects, like diffration or scattering of the beam, but it is sufficient for our purposes here. Moreover, the usual assumption is that the ray is *monochromatic*, i.e., it has only a certain well-defined wavelength; also this is not really refleting the reality.

If we denote the material dependend absorption rate at x by $f(x)$, $x \in \mathbb{R}^2$, and by $I(x)$ the intensity there, then for another point on the beams, say $x + \delta$ the energy

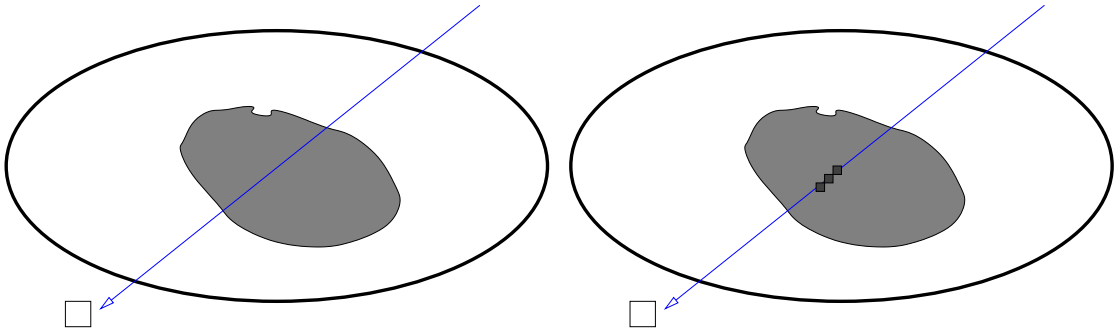

Figure 3.1.1: The basic idea of Computed Tomography: a ray (usually X-ray) is sent through an object and the intensity at the other end is measured by a detector. The absorption of energy can then be related to a line integral *(left)*. In the approximation, the ray is decomposed into small pieces *(right)*.

approximately satisfies

$$I(x+\delta) - I(x) \approx -f(x)\,\delta\, I(x), \qquad \text{i.e.,} \qquad I(x+\delta) \approx I(x)\,(1 - f(x)\,\delta)\,,$$

cf. (Olafsson and Quinto, 2006). To turn this multiplicative relationship into an additive one, we take the logarithm of both sides and apply the Taylor expansion

$$\log(1-ax) = -\sum_{j=1}^{\infty} \frac{a^j}{(1-ay)}\bigg|_{y=0} x^j = -\sum_{j=1}^{\infty} (ax)^j\,.$$

of the logarithm with respect to δ at $\delta = 0$ to obtain

$$\begin{aligned} \log I(x+\delta) &= \log I(x) + \log\left(1 - f(x)\,\delta\right) = \log I(x) - f(x)\,\delta + O\left(\delta^2\right) \\ &\approx \log I(x) - f(x)\,\delta, \end{aligned}$$

that is, we have the approximate identity

$$\log I(x+\delta) \log I(x) - f(x)\,\delta. \tag{3.1.1}$$

Next, we decompose the line from the source x_S to the detector x_D into $N+1$ pieces of length δ and apply (3.1.1) iteratively to get

$$\begin{aligned} \log \frac{I\,(x_S)}{I\,(x_D)} &= -\left(\log I\,(x_D) - \log I\,(x_S)\right) \\ &= -\sum_{j=0}^{N} \log I\,(x_S + (j+1)\delta) - \log I\,(x_S + j\delta) \approx \sum_{j=0}^{n} f\,(x_S + j\delta)\;\delta, \end{aligned}$$

which is a **quadrature formula** or **Riemann sum**, cf. (Heuser, 1984; Sauer, 2014) for the *line integral*

$$\int_{[x_S,x_D]} f(x)\,dx := \|x_D - x_S\| \int_0^1 f\,(\lambda x_S + (1-\lambda)x_D)\;d\lambda,$$

which we normalized such that $\int_{[x_S,x_D]} 1\,dx = \|x_D - x_S\|$ reproduces the length of the line. This already is the Radon transform and asks for a formal definition.

3.2 Definition and Invertibility

Despite the two dimensional "practical" application in Section 3.1, we will define the Radon transform on $\mathbb{R}^d$, in particular as this will guide us to other transforms[1] for $d > 2$.

Definition 3.2.1 (Hyperplanes and Radon transform).

1. The **unit ball** and the **unit sphere** in $\mathbb{R}^d$ are denoted by

$$\mathbb{B}^d := \left\{x \in \mathbb{R}^d \;:\; |x|_2 \le 1\right\} \qquad \text{and} \qquad \mathbb{S}^{d-1} := \left\{x \in \mathbb{R}^d \;:\; |x|_2 = 1\right\}, \tag{3.2.1}$$

respectively.

[1] *Hint:* the case $d = 2$ is very special since lines and hyperplanes are the same.

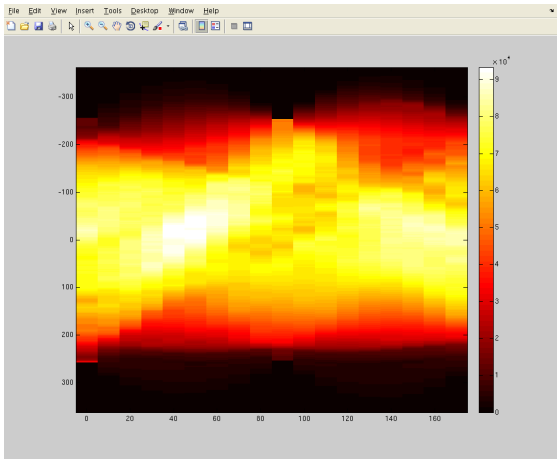

Figure 3.2.1: The sinogram of some object, computed with Matlabs `radon` function.

2. For $v \in \mathbb{S}^{d-1}$ and $s \in \mathbb{R}$, the **hyperplace** $H(v,s)$, is defined as

$$H(v,s) = \left\{x \in \mathbb{R}^d : v^T x = s\right\}. \tag{3.2.2}$$

3. For $v \in \mathbb{S}^d$ and $s \in \mathbb{R}$ the **Radon transform** of $f \in \mathscr{S}(\mathbb{R}^d)$ is defined as

$$Rf(v,s) := \int_{v^T x = s} f(x)\, dx, \tag{3.2.3}$$

where dx denotes the $(d-1)$ dimensional *area integral*, i.e., integration over a set of measure zero.

Remark 3.2.2. The "$d-1$" for the sphere in $\mathbb{R}^d$ may appear a bit strange, but it is common use, a tradition motivated by the fact that the sphere is as $(d-1)$ dimensional *manifold*.

In the plane, i.e., for $d = 2$, the normalized vector can be parametrized *uniquely* as $v = (\cos\theta, \sin\theta)$ for some $\theta \in [0, \pi]$, then the Radon transform can be written as $Rf(\theta, s)$ and we can arrange the two parameters as a two dimensional coordinate grid. Such a **sinogram** which color codes the values of the Radon trasnform can be seen in Fig. 3.2.1.

The definition of the integral in (3.2.3) may appear a bit strange, but we can define in a more formal way as follows. Any *orthogonal matrix*[2] $V \in \mathbb{R}^{d\times d-1}$ with $V^T v = 0$ yields a *parametrization*

$$\int_H f(x)\, dx = \int_{v^T x = s} f(x)\, dx = \int_{\mathbb{R}^{d-1}} f\left(sv + Vy\right)\, dy \tag{3.2.4}$$

of the integral. Indeed, each point of the form $z = sv + Vy$ satisfies

$$v^T z = s\, \underbrace{v^T v}_{=1} + \underbrace{v^T V}_{=V^T v = 0}\, y = s,$$

[2] An **orthogonal matrix** $V \in \mathbb{R}^{d\times d-1}$ is a matrix with orthogonal columns which can conveniently be written as $V^T V = I$. Note, however, that $VV^T \neq I$ since V is not a *square matrix*.

and lies on the hyperplane. The representation is independent of the chosen V. Indeed, if V_1, V_2 are two such matrices, then their columns span the orthogonal complement of v in $\mathbb{R}^d$, hence there exists a square matrix $Q \in \mathbb{R}^d$ such that $V_2 = V_1 Q$. Since

$$I = {V_2}^T V_2 = Q^T V_2^T V_2 Q = Q^T Q,$$

Q is a ortgonal matrix with $Q^{-1} = Q^T$ and also $V_1 = V_2 Q^T$ as well as

$$I = V_1^T V_1 = V_1^T V_2 Q^T \qquad \Leftrightarrow \qquad Q = V_1^T V_2.$$

Then,

$$\int_{\mathbb{R}^{d-1}} f\left(sv + V_2 y\right)\, dy = \int_{\mathbb{R}^{d-1}} f\left(sv + V_1 Q y\right)\, dy = \underbrace{|\det Q|^{-1}}_{=1} \int_{\mathbb{R}^{d-1}} f\left(sv + V_1 y\right)\, dy,$$

verifies that the definition in (3.2.4) is indeed independent of V.

Remark 3.2.3 (Hyperplane trouble).

1. The hyperplane $H(v,s)$ is a set of measure zero and since L_p functions are only defined up to a set of measure zero, we had to restrict the definition to continous functions for which such integrals can be formed consistently.

2. The association $(v,s) \mapsto H(v,s)$ is ambiguous! Since $v^T x = s$ is equivalent to $(-v)^T x = -s$, we also have that

$$H(v,s) = H(-v,-s), \qquad \text{that is,} \qquad Rf(v,s) = Rf(-v,-s) \tag{3.2.5}$$

which has to be considered once we ask questions about invertibility and in particular injectivity of Rf.

Lemma 3.2.4 (**Projection Slice Theorem**). *For $f \in \mathscr{S}(\mathbb{R}^d)$ and $g \in L_\infty(\mathbb{R})$ we have*

$$\int_{\mathbb{R}} g(s)\, Rf\left(v,s\right)\, ds = \int_{\mathbb{R}^d} f(x)\, g(v^T x)\, dx, \qquad v \in \mathbb{S}^{d-1}. \tag{3.2.6}$$

Proof: First, we substitute all definitions and choose V such that $V^T v = 0$ to get

$$\begin{aligned}
\int_{\mathbb{R}} g(s)\, Rf\left(v,s\right)\, ds &= \int_{\mathbb{R}} g(s) \int\limits_{v^T x = s} f(x)\, dx\, ds \\
&= \int_{\mathbb{R}} g(s) \int_{\mathbb{R}^{d-1}} f\left(sv + Vy\right)\, dy\, ds = \int_{\mathbb{R}} \int_{\mathbb{R}^{d-1}} g(s)\, f\left(\left[v, V\right] \begin{bmatrix} s \\ y \end{bmatrix}\right) dy ds \\
&= \int_{\mathbb{R}^d} g(e_1^T x)\, f\left(\left[v,V\right] x\right)\, dx = \underbrace{|\det\left[v,V\right]|^{-1}}_{=1} \int_{\mathbb{R}^d} g\left(e_1^T \left[v,V\right]^{-1} x\right)\, f(x)\, dx \\
&= \int_{\mathbb{R}^d} g(v^T x)\, f(x)\, dx,
\end{aligned}$$

since the matrix $[v, V]$ is also orthogonal,

$$\begin{pmatrix} v^T \\ V^T \end{pmatrix} \begin{pmatrix} v & V \end{pmatrix} = \begin{pmatrix} v^T v & v^T V \\ V^T v & V^T V \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix},$$

hence

$$e_1^T\,[v,V]^{-1} = e_1^T\,[v,V]^T = ([v,V]\,e_1)^T = v^T.$$

□

Note that Schwartz functions are also rapidly decaying and therefore the integral

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} g(e_1^T x)\, f\left([v,V]\,x\right)\, dx$$

would even be defined for functions g that increase a most polynomially, i.e. for functions for which there exists a polynomial $p \in \Pi$ such that $|g(x)| \leq |p(x)|$, $x \in \mathbb{R}^d$.

Lemma 3.2.4 gives us a result for the Fourier transform of the Radon transform[3] with respect to the offset s by simply specializing $g(x) = e^{-i\sigma x}$ which turns (3.2.6) into

$$\left(Rf\,(v,\cdot)\right)^\wedge(\sigma) = \int_{\mathbb{R}} e^{-i\sigma s} Rf\,(v,s)\, ds = \int_{\mathbb{R}^d} f(x)\, e^{-i\sigma v^T x}\, dx = \widehat{f}\,(\sigma v)\,.$$

Corollary 3.2.5 (**Fourier Slice Theorem**). *For $f \in \mathscr{S}(\mathbb{R}^d)$ one has*

$$\left(Rf\,(v,\cdot)\right)^\wedge(\sigma) = \widehat{f}\,(\sigma v)\,, \qquad v \in \mathbb{S}^{d-1}, \quad \sigma \in \mathbb{R}. \tag{3.2.7}$$

A side effect effect of the Fourier slice theorem is that it gives us a consistent embedding on the Radon transform into $L_1(\mathbb{R}^d)$.

Corollary 3.2.6. *The mapping $f \mapsto Rf$ is injective on $\mathscr{S}(\mathbb{R}^d)$.*

Proof: Suppose that $Rf = Rf'$, i.e., $R\,(f - f') = 0$, for some $f \neq f' \in L_1(\mathbb{R}^d)$. Since we can write any $\xi \in \mathbb{R}^d$ as $\xi = \sigma v$, $\sigma \in \mathbb{R}$, $v \in \mathbb{S}^d$, (3.2.7) implies that

$$(f - f')^\wedge(\xi) = 0, \qquad \xi \in \mathbb{R}^d,$$

which is a contradiction. □

Exercise 3.2.1 Is the representation of ξ as $\xi = \sigma v$, $\sigma \in \mathbb{R}$, $v \in \mathbb{S}^d$ unique? ◊

Exercise 3.2.2 Prove that the Fourier transform is injective. ◊

The fact that the Radon transform is injective gives us hope that there may be an inverse of the Radon transform; and even if injectivity is, of course, only a *necessary* condition for the existence of an inverse, this hope is justified.

Theorem 3.2.7 (**Inverse Radon transform**). *For $f, \widehat{f} \in \mathscr{S}\left(\mathbb{R}^d\right)$,*

$$R^* I^{1-d} R f = f \tag{3.2.8}$$

where

$$R^* f(x) = \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^{d-1}} f\left(v, x^T v\right)\, dv, \qquad x \in \mathbb{R}^d, \tag{3.2.9}$$

*denotes the **dual Radon transform** or **backprojection**.*

[3] In accordance with the "engineering approach" to convert everything into its Fourier transform from the very beginning.

Proof: Each point $0 \neq \xi \in \mathbb{R}^d$ can be written in two ways as $\xi = \sigma v$, namely, $\sigma = |\xi|_2$, $v = \xi/|\xi|_2$ and $\sigma = -|\xi|_2$, $v = -\xi/|\xi|_2$, which implies that

$$\int_{\mathbb{R}^d} g(\xi) = \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} g(\sigma v)\, |\sigma|^{d-1}\, dv d\sigma.$$

Taking this into account, we can expand the inverse Fourier transform of $\hat{f}$ to obtain

$$\begin{aligned}
f(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi = \frac{1}{2(2\pi)^d} \int_{\mathbb{R}} \int_{\mathbb{S}^d} \widehat{f}\,(\sigma v)\, e^{ix^T(\sigma v)}\, |\sigma|^{d-1}\, dv\, d\sigma \\
&= \frac{1}{2(2\pi)^d} \int_{\mathbb{R}} \int_{\mathbb{S}^d} (Rf(v,\cdot))^\wedge(\sigma)\, e^{ix^T(\sigma v)}\, |\sigma|^{d-1}\, dv\, d\sigma \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \frac{1}{2\pi} \int_{\mathbb{R}} (Rf(v,\cdot))^\wedge(\sigma)\, e^{i(x^Tv)\sigma}\, |\sigma|^{d-1}\, d\sigma\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \frac{1}{2\pi} \int_{\mathbb{R}} \left(I^{1-d} Rf(v,\cdot)\right)^\wedge(\sigma) e^{i(x^Tv)\sigma}\, d\sigma\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \left(\left(I^{1-d} Rf(v,\cdot)\right)^\wedge\right)^\vee (x^Tv)\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} I^{1-d} Rf\left(v, x^Tv\right)\, dv = R^* I^{1-d} Rf(x),
\end{aligned}$$

which already is (3.2.8). □

The name "dual Radon transform" needs some explanation; looking at the projection slice theorem, that is, at equation (3.2.6), we see that

$$\begin{aligned}
\int_{\mathbb{S}^d} \int_{\mathbb{R}} g(s)\, Rf(v,s)\, ds\, dv &= \int_{\mathbb{S}^d} \int_{\mathbb{R}^d} g\left(v^Tx\right)\, f(x)\, dx\, dv \\
&= \int_{\mathbb{R}^d} f(x) \underbrace{\int_{\mathbb{S}^d} g\left(v^Tx\right)\, dv}_{=R^*g(x)}\, dx = \int_{\mathbb{R}^d} R^*g(x)\, f(x) dx.
\end{aligned}$$

In other words, R^* is **adjoint** of R if we consider the function g, defined on $\mathbb{R}$, as a function on $\mathbb{S}^d \times \mathbb{R}$ that is constant in the first variable, $g(v,s) := g(s)$. This is a natural and canonical way to introduce additional variables into a function.

The formula

$$R^{-1} = R^* I^{1-d} \tag{3.2.10}$$

for the inverse Radon transform in d variables is definitely beautiful and elegant, it only has one disadvantage: it does not work numerically or in practice, at least according to (Natterer and Wübbeling, 2001). But the reason is not hard to understand: the Riesz potential I^{1-d} multiplies the Fourier transform with $|\sigma|^{d-1}$ which amplifies the *high frequency* oscillatory content in the signal Rf and thus increases the noise in the measurements.

Exercise 3.2.3 Implement the Radon inversion by means of (3.2.10) and test it. ◊

To overcome this problem, we have to include some denoising or smoothing into the inversion, and this is where the convolution comes for our rescue.

Theorem 3.2.8 (**Filtered backprojection**). *For $f \in L_1\left(\mathbb{R}^d\right)$ and $g \in L_1\left(\mathbb{S}^d \times \mathbb{R}\right)$,*

$$(R^* g) * f = R^* (g * Rf) . \tag{3.2.11}$$

Proof: By definition, we have for $x \in \mathbb{R}^d$,

$$\begin{aligned}
((R^* g) * f)(x) &= \int_{\mathbb{R}^d} (R^* g)(t)\, f(x-t)\, dt \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{R}^d} \int_{\mathbb{S}^d} g\left(v, v^T t\right)\, f(x-t)\, dv\, dt \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{R}^d} \int_{\mathbb{S}^d} g\left(v, v^T (x-t)\right)\, f(t)\, dv\, dt \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \int_{\mathbb{R}} \int_{\mathbb{R}^{d-1}} g\left(v, v^T (x - sv - Vy)\right)\, f(sv + Vy)\, dy\, ds\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \int_{\mathbb{R}} g\left(v, v^T x - s\right) \underbrace{\int_{\mathbb{R}^{d-1}} f(sv + Vy)\, dy}_{= \int_{v^T y = s} f(y)\, dy = Rf(v,s)}\, ds\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} (g(v,\cdot) * Rf(v,\cdot))\left(v, v^T x\right)\, dv = R^* (g * Rf) ,
\end{aligned}$$

which also shows how the convolutions in (3.2.11) have to be understood. □

Remark 3.2.9. Let us recall once more how the convolutions work here: if it involves a function defined on $\mathbb{R}^d$, we use the "normal" convolution, if the function is defined on $\mathbb{S}^d \times \mathbb{R}$, on the other hand, the convolution only accesses the second argument which belongs to $\mathbb{R}$. This makes sense and is only reasonable since the convolution involves a translation and this is not so clear on $\mathbb{S}^{d-1}$, in contrast to $\mathbb{T}^d$, by the way.

Remark 3.2.10. In principle or on an abstract level, the filtered backprojection (3.2.11) determines f as the solution of a linear system, namely

$$Tf = b, \qquad T : f \mapsto (R^* g) * f, \qquad b = R^* (g * y) . \tag{3.2.12}$$

where $y = Rf$ is the measurement. However, this is an interaction of operators and functions and not of matrices and vectors[4] in finite dimensional spaces. To solve this numerically, the equation has to be *discretized*.

The **filter** g in (3.2.12) which gives the method its name, can be chosen freely, at least in principle. In practice, the concrete choice of the filter is a fundamental issue that strongly influences the behavior of the algorithm and usually requires some experience. The filtered backprojection has to advantages:

1. We can avoid the Riesz potential which corresponds to a differentiation and only have to compute backprojections.

2. To solve linear *operator equations* of the form $Tf = b$, a lot of methods are known that can even be adapted to over- and underdetermined problems. Usually, we will face the latter here since normally it is practically and technically impossible to measure $Rf(v,s)$ for *all* $v \in \mathbb{S}^{d-1}$ and $s \in \mathbb{R}$.

[4] Elements of function spaces are vectors as well.

3.3 Further Transforms

In two dimensions, the case $d = 2$, the Radon transform is a model of X-ray tomography. For $d > 2$, this is no more the case as we integrate over affine hyperplanes then. To that end, we define two other transforms as in (Natterer, 1986).

Definition 3.3.1. For $f \in \mathscr{S}(\mathbb{R}^d)$ and $v \in \mathbb{S}^{d-1}$, we define

1. the **X-ray transform**

$$Xf(v,x) := \int_{\mathbb{R}} f(x+tv)\,dt, \qquad x \in \mathbb{R}^d, \tag{3.3.1}$$

2. the **divergent beam transform**

$$Df(v,x,) := \int_0^\infty f(x+tv)\,dt, \qquad x \in \mathbb{R}^d. \tag{3.3.2}$$

These transforms have some obvious properties, for example,

1. $Xf(v,x) = Xf(-v,x)$,
2. $Xf(v,x+\lambda v) = Xf(v,x)$, $\lambda \in \mathbb{R}$,
3. $Xf(v,x) = Df(v,x) + Df(-v,x)$,

and, in addition,

$$\lim_{\lambda\to\infty} Df(v,x-\lambda x) = Xf(v,x).$$

The second property 2) makes it reasonable to restrict the X-ray transform to pairs

$$T\mathbb{S}^{d-1} := \left\{(v,x) : v \in \mathbb{S}^{d-1}, x \in \mathbb{R}^d, v^T x = 0\right\} \simeq \mathbb{S}^{d-1} \times \mathbb{R}^{d-1}, \tag{3.3.3}$$

the set of which consists of a point on the sphere and the tangent plane there[5].

Moreover, the Radon transform can be written in terms of the X-ray transform in the following way. Given $v \in \mathbb{S}^{d-1}$, we choose any $w \in \mathbb{S}^{d-1}$ such that $v^T w = 0$ and integrate the X-ray transform in this direction over the hyperplane,

$$Rf(v,s) = \int_{x\in w^\perp, v^T x = s} Xf(w,x)\,dx$$

Here we take into account (3.3.3) to factor out ambiguities.

The geometrical difference between the Radon transform and the X-ray transform are shown in Fig. 3.3.1. In both cases, we have the direction v, depicted in red and the orthogonal plane $v^\perp$. In the Radon transform, integration is performed over the hyperplane $v^\perp$ and the region of integration is shifted parallel to v by the offset parameter $s \in \mathbb{R}$. In the X-ray transform, on the other hand, we integrate along the red ray and shift the ray; since shifts along the ray make no sense, we can restrict these shifts to shifts within the hyperplane.

[5] This is called the **tangent bundle**.

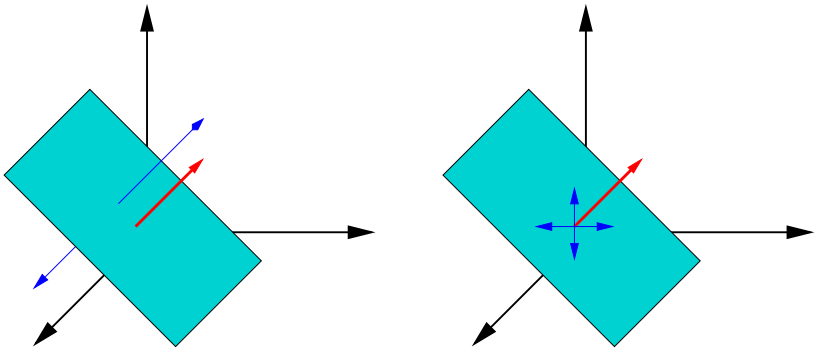

Figure 3.3.1: Comparison of Radon and X-ray transform. In the Radon transform *(left)* the plane is moved along the direction v *(red)* while in the X-ray transform the direction is moved in the plane $v^\perp$ *(right)*.

Remark 3.3.2. The direction v and the shifts $s \in \mathbb{R}$ and $x \in v^\perp$ can be seen as a parameter and another variable that depends on this parameter. This is why the transforms are also written as

$$R_v f = Rf(v,\cdot) : \mathbb{R} \to \mathbb{R} \qquad \text{and} \qquad X_v f = Xf(v,\cdot) : v^\perp \to \mathbb{R} \tag{3.3.4}$$

with $R_v f \in \mathscr{S}(\mathbb{R})$ and $X_v f \in \mathscr{S}(v^\perp)$. Convolutions and Fourier transforms are then applied to these functions.

There is also an analog of the Fourier Slice theorem for the X-ray transform.

Theorem 3.3.3. *For $f \in \mathscr{S}(\mathbb{R}^d)$ we have*

$$(Xf(v,\cdot))^\wedge(\xi) = \hat{f}(\xi), \qquad v \in \mathbb{S}^{d-1}, \quad \xi \in v^\perp. \tag{3.3.5}$$

Proof: We again define $V \in \mathbb{R}^{d\times d-1}$ as the orthogonal matrix such that $v^\perp = V\mathbb{R}^{d-1}$, i.e., $V^T v = 0$, write $\xi = V\eta$ as well as $\hat{V} = \begin{pmatrix} v & V \end{pmatrix} \in \mathbb{R}^{d\times d}$, and compute

$$\begin{aligned}
(Xf(v,\cdot))^\wedge(\xi) &= \int_{v^\perp} e^{-i\xi^T x} \int_{\mathbb{R}} f(x+vt)\,dt\,dx \\
&= \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} e^{-i\xi^T \begin{pmatrix} v & V \end{pmatrix}\begin{pmatrix} 0 \\ y \end{pmatrix}} f\left(\begin{pmatrix} v & V \end{pmatrix}\begin{pmatrix} t \\ y \end{pmatrix}\right) dt\,dy = \int_{\mathbb{R}^d} e^{-i\xi^T \hat{V} x} f(\hat{V}x)\,dx \\
&= \int_{\mathbb{R}^d} e^{-i\xi^T x} = \hat{f}(\xi) = \hat{f}(V\eta),
\end{aligned}$$

which is correct since

$$\xi^T \begin{pmatrix} v & V \end{pmatrix}\begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 & \xi^T V \end{pmatrix}\begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 & \xi^T V \end{pmatrix}\begin{pmatrix} t \\ y \end{pmatrix} = \xi^T \begin{pmatrix} v & V \end{pmatrix}\begin{pmatrix} t \\ y \end{pmatrix},$$

which shows that $\xi \in v^\perp$ is essential for the validity of (3.3.5). □

Also the X-ray transform has an inversion formula whose proof follows in some way dual lines to the one for the Radon transform.

Theorem 3.3.4 (**Inverse X-ray transform**). *For $f \in \mathscr{S}(\mathbb{R}^d)$,*

$$X^* I^{-1} X f = f, \tag{3.3.6}$$

*with the **dual X-ray transform***

$$X^* f = \frac{1}{2\pi\,|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} f(v, VV^T \cdot)\, dv. \tag{3.3.7}$$

Remark 3.3.5. Note that $VV^T x$ is the **orthogonal projection** of x on the hyperplane spanned by the complement matrix V, i.e. to $v^\perp$. Indeed,

$$v^T P_V x := \underbrace{v^T V}_{=0} V^T x = 0, \qquad x \in \mathbb{R}^d,$$

hence $P_v \mathbb{R}^d = v^\perp$ since rank $P_V = d-1$; P_V is also easily verified to be a projection:

$$P_V^2 = V \underbrace{V^T V}_{=I} V^T = VV^T = P_V.$$

Nevertheless, keep in mind that $VV^T \neq I$, only $V^T V = I$.

Proof: We use the integral formula

$$\int_{\mathbb{R}^d} g(\xi)\, d\xi = \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{v^\perp} |\eta|_2\, g(\eta)\, d\eta dv \tag{3.3.8}$$

from (Natterer, 1986, VII.(2.8)) for the inverse Fourier transform

$$\begin{aligned}
f(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \underbrace{e^{ix^T\xi} \hat{f}(\xi)}_{=g(\xi)}\, d\xi = \frac{1}{(2\pi)^d} \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{v^\perp} |\eta|_2\, e^{ix^T\eta} \hat{f}(\eta)\, d\eta dv \\
&= \frac{1}{(2\pi)^d} \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{v^\perp} |\eta|_2\, e^{ix^T\eta}\, (Xf(v,\cdot))^\wedge(\eta)\, d\eta dv \\
&= \frac{1}{(2\pi)^d} \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^{d-1}} |Vy|_2\, e^{ix^T Vy}\, (Xf(v,\cdot))^\wedge(Vy)\, dy dv \\
&= \frac{1}{(2\pi)^d} \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^{d-1}} |Vy|_2\, e^{i(sv+VV^Tx)^T Vy}\, (Xf(v,\cdot))^\wedge(Vy)\, dy dv \\
&= \frac{1}{(2\pi)^d} \frac{1}{|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^{d-1}} |z|_2\, e^{i(VV^Tx)^T z}\, (Xf(v,\cdot))^\wedge(z)\, dz dv \\
&= \frac{1}{2\pi\,|\mathbb{S}^{d-2}|} \int_{\mathbb{S}^{d-1}} I^{-1} Xf(v, VV^T x)\, dv,
\end{aligned}$$

which completes the proof. □

Again, up to normalization[6], the dual X-ray transform is indeed a dual to the

[6] That normalization is sometimes an unavoidable issue, we already learned at the beginning from the Fourier transform.

X-ray transform as the computation

$$\begin{aligned}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^{d-1}} Xf(v,Vy)g(Vy)\,dydv &= \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} f(tv+Vy)g(Vy)\,dtdydv\\ &= \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^d}\int_{\mathbb{R}} f(\hat{V}x)g(VV^T\hat{V}x)\,dxdv = \int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^d} f(x)g(VV^Tx)\,dxdv\\ &= \int_{\mathbb{R}^d} f(x)\int_{\mathbb{S}^{d-1}} g(v,VV^Tx)\,dvdx = 2\pi\left|\mathbb{S}^{d-2}\right|\int_{\mathbb{R}^d} f(x)X^*g(x)\,dx\end{aligned}$$

shows, where we set $x = \begin{pmatrix} t \\ y \end{pmatrix}$ and note that

$$V^T\hat{V}x = V^T\begin{pmatrix} v & V\end{pmatrix}\begin{pmatrix} t \\ y\end{pmatrix} = \begin{pmatrix} V^Tv & V^TV\end{pmatrix}\begin{pmatrix} t \\ y\end{pmatrix} = \begin{pmatrix} 0 & I\end{pmatrix}\begin{pmatrix} t \\ y\end{pmatrix} = y.$$

Finally, there is also a filtered backprojection for the X-ray transform and the proof follows the same computational lines as the one for the Radon transform.

Theorem 3.3.6. *For $f \in \mathscr{S}(\mathbb{R}^d)$ and $g \in L_1(\mathbb{S}^{d-1}\times\mathbb{R}^d)$ we have*

$$(X^*g) * f = X^*(g * Xf), \tag{3.3.9}$$

Proof: We compute

$$\begin{aligned}(X^*g * f)(x) &= \int_{\mathbb{R}^d} X^*g(x-t)f(t)\,dt\\ &= \frac{1}{2\pi\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{S}^{d-1}} g\left(v, VV^T(x-t)\right)f(t)\,dvdt\\ &= \frac{1}{2\pi\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}} g\left(v, VV^T(x-sv-Vy)\right)f(sv+Vy)\,dsdydv\\ &= \frac{1}{2\pi\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^{d-1}} g\left(v, VV^Tx-Vy\right)\int_{\mathbb{R}} f(sv+Vy)\,ds\,dydv\\ &= \frac{1}{2\pi\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{S}^{d-1}}\int_{\mathbb{R}^{d-1}} g\left(v, VV^Tx-Vy\right)Xf(v,Vy)\,dydv\\ &= \frac{1}{2\pi\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{S}^{d-1}} (g*Xf)\left(v, VV^Tx\right)\,dv = X^*(g*Xf),\end{aligned}$$

where we note that the convolution $g * Xf$ ran only over $x \in v^\perp$ in accordance to what we said before. □

It follows that the method of applying the filtered backprojection sketched in Remark 3.2.10 can also be applied to the X-ray transform, but of course also the same practical challenges are present.

Remark 3.3.7. In our computations we frequently used the substitution trick

$$\int_{\mathbb{R}}\int_{v^\perp} f(tv+w)dwdt = \int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}} f(tv+Vy)dydt = \int_{\mathbb{R}^d} f(\hat{V}x)\,dx = \int_{\mathbb{R}^d} f(x)\,dx$$

with $x = \begin{pmatrix} t \\ y\end{pmatrix}$ and using the fact that $\hat{V}$ is orthogonal, hence $|\det\hat{V}| = 1$. In the future, we may keep computations simpler and do this transform implicitly. It is a good exercise, however, to work out the explicit form.

3.4 Uniqueness for Compactly Supported Functions

We have already seen in Corollary 3.2.6 that the Radon transform is injective, hence, as a consequence $Rf = 0$ implies that $f = 0$. In this section, we collect some more results of this type. Such results are especially nice if f is compactly supported which is the standard assumption in Computed Tomography – objects without compact support would not fit into the machine. Many of the results are based on the fact that the Fourier transform of a compactly supported continous function is **analytic**, i.e., infinitely differentiable and its Taylor series converges everywhere. Let us recall the basics.

Definition 3.4.1. 1. A **polynomial** p is a function of the form

$$p(x) = \sum_{\alpha \in \mathbb{N}_0^d} p_\alpha x^\alpha = \sum_{\alpha \in \mathbb{N}_0^d} p_\alpha x_1^{\alpha_1} \cdots x_d^{\alpha_d}, \qquad \#\{\alpha : p_\alpha \neq 0\} < \infty;$$

if the number of nonzero terms is infinite, we speak of a **power series**. By Π we denote the space of all polynomials.

2. A polynomial is called **homogeneous** of degree n if it is of the form

$$p(x) = \sum_{|\alpha|=n} p_\alpha x^\alpha, \qquad |\alpha| := \sum_{j=1}^{d} \alpha_j.$$

We denote the vector space of all homogeneous polynomials of degree n by Π_n^0 and write

$$\Pi_n = \bigoplus_{j=1}^{n} \Pi_j^0$$

fot the space of all polynomials of total degree (at most) n.

Any polynomial and any power series can be written in terms of its homogeneous components as

$$p(x) = \sum_{n=0}^{\infty} p_n(x), \qquad p_n \in \Pi_n^0.$$

Moreover, any homogeneous polynomial of degree n has the property that

$$p(\lambda x) = \sum_{|\alpha|=n} p_\alpha \lambda^{|\alpha|} x^\alpha = \lambda^n p(x). \tag{3.4.1}$$

Being equipped with this machinery, we can state the first result which defines certain sets of directions that yield uniqueness via the Radon transform.

Proposition 3.4.2. *Let $\mathcal{V} \subseteq \mathbb{S}^{d-1}$ be a set of directions such that $p(\mathcal{V}) = 0$ implies $p = 0$ for any homogeneous polynomial. If $f \in C_{00}^\infty(\mathbb{R}^d)$ and $Rf(\mathcal{V}, \cdot) = 0$, then $f = 0$.*

Proof: Since $f \in C_{00}^{\infty}(\mathbb{R}^d)$, the Fourier transform $\hat{f}$ is analytic and can thus be written as

$$\hat{f}(\xi) = \sum_{n=0}^{\infty} a_n(\xi), \qquad a_n \in \Pi_n^0.$$

The assumption on f yields that

$$0 = (Rf(v,\cdot))^{\wedge}(\sigma) = \hat{f}(\sigma v) = \sum_{n=0}^{\infty} \sigma^n a_n(v), \qquad v \in \mathscr{V}, \quad \sigma \in \mathbb{R}.$$

Hence, $a_n = 0$, $n \in \mathbb{N}_0$, yielding $\hat{f} = 0$ and therefore $f = 0$. □

We also get a nice result for the X-ray transform.

Proposition 3.4.3. *If $f \in C_{00}^{\infty}(\mathbb{R}^d)$ and $Xf(v,\cdot) = 0$ for infinitely many $v \in \mathbb{S}^{d-1}$, then $f = 0$.*

Proof: The assumption implies by Theorem 3.3.3 that $0 = \hat{f}(\xi)$ for $\xi \in v^{\perp}$ for an infinite set $\mathscr{V} \subseteq \mathbb{S}^{d-1}$ of directions. Hence, for any such v, and $\eta \in \mathbb{R}^d$, $\eta = v^T\eta\, v + \xi$, $\xi \in v^{\perp}$ we have that

$$\hat{f}(\eta) = \hat{f}(\eta) - \hat{f}(\xi) = \sum_{\alpha \in \mathbb{N}_0^d} \frac{1}{\alpha!} \frac{\partial^{|\alpha|}}{\partial x^{\alpha}} \hat{f}(\xi) \left(v^T\eta\right)^{|\alpha|} v^{\alpha} = \sum_{n=1}^{\infty} \left(v^T\eta\right)^n p_n(v)$$

for some $p_n \in \Pi_n^0$, $n \in \mathbb{N}$, hence $\hat{f}(\eta) = (v^T\eta) g_v$ for any $v \in \mathscr{V}$. Thus

$$\hat{f}(\eta) = g(\eta) \prod_{v \in \mathscr{V}} v^T\eta$$

has a zero of infinite order at $\eta = 0$ from which we conclude that $\hat{f} = 0$ and therefore $f = 0$. □

Unfortunately, a finite number of directions will never do the job. To that end, we consider the following result which is of a slightly more technical nature than Proposition 3.4.3.

Proposition 3.4.4. *Let $v_1, \ldots, v_n \in \mathbb{S}^{d-1}$, $\Omega \subset \mathbb{R}^d$ compact and[7] $f \in C_0^{\infty}(\Omega)$. For each $\Omega' \subset \Omega^{\circ}$ there exists a function $g \in C_0^{\infty}(\Omega)$ that coincides with f on Ω' but satisfies $Xg(v_j,\cdot) = 0$, $j = 1, \ldots, n$.*

Proof: This time we enforce a factor of the form $\prod v_j^T\eta$ in the Fourier transform. To that end, set

$$q(x) = \prod_{j=1}^{n} v_j^T x = \sum_{|\alpha| \le n} q_{\alpha}\, x^{\alpha},$$

and set

$$q(-iD) = q\left(-i\frac{\partial}{\partial x}\right) := \sum_{|\alpha| \le n} q_{\alpha} (-i)^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial x^{\alpha}}.$$

[7] The set $C_0^{\infty}(\Omega)$ consists of all infinitely differentiable functions on Ω that vanish on the boundary $\partial\Omega$ of Ω.

Now, we construct a function h such that

$$q(-iD)h = f. \tag{3.4.2}$$

For $n = 1$, the equation takes the form

$$-iv_1^T \nabla h = f, \qquad \nabla h = \left(\tfrac{\partial}{\partial x_j} h : j = 1, \dots, d\right),$$

and a solution is the line integral

$$h(x) = i \int_0^s f(t\, v_1 + y)\, dt, \qquad x = s v_1 + y, \quad y \in v^\perp;$$

this is the fundamental theorem of integration and differentiation. For $n > 1$, we repeat the construction inductively taking into account that

$$q(D)h = \left(-iv_1^T \nabla\right) \cdots \left(-iv_n^T \nabla\right) h.$$

Finally, choose any "mask" function $\psi \in C_0^\infty(\Omega)$ which is 1 on Ω' and set $g = q(-iD)\,(\psi h)$, so that, for $x \in \Omega'$,

$$g(x) = q(-iD)(\underbrace{\psi}_{\equiv 1} h)(x) = q(-iD)h(x) = f(x).$$

Then, by Theorem 2.1.5, 4)

$$\hat{g}(\xi) = q(\xi)\,(\psi h)^\wedge(\xi),$$

and therefore, by Theorem 3.3.3, for $j = 1, \dots, n$ and $\xi \in v_j^\perp$,

$$\left(Xg(v_j, \cdot)\right)^\wedge(\xi) = \hat{g}(\xi) = q(\xi)\,(\psi h)^\wedge(\xi) = (v_1^T\xi) \cdots \underbrace{(v_j^T\xi)}_{=0} \cdots (v_n^T\xi)\,(\psi h)^\wedge(\xi) = 0,$$

which gives $Xg(v_j, \cdot) = 0$ as claimed since the X-ray transform is only considered on the complement of the respective direction. □

In other words: *infinitely many directions yield uniqueness in the X-ray transform, finitely many don't.* Here, we have a clear cut and, first of all, this is not very promising. Obiously, we can only measure *finitely many* rays in practice as the number of pixels on a detector and the number of measurements are clearly finite. Note, however, $h = f - g$ satisfies $h(\Omega') = 0$ but still $Xh(v, \cdot) = Xf(v, \cdot)$, so all information on f is squeezed into the possibly small domain $\Omega \setminus \Omega'$.

The Fourier transform $\hat{g}$ of g vanishes of high order at the origin. Since

$$\hat{g}(0) = \int_{\mathbb{R}^d} e^{-i0^T x} g(x)\, dx = \int_\Omega g(x)\, dx,$$

the function g must oscillate on Ω. Higher orders of the zero yield even more oscillations as the following argument shows. Let

$$q_J(\xi) = \prod_{k \in J} v_j^T \xi, \qquad J \subset \{1, \dots, n\},$$

then

$$0 = q_J(D)\hat{g}(0) = (q_J(-i\cdot)g)^\wedge(\xi) = \int_\Omega g(x) q_J(-ix)\,dx,$$

hence g is perpendicular to the space spanned by the q_J. These **vanishing moments** are known in wavelet theory, cf. (Mallat, 2009) and imply a certain amount of oscillarion of g on Ω. This makes g a "noisy" function.

We finish this section with a result on the (realistic) divergent beam transform that we have not considered yet. Here, we restrict the support of the function to the unit ball and locate the X-ray source outside the unit ball.

Theorem 3.4.5. *Let $\mathscr{X} \in \mathbb{R}^d \setminus \mathbb{B}^d$ be an infinite set and $f \in C_{00}^\infty(\mathbb{B}^d)$. If $Df(\cdot, x) = 0$, $x \in \mathscr{X}$, then $f = 0$.*

Proof: Let[8] v^* be an accumulation point of $\left\{\frac{x}{|x|_2} : x \in \mathscr{X}\right\}$ and choose $\varepsilon > 0$ such that f is supported on $B_{1-2\varepsilon}(0) = \{x : |x|_2 \leq 1 - 2\varepsilon\}$; this is possible due to the assumption that the compact support of f is contained in the interior[9] of $\mathbb{B}^d$. Since

$$x^T v^* = \underbrace{|x|_2}_{>1} \underbrace{\left(\frac{x}{|x|_2}\right)^T v_*}_{\to 1} \geq 1 - \frac{\varepsilon}{2}$$

for infinitely many points sufficiently close to the accumulation point v^*, we can find an open neighborhood $U(v^*) \subset \mathbb{S}^{d-1}$ of v^* such that

$$\#\mathscr{X}' = \left\{x : x^T v > 1 - \varepsilon, v \in U(v^*)\right\} = \infty$$

and we have infinitely many footpoints for infinitely many rays. Now, for $x \in \mathscr{X}'$ and $v \in U(v^*)$,

$$\begin{aligned}
\int_{\mathbb{S}^{d-1}} Df(w,x)(v^T w)^{1-d}\,dw &= \int_{\mathbb{S}^{d-1}} \int_0^\infty f(x+tw)(v^T w)^{1-d}\,dt dw \\
&= \int_{\mathbb{R}^d} f(x+y)(v^T y)^{1-d}\,dy = \int_{\mathbb{R}^d} f(y)\left(v^T(y-x)^T\right)^{1-d}\,dy \\
&= \int_{|y|_2<1-2\varepsilon} f(y)\left(v^T y - v^T x\right)^{1-d}\,dy, \qquad (3.4.3)
\end{aligned}$$

due to the support of f, where

$$v^T y - v^T x < \underbrace{|v|_2}_{=1}|y|_2 - \underbrace{v^T x}_{>1-\varepsilon} \leq (1-2\varepsilon) - (1-\varepsilon) = -\varepsilon$$

whenever $f(y) \neq 0$, hence $\left(v^T y - v^T x\right)^{1-d}$ is well defined as is the integral itself.

[8] This is correct: we choose the direction according to the points from where they emerge.

[9] Recall that, by definition, $C_{00}(\Omega)$ is the set of continuous functions with a compact support $\Omega' \subset \Omega^\circ$.

Replacing[10] y in (3.4.3) by $y = sv + w$, $w \in V^\perp$, we this get

$$\begin{aligned}\int_{\mathbb{S}^{d-1}} Df(w,x)(v^T w)^{1-d}\,dw &= \int_{\mathbb{R}^d} f(y)\left(v^T y - v^T x\right)^{1-d}\,dy \\ &= \int_{\mathbb{R}}\int_{v^\perp} f(sv+w)\left(v^T(sv-w) - v^T x\right)^{1-d}\,dw\,ds \\ &= \int_{\mathbb{R}}\left(s - v^T x\right)^{1-d}\underbrace{\int_{v^\perp} f(sv+w)dw}_{=Rf(v,s)}\,ds = \int_{\mathbb{R}}\left(s - v^T x\right)^{1-d} Rf(v,s)\,ds,\end{aligned}$$

and the assumption that $Df(\cdot, x) = 0$, $x \in \mathscr{X}$, allows us to conclude that

$$0 = \int_{\mathbb{R}}\left(s - v^T x\right)^{1-d} Rf(v,s)\,ds, \qquad v \in U(v^*), \quad x \in \mathscr{X}'. \tag{3.4.4}$$

The power series expansion

$$\begin{aligned}\left(s - v^T x\right)^{1-d} &= \sum_{n=0}^{\infty}(-1)^n \frac{(n+d-2)!}{(d-2)!}\frac{(v^T x)^{1-d-n}}{n!}\,s^n \\ &= \sum_{n=0}^{\infty}\underbrace{(-1)^n\binom{n+d-2}{n}}_{=:c_n\neq 0}(v^T x)^{1-d-n}\,s^n\end{aligned}$$

converges uniformly in $U(v^*) \times \mathscr{X}'$ for $|s| \le 1 - 2\varepsilon$, so that

$$0 = \sum_{n=0}^{\infty} c_n (v^T x)^{1-d-n}\int_{\mathbb{R}} s^n Rf(v,s)\,ds, \qquad v \in U(v^*), \quad x \in \mathscr{X}'. \tag{3.4.5}$$

As we will see in Proposition 3.5.1, there exist *homogeneous* polynomials $p_n \in \Pi_n^0$ such that

$$\int_{\mathbb{R}} s^n Rf(v,s)\,ds = p_n(v), \qquad v \in \mathbb{S}^{d-1},$$

and therefore

$$0 = \sum_{n=0}^{\infty} c_n \frac{p_n(v)}{(x^T v)^n}, \qquad v \in U(v^*), \quad x \in \mathscr{X}'.$$

The set $\{w \in \mathbb{R}^d : \#w^T\mathscr{X}' < \infty\}$ is at most a $(d-1)$-dimensional subspace of $\mathbb{R}^d$, see Exercise 3.4.2. Since $U(v^*)$ is open, there exists for almost any $v \in U(v^*)$, a sequence $x_j \in \mathscr{X}'$, $j \in \mathbb{N}$, such that all the numbers

$$t_j := (v^T x_j)^{-1} \in \left(0, \frac{1}{1-\varepsilon}\right] \qquad j \in \mathbb{N},$$

are disjoint and therefore the power series

$$g_v(t) := \sum_{n=0}^{\infty} c_n p_n(v)\,t^n$$

[10] Our standard trick.

has infinitely many zeros[11]. This implies that all p_n vanish on a dense subset of $U(v^*)$, hence, being their polynomials, $p_n = 0$, $n \in \mathbb{N}_0$, and the density of monomials yields that $Rf(v,\cdot) =$ for *any* $v \in \mathbb{S}^{d-1}$, hence $f = 0$. □

Exercise 3.4.1 Show *in one line* that a homogeneous polynomial is uniquely defined by its values on $\mathbb{S}^{d-1}$. ◊

Exercise 3.4.2 To complete the proof of Theorem 3.4.5 assume that $\mathscr{X} \subset \mathbb{R}^d \setminus \mathbb{B}^d$ and show that:

1. $\{w : \#w^T\mathscr{X} < \infty\}$ is a linear subspace,
2. for any finite $\mathscr{Y} \subset \mathbb{R}^d$ there exist only *finitely many* $w \in \mathbb{R}^d$ such that $\#w^T\mathscr{Y} < \#\mathscr{Y}$.
   Hint: consider $\{y - y' : y, y' \in \mathscr{Y}\}$.

Use this to show that $\{w : \#w^T\mathscr{X} < \infty\}$ is a linear subspace of $\mathbb{R}^d$ of dimension at most $d-1$. ◊

Remark 3.4.6. The proof of Theorem 3.4.5 is quite remarkable. It tells us that the principal direction of the beams is the one determined by the[12] accumulation point of the projections of $\mathscr{X}$ on the unit sphere. The open neighborhood $U(v^*)$ then represents a small cone around this principal direction to get an infinity of directions.

3.5 Consistency

In the proof of Theorem 3.4.5 we already used the fact that the integral of the Radon transform against a homogeneous polynomial gives a homogeneous polynomial again. We will prove that important fact and show that in fact it even *characterizes* the Radon transform and the X-ray transform.

Proposition 3.5.1. *For $f \in \mathscr{S}(\mathbb{R}^d)$ and $n \in \mathbb{N}_0$, there exist homogeneous polynomials p_n and q_n of degree n such that*

$$\int_{\mathbb{R}} s^n\, Rf(v,s)\, ds = p_n(v), \qquad v \in \mathbb{S}^{d-1}, \tag{3.5.1}$$

and

$$\int_{v^\perp} (x^T y)^n Xf(v,x)\, dx = q_n(y), \qquad y \in v^\perp; \tag{3.5.2}$$

moreover, the polynomial q_n is independent of v.

Proof: For (3.5.1), we compute

$$\begin{aligned}\int_{\mathbb{R}} s^n\, Rf(v,s)\, ds &= \int_{\mathbb{R}} s^n \int_{v^\perp} f(sv+w)\, dw ds \\ &= \int_{\mathbb{R}}\int_{v^\perp} \Big((sv+w)^T v\Big)^n f(sv+w)\, dw ds = \int_{\mathbb{R}^d} (x^T v)^n f(x)\, dx,\end{aligned}$$

[11] And these zeros have an accumulation point if $\mathscr{X}'$ is bounded.
[12] If there are several ones, the argument holds for each of them, pick your favorite.

which is a homogeneous polynomial with respect to v. (3.5.2) is obtained in almost the same way by

$$\begin{aligned}\int_{v^\perp} (x^T y)^n Xf(v,x)\,dx &= \int_{v^\perp} (x^T y)^n \int_{\mathbb{R}} f(x+tv)\,dtdx \\ &= \int_{\mathbb{R}^d} \left((z-tv)^T y\right)^n f(z)dz = \int_{\mathbb{R}^d} (z^T y)^n f(z)dz,\end{aligned}$$

which is independent of v as claimed. □

What makes Proposition 3.5.1 remarkable is the fact that is also has a converse. This also motivates the following definition which we give before we state and proof the converses.

Definition 3.5.2. The conditions (3.5.1) and (3.5.2) are called **Helgasson consistency conditions** for the Radon transform and the X-ray transform, respectively.

Theorem 3.5.3 (Consistency of the Radon transform). *If $g \in \mathscr{S}(\mathbb{S}^{d-1} \times \mathbb{R})$ is an even function, i.e., $g(-v,-s) = g(v,s)$ and*

$$\int_{\mathbb{R}} g(v,s)s^n\,ds = p_n(v), \qquad p_n \in \Pi_n^0, \qquad n \in \mathbb{N}_0, \tag{3.5.3}$$

then there exists $f \in \mathscr{S}(\mathbb{R}^d)$ such that $g = Rf$.

Proof: The first step is straightforward. In view of Fourier Slice Theorem 3.2.5, we define f by its Fourier transform

$$\hat{f}(\sigma v) = (g(v,\cdot))^\wedge(\sigma), \qquad v \in \mathbb{S}^d, \quad \sigma \in \mathbb{R}, \tag{3.5.4}$$

which is well-defined since g is even and hence

$$\hat{f}\left((-\sigma)(-v)\right) = (g(-v,\cdot))^\wedge(-\sigma) = (g(-v,\cdot))^\wedge(-\sigma) = (g(-v,-\cdot))^\wedge(\sigma) = \hat{f}(\sigma v).$$

The main task is to show that the f defined this way really belongs to $\mathscr{S}(\mathbb{R}^d)$. To encorporate polynomials, we write the exponential as

$$e^{-it} = \sum_{j=0}^{n-1} \frac{(-it)^j}{j!} + e_n(-it), \qquad e_n(t) = t^n \sum_{j=0}^{\infty} \frac{t^j}{(j+n)!}, \tag{3.5.5}$$

where $e : \mathbb{R} \to \mathbb{R}$ is an analytic function with a zero of order n at the origin. Therefore, by (3.5.3),

$$\begin{aligned}\hat{f}(\sigma v) &= (g(v,\cdot))^\wedge(\sigma) = \int_{\mathbb{R}} e^{-i\sigma s} g(v,s)\,ds \\ &= \sum_{j=0}^{n-1} \frac{(-i\sigma)^j}{j!} \int_{\mathbb{R}} s^j g(v,s)\,ds + \int_{\mathbb{R}} e_n(-i\sigma s) g(v,s)\,ds \\ &= \sum_{j=0}^{n-1} \frac{(-i\sigma)^j}{j!} p_j(v) + \int_{\mathbb{R}} e_n(-i\sigma s) g(v,s)\,ds \\ &= \sum_{j=0}^{n-1} \frac{(-i)^j}{j!} p_j(\sigma v) + \int_{\mathbb{R}} e_n(-i\sigma s) g(v,s)\,ds,\end{aligned}$$

hence

$$\hat{f}(\xi) = \sum_{j=0}^{n-1} \frac{(-i)^j}{j!} p_j(\xi) + \int_{\mathbb{R}} e_n(-i\sigma s) g(v,s)\, ds, \qquad \xi \in \mathbb{R}^d. \tag{3.5.6}$$

Differentiablity of the polynomial part is no problem, for the integral, we have to express $\frac{\partial}{\partial \xi}$ by $\frac{\partial}{\partial v}$ and $\frac{\partial}{\partial \sigma}$, which is getting a bit technical. Since $\sigma = |\xi|_2$ and $v = \frac{\xi}{\sigma}$, the chain rule gives[13]

$$\frac{\partial}{\partial \xi_j} = \frac{\partial \sigma}{\partial \xi_j}\frac{\partial}{\partial \sigma} + \sum_{k=1}^{d} \frac{\partial v_k}{\partial \xi_j}\frac{\partial}{\partial v_k} = \frac{\xi_j}{|\xi|_2}\frac{\partial}{\partial \sigma} + \frac{1}{\sigma}\sum_{k=1}^{d} \delta_{jk}\frac{\partial}{\partial v_k} = v_j \frac{\partial}{\partial \sigma} + \frac{1}{\sigma}\frac{\partial}{\partial v_j}.$$

Taking into accout that $e_n(t) = t^n b_n(t)$ with $b \in C^\infty(\mathbb{R})$ and differentiating under the integral sign[14], we get that

$$\begin{aligned}
\frac{\partial}{\partial \xi_j}(-i\sigma s)^n b_n(-i\sigma s) g(v,s) &= \left(v_j \frac{\partial}{\partial \sigma} + \frac{1}{\sigma}\frac{\partial}{\partial v_j}\right) \sigma^n \left((-is)^n b_n(-i\sigma s)\right) g(v,s) \\
&= \sigma^{n-1} v_j \left((-is)^n b(-i\sigma s) + \sigma(-is)^{n+1} b_n'(-i\sigma s)\right) g(v,s) \\
&\quad + \sigma^{n-1}\left((-is)^n b(-i\sigma s)\frac{\partial}{\partial v_j} g(v,s)\right).
\end{aligned}$$

It follows that if a function h is of the form

$$h(v,\sigma) = \sigma^n b(\sigma) a(v), \qquad b \in C^\infty(\mathbb{R}), \quad a \in \mathscr{S}(\mathbb{S}^{d-1}),$$

then $\frac{\partial^{|\alpha|}}{\partial \xi^\alpha} h$ is of the same form with n replaced by $n - |\alpha|$, and in particular with $|\alpha| = n$ we have that

$$\frac{\partial^{|\alpha|}}{\partial \xi^\alpha}\hat{f}(\xi) = \int_{\mathbb{R}} b(-i\sigma s) a(v,s)\, dx, \qquad b \in C^\infty(\mathbb{R}), \quad a \in \mathscr{S}(\mathbb{S}^{d-1}), \tag{3.5.7}$$

and the function $\hat{f}$ is indeed C^∞. Rapid decay is obtained by muliplying with $\xi^\alpha = \sigma^{|\alpha|} v^\alpha$ and taking into account that g is a Schwartz function. □

Theorem 3.5.4 (Consistency of the X-ray transform). *If $g \in \mathscr{S}(T\mathbb{S}^{d-1}\times)$ satisfies $g(\cdot, x) = 0$ for $|x|_2 \geq c$ and*

$$\int_{v^\perp} (x^T y)^n g(v,x)\, dx = q_n(y), \qquad y \in v^\perp, \quad q_n \in \Pi_n^0, \qquad n \in \mathbb{N}_0, \tag{3.5.8}$$

with q_n being independent of v, then there exists $f \in \mathscr{S}(\mathbb{R}^d)$ such that $g = Xf$.

Proof: We will eventually use Theorem 3.5.3 for the proof. To that end, we set for $v \in \mathbb{S}^{d-1}$,

$$h_v(w,s) := \int\limits_{\{x \in v^\perp : w^T x = s\}} g(v,x)\, dx, \qquad w \in v^\perp, \tag{3.5.9}$$

[13] There is a little bit more to it, we take the "physicist approach" here

[14] Check why we are permitted to do so.

which is well-defined due to the compact support of g with respect to the second variable, and compute for $n \in \mathbb{N}_0$

$$\begin{aligned}\int_{\mathbb{R}} s^n h_v(w,s)\,ds &= \int_{\mathbb{R}} \int_{\{x\in v^\perp : w^T x = s\}} s^n g(v,x)\,dx\,ds \\ &= \int_{\mathbb{R}} \int_{\{x\in v^\perp : w^T x = s\}} (x^T w)^n g(v,x)\,dx\,ds = \int_{v^\perp} (x^T w)^n g(v,x)\,dx = q_n(w),\end{aligned}$$

which implies by Theorem 3.5.3 that $h_v = Rf$ for some f. Since q_n does not depend on v by assumption, so does h_v. For $d = 2$ this is sufficient since then (3.5.9) says that $g = h_v = Rf = Xf$. For $n > 2$, we consider, for fixed v, $w \in v^\perp$ and $s \in \mathbb{R}$, the integral

$$\begin{aligned}\int_{\{x\in v^\perp : w^T x = s\}} Xf(v,x)\,dx &= \int_{\{x\in v^\perp : w^T x = s\}} \int_{\mathbb{R}} f(x+tv)\,dt\,dx \\ &= \int_{v^\perp \cap w^\perp} \int_{\mathbb{R}} f(sw+y+tv)\,dt\,dy = \int_{w^\perp} f(sw+x)\,dx = Rf(w,s) = h(w,s).\end{aligned}$$

In other words, the Radon transforms of $g(v,\cdot)$ and $Xf(v,\cdot)$ coincide on $v^\perp$ for any $v \in \mathbb{S}^{d-1}$ and since the Radon transform is injective, it follows that $g = Xf$. □

Remark 3.5.5. As shown in (Natterer, 1986, II.4), there is even a support property for f in Theorem 3.5.3 and Theorem 3.5.4: if $g(v,s) = 0$ for $|s| \geq a$ or $g(v,x) = 0$ for $|x|_2 \geq a$, respectively, then $f(x) = 0$ for $|x| \geq a$. This is due to an explicit inversion formula based on Gegenbauer polynomials, see (Natterer, 1986, II.2, Theorem 2.3).

3.6 Stability

The results in Section 3.4 about uniqueness showed that for compactly supported functions, the Radon transform and the X-ray transform are injective, that is, if the transforms as *linear* operators yield zero, them the function must be zero. However, there exist even quantitative statements in terms of proper norms which we are going to consider now. Let us start with the norms, as an extension of what we encountered in (2.3.17).

Definition 3.6.1. The **Sobolev norm** $\|\cdot\|_{H^r(\mathbb{R}^d)}$ is defined as

$$\|\cdot\|^2_{H^r(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \left(1 + |\xi|_2^2\right)^r \left|\hat{f}(\xi)\right|^2 d\xi, \tag{3.6.1}$$

and the **Sobolev space** $H^r(\mathbb{R}^d)$ consists of all functions for which this norm is finite. Moreover, we define the Sobolev spaces $H^r(\mathbb{S}^{d-1}\times\mathbb{R})$ and $H^r(T\mathbb{S}^{d-1})$ for the Radon transform and the X-ray transform by

$$\|f\|^2_{H^r(\mathbb{S}^{d-1}\times\mathbb{R})} := \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} (1+\sigma^2)^r \left|\hat{f}_v(\sigma)\right|^2 d\sigma\,dv \tag{3.6.2}$$

and

$$\|f\|^2_{H^r(T\mathbb{S}^{d-1})} := \int_{\mathbb{S}^{d-1}} \int_{v^\perp} \left(1+|\eta|_2^2\right)^r \left|\hat{f}_v(\eta)\right|^2 \, d\eta dv, \tag{3.6.3}$$

respectively.

A stronger and quantitative form of invertibility is given by the following concept. In fact, an operator is stable in the sense of Definition 3.6.2 if it is continuous and has a continuous inverse.

Definition 3.6.2. A linear operator $T : X \to Y$ between two Banach spaces is called **stable** if there exist two constants $0 < A < B < \infty$ such that

$$A\|f\|_X \leq \|f\|_Y \leq B\|f\|_X, \qquad f \in X. \tag{3.6.4}$$

Since it will become relavant in the proof of the following theorem, let us recall the following important concept.

Definition 3.6.3. For $\Omega \subseteq \mathbb{R}^d$ we denote by $C^\infty_{00}(\Omega)$ the set of all compactly supported C^∞ functions f such that $\operatorname{supp} f \subset \Omega^\circ$, i.e., the compact support is a subset of the *interior* of Ω.

With respect to the correct spaces[15], the Radon transform and the X-ray transform are indeed stable.

Theorem 3.6.4. *For any compact $\Omega \subset \mathbb{R}^d$ and $r > 0$ there exist constants $0 < A_r < B_r(\Omega) < \infty$ such that*

$$A_r\|f\|_{H^r(\mathbb{R}^d)} \leq \|Rf\|_{H^{r+(d-1)/2}(\mathbb{S}^{d-1}\times\mathbb{R})} \leq B_r(\Omega)\|f\|_{H^r(\mathbb{R}^d)} \tag{3.6.5}$$

and

$$A_r\|f\|_{H^r(\mathbb{R}^d)} \leq \|Xf\|_{H^{r+1/2}(T\mathbb{S}^{d-1})} \leq B_r(\Omega)\|f\|_{H^r(\mathbb{R}^d)} \tag{3.6.6}$$

hold for any $f \in C^\infty_{00}(\mathbb{R}^d)$.

Remark 3.6.5. The notation in (3.6.5) and (3.6.6) is deliberate and correct: the proof will show that only the upper bound depends on Ω.

Proof: To prove (3.6.5), we use the Fourier Slice Theorem for the Radon transform to compute

$$\begin{aligned}
\|Rf\|^2_{H^{r+(d-1)/2}(\mathbb{S}^{d-1}\times\mathbb{R})} &= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} (1+\sigma^2)^{r+(d-1)/2} \left|(Rf(v,\cdot))^\wedge(\sigma)\right|^2 \, d\sigma dv \\
&= \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} (1+\sigma^2)^{r+(d-1)/2} \left|\hat{f}(v\sigma)\right|^2 \, d\sigma dv \\
&= 2\int_{\mathbb{S}^{d-1}} \int_0^\infty (1+\sigma^2)^{r+(d-1)/2} \left|\hat{f}(v\sigma)\right|^2 \, d\sigma dv \\
&= 2\int_{\mathbb{R}^d} |\xi|_2^{1-d}(1+|\xi|_2^2)^{r+(d-1)/2} \left|\hat{f}(\xi)\right|^2 \, d\xi,
\end{aligned}$$

[15] And these will be ones we just introduced in Definition 3.6.1 – what a coincidence.

while for (3.6.6) we use

$$\begin{aligned}\|Xf\|^2_{H^{r+1/2}(T\mathbb{S}^{d-1})} &= \int_{\mathbb{S}^{d-1}}\int_{v^\perp}\left(1+|\eta|_2^2\right)^{r+\frac{1}{2}}\left|(Xf(v,\cdot))^\wedge(\eta)\right|^2\,d\eta dv\\ &= \int_{\mathbb{S}^{d-1}}\int_{v^\perp}\left(1+|\eta|_2^2\right)^{r+\frac{1}{2}}\left|\hat f(\eta)\right|^2\,d\eta dv\\ &= \left|\mathbb{S}^{d-2}\right|\int_{\mathbb{R}^d}|\xi|_2^{-1}(1+|\xi|_2^2)^{r+\frac{1}{2}}\left|\hat f(\xi)\right|^2\,d\xi,\end{aligned}$$

which are both of the form

$$F(\alpha) = C_\alpha\int_{\mathbb{R}^d}|\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat f(\xi)\right|^2\,d\xi,\qquad 1\le\alpha\le d-1,\tag{3.6.7}$$

for $\alpha = d-1$ and $\alpha = 1$. Since $\alpha > 0$,

$$|\xi|_2^{-\alpha} = \left(|\xi|_2^2\right)^{-\alpha/2} \ge \left(1+|\xi|_2^2\right)^{-\alpha/2},$$

it follows that

$$F(\alpha)\ge C_\alpha\int_{\mathbb{R}^d}(1+|\xi|_2^2)^r\left|\hat f(\xi)\right|^2\,d\xi = C_\alpha\,\|f\|^2_{H^r(\mathbb{R}^d)},\tag{3.6.8}$$

which is the lower estimate in (3.6.5) and (3.6.6).

For the upper estimate we note that for $|\xi|_2\ge 1$,

$$|\xi|_2^2\ge\frac{1}{2}\left(1+|\xi|_2^2\right)\qquad\Rightarrow\qquad |\xi|_2^{-\alpha}\le 2^{\alpha/2}\left(1+|\xi|_2^2\right)^{-\alpha/2}$$

and therefore

$$\begin{aligned}&\int\limits_{|\xi|_2\ge1}|\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat f(\xi)\right|^2\,d\xi\\ &\quad\le\ 2^{\alpha/2}\int\limits_{|\xi|_2\ge1}\left(1+|\xi|_2^2\right)^{-\alpha/2}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat f(\xi)\right|^2\,d\xi\\ &\quad\le\ 2^{\alpha/2}\int_{\mathbb{R}^d}(1+|\xi|_2^2)^r\left|\hat f(\xi)\right|^2\,d\xi = 2^{\alpha/2}\|f\|^2_{H^r(\mathbb{R}^d)}.\end{aligned}$$

To estimate the remaining integral, we first note that

$$\int\limits_{|\xi|_2\le1}|\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat f(\xi)\right|^2\,d\xi\le\max_{|\xi|_2\le1}\left|\hat f(\xi)\right|^2\int\limits_{|\xi|_2\le1}|\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\,d\xi$$

and the integral is some constant independent of f, cf. Exercise 3.6.1. We choose a function $\psi\in C^\infty_{00}(\mathbb{R}^d)$ that is equal to one on $\Omega = \operatorname{supp} f$ and set $\psi_\xi := e^{-i\xi^T\cdot}\psi$.

Then, by Theorem 2.1.8,

$$\begin{aligned}
\left|\hat{f}(\xi)\right|^2 &= \left|\int_\Omega e^{-i\xi^T x} f(x)\,dx\right|^2 = \left|\int_{\mathbb{R}^d} \psi_\xi(x) f(x)\,dx\right|^2 = \frac{1}{(2\pi)^d}\left|\int_{\mathbb{R}^d} \hat{\psi}_x(\eta)\hat{f}(\eta)\,d\eta\right|^2 \\
&= \frac{1}{(2\pi)^d}\left|\int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{-r/2} \hat{\psi}_\xi(\eta)\,\left(1+|\eta|^2\right)^{r/2} \hat{f}(\eta)\,d\eta\right|^2 \\
&\le \frac{1}{(2\pi)^d}\left(\int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{-r} \left|\hat{\psi}_\xi(\eta)\right|^2\,d\eta\right)\left(\int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{r} \hat{f}(\eta)\,d\eta\right) \\
&= \frac{1}{(2\pi)^d}\left(\int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{-r} \left|\hat{\psi}_\xi(\eta)\right|^2\,d\eta\right)\|f\|^2_{H^r(\mathbb{R}^d)},
\end{aligned}$$

and since

$$\hat{\psi}_\xi(\eta) = \int_{\mathbb{R}^d} e^{-i\eta^T x} e^{-i\xi^T x}\psi(x)\,dx = \hat{\psi}(\eta+\xi),$$

it follows that

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{-r} \left|\hat{\psi}_\xi(\eta)\right|^2\,d\eta &= \int_{\mathbb{R}^d} \left(1+|\eta|^2\right)^{-r} \left|\hat{\psi}(\eta+\xi)\right|^2\,d\eta \\
&= \int_{\mathbb{R}^d} \left(1+|\eta-\xi|_2^2\right)^{-r} \left|\hat{\psi}(\eta)\right|^2\,d\eta
\end{aligned}$$

is a continuous function in ξ and therefore

$$\max_{|\xi|_2\le 1}\left|\hat{f}(\xi)\right|^2 \le \|f\|_{H^r(\mathbb{R}^d)} \max_{|\xi|\le 1}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(1+|\eta|_2^2\right)^{-r}\left|\hat{\psi}_\xi(\eta)\right|^2\,d\eta =: C_2\|f\|^2_{H^r(\mathbb{R}^d)}. \tag{3.6.9}$$

In summary,

$$\begin{aligned}
&F(\alpha) \\
&\le \int_{|\xi|_2\ge 1} |\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat{f}(\xi)\right|^2\,d\xi + \int_{|\xi|_2\le 1} |\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\left|\hat{f}(\xi)\right|^2\,d\xi \\
&\le \left(2^{\alpha/2}+C_2\right)\|f\|^2_{H^r(\mathbb{R}^d)},
\end{aligned}$$

which gives the upper estimates and completes the proof. □

Exercise 3.6.1 Show that

$$\int_{|\xi|_2\le 1} |\xi|_2^{-\alpha}(1+|\xi|_2^2)^{r+\frac{\alpha}{2}}\,d\xi < \infty, \qquad 1\le\alpha\le d-1.$$

Hint: Use polar coordinates. ◊

The constant C_2 in (3.6.9) depends on Ω and grows with the size of Ω. However, in many situations this is not that bad.

Example 3.6.6. Let us consider a simple example where $\Omega_\lambda := \lambda\Omega$, $\lambda \geq 1$, and accordingly $\psi^\lambda = \psi(\cdot/\lambda)$ and $\psi^\lambda_\xi = e^{-i\xi^T\cdot}\psi(\cdot/\lambda)$ Thus, by (2.1.7), $\hat{\psi}^\lambda_\xi = \lambda^d\hat{\psi}\,(\lambda(\cdot+\xi))$ and therefore

$$\begin{aligned}
&\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(1+|\eta|_2^2\right)^{-r}\left|\hat{\psi}^\lambda_\xi(\eta)\right|^2\,d\eta\\
&= \frac{\lambda^d}{(2\pi)^d}\int_{\mathbb{R}^d}\left(1+|\eta-\xi|_2^2\right)^{-r}\left|\hat{\psi}(\lambda\eta)\right|^2\,d\eta\\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(1+|\eta/\lambda-\xi|_2^2\right)^{-r}\left|\hat{\psi}(\eta)\right|^2\,d\eta.
\end{aligned}$$

Although this is monotonically increasing in λ, the above term is bounded independently of λ and ξ by

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left|\hat{\psi}(\eta)\right|^2\,d\eta = \frac{1}{(2\pi)^d}\|\hat{\psi}\|_2^2 = \|\psi\|_2^2$$

which therefore gives an upper bound of $B_r(\Omega_\lambda)$ independently of λ.

3.7 The Limited Source Problem

In many practical applications, especially in more than two dimensions, the X-ray source is not located at *all* positions around the object, for example on $B_\rho(0)$ for some $\rho > 1$ if f is supported on $\mathbb{B}^d$, but one can only move the source along some curve. Let's make this formal.

Definition 3.7.1. Let $a : I \to \mathbb{R}^d \setminus \mathbb{B}^d$ be a differentiable **curve**, then the associated measurements along the curve are

$$Df(v, a(u)), \qquad v \in \mathbb{S}^{d-1}, \quad u \in I. \tag{3.7.1}$$

Remark 3.7.2. It is a standard assumption in Differential Geometry of curves that the curve a has a C^1 **parametrization**. Then the **arc length**

$$\ell_a(x) := \int_{u\leq x}|\dot{a}(u)|_2\,du, \qquad x \in I,$$

with $\dot{a} = \frac{d}{du}a$, is well-defined. Note that there is a subtle difference between the curve $a(I)$ as a subset of $\mathbb{R}^d$, and the curve as a function $I \to \mathbb{R}^d$. In fact, there are three classical ways to parametrize a curve:

1. with a free parameter $u \in I$, usually to describe the curve a in terms of simple functions,
2. with respect to the arc length which means that $|\dot{a}|_2 = 1$ and yields the *intrinsic* geometric parametrization,
3. with respect to the time t which gives a physical meaning to the derivatives, namely $\dot{a}$ as the velocity, $\ddot{a}$ as the acceleration and $a^{(3)}$ as the so-called jerk.

By means of appropriate **reparametrization** one can, of course, switch between the different types of parametrization.

The obvious question is whether this information is sufficient to reconstruct am function $f \in C_{00}^{\infty}(\mathbb{B}^d)$. And indeed, the answer is positive as soon as the curve is *nontrivial*, i.e., $\#a(I) \neq 1$. Since a is differentiable[16], it must assume infinitely many different values since with any two disjoint points there is a whole curve connecting them. Theorem 3.4.5 then tells us that the divergent beam transform is injective. The situation become worse, however, if we ask for *stable* reconstruction, that is, an estimate as in Theorem 3.6.4. This is impossible in general.

Example 3.7.3. Suppose there exists a hyperplane $H = (v, s)$ such that $H \cap a([0,1]) = \emptyset$ but $H \cap \mathbb{B}^d \neq \emptyset$. Then choose any $f \in C_{00}^{\infty}(\mathbb{B}^d)$ such that $f(H) \neq 0$, i.e., there existst some $x \in H$ such that $f(x) \neq 0$. The hyperplane H cuts $\mathbb{R}^d$ into the **halfspace** $H_+ = \{x : v^T x - s > 0\}$, the halfspace $H_- = \{x : v^T x - s < 0\}$ and H itself. Finally, define

$$f_+(x) = \begin{cases} f(x), & x \in H_+ \\ 0, & x \notin H_+, \end{cases}$$

which leads to a discontinuous function. However, the function

$$(v, t) \mapsto Df(v, a(t))$$

is C^{∞} provided that a is a smooth curve[17] since any ray from $a(t)$ meets f transversally only. So, f can be reconstructed, but since it has no finite Sobolev norm[18], there cannot be a stable embedding.

The example suggests that the source of the trouble may be the fact that there exists a hyperplane through the object that does not meet the curve a. This suggests the following definition due to Tuy (Tuy, 1983).

Definition 3.7.4 (**Tuy's condition**). A differentiable curve $a : I \to \mathbb{R}^d \setminus \mathbb{B}^d$ is said to satisfy **Tuy's condition** if for any $x \in \mathbb{B}^d$ and any $w \in \mathbb{S}^{d-1}$ there exists $u = u(x, w) \in I$, such that

$$w^T (x - a(u)) = 0 \qquad \text{and} \qquad w^T \dot{a}(u) \neq 0. \tag{3.7.2}$$

This condition means that *any* hyperplane through our "measurement region" $\mathbb{B}^d$ intersects with a and that the intersection is *transversally*, i.e., that a does not run within the plane, but "pierces" the plane. If Tuy's condition holds true, we get a stable analytic inverse formula.

Example 3.7.5. A circle around $\mathbb{B}^d$ does obviously not satisfy Tuy's connection while a helical curve around the cylindery containing the ball satisfies it.

[16] And therefore continuous.

[17] Sometimes being at least differentiable is even part of the definition of a **curve**.

[18] The Sobolev embedding theorem says that all functions of sufficiently high smoothness must be differentiable.

Theorem 3.7.6. *If a satisfies Tuy's condition and $f \in C_{00}^{\infty}(\mathbb{B}^d)$, then*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{S}^{d-1}} \frac{1}{v^T \dot{a}\left(u(x,v)\right)} \frac{\partial G}{\partial u}\left(u(x,v), v\right)\, dv, \qquad x \in \mathbb{B}^d, \tag{3.7.3}$$

where

$$G(u,v) = \left(Df(\cdot, a(u))\right)^{\wedge}(v), \qquad u \in [0,1],\ v \in \mathbb{S}^{d-1}.$$

Proof: We first extend Df to $y \in \mathbb{R}^d$ by[19]

$$Df(y,x) := \int_0^{\infty} f(x+ty)\, dt = |y|_2^{-1} \int_0^{\infty} f\left(x + t\frac{y}{|y|_2}\right) dt = \frac{1}{|y|_2} Df\left(\frac{y}{|y|_2}, x\right).$$

This function has a Fourier transform $(D_x f)^{\wedge} := (Df(\cdot,x))^{\wedge}$ defined as[20]

$$\begin{aligned}
(D_x f)^{\wedge}(\xi) &= \int_{\mathbb{R}^d} e^{-i\xi^T y} D_x f(y)\, dy = \int_{\mathbb{R}^d} e^{-i\xi^T y} \int_0^{\infty} f(x+ty)\, dt dy \\
&= \int_0^{\infty} \int_{\mathbb{R}^d} e^{-i\xi^T y} f(x+ty)\, dt dy = \int_0^{\infty} t^{-d} \int_{\mathbb{R}^d} e^{-i\xi^T (y-x)/t} f(y)\, dt dy \\
&= \int_0^{\infty} t^{-d} e^{i\xi^T x/t} \int_{\mathbb{R}^d} e^{-i\xi^T y/t} f(y)\, dt dy = \int_0^{\infty} t^{-d} e^{i\xi^T x/t} \hat{f}\left(\frac{\xi}{t}\right) dt \\
&= \int_0^{\infty} t^{d-2} e^{it\xi^T x} \hat{f}(t\xi)\, dt.
\end{aligned}$$

This looks similar to the inverse Fourier transform in polar coordinates,

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{S}^{d-1}} \int_0^{\infty} t^{d-1} e^{itx^T \xi} \hat{f}(t\xi)\, dt d\xi, \tag{3.7.4}$$

except the incorrect power of t. Therefore, we consider $x = a(u)$. i.e.,

$$\left(D_{a(u)} f\right)^{\wedge}(\xi) = \int_0^{\infty} t^{d-2} e^{it\xi^T a(u)} \hat{f}(t\xi)\, dt$$

and take the derivative with respect to u,

$$\begin{aligned}
\frac{d}{du}\left(D_{a(u)} f\right)^{\wedge}(\xi) &= \int_0^{\infty} t^{d-2} e^{it\xi^T a(u)} \left(it\xi^T \dot{a}(u)\right) \hat{f}(t\xi)\, dt \\
&= \left(i\xi^T \dot{a}(u)\right) \int_0^{\infty} t^{d-1} e^{it\xi^T a(u)} \hat{f}(t\xi)\, dt,
\end{aligned}$$

that is,

$$\int_0^{\infty} t^{d-1} e^{itx^T a(u)} \hat{f}(t\xi)\, dt d\xi = \frac{1}{i\xi^T \dot{a}(u)} \frac{d}{du}\left(D_{a(u)} f\right)^{\wedge}(\xi), \qquad \xi \in \mathbb{S}^{d-1}. \tag{3.7.5}$$

[19] This a homogeneous extension of order −1.

[20] Since $D_x f(0)$ is not well-defined, a careful proof would first use

$$(D_x f)^{\wedge}(\xi) = \int_{|y|>\varepsilon} e^{-i\xi^T y} D_x f(y)\, dy, \qquad \varepsilon > 0,$$

and then consider the limit $\varepsilon \to 0$.

No we make use of Tuy's condition: For each x and ξ there exists $u(x,\xi)$ such that $\xi^T x = \xi^T a(x,\xi)$, so that we can substitute (3.7.5) into (3.7.4) to obtain (3.7.3) after replacing ξ by v. □

The point of the inversion formula (3.7.3) is that all terms appearing there are nice and well-defined. If f is smooth, then $t^{d-1}\hat{f}(tv)$ decays rapidly with t and the integral

$$(D_a(u)f)^\wedge(v) = \int_0^\infty t^{d-1} e^{itv^T a(u)} \hat{f}(tv)\, dt$$

is well defined and smooth with respect to u and v. Averaging that over v only makes ist smoother. In other words: if Tuy's condition is satisfied, a *stable* reconstruction can be expected.

Let us try to understand the geometric meaning of (3.7.3). One starts with the point x for which one wants to reconstruct[21] f and looks, for any $v \in \mathbb{S}^{d-2}$, at the hyperplane with normal v that passes through x. By Tuy's condition, the hyperplane intersects the curve in at least one point with parameter $u = u(x,v)$ such that $v^T\dot{a}(u) \neq 0$. Then the function $(D_a(u)f)^\wedge(v)$ is tracked along the curve and this derivative is evaluated at $u(x,v)$.

Note that the intersection parameter $u(x,v)$ can be chosen in a continous way despite the fact that it need not be unique. In particular, if we fix x and vary the direction v along a continuous curve on $\mathbb{S}^{d-1}$, then the intersections from a subcurve of a.

Remark 3.7.7.

1. If we assume that a is intrinsically parametrized with respect to the arc length, then $v^T\dot{a}(u)$ is the (cosine of the) *angle* between the normal to hte hyperplane and the tangent to the curve. In other words: the flatter the curve intersects the plane, the larger the number $\left(v^T\dot{a}(u)\right)^{-1}$ and the smaller the derivative of G must be in order to obtain a finite value $f(x)$. This of course amplifies any small errors in the measurements and leads to numerical instability.

2. If a is parametrized with respect to time, then the factor $\left(v^T\dot{a}(u)\right)^{-1}$ becomes smaller if the speed increased along the curve, but of course the increased speed also increases the derivative with respect to u of G.

3. More precisely, let $\varphi : I' \to I$ be a *reparametrization* of the curve a, that is, a surjective function with $\dot{\varphi} > 0$, then, with $a_\varphi = a \circ \varphi$, the chain rule gives

$$\dot{a}_\varphi(u) = \dot{a}\left(\varphi(u)\right)\, \dot{\varphi}(u) \qquad \text{and} \qquad \frac{\partial G_\varphi}{\partial u}(u,v) = \frac{\partial G}{\partial u}\left(\varphi(u),v\right)\, \dot{\varphi}(u),$$

   and the formula (3.7.3) turns out to be invariant under reparametrizations.

4. In other words: the formula (3.7.3) depends on the *curve*, not on its *parametrization.*

[21] Yes, in this respect the reconstruction is a pointwise one.

Signal Processing and the Filtered Backprojection

4

Nothing spoils numbers faster than a lot of arithmetic.

((Peppermint Patty, *The Peanuts, 4.12.1968*))

We already encountered the filtered backprojection in Theorem 3.2.8, but so far it is a result in terms of convolutions; to relate it to filters and to build reconstruction algorithms based on the filtered backprojection, we will give a short introduction to continuous and discrete signal processing and then return to the filtered backprojection itself.

4.1 Signals and Filters

While "backprojection" is just another name for the dual Radon transform and thus easily explained, the meaning of "filtered" needs some more concepts; digital or discrete signal processing is explained for example in (Hamming, 1989; Schüßler, 1992).

In the general sense of Signal Processing, a **signal** is simply an element from a function space, for example, $f \in L_p(\mathbb{R}^d)$ or $f \in \mathscr{S}(\mathbb{R}^d)$, or maybe the images of the Radon or the X-ray transform.

Definition 4.1.1. A **filter** is an Operator from a signal space X to a signal space Y.

Of course, on this level of generality almost everything is signal and filter; therefore, the common understanding of a filter means an important subclass of filters.

Definition 4.1.2 (Filter). An operator $F : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ is called a *linear and time invariant filter* or **LTI filter** or simply a **filter** if

1. F is **linear**, i.e., it is a linear operator,
2. F is **time invariante**, that is, ist:

$$(Ff)(\cdot + y) = F(f(\cdot + y)), \qquad y \in \mathbb{R}^d.$$

Usually one also assumes that F is continuous or, equivalently, bounded.

Exercise 4.1.1 Show that an operator $F : X \to Y$ on Banach spaces X, Y is continuous if and only if it is bounded. (Riesz and Sz.-Nagy, 1955, S. 149) ◊

Using the translation operator from (2.1.3) in Remark 2.1.2, time invariance can also be written as the commuting property

$$F\tau_y = \tau_y F, \qquad y \in \mathbb{R}^d. \tag{4.1.1}$$

Moreover, LTI filters have a very nice representation which is based on the following concept.

Definition 4.1.3. The **impulse response** of a filter F is the distribution $f = F\delta$ obtained by applying the filter to the **Dirac distribution** δ, definiert by $\delta\phi = \phi(0)$.

Remark 4.1.4. The distributional definition of the impulse response is obtained by applying the filter to the sequence of functions or regular distributions $f_n := \frac{n}{2}\chi_{[-1/n,1/n]} \in L_p(\mathbb{R})$ which yields a sequence of functions Ff_n. This sequence may not converge in $L_p(\mathbb{R}^d)$, but it converges in the distributional sense.

The following result is presented and "proven" in a handwaving fashion by mixing functions and distributions in a very naive and incorrect way; it can also be done correctly, however, with more effort, see (Gasquet and Witomski, 1998). Nevertheless, it explains the particular role of convolutions in signal processing.

Proposition 4.1.5. *An operator F is a filter if and only $Fg = f * g$, where f is the impulse response of F.*

Proof: To write the filter as a convolution, we use the trivial[1] identity

$$g(x) = (g * \delta)(x) = \int_{\mathbb{R}^d} g(\cdot - t)\ \delta(t)\, dt = \int_{\mathbb{R}^d} g(t)\ \delta(x - t)\, dt,$$

and generously change the order of integration to obtain

$$Fg = F\int_{\mathbb{R}^d} g(t)\ \delta(\cdot - t)\, dt = \int_{\mathbb{R}^d} g(t)\ F\delta(\cdot - t)\, dt = \int_{\mathbb{R}^d} g(t)\ f(\cdot - t)\, dt = f * g.$$

For the converse, linearity is obvious and

$$\begin{aligned}\left(\tau_y Fg\right)(x) &= (Fg)(x + y) = \int_{\mathbb{R}^d} f(t)\ g(x + y - t)\ dt \\ &= \int_{\mathbb{R}^d} f(t)\ \tau_y g(x - t)\ dt = \left(f * \tau_y g\right)(x)\end{aligned}$$

implies time invariance. □

Ignoring such banalities as mathematical correctness, nasty details and existence of objects, the concept of a filter is equivalent to that of a convolution or, when passing to the Fourier transform, to a multiplication:

$$Fg = f * g \qquad \Leftrightarrow \qquad (Fg)^\wedge = \widehat{f}\,\widehat{g}.$$

The Fourier transform $\widehat{f}$ of the impulse response f is called the **transfer function** of the filter F and due to the linear relation between the filtered function and the function itself, sometimes people speak of a *linear system.*

Let us look at some not so standard example of filters.

[1] And formally incorrect ...

Example 4.1.6 (Partial differential operators). Any partial differential operator with constant coefficients is a filter. ein Filter! Indeed, writing the **partial differential operator** F of order n as

$$F = \sum_{|\alpha|\le n} a_\alpha \frac{\partial^{|\alpha|}}{\partial x^\alpha}, \qquad a_\alpha \in \mathbb{R}, \quad |\alpha| \le n,$$

and applying it to a sufficiently smooth function g, we get for any $\xi \in \mathbb{R}^d$ due to (2.1.9) that

$$(Fg)^\wedge(\xi) = \sum_{|\alpha|\le n} a_\alpha \left(\frac{\partial^{|\alpha|}}{\partial x^\alpha} g\right)^\wedge(\xi) = \sum_{|\alpha|\le n} a_\alpha \, (-i\xi)^\alpha \, \widehat{g}(\xi) = \widehat{f}(\xi)\,\widehat{g}(\xi)$$

with the polynomial transfer function

$$\widehat{f}(\xi) = \sum_{|\alpha|\le n} (-i)^{|\alpha|} a_\alpha \, \xi^\alpha, \qquad \xi \in \mathbb{R}^d.$$

This example also highlights the problem that we are facing here: the impulse response $f = \widehat{f}^\vee$ is not well-defined as a function, only as a distribution since the only polynomial belonging to some L_p space or to the Schwartz space is the zero polynomial. This is the reason why we have to switch to distributions eventually.

Particular classes of filters are described by means of their **frequency response**, i.e., by the way how they act on certain frequencies in the sense of parameters of the Fourier transform.

Definition 4.1.7 (High, Low & Band). Let $0 < C_1 < C_2 \in \mathbb{R}$. A filter F is called

1. **low pass** filter if $\hat{f}(\xi) = 0$ for $|\xi| > C_1$,
2. **high pass** filter if $\hat{f}(\xi) = 0$ for $|\xi| < C_1$,
3. **band pass** filter if $\operatorname{supp} \widehat{f} \subseteq [C_1, C_2]^d$, i.e., $\hat{f}(\xi) = 0$ for $\xi \notin [C_1, C_2]^d$.

The ideal low pass filter would naturally be the characteristic function of some cube centered around the origin, for example[2] with the transfer function $\widehat{f} = \chi_{[-1,1]^d}$, which leads to the **sinc function** which is an abbreviation for **sinus cardinalis**; the function is

$$f(x) = \frac{1}{\pi^d} \operatorname{sinc} x, \qquad \operatorname{sinc} x := \prod_{j=1}^{d} \frac{\sin x_j}{x_j},$$

and can be seen in Fig. 4.1.1. As we will see soon, the sinc function is of great theoretical importance, but has significant deficits in practice, cf. (Gasquet and Witomski, 1998).

Exercise 4.1.2 Show that

$$\chi^\vee_{[-C,C]^d}(x) = \left(\frac{C}{\pi}\right)^d \operatorname{sinc} Cx, \qquad x \in \mathbb{R}^d. \tag{4.1.2}$$

◊

[2] The we normalieze the bandwidth to $[-1, 1]$ is no conceptual restriction.

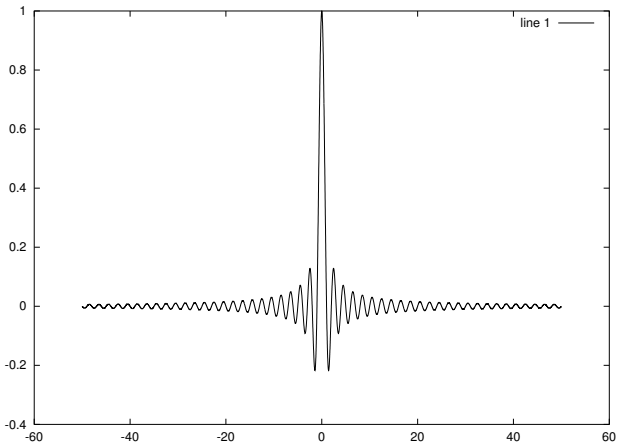

Figure 4.1.1: The function $f(x) = \operatorname{sinc} x$.

4.2 Bandlimited Functions and Shannon

If we imagine for a moment a sound record, then we realize that the technical equipment like microphones can only record a certain frequency range, in many cases 20-20000 Hz which is related to the ability of human hearing[3]. Thus, the spectrum $\hat{f}$ of such a recording has a compact support – it is *bandlimited.*

Definition 4.2.1. A function $f \in L_1\left(\mathbb{R}^d\right)$ is called **bandlimited** if there exist some bounded set $\Omega \subset \mathbb{R}^d$ such that $\operatorname{supp} \widehat{f} \subseteq \Omega$, i.e., $\operatorname{supp} \hat{f}$ is a compact set[4].

From the above perspective, bandlimited functions seem to be the only ones that exist in a "real" world of recording data. And they are even nicer, they are as smooth as we want. Let us recall the argument.

Proposition 4.2.2. *If $f \in L_1\left(\mathbb{R}^d\right)$ is bandlimited, then $f \in C^\infty\left(\mathbb{R}^d\right)$.*

Proof: Since $\hat{f}$ is uniformly continuous, it belongs to $L_1(\mathbb{R}^d)$ if it is compactly supported and the same holds true for the functions $g_\alpha := (-i\cdot)^\alpha \hat{f}$, $\alpha \in \mathbb{N}_0^d$, hence they have an inverse Fourier transform

$$f_\alpha := g_\alpha^\vee = \left((-i\cdot)^\alpha \widehat{f}\right)^\vee \in C_u(\mathbb{R}^d).$$

But by Theorem 2.1.5, $f_\alpha = \frac{\partial^{|\alpha|} f}{\partial x^\alpha}$, hence f has derivatives of any order. □

Going beyond $f \in L_1$ and requesting that f and $\widehat{f}$ are *both* smooth, i.e., $f, \widehat{f} \in C^\infty\left(\mathbb{R}^d\right)$, one embeds everything into $\mathbb{C}^d$ and obtians a classic from functional analysis, cf. (Yosida, 1965, S. 161), where it is proved of tempered distributions. Although the proof is not too hard, we will only quote the result here without a proof.

[3] And equipment that exceeds these restrictions in expensive like hell.

[4] By definition, the **support** $\operatorname{supp} f$ of a function f is the *closure* of $\{x : f(x) \neq 0\}$, hence always closed and therefore compact if and only if it is bounded.

Theorem 4.2.3 (Paley–Wiener). *A holomorphic[5] function* $F : \mathbb{C}^d \to \mathbb{C}$ *is the Fourier transform*

$$F(z) = \int_{\mathbb{R}^d} f(x)\, e^{-iz^T x}\, dx, \qquad z \in \mathbb{C}^d,$$

of a function $f \in C^\infty\left(\mathbb{R}^d\right)$ *whose support is contained in* $[-M, M]^d$*, if for any* $n > 0$ *there exists a constant* $C > 0$ *such that*

$$|F(z)| \le C\,(1 + |z|)^{-n}\; e^{M\,|\Im z|}.$$

In the "reality" of thechnical applications, we are usually not dealing with functions defined on a continuum, but with discrete measurements, the functions are *sampled.* The fundamental theorem that connects samples and functions is the famous *sampling theorem* that we will consider next.

Definition 4.2.4. For $h > 0$, the **sampling operator** S_h maps a function $f : \mathbb{R}^d \to \mathbb{R}$ to the sequence

$$S_h f = \left(f\,(h\alpha) \;:\; \alpha \in \mathbb{Z}^d\right). \tag{4.2.1}$$

Given a function ϕ, the associated **quasi interpolant** is defined as

$$Q_h f = S_h f * \phi = \sum_{\alpha \in \mathbb{Z}^d} S_h f(\alpha)\, \phi\,(\cdot - \alpha)\,, \tag{4.2.2}$$

provided that the sum exists.

The question is now whether, when and how we can reconstruct f from its samples $S_h f$ and what the influence of the function ϕ is in this process. The answer is the celebrated Shannon[6] Sampling Theorem that shows that reconstruction is possible and this even with a universal choice of ϕ, namely,

$$\operatorname{sinc}_\pi := \operatorname{sinc}\,(\pi\cdot)\,.$$

Theorem 4.2.5 (**Shannon Sampling Theorem**). *If* $f \in L_1\left(\mathbb{R}^d\right)$ *is a bandlimited function with* $\operatorname{supp} \widehat{f} \in [-T, T]^d$ *and* $h < h^* = \frac{\pi}{T}$*, then*

$$f = (S_h f * \operatorname{sinc}_\pi)\left(h^{-1}\cdot\right) = \sum_{\alpha \in \mathbb{Z}^d} f\,(h\alpha) \prod_{j=1}^{d} \frac{\sin \pi\left(x_j/h - \alpha_j\right)}{\pi\left(x_j/h - \alpha_j\right)}. \tag{4.2.3}$$

Hence, f *can be reconstructed* exactly *from the samples.*

Remark 4.2.6. The coupling between the **bandwidth** T of f and the **sampling distance** h is on of the most important concepts of digital signal processing and provides the link between the discrete and the continous world. Therefore a few remarks on this fundamental result are in order:

[5] Complex differentiation is not so obvious and the driving force behind Function Theory, cf. (Fisher, 1990; Hille, 1982).

[6] Claude Elwood Shannon, 1916–2001, electrical engineer and mathematician, one of the fathers of the theory of digital signal processing, invented the word "bit" and developed chess programms (before 1950).

1. In the literature, the **sampling frequency** $1/h$, also called the **Nyquist frequency** is said to be half the bandwidth. This is due to to some normalizations, but also due to the fact that the bandwidth is often defined as the *width* of the frequency support which is $2T$ in our case.

2. The sinc function is no good practical way to reconstruct functions from signals as it has no finite support, decays quite slowly, namely like $1/x$ and is just an L_1 function. Hence cutting of the expansion is not easy and prone to errors.

3. In many cases the sampling distance h is chosen as h^*/k for some $k \in \mathbb{N}$ which is then called k-fold **oversampling**.

4. Some proofs of the sampling theorem in the engineering literature (Grüningen, 1993; Schüßler, 1992) and the arguments there are not formally correct and only provide intuition. A good example for the opposite is (Mallat, 1999). The proof in this lecture is a modification of the one from (Kammeyer and Kroschel, 1998).

5. While Shannon published his theorem in 1949 in (Shannon, 1949), the result itself was given earlier as a result on infinite cardinal by Whittaker (Whittaker, 1935) in 1935, and Kotelnikov (Kotelnikov, 1933) gave a similar result even earlier. This is the reason why it is frequently referred to as Shannon-Whittaker-Kotelnikov theorem nowadays.

Proof of Theorem 4.2.5: Since f is bandlimited, $\hat{f} \in L_1\left(\mathbb{R}^d\right)$ and the orthgonality of complex exponentials on $\mathbb{T}^d$ yields for $h > 0$ and $\alpha \in \mathbb{Z}^d$

$$S_h f(\alpha) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} (S_h f)^\wedge(\theta)\, e^{i\alpha^T\theta}\, d\theta, \qquad (S_h f)^\wedge(\theta) := \sum_{\beta\in\mathbb{Z}^d} S_h f(\beta)\, e^{-i\beta^T\theta}, \tag{4.2.4}$$

but also

$$\begin{aligned}
S_h f(\alpha) = f(h\alpha) = \widehat{f}^\vee(h\alpha) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ih\alpha^T\xi}\, d\xi \\
&= \frac{1}{(2\pi)^d} \sum_{\beta\in\mathbb{Z}^d} \int_{h^{-1}\left(2\pi\beta+[-\pi,\pi]^d\right)} \widehat{f}(\xi)\, e^{ih\alpha^T\xi}\, d\xi \\
&= \frac{1}{(2\pi)^d} \sum_{\beta\in\mathbb{Z}} h^{-d} \int_{2\pi\beta+[-\pi,\pi]^d} \widehat{f}\left(h^{-1}\xi\right)\, e^{i\alpha^T\xi}\, d\xi \\
&= \frac{1}{(2\pi h)^d} \sum_{\beta\in\mathbb{Z}^d} \int_{\mathbb{T}^d} \widehat{f}\left(h^{-1}(\xi+2\pi\beta)\right)\, e^{i\alpha^T\xi}\, d\xi \\
&= \frac{1}{(2\pi h)^d} \int_{\mathbb{T}^d} \left(\sum_{\beta\in\mathbb{Z}^d} \widehat{f}\left(h^{-1}(\xi+2\pi\beta)\right) \right) e^{i\alpha^T\xi} d\xi, =: \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} F(\xi)\, e^{i\alpha^T\xi},
\end{aligned}$$

which means that

$$S_h f(\alpha) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} F(\xi)\, e^{i\alpha^T \xi}, \qquad F := \frac{1}{h^d} \sum_{\beta \in \mathbb{Z}^d} \widehat{f}\left(\frac{\cdot + 2\pi\beta}{h}\right); \tag{4.2.5}$$

the function F is a periodization of the Fourier transform and up to the factor h^{-1} exactly the same as (2.2.2). Moreover, $F \in C\left(\mathbb{T}^d\right) \subset L_1\left(\mathbb{T}^d\right)$ is 2π periodic and since f is bandlimted, the sum in (4.2.5) is even finite for any $x \in \mathbb{R}^d$. Since the exponentials $e^{i\alpha^T \cdot}$, $\alpha \in \mathbb{Z}^d$, form a *complete* orthonormal system[7], we can conclude from (4.2.4) and (4.2.5) that

$$h^{-d} \sum_{\beta \in \mathbb{Z}^d} \widehat{f}\left(\frac{\xi + 2\pi\beta}{h}\right) = F(\xi) = (S_h f)^\wedge(\xi) = \sum_{\beta \in \mathbb{Z}^d} f(h\beta)\, e^{-i\beta^T \xi}, \qquad \xi \in \mathbb{T}^d. \tag{4.2.6}$$

If h is so small that

$$\begin{aligned} h^{-1}[-\pi,\pi]^d \supseteq [-T,T]^d \quad &\Longleftrightarrow \quad [-\pi,\pi]^d \supseteq [-Th,Th]^d \quad \Longleftrightarrow \quad Th < \pi \\ &\Longleftrightarrow \quad h < \frac{\pi}{T}, \end{aligned}$$

we get for $j = 1, \ldots, d$ and $\xi \in \mathbb{T}^d$ that

$$h^{-1}(\xi + 2\pi\beta)_j > \frac{T}{\pi}\left(-\pi + 2\pi\beta_j\right) \geq T\left(-1 + 2\beta_j\right) \geq T,$$

whenever $\beta_j > 0$, and that $h^{-1}(\xi + 2\pi\beta)_j < -T$ whenever $\beta_j < 0$. This, however, means that the sum of the left hand side of (4.2.6) consists only of the term with $\beta = 0$ and replacing ξ by $h\xi$ we get

$$\widehat{f}(\xi) = h^d \sum_{\alpha \in \mathbb{Z}^d} f(h\alpha)\, e^{-ih\alpha^T \xi}.$$

[7] A standard result in Analysis.

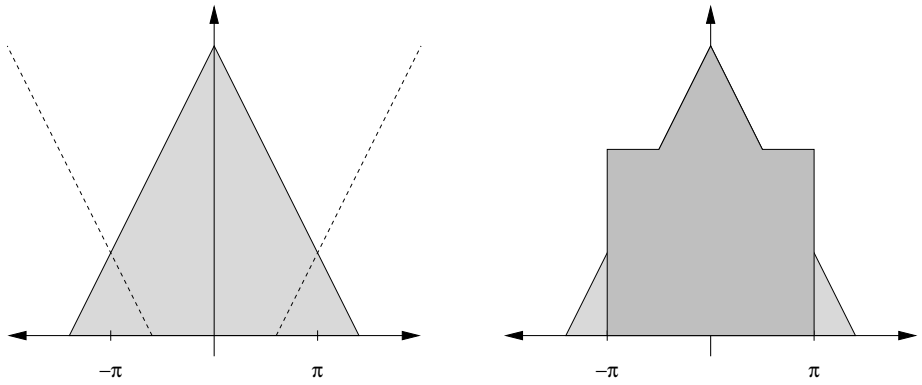

Figure 4.2.1: Periodization of a function with too large support. The overlap makes it impossible to reconstruct the function uniquely from the periodization; of course the dark gray function in the right hand image has to be extended periodically.

Therefore, since $T < \pi/h$, the T-bandlimitedness of f yields

$$\begin{aligned}
f(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi \\
&= \frac{1}{(2\pi)^d} \int\limits_{[-T,T]^d} \widehat{f}(\xi)\, e^{ix^T\xi}\, d\xi = \left(\frac{h}{2\pi}\right)^d \int\limits_{[-T,T]^d} \sum_{\alpha\in\mathbb{Z}} f(h\alpha)\, e^{i(x-h\alpha)^T\xi}\, d\xi \\
&= \left(\frac{h}{2\pi}\right)^d \sum_{\alpha\in\mathbb{Z}} f(h\alpha) \int\limits_{h^{-1}\mathbb{T}^d} e^{i(x-h\alpha)^T\xi}\, d\xi \\
&= \left(\frac{h}{2\pi}\right)^d \sum_{\alpha\in\mathbb{Z}^d} f(h\alpha) \prod_{j=1}^{d} \left[\frac{e^{i(x_j-h\alpha_j)\xi}}{i\,(x_j-h\alpha_j)} \Bigg|_{\xi=-\pi/h}^{\pi/h} \right] \\
&= \sum_{\alpha\in\mathbb{Z}} f(h\alpha) \prod_{j=1}^{d} \underbrace{\frac{e^{i(x_j-h\alpha_j)\pi/h} - e^{-i(x_j-h\alpha_j)\pi/h}}{2i}}_{=\sin\pi(x_j/h-\alpha_j)} \underbrace{\frac{h}{\pi}\frac{1}{(x_j-h\alpha_j)}}_{=(\pi(x_j/h-\alpha_j))^{-1}} \\
&= \sum_{\alpha\in\mathbb{Z}^d} f(h\alpha) \prod_{j=1}^{d} \frac{\sin\pi\,(x_j/h-\alpha_j)}{\pi\,(x_j/h-\alpha_j)} = (S_h f * \operatorname{sinc}_\pi)\,(\cdot/h),
\end{aligned}$$

which is (4.2.3). □

The clou in the proof of Theorem 4.2.5 is to consider the "Poisson sum"

$$\sum_{\alpha\in\mathbb{Z}^d} \widehat{f}\left(\frac{\xi+2\pi\alpha}{h}\right) = h^d \sum_{\alpha\in\mathbb{Z}^d} f(h\alpha)\, e^{-i\alpha^T\xi}, \qquad \xi\in\mathbb{T}^d, \tag{4.2.7}$$

cf. (2.2.5), that combines the periodization of the *Fourier transform* of a function with the *Fourier series* of the sampling sequence; in principle, this works *without* the

assumption that f is bandlimited[8] or that the sampling is sufficiently fine. But if h is so large that the function $\widehat{f}\left(h^{-1}\cdot\right)$ has a support larger than the invterval $[-\pi,\pi]$, the overlapping parts of the periodization contaminate the function, see Fig.4.2.1, and a reconstruction of $\hat{f}$ and therefore of f becomes impossible.

But things are even worse: since we combine frequencies that have nothing to do with each other and which are now considered modulo 2π, the reconstruction of an undersampled signal can lead to unwanted effects that are called **aliasing**. Their removal, called antialiasing, is an important task in many applications of signal and image processing.

4.3 More on the Filtered Backprojection

Once we know at least some basics of signal processing and know what the "filtered" in the name means, it is getting time to have a look at the filtered backprojection from a numerical perspective. To that end, we start with the **measurement** $y := Rf \in \mathscr{S}\left(\mathbb{S}^d \times \mathbb{R}\right)$ and also precompute the dual Radon transform $G := R^* g \in L_1(\mathbb{R}^d)$ of the filter $g \in L_1(\mathbb{R})$, more precisely, the impulse response of the filter that we are considering. This **reconstruction function** g is a free parameter in our reconstruction process and it usually can be chosen from a full library of possible filters, according to the concrete situation at hand.

With the definitions above, (3.2.11) takes the form

$$G * f = R^* (g * y) . \tag{4.3.1}$$

The ideal filter would be $G = \delta$ as then we can directly write down the solution

$$f = R^* (g * y) . \qquad g = (R^*)^{-1} \delta.$$

But whenever something becomes so simple, there might be a catch to it. Of course, the problem here is that, although δ is a formal identity for the "convolution product"

$$\delta * f(x) = f(x),$$

this identity only holds in the sense of distributions, not of functions, and we should not expect that there are functions such that $R^* g = \delta$.

Exercise 4.3.1 What are the Fourier transform $\widehat{\delta}$ and the inverse Fourier transform $\delta^\vee$ of the δ distribution, respectively. ◊

To simplify notation a bit in the following, we write for $g \in L_1\left(\mathbb{S}^d \times \mathbb{R}\right)$,

$$\hat{g}\,(v,\sigma) = \hat{g}_v(\sigma)\,(g(v,\cdot))^\wedge\,(\sigma), \qquad v \in \mathbb{S}^d, \quad \sigma \in \mathbb{R}. \tag{4.3.2}$$

In particular for numerical applications, it makes a lot of sense, to choose the filter as *simple* as possible and a good approach in this direction is to exploit suitable symmetries that are in accordance with the underlying transformation. In the case of the Radon transform this is of course the ambiguity

$$Rf(v,s) = Rf(-v,-s),$$

[8] Of course, in order for the left hand side of (4.2.7) to exist, some sort of decay of $\hat{f}$ should be incorporated.

that we noticed quite early in Remark 3.2.3. Resepcting this ambiguity, we get a usefuly formula for the dual Radon transform of our filter.

Theorem 4.3.1. *If $g \in \mathscr{S}\left(\mathbb{S}^{d-1} \times \mathbb{R}\right)$ is an **even function**, i.e., $g(v,s) = g(-v,-s)$, then*

$$(R^* g)^\wedge(\xi) = |\xi|_2^{1-d}\, \widehat{g}\left(\frac{\xi}{|\xi|_2}, |\xi|_2\right), \qquad \xi \in \mathbb{R}^d \setminus \{0\}. \tag{4.3.3}$$

To prove the theorem, we need a little computation that gives yet another relationship between the Radon transform, the dual Radon transform and the Riesz potential.

Lemma 4.3.2. *For $f \in \mathscr{S}\left(\mathbb{R}^d\right)$,*

$$(R^* R f)^\wedge(\xi) = |\xi|_2^{1-d}\ \widehat{f}(\xi), \tag{4.3.4}$$

that is, $R^* R = I^{d-1}$.

Remark 4.3.3.

1. As we have already seen, the Riesz potential I^r can be understood as some sort of differential operator, at least if $r < 0$, otherwise it takes the form of a (smoothing) integral operator. Usually, such operators that are defined by multplication of the Fourier transform with the norm of the argument are called *pseudodifferential operators.*

2. Being a multiplication of the Fourier transform, (4.3.4) is nothing but a *convolution* of f with the filter $\left(|\cdot|_2^{1-d}\right)^\vee$. The only drawaback here is that in the case $d = 2$ we take the inverse Fourier transform of a function that does not belong to $L_1(\mathbb{R}^2)$ as it decays only like $|\xi|_2^{-1}$.

Exercise 4.3.2 Compute the impulse response of the transfer function $F(\xi) = \hat{f}(\xi) = \|\xi\|_2^{1-d}$. ◊

Proof: The proof is similar to the one of the inversion formula in Theorem 3.2.7, but this time we consider

$$\begin{aligned}
\left(I^{d-1} f\right)(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\xi|_2^{1-d}\ \widehat{f}(\xi)\, e^{i\xi^T x}\, d\xi \\
&= \frac{1}{2(2\pi)^d} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} \underbrace{|\sigma v|^{1-d}}_{=|\sigma|^{1-d}} \widehat{f}(\sigma v)\ e^{ix^T(\sigma v)} |\sigma|^{d-1}\, dv\, d\sigma \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} \frac{1}{2\pi} \int_{\mathbb{R}} (Rf(v,\cdot))^\wedge(\sigma)\, e^{i\sigma x^T v}\, d\sigma\, dv \\
&= \frac{1}{2(2\pi)^{d-1}} \int_{\mathbb{S}^d} Rf\left(v, v^T x\right) dv = (R^* R)\, f(x)
\end{aligned}$$

to obtain (4.3.4). □

Proof of Theorem 4.3.1: We recall from equation (3.5.5) in the proof of Theorem 3.5.3 that whenever g is an even function such that $\hat{g}_v \in L_1(\mathbb{R})$ for any $v \in \mathbb{S}^{d-1}$,

then the inverse Fourier transform f of the function such that $\hat{g}_v(\sigma)$ for $\xi = \sigma v$ satisfies[9] $g = Rf$. Supposing that g is "sufficiently nice" such that this is possible, we then get that $R^*g = R^*Rf = I^{d-1}f$, and because of (3.2.7) we have for $\xi \neq 0$,

$$\begin{aligned}(R^*g)^\wedge(\xi) &= |\xi|_2^{1-d}\,\widehat{f}(\xi) = |\xi|_2^{1-d}\,\widehat{f}\left(|\xi|\frac{\xi}{|\xi|_2}\right)\\ &= |\xi|_2^{1-d}\,Rf\left(\frac{\xi}{|\xi|_2},\cdot\right)^\wedge(|\xi|) = |\xi|_2^{1-d}\,\widehat{g}\left(\frac{\xi}{|\xi|_2},|\xi|_2\right),\end{aligned}$$

which completes the proof. □

Remark 4.3.4. The extension of (4.3.6) to $\xi = 0$ is *not* obvious since $|\xi|^{1-d}$ leads to a pole of order $d-1$ there; this is the price to be payed for the smoothing effect of I^{d-1} with respect to the high frequencies. So, in order for (4.3.6) to be well-defined at $\xi = 0$, we request $\hat{g}(v,\sigma)$ to have a zero of order $d-1$ at $\sigma = 0$, that is,

$$0 = \int_{\mathbb{R}} g_v(s)s^k\,ds, \qquad k = 0,\dots,d-1,$$

see the discussion after Proposition 3.4.4. Once more, this is a wavelet-like property and requires that g has a certain oscillation property that depends on the dimension of the space in which it is applied.

Since the inversion formula also exists for the X-ray transform, it comes as no surprise that similar results exist there as well. The counterpart of Lemma 4.3.2 is the following result.

Lemma 4.3.5. *For $f \in \mathscr{S}\left(\mathbb{R}^d\right)$,*

$$(X^*Xf)^\wedge(\xi) = |\xi|_2^{-1}\,\widehat{f}(\xi), \tag{4.3.5}$$

that is, $X^*X = I$.

Proof: We compute, again with the help of (3.3.8),

$$\begin{aligned}If(x) &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}|\xi|_2^{-1}\hat{f}(\xi)e^{i\xi^Tx}\,d\xi\\ &= \frac{1}{(2\pi)^d}\frac{1}{|\mathbb{S}^{d-2}|}\int_{\mathbb{S}^{d-1}}\int_{v^\perp}|\eta|_2|\eta|_2^{-1}\hat{f}(\eta)\,e^{i\eta^Tx}\,d\eta dv\\ &= \frac{1}{(2\pi)^d}\frac{1}{|\mathbb{S}^{d-2}|}\int_{\mathbb{S}^{d-1}}\int_{v^\perp}(Xf(v,\cdot))^\wedge(\eta)\,e^{i\eta^Tx}\,d\eta dv\\ &= \frac{1}{2\pi}\frac{1}{|\mathbb{S}^{d-2}|}\int_{\mathbb{S}^{d-1}}Xf\left(v,VV^Tx\right)\,dv = X^*Xf(x),\end{aligned}$$

which gives (4.3.5). □

This again allows us to conveniently compute the Fourier transform of X^*g for some filter g.

[9] The consistency conditions (3.5.3) "only" server the purpose of ensuring that $f \in \mathscr{S}(\mathbb{R}^d)$, i.e., to prove smoothness and decay.

Theorem 4.3.6. *If* $g \in \mathscr{S}\left(T\mathbb{S}^{d-1}\right)$ *satisfies* $g(\cdot, x) = 0$ *for* $|x|_2 \geq C$, *then*

$$(X^* g)^\wedge(\xi) = |\xi|_2^{-1}\, \hat{g}(v, VV^T\xi) \qquad v \in \mathbb{S}^{d-1}, \quad \xi \in \mathbb{R}^d \setminus \{0\}. \tag{4.3.6}$$

Proof: Again we make use of the fact that $g = Xf$ for some $f : \mathbb{R}^d \to \mathbb{R}$, see Theorem 3.5.4, and so, for any $v \in \mathbb{S}^{d-1}$,

$$\begin{aligned}(X^* g)^\wedge(\xi) &= (X^* X f)^\wedge(\xi) = |\xi|_2^{-1} \hat{f}(\xi) = |\xi|_2^{-1} \hat{f}\left(vv^T\xi + VV^T\xi\right) \\ &= |\xi|_2^{-1} (X_v f)^\wedge(VV^T\xi) = |\xi|_2^{-1}\, \hat{g}(v, VV^T\xi).\end{aligned}$$

as in (4.3.5). □

4.4 Radial Filters for the Radon Transform

In practice, one makes an even more fundamental simplification in terms of symmetry by requesting that g is *independent* of v and defines it as a **radial function**

$$g(v, s) = \phi(s), \qquad v \in \mathbb{S}^d, \quad s \in \mathbb{R}. \tag{4.4.1}$$

Of course, in view of Theorem 4.3.1 we choose g as an *even* radial function which means that ϕ has to be an even function, i.e., $\phi(s) = \phi(-s)$. Theorem 4.3.1,

$$\widehat{G}(\xi) = (R^* g)^\wedge(\xi) = |\xi|_2^{1-d}\, \widehat{\phi}\,(|\xi|_2), \qquad \xi \in \mathbb{R}^d,$$

is also a radial function and according to Remark 4.3.4, ϕ should also vanish of some order at the origin.

Our goal is the reconstruction of bandlimited functions as those are the ones that are, according to Shannon, accessible from discrete samples[10]. Let us denote, like in Definition 4.2.1 the compact support of the Fourier transform by Ω, then the ideal filter for the reconstruction would be

$$(R^* g)^\wedge(x) = \hat{G}(\xi) = \chi_\Omega(\xi) = \begin{cases} 1, & \xi \in \Omega, \\ 0, & \xi \notin \Omega, \end{cases} \qquad \xi \in \mathbb{R}^d,$$

i.e., $\hat{g} = |\cdot|^{d-1}\chi_\Omega$, as then

$$(G * f)^\wedge = \hat{G}\,\hat{f} = \chi_\Omega \hat{f} = \hat{f}$$

and therefore

$$G * f = (G * f)^{\wedge\vee} = \widehat{f}^\vee = f.$$

The function $\hat{G}$ is radial if Ω has radial symmetry, and we can choose $\Omega = B_\omega(0) = \omega\,\mathbb{B}^d$ as the ball with radius ω. For a little more liberty in the choice of the filter, we rely on a *univariate* transfer function ψ with the property

$$\psi(\xi) \sim 1, \quad |\xi| \leq 1, \qquad \psi(\xi) = 0, \qquad |\xi| > 1.$$

[10] But we only *proved* that bandlimited is a *sufficient* condition for reconstruction, we did not say anything about the necessity.

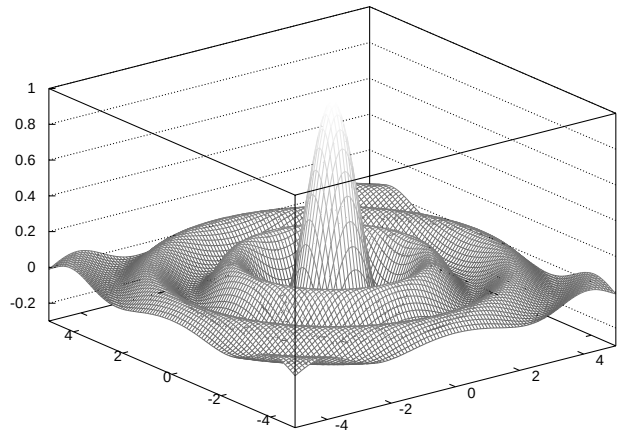

Figure 4.4.1: The radial sinc function, $G(x) = \operatorname{sinc}_\pi(|x|_2)$ plotted in 2D as an example of a radial function. Later we will only show the univariate "distance" part of the filters in the images.

Then we set

$$\widehat{G}(\xi) := \psi\left(\frac{|\xi|_2}{\omega}\right), \qquad \text{i.e.} \qquad \hat{g}(\xi) =: \hat{\rho}(|\xi|_2) = |\xi|_2^{d-1}\psi\left(\frac{|\xi|_2}{\omega}\right), \tag{4.4.2}$$

and now the impulse response g also depends on the dimension d. Note that this construction *automatically* deals with the problem mentioned in Remark 4.3.4.

In practice, the band one wants to reconstruct is of fundamental importance: cutting off too many high frequencies results in a loss of details, keeping to many high frequencies results in noise. And dropping low frequency content can lead to a more or less constant shift in the values that affects the contrast.

For a first example, we keep it simple and choose $\omega = 1$ and the ideal lowpass $\psi = \chi_\Omega$ for $\Omega = \mathbb{B}^d$. For $d = 2$ this leads to

$$G = \operatorname{sinc}|x|_2 \qquad \text{and} \qquad \widehat{g}(\xi) = \left((\cdot)\chi_{[-1,1]}\right)(|\xi|_2) = \phi(|\xi|_2), \quad \phi = |\cdot|\chi_{[-1,1]},$$

see Fig.4.4.1. According to (Natterer and Wübbeling, 2001), this filter was proposed by Ramachandran und Lakshminarayanan which was abbreviated to the more catchy and easier to remember name of **Ram-Lak filter**.

The function $\rho(s)$ can now be computed from its Fourier transform

$$\widehat{\rho}(\sigma) = |\sigma|^{d-1}\chi_{[-1,1]}(\sigma), \qquad \sigma \in \mathbb{R},$$

where for $d = 2$ we obtain the classical Ram-Lak filter as

$$\begin{aligned}
\rho(s) &= \frac{1}{2\pi}\int_{\mathbb{R}} |\sigma|\,\chi_{[-1,1]}(\sigma)\,e^{i\sigma s}\,d\sigma = \frac{1}{2\pi}\int_{-1}^{1} |\sigma| e^{i\sigma s}\,d\sigma \\
&= -\int_{-1}^{0} \sigma e^{i\sigma s}\,d\sigma + \int_{0}^{1} \sigma e^{i\sigma s}\,d\sigma \\
&= -\left[\sigma\,\frac{e^{i\sigma s}}{is}\right]_{-1}^{0} + \int_{-1}^{0}\frac{1}{is}e^{i\sigma s}\,d\sigma + \left[\sigma\,\frac{e^{i\sigma s}}{is}\right]_{0}^{1} - \int_{0}^{1}\frac{1}{is}e^{i\sigma s}\,d\sigma \\
&= \underbrace{\frac{e^{is}-e^{-is}}{is}}_{=2s^{-1}\sin s} + \left[\frac{1}{-s^2}\,e^{i\sigma s}\right]_{-1}^{0} - \left[\frac{1}{-s^2}\,e^{i\sigma s}\right]_{0}^{1} = 2\,\mathrm{sinc}\,s - \frac{e^{is}-2+e^{-is}}{-s^2} \\
&= 2\mathrm{sinc}\,s - \frac{\left(e^{is/2}-e^{-is/2}\right)^2}{(2is/2)^2} = 2\mathrm{sinc}\,s - \left(\mathrm{sinc}\,\frac{s}{2}\right)^2.
\end{aligned}$$

The sharp edges of Ram-Lak in the Fourier domain, i.e., its low order of smoothness, lead to a bad decay rate of ρ itsel, cf. Theorem 2.3.8, and lead to artifacts in practical applications, see (Olafsson and Quinto, 2006). These artifacts can be reduced by using the filter based ony $\hat{\phi} = \chi_{[-1,1]}\cos\frac{\pi\cdot}{2}$ which is globally continuous and whose associated ρ decays much faster, see Fig. 4.4.2. In 1974, Shepp and Logan proposed to use $\hat{\phi} = \chi_{[-1,1]}\,\mathrm{sinc}\,\frac{\pi\cdot}{2}$ which switches the role of Fourier transform and function since the sinc function is the (inverse) Fourier transform of the characteristic function.

Exercise 4.4.1 Compute the function ρ for $\widehat{\phi} = \chi_{[-1,1]}\,\cos\frac{\pi\cdot}{2}$. ◊

Unfortunately, an explicit computation of g is not easy already in the case of the Ram-Lak filter since

$$\begin{aligned}
g(x) &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\hat{g}(\xi)\,e^{i\xi^T x}\,d\xi = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}|\xi|_2^{d-1}\,\chi_{[-1,1]}\left(|\xi|_2\right)\,e^{i\xi^T x}\,d\xi \\
&= \frac{1}{2(2\pi)^d}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}}|\sigma|^{d-1}\chi_{[-1,1]}(\sigma)\;e^{i\sigma v^T x}\,|\sigma|^{d-1}\,dv\,d\sigma \\
&= \frac{1}{2(2\pi)^d}\int_{\mathbb{R}}\underbrace{\left(\int_{\mathbb{S}^{d-1}}e^{i\sigma v^T x}\,dv\right)}_{=:J_0(\sigma|x|_2)}|\sigma|^{2d-2}\chi_{[-1,1]}(\sigma)\,d\sigma,
\end{aligned}$$

where J_0 is a so-called **Bessel function**[11]; these very particular *special functions* ask for a separate treatment which will be provided in the next section.

[11] Friedrich Wilhelm Bessel, 1784–1846, German astronomer, contemporary and friend of Gauss. Being director of the Prussian Observatory at Königsberg, he developed these functions in his studies of planetary orbits.

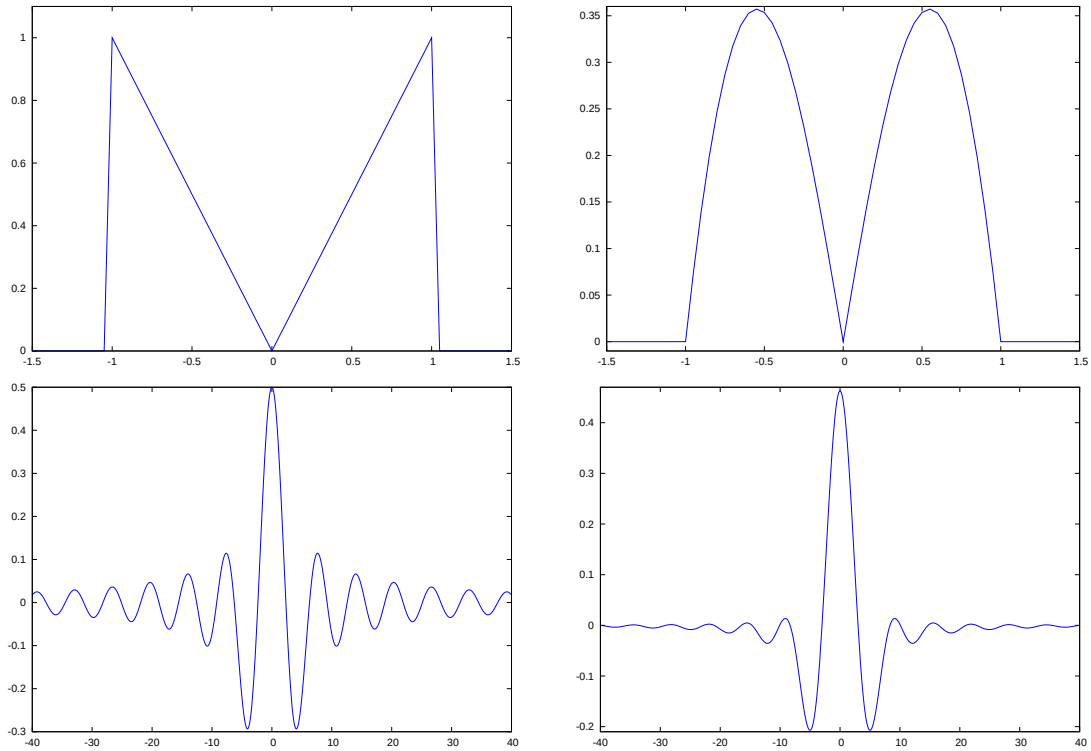

Figure 4.4.2: The functions $\hat{\rho}$ *(upper row)* and ρ *(lower row)* for Ram-Lak *(left)* and the cos-filter *(right)*. As suggestet by the theory, the cos-filter decays much faster.

4.5 Bessel Functions

When computing the Fourier transform of a radial function, i.e., of a function of the form $f(x) = \phi(|x|_2)$, some strange effect happens, namely

$$\begin{aligned}\hat{f}(\xi) &= \int_{\mathbb{R}^d} f(x) e^{-i\xi^T x}\, dx = \frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} f(\sigma v)\; e^{i\sigma\xi^T v}|\sigma|^{d-1}\, dv\, d\sigma \\ &= \frac{1}{2}\int_{\mathbb{R}}\left(\int_{\mathbb{S}^{d-1}} e^{-i\sigma\xi^T v}\, dv\right)\phi(\sigma)\,|\sigma|^{d-1}\, d\sigma,\end{aligned}$$

so that the integral

$$\Psi(\sigma,\xi) = \int_{\mathbb{S}^{d-1}} e^{-i\sigma\xi^T v}\, dv \tag{4.5.1}$$

can be expected to have a special meaning. We can expect Ψ to be a radial function with respect to ξ since we can write the inner product as

$$\xi^T v = |\xi|_2 \underbrace{|v|_2}_{=1} \sin\theta = |\xi|_2 \sin\theta, \qquad \theta \in [-\pi,\pi],$$

and if v varies over $\mathbb{S}^{d-1}$, the integral runs over all values of θ, independently of ξ; this is simply the symmetry of the unit sphere. For example, in $d = 2$ we can paramaterize v as

$$v = \frac{\xi^\perp \cos\theta + \xi \sin\theta}{\|\xi\|},$$

and obtain that

$$\Psi(\sigma,\xi) = \Psi\left(\sigma,|\xi|_2\right) = \int_{-\pi}^{\pi} e^{-i\sigma\,|\xi|_2\,\sin\theta}\,d\theta = J\left(\sigma|\xi|_2\right) = J\left(\sigma\,|\xi|_2\right)$$

for

$$J(x) := \int_{-\pi}^{\pi} e^{-ix\sin\theta}\,d\theta, \qquad x\in\mathbb{R}. \tag{4.5.2}$$

The general case is obtained by writing v as

$$v = \frac{\xi\sin t + w\cos\theta}{|\xi|_2}, \qquad w\in\mathbb{S}^{d-1}\cap\xi^{\perp} = \left\{y\in\mathbb{S}^{d-1}\ :\ \xi^T y = 0\right\}\sim\mathbb{S}^{d-2},$$

and decomposing the integral into

$$\begin{aligned}\int_{\mathbb{S}^{d-1}} e^{-i\xi^T v}\,dv &= \int_{\mathbb{T}}\int_{\mathbb{S}^{d-2}} e^{-i|\xi|_2^{-1}\xi^T(\xi\sin\theta+w\cos\theta)}\,dw\,d\theta\\ &= \underbrace{\left(\int_{\mathbb{S}^{d-2}} dw\right)}_{=\left|\mathbb{S}^{d-2}\right|}\int_{\mathbb{T}} e^{-i|\xi|_2\,\sin\theta}\,d\theta = \left|\mathbb{S}^{d-2}\right|\,J\left(|\xi|_2\right).\end{aligned}$$

with the function J of (4.5.2). In summary, the **spherical integral**

$$\int_{\mathbb{S}^{d-1}} e^{-i\xi^T v}\,dv$$

yields always the **Bessel function** J as a radial function, the ambient dimension d only affects this via the constant $\left|\mathbb{S}^{d-2}\right|$ that we already know from the transformation (3.3.8). Thus, a radial function $f=\phi\left(|\xi|_2\right)$ always has the Fourier transform

$$\hat{f}(\xi) = \left|\mathbb{S}^{d-2}\right|\int_{\mathbb{R}}\phi(\sigma)\,J\left(\sigma\,|\xi|_2\right)\,|\sigma|^{d-1}\,d\sigma. \tag{4.5.3}$$

The operation in (4.5.3) is an integral transform for ϕ in which the role of the exponential is taken by the Bessel function J; lots of formulas for Bessel functions and different ways of representing them can be found in (Abramowitz and Stegun, 1972).

What can we say about the function J? It is real valued since

$$\begin{aligned}\int_{-\pi}^{\pi} e^{-ix\sin\theta}d\theta &= \int_{-\pi}^{0} e^{-ix\sin\theta}d\theta + \int_{0}^{\pi} e^{-ix\sin\theta}d\theta = \int_{0}^{\pi} e^{-ix\sin\theta} + e^{ix\sin\theta}\,d\theta\\ &= 2\int_{0}^{\pi}\cos\left(x\sin\theta\right)\,d\theta = 2\int_{0}^{\pi}\cos\left(x\cos(\theta-\pi)\right)\,d\theta\\ &= 2\int_{-\pi}^{0}\cos\left(x\cos\theta\right)\,d\theta = 2\int_{0}^{\pi}\cos\left(x\cos\theta\right)\,d\theta.\end{aligned}$$

and this formula also suggests a proper normalization via $J(0)=1$.

Definition 4.5.1. The **Bessel function** of the first kind[12] is defined as

$$J_0(x) = \frac{1}{\pi}\int_{0}^{\pi}\cos\left(x\cos\theta\right)\,d\theta, \qquad x\in\mathbb{R}_+. \tag{4.5.4}$$

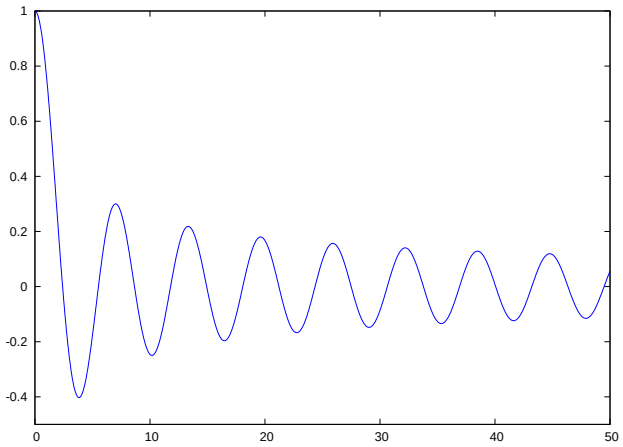

Figure 4.5.1: The Bessel function J_0 in the interval $[0, 50]$, looks like a sine-like damped oscillation.

J_0 is an even function and therefore can be easily extended to $x < 0$ although it suffices to consider J_0 on $\mathbb{R}_+$. Some important basic propeties of the Bessel function, taken from (Jackson, 1941), are as follows.

Theorem 4.5.2 (Properties of the Bessel function). *The Bessel function J_0 from (4.5.4)*

1. *satisfied the second order differential equation*

$$f''(x) + \frac{1}{x} f'(x) + f(x) = 0, \qquad x \in \mathbb{R}. \tag{4.5.5}$$

2. *The zeros $\lambda_1 < \lambda_2 < \dots$ of J_0 fulfill*

$$\left(k - \frac{1}{2}\right)\pi < \lambda_k < \left(k + \frac{1}{2}\right)\pi, \qquad k \in \mathbb{N}. \tag{4.5.6}$$

3. *For any two zeros $\lambda \neq \mu$ of J_0 one has*

$$\int_0^1 x\, J_0(\lambda x)\, J_0(\mu x)\, dx = 0; \tag{4.5.7}$$

*the functions $J_0(\lambda_j \cdot)$, form an **orthonormal system** of functions.*

Proof: For (4.5.5) we take derivatives of the integrand,

$$\frac{d}{dx} \cos(x \cos\theta) = -\sin(x \cos\theta)\, \cos\theta$$

as well as

$$\frac{d^2}{dx^2} \cos(x \cos\theta) = -\cos(x \cos\theta)\, \cos^2\theta$$

[12]There is also a second kind and the parameter can be varied as J_α, $\alpha \in \mathbb{R}$.

and obtain with a little bit of partial integration that

$$\begin{aligned} J_0''(x) + \frac{1}{x} J_0'(x) + J_0(x) & \\ = \; & \frac{1}{\pi} \int_0^\pi \left(1 - \cos^2\theta\right) \cos\left(x \cos\theta\right) - \frac{1}{\pi\, x} \int_0^\pi \cos\theta \, \sin\left(x\cos\theta\right) \, d\theta \\ = \; & \frac{1}{\pi} \int_0^\pi \left(1 - \cos^2\theta\right) \cos\left(x\cos\theta\right) \\ & + \frac{1}{\pi x} \underbrace{\left[\sin\theta \sin\left(x\cos\theta\right)\right]_0^\pi}_{=0} - \frac{1}{\pi x} \int_0^\pi x \, \sin^2\theta \, \cos\left(x\cos\theta\right) \, d\theta \\ = \; & \frac{1}{\pi} \int_0^\pi \underbrace{\left(1 - \cos^2\theta - \sin^2\theta\right)}_{=0} \cos\left(x\cos\theta\right) = 0. \end{aligned}$$

Next we use the transform of variable $\theta \to \pi - \theta$ in

$$\int_0^{\pi/2} \cos\left(x\cos\theta\right) \, d\theta \;\; = \;\; \int_{\pi/2}^{\pi} \cos\Big(x \underbrace{\cos(\pi - \theta)}_{=\cos(-\theta)=\cos\theta}\Big) \, d\theta = \int_{\pi/2}^{\pi} \cos\left(x\cos\theta\right) \, d\theta,$$

which yields, together with (4.5.4), the alternative formula

$$J_0(x) = \frac{2}{\pi} \int_0^{\pi/2} \cos\left(x\cos\theta\right) \, d\theta. \tag{4.5.8}$$

Another transform $\tau = x\cos\theta$ with

$$d\tau = \frac{d\tau}{d\theta} d\theta = -x\sin\theta \, d\theta = -\sqrt{x^2 - x^2\cos^2\theta} \, d\theta = -\sqrt{x^2 - \tau^2} \, d\theta,$$

yields

$$J_0(x) = \frac{2}{\pi} \int_0^x \frac{\cos\tau}{\sqrt{x^2 - \tau^2}} \, d\tau. \tag{4.5.9}$$

The denominator can be written as $\sqrt{(x+\tau)(x-\tau)}$ which behaves for $\tau \to x$ as $t^{\frac{1}{2}}$ for $t \to 0$ and thus constitutes an *integrable* singularity. If $x = \left(k + \frac{1}{2}\right)\pi$, $k \in \mathbb{N}$, the integrand has its zeros at $\left(j + \frac{1}{2}\right)\pi$, $j = 0, \ldots, k$, which is clear for $j < k$ and an application of the l'Hospital rule for $j = k$. Now we fix $k \in \mathbb{N}_0$, set $x = \left(k + \frac{1}{2}\right)\pi$ and define

$$c = \int_0^{\pi/2} \frac{\cos\tau}{\sqrt{x^2 - \tau^2}} \, d\tau,$$

as well as

$$d_j = \left| \int_{j\pi + \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]} \frac{\cos\tau}{\sqrt{x^2 - \tau^2}} \, d\tau \right|, \qquad j = 1, \ldots, k. \tag{4.5.10}$$

We have that $0 < c < d_1 < \cdots < d_k$ since the denominator in (4.5.10) decreases with increasing j by the definition of x. Moreover,

$$\begin{aligned}
&\frac{\pi}{2} J_0(x) \\
&= \frac{\pi}{2} J_0\left(\left(k+\frac{1}{2}\right)\pi\right) = \int_0^{\pi/2} \frac{\cos\tau}{\sqrt{x^2-\tau^2}}\, d\tau + \sum_{j=1}^{k} \int_{j+[-\pi/2,\pi/2]} \frac{\cos\tau}{\sqrt{x^2-\tau^2}}\, d\tau \\
&= c + \sum_{j=1}^{k} (-1)^j d_j = \begin{cases} c + \sum_{j=1}^{m} \left(d_{2j} - d_{2j-1}\right), & k = 2m, \\ (c-d_1) + \sum_{j=1}^{m} \left(d_{2j} - d_{2j+1}\right), & k = 2m+1, \end{cases}
\end{aligned}$$

and since all summands are positive for even k and negative for odd k, the sign of $J_0\left(\left(k+\frac{1}{2}\right)\pi\right)$ behaves like $(-1)^k$, $k \in \mathbb{N}_0$. Since the Bessel function is continuous, this eventually yields (4.5.6).

The orthogonality (4.5.7) follows from the differential equation[13] (4.5.5). Indeed,

$$\begin{aligned}
&\int_0^1 x J_0(\lambda x)\ J_0(\mu x)\ dx = -\int_0^1 x\left(J_0''(\lambda x) + \frac{1}{\lambda x} J_0'(\lambda x)\right) J_0(\mu x)\ dx \\
&= -\int_0^1 J_0''(\lambda x)\, x J_0(\mu x)\ dx - \frac{1}{\lambda}\int_0^1 J_0'(\lambda x)\, J_0(\mu x)\, dx \\
&= -\underbrace{\left[\frac{1}{\lambda} J_0'(\lambda x) x J_0(\mu x)\right]_0^1}_{=0} + \frac{1}{\lambda}\int_0^1 J_0'(\lambda x) J_0(\mu x)\, dx \\
&\quad + \frac{1}{\lambda}\int_0^1 J_0'(\lambda x)\, x J_0'(\mu x)\, dx - \frac{1}{\lambda}\int_0^1 J_0'(\lambda x)\, J_0(\mu x)\, dx \\
&= \frac{1}{\lambda}\int_0^1 J_0'(\lambda x)\, x J_0'(\mu x)\, dx.
\end{aligned}$$

The same computation can be done by substituting the differential equation for $J_0(\mu x)$ which analogously yields

$$\int_0^1 x J_0(\lambda x)\ J_0(\mu x)\ dx = \frac{1}{\mu}\int_0^1 J_0'(\lambda x)\, x J_0'(\mu x)\, dx,$$

hence also

$$\begin{aligned}
&(\lambda-\mu)\int_0^1 x J_0(\lambda x)\ J_0(\mu x)\ dx \\
&= \int_0^1 J_0'(\lambda x)\, x J_0'(\mu x)\, dx - \int_0^1 J_0'(\lambda x)\, x J_0'(\mu x)\, dx = 0,
\end{aligned}$$

[13] This is a fairly standard trick that also finds its application in the world of orthogonal polynomials, cf. (Szegö, 1939).

and this proves orthogonality of Bessel functions with different zeros λ, μ. And

$$\int_0^1 x\, J_0^2(\lambda x)\, dx > 0$$

is obvious. □

With the Bessel functions we can compute decompositions into orthogonal series just like the Fourier series do with the exponentials $e^{ik\cdot}$; this is done by expanding f into

$$f \sim \sum_{k\in\mathbb{N}} f_k\, J_0(\lambda_k \cdot), \qquad f_k = \frac{\int_0^1 x\, f(x)\, J_0(\lambda_k x)\, dx}{\int_0^1 x\, J_0^2(\lambda_k x)\, dx}$$

which is a **discrete Bessel transform**.

An interesting aspect are the *asymptotics* of Bessel functions, i.e., the question how $J_0(x)$ behaves for $x \to \infty$. This is neither trivial nor simple. According to (Jackson, 1941), one method for such results is to show that $f(x) = \sqrt{x}\, J_0(x)$ satisfies the differential equation

$$f''(x) + \left(1 - \frac{1}{4x^2}\right) f(x) = 0,$$

which yields asymptotics of the form

$$J_0(x) = \alpha \frac{\sin(x+\varphi)}{\sqrt{x}} + \frac{r(x)}{x^{3/2}}, \qquad \sup_{x\in\mathbb{R}_+} |r(x)| < \infty;$$

Jackson refers to (Courant and Hilbert, 1931) there. A behavior like $x^{-1/2}$ is not great, but at least the function decays at a defined rate, other than for example sine and cosine functions.

The index "0" in J_0 suggests that there may be other Bessel functions *of order* $n \in \mathbb{N}_0$ and, yes, those exist, satisfy the differential equation

$$f''(x) + \frac{1}{x} f'(x) + \left(1 - \frac{n^2}{x^2}\right) f(x) = 0,$$

and can be written as

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos\left(x\cos\theta - n(\pi - \theta)\right)\, d\theta, \tag{4.5.11}$$

or, with proper normalization,

$$J_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ix\sin\theta - in\theta}\, d\theta. \tag{4.5.12}$$

geschrieben werden. They have properties like the ones we described in Theorem 4.5.2 for order 0, cf. (Benson, 2007; Jackson, 1941).

A nice application of Bessel functions is to solve the euqation

$$\phi = \sin(\omega t + x\phi), \tag{4.5.13}$$

for ϕ, which has the explicit solution

$$\phi = \sum_{n=1}^{\infty} \frac{2J_n(nx)}{nx} \sin n\omega t. \tag{4.5.14}$$

And (4.5.13) is no artificial problem, the equation apppears naturally in planetary motions but also with frequency modulated sythesizers, see (Benson, 2007), where one can find quite some material about the role of Bessel functions in music synthesis.

4.6 Fourier Reconstruction and the FFT

The conceptionally simplest way to compute an inverse Radon transform, even simpler than filtered backprojection, is to just use the Fourier Slice Theorem, stated as Corollary 3.2.5. Recall that it says that

$$(Rf(v,\cdot))^{\wedge}(\sigma) = \hat{f}(\sigma v), \tag{4.6.1}$$

hence, if $y \in \mathscr{S}(\mathbb{S}^{d-1} \times \mathbb{R})$ is the measurement function, then we can reconstruct f from its Fourier transform

$$\hat{f}(\sigma v) = \frac{1}{2}\left((y(v,\cdot))^{\wedge}(\sigma) + (y(-v,\cdot))^{\wedge}(-\sigma)\right); \tag{4.6.2}$$

here, we symmetrize the measurement into an even function[14] which just reproduces y if it has the proper structure. What we get is, no surprise,

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ix^T\xi} \hat{f}(\xi)\, d\xi \\ &= \frac{1}{2(2\pi)^d} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} e^{i\sigma v^T x} |\sigma|^{d-1} \left((y(v,\cdot))^{\wedge}(\sigma) + (y(-v,\cdot))^{\wedge}(-\sigma)\right) dv d\sigma. \end{aligned}$$

But: how shall we compute this numerically as there are two integrals involved and numerical integration is a topic by its own, especially in several variables.

This brings us to the question how the Fourier transform is implemented numerically, and here we usually rely on the Fast Fourier Transform, also known as FFT. To that end, it is reasonable to first define what the fast Fourier transform does fast.

Definition 4.6.1 (DFT). 1. For $n \in \mathbb{N}$ we use $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$ to denote the integers modulo n which we usually represent[15] as $\{0, \dots, n-1\}$.

2. For a vector

$$c \in \ell(\mathbb{Z}_n^d) := \left\{c : \mathbb{Z}_n^d \to \mathbb{C}\right\}$$

we define the **discrete Fourier transform** of **DFT** of c as

$$\hat{c} := \left(\sum_{\beta \in \mathbb{Z}_n^d} c(\beta) e^{2\pi i \alpha^T \beta / n} : \alpha \in \mathbb{Z}_n^d \right). \tag{4.6.3}$$

[14] Nice exercise: any function $f : \mathbb{R} \to \mathbb{R}$ can be written as the sum of an odd and an even function. And: what is the problem with odd functions in several variables?

[15] In the same way as the torus $\mathbb{R}/2\pi\mathbb{Z}$ is represented by $[0, 2\pi]$ or $[-\pi, \pi]$.

Remark 4.6.2. 1. The DFT maps $\ell(\mathbb{Z}_n^d)$ to itself and depends on the dimension of the vector c we apply it to.

2. The intuition of the DFT is that we first form the trigonometric polynomial

$$f(\xi) = \sum_{\beta \in \mathbb{Z}_n^d} c(\beta) e^{i\beta^T \xi}$$

and the *sample* this function uniformly over the torus at the points $2\pi\frac{\alpha}{n}$ for $\alpha \in \mathbb{Z}_n^d$.

3. This is by far not the most general way to define a DFT. One simple extension which is relevant in practice is to consider a different length in each coordinate, i.e., to consider $\mathbb{Z}_\nu = \mathbb{Z}_{\nu_1} \times \cdots \times \mathbb{Z}_{\nu_d}$ for some $\nu \in \mathbb{N}^d$. The situation from (4.6.3) is the special case $\nu = (n, \ldots, n)$. Since the extension is pretty straightforward but notationally intricate, we will stick with the "uniform" case here.

The DFT is closely related to the **periodic convolution**

$$c * d := c *_n d := \sum_{\alpha \in \mathbb{Z}_n^d} c(\cdot - \alpha)\, d(\alpha) \in \ell(\mathbb{Z}_n^d), \qquad c, d \in \ell(\mathbb{Z}_n^d), \tag{4.6.4}$$

where the subtraction in the argument has to be performed modulo $\mathbb{Z}_n^d$. It is straightforward to show that

$$\hat{c}\,\hat{d} = (c *_n d)^\wedge \tag{4.6.5}$$

holds for the DFT.

Exercise 4.6.1 Prove (4.6.5). ◊

The DFT can be seen as a matrix-vector multiplication

$$\hat{c} = V_n c, \qquad V_n := V_n^d := \left(e^{2\pi i \alpha^T \beta / n} : \begin{array}{l} \alpha \in \mathbb{Z}_n^d \\ \beta \in \mathbb{Z}_n^d \end{array} \right), \tag{4.6.6}$$

with the **Fourier matrix** V_n.

Remark 4.6.3. Note that in (4.6.6) we index the matrix with *multiindices*, not with "normal" linear indices. This reflects the structure of such a matrix but in order to use, for example, standard matrix software, one would have to impose a total ordering on $\mathbb{Z}_n^d$. This could be a *lexicographical ordering*

$$\alpha \prec_l \beta \qquad \text{iff} \qquad \alpha_j = \beta_j,\ j < k, \quad \alpha_k = \beta_k,$$

or a *graded lexicographical ordering*

$$\alpha \prec_l \beta \qquad \text{iff} \qquad |\alpha| < |\beta| \text{ or } |\alpha| = |\beta|, \alpha \prec_l \beta.$$

These "real" matrices, however, will not show a visible structure any more

Exercise 4.6.2 Compute the lexicographically and graded lexicographically ordered Fourier matrices for $d = 3$ and $n = 3$. ◊

The Fourier matrix has the interesting property that its entries

$$(V_n)_{\alpha,\beta} = e^{2\pi i \alpha^T \beta/n} = e^{2\pi i(\alpha_1\beta_1+\cdots+\alpha_d\beta_d)/n} = \prod_{j=1}^{d} e^{2\pi i \alpha_j\beta_j/n}, \qquad \alpha,\beta \in \mathbb{Z}_n^d, \tag{4.6.7}$$

are the products of the entries of univariate Fourier matrices. Such matrices are known as a particular structure that plays a fundamental role in (numerical) Linear Algebra, cf. (Horn and Johnson, 1985; Marcus and Minc, 1969; Van Loan, 2000; Van Loan, 2009).

Definition 4.6.4. The **Kronecker product** $A \otimes B$ of two matrices $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{m'\times n'}$ is the matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mm'\times nn'}. \tag{4.6.8}$$

It is immediate from the definition (4.6.8) that the components of $A \times B$ run exactly over all products $a_{jk}b_{j'k'}$, $1 \le j \le m$, $1 \le k \le n$, $1 \le j' \le m'$, $1 \le k' \le n'$, hence, if we write

$$\begin{pmatrix} c_{\alpha,\beta} : & \alpha \in \mathbb{Z}_m \times \mathbb{Z}_{m'} \\ & \beta \in \mathbb{Z}_n \times \mathbb{Z}_{n'} \end{pmatrix} = C := A \otimes B = \begin{pmatrix} a_{\alpha_1,\beta_1}b_{\alpha_2,\beta_2} : & \alpha \in \mathbb{Z}_m \times \mathbb{Z}_{m'} \\ & \beta \in \mathbb{Z}_n \times \mathbb{Z}_{n'} \end{pmatrix}, \tag{4.6.9}$$

then we immediately see that the multiindex notation reveals the structure of the Kronecker product as a proper natural ordering. By induction, this formula is easily to higher order Kronecker products.

Lemma 4.6.5. *For $A_p = \left(a^p_{jk}\right) \in \mathbb{R}^{m_j\times n_j}$, $p = 1,\dots,q$, the product can be indexed such that*

$$A_1 \otimes \cdots \otimes A_q = \begin{pmatrix} a_{\alpha,\beta} : & \alpha \in \mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_q} \\ & \beta \in \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_q} \end{pmatrix}, \qquad a_{\alpha,\beta} = \prod_{p=1}^{q} a^p_{\alpha_p,\beta_p}. \tag{4.6.10}$$

Proof: For $q = 2$, the claim is (4.6.9), while otherwise we write

$$\begin{aligned} \bigotimes_{p=1}^{q+1} A_p &= A_1 \otimes \left(\bigotimes_{p=2}^{q+1} A_p\right) \\ &= \begin{pmatrix} a^1_{11}\bigotimes_{p=2}^{q+1} A_p & \dots & a^1_{1,n_1}\bigotimes_{p=2}^{q+1} A_p \\ \vdots & \ddots & \vdots \\ a^1_{m_1,1}\bigotimes_{p=2}^{q+1} A_p & \dots & a^1_{m_1,n_1}\bigotimes_{p=2}^{q+1} A_p \end{pmatrix} \end{aligned}$$

and since, by the induction hypothesis,

$$a^1_{jk}\bigotimes_{p=2}^{q+1} A_p = \begin{pmatrix} a^1_{jk}\prod_{p=2}^{q+1} a^p_{\alpha_p,\beta_p} : & \alpha \in \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_{q+1}} \\ & \beta \in \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_{q+1}} \end{pmatrix},$$

we can simply set[16]

$$a_{\alpha,\beta} = a^1_{\alpha_1,\beta_1}\left(\bigotimes_{p=2}^{q+1} A_p\right)_{(\alpha_2,\ldots,\alpha_{q+1}),(\beta_1,\ldots,\beta_{q+1})} = a^1_{\alpha_1,\beta_1}\prod_{p=2}^{q+1} a^p_{\alpha_p,\beta_p}, \qquad \begin{array}{l}\alpha \in \mathbb{Z}_{m_1}\times\cdots\times\mathbb{Z}_{m_{q+1}},\\ \beta \in \mathbb{Z}_{n_1}\times\cdots\times\mathbb{Z}_{n_{q+1}},\end{array}$$

to advance the induction hypothesis. □

Now, (4.6.7) and Lemma 4.6.5 tell us that Fourier matrices have a simple structure.

Corollary 4.6.6. *The Fourier matrix in d variables is a Kronecker product of d univariate Fourier matrices:*

$$V_n^d = \underbrace{V_n^1 \otimes \cdots \otimes V_n^1}_{d}. \tag{4.6.11}$$

Remark 4.6.7. If we would work with different n in different dimensions, i.e., $c \in \ell\left(\mathbb{Z}_{n_1}\times\cdots\mathbb{Z}_{n_d}\right)$, then the DFT is given by the matrix

$$V_n^d = \bigotimes_{j=1}^{d} V^1_{n_j},$$

and only the bookkeeping becomes a bit more intricate.

Finally, also the inverse of a Kronecker product is simple.

Lemma 4.6.8. *If $A_1,\ldots,A_q$ are nonsingular, then so is $A_1\otimes\cdots\otimes A_q$ and*

$$\left(\bigotimes_{p=1}^{q} A_p\right)^{-1} = \bigotimes_{p=1}^{q} A_p^{-1}. \tag{4.6.12}$$

Proof: Again we start with $q = 2$, i.e.,

$$\begin{aligned}
&(A\otimes B)\left(A^{-1}\otimes B^{-1}\right)\\
&= \begin{pmatrix} A\,b_{11} & \ldots & A\,b_{1,n'}\\ \vdots & \ddots & \vdots\\ A\,b_{m',1} & \ldots & A\,b_{m',n'}\end{pmatrix}\begin{pmatrix} A^{-1}\,b_{11}^{-1} & \ldots & A\,b_{1,n'}^{-1}\\ \vdots & \ddots & \vdots\\ A\,b_{m',1}^{-1} & \ldots & A\,b_{m',n'}^{-1}\end{pmatrix}\\
&= \left(\sum_{r=1}^{n'} AA^{-1}b_{jr}b_{rk}^{-1} : \begin{array}{l} j=1,\ldots,m'\\ k=1,\ldots,n'\end{array}\right) = \left(I\,\delta_{jk} : \begin{array}{l} j=1,\ldots,m'\\ k=1,\ldots,n'\end{array}\right) = I\otimes I = I,
\end{aligned}$$

and use induction for the general case. □

Now we have all tools to show that the DFT is invertible.

[16] To be honest: this proof is not mathematics but bookkeeping, but that is worth the while.

Proposition 4.6.9. *Up to normalization, the Fourier matrix is* ***unitary***[17]*, i.e.,*

$$V_n^{-1} = \frac{1}{n^d} V_n^H. \tag{4.6.13}$$

Proof: By Corollary 4.6.6, Lemma 4.6.8 and Exercise 4.6.3, it suffices to compute the inverse of the univariate DFT V_n^1. Denoting by

$$W_n := \frac{1}{n}(V_n^1)^H = \frac{1}{n}\left[e^{-2\pi i jk/n} \ : \ j,k \in \mathbb{Z}_n\right] =: \frac{1}{n}\left[\omega^{-jk} \ : \ j,k \in \mathbb{Z}_n\right],$$

with $\omega := e^{2\pi i/n}$, we observe that

$$(V_n W_n)_{jk} = \frac{1}{n}\sum_{\ell\in\mathbb{Z}_n} \omega^{j\ell}\,\omega^{-\ell k} = \frac{1}{n}\sum_{\ell\in\mathbb{Z}_n}\left(\omega^{j-k}\right)^{\ell}.$$

Hence, the diagonal elements are

$$(V_n W_n)_{jj} = \frac{1}{n}\sum_{\ell\in\mathbb{Z}_n}\left(\omega^{0}\right)^{\ell} = \frac{n}{n} = 1$$

while for the off-diagonal elements with $j \neq k$ we obtain

$$(V_n W_n)_{jk} = \frac{1}{n}\sum_{\ell=0}^{n-1}\left(\omega^{j-k}\right)^{\ell} = \frac{1}{n}\,\frac{\omega^0 - \left(\omega^{j-k}\right)^n}{1-\omega^{j-k}} = \frac{1}{n}\,\frac{1-(\omega^n)^{j-k}}{1-\omega^{j-k}};$$

since $\omega^n = 1$ and $-n < j-k < n$, hence $\omega^{j-k} \neq 1$, we obtain that

$$(V_n W_n)_{jk} = \frac{1}{n}\,\frac{1-1^{j-k}}{1-\omega^{j-k}} = 0$$

hence $V_n W_n = I$ and therefore $W_n = V_n^{-1}$ since V_n is a square matrix[18]. □

Exercise 4.6.3 Show that $(A\otimes B)^T = A^T \otimes B^T$ and $(A\otimes B)^H = A^H \otimes B^H$. ◊

The computation of a univariate DFT appears to have a complexity of $O(n^2)$, since it is the multiplication of a matrix and a vector and the usual effort for this process is $O(n^2)$. This can be improved and the idea behind the FFT is strikingly simple: let us suppose that $n = 2m$ is an even number and remark that

$$\omega^2 = e^{2\pi i 2/n} = e^{2\pi i/m} =: \omega_m, \qquad \omega_n := \omega,$$

then we get for any $c \in \ell(\mathbb{Z}_n)$ and $j \in \mathbb{Z}_n$ that

$$\begin{aligned}
\hat{c}_n(j) &= \sum_{k\in\mathbb{Z}_n} c(k)\,\omega^{jk} = \sum_{k\in\mathbb{Z}_m} c(2k)\,\omega^{2jk} + \sum_{k\in\mathbb{Z}_m} c(2k+1)\,\omega^{j(2k+1)} \\
&= \sum_{k\in\mathbb{Z}_m} c(2k)\,\omega_m^{jk} + \omega^j \sum_{k\in\mathbb{Z}_m} c(2k+1)\,\omega_m^{jk} \\
&= (c(2\cdot))_m^{\wedge}(j) + \omega^j\ (c(2\cdot+1))_m^{\wedge}(j),
\end{aligned}$$

[17] This is the complex version of *orthogonality*, namely $AA^H = I$ where $A^H = \overline{A}^T$.

[18] Keep in mind that for non-square matrices there is a difference between a *left inverse* and a *right inverse*.

which is

$$\hat{c}_n = c_n^\wedge = (c(2\cdot))_m^\wedge + \omega^{\cdot}\,(c(2\cdot+1))_m^\wedge\,. \tag{4.6.14}$$

The **FFT** simply consists of applying this computational rule recursively, which of course requires that $n = 2^k$ for some k, at least in our naive form.

What is now the value of this representation? If we assume that the values $\omega, \ldots, \omega^{n-1}$ are *precomputed* and available in a table[19], the naive version of the DFT as a matrix-vector mulitplication of an $n \times n$ matrix with an n-vector would require $O\left(n^2\right)$ operations. So let us check what a computation via (4.6.14) would need and call that number $F(n)$. Then (4.6.14) tells us that for the computation of $\hat{c}_n$ we first have to compute the two DFTs of length $m = n/2$ on the right hand side (cost of $2F(n/2)$), then multiply the second result componentwise with the vector $\left(\omega^j \;:\; j \in \mathbb{Z}_n\right)$ (cost of n) and the add them componentwise[20] (cost of n). Hence, the total effort in (4.6.14) is $2\,(F(n/2)+n)$, which leads to the *recurrence relation*

$$F(n) = 2\,(F(n/2)+n)\,. \tag{4.6.15}$$

For a dyadic $n = 2^\ell$ with $\ell \in \mathbb{N}$, we thus get

$$F(n) = 2^k F\left(2^{\ell-k}\right) + k\,2^{\ell+1}, \qquad k = 1, \ldots, \ell, \tag{4.6.16}$$

which follows by induction from the fact that for $k = 1$ (4.6.16) is just (4.6.15) while the inductive step is

$$\begin{aligned} F(n) &= 2^k F\left(2^{\ell-k}\right) + k\,2^{\ell+1} = 2^k\,2\left(F\left(2^{\ell-k-1}\right) + 2^{\ell-k}\right) + k\,2^{\ell+1} \\ &= 2^{k+1} F\left(2^{\ell-k-1}\right) + 2^{\ell+1} + k\,2^{\ell+1} = 2^{k+1} F\left(2^{\ell-k-1}\right) + (k+1)\,2^{\ell+1}. \end{aligned}$$

Considering (4.6.16) for the special case $k = \ell = \log_2 n$, this becomes

$$F(n) = \underbrace{2^\ell}_{=n} F(1) + \underbrace{\ell\,2^{\ell+1}}_{=2n\,\log_2 n} = n\left(2\,\log_2 n + F(1)\right) = O\left(n\,\log_2 n\right),$$

which is significantly better than the $O\left(n^2\right)$ of the naive matrix-vector multiplication. Indeed, $O\left(n\,\log_2 n\right)$ is a typical asymptotic complexity[21] for methods based on the principle of *divide and conquer*, and is often derived as a consequence of the so-called *master theorem*, (Steger, 2001).

This complexity results holds not only for numbers n that are powers of 2. If ℓ is chosen such that $2^{\ell-1} < n \le 2^\ell$, then we can simply replace n by 2^ℓ and embed the original signal into a larger one, for example by *padding* it, i.e., adding zeros. Then the computational cost

$$\begin{aligned} 2^\ell F(1) + 2\ell\,2^\ell &\le 2n\,F(1) + 2\log_2(2n)\,2n = 2n\,F(1) + 4n\left(\log_2 n + 1\right) \\ &= 2n\left(2\log_2 n + F(1) + 2\right), \end{aligned}$$

[19] They are the same for all $c \in \ell\,(\mathbb{Z}_n)$ and could be held in some cache; and even if they were not precomputed, the effort for that is $O(n)$.

[20] The vectors are m-periodic and will simply be extended by periodicity.

[21] And the constant here is *very* moderate!

is increased by a factor of two, hence the *asymptotic* complexity of $O\left(n \log_2 n\right)$ still remains valid.

To extend the FFT to multivariate data, we again make use of a computational operation for Kronecker products; to prove it, we need another definition.

Definition 4.6.10. The **vectorization** $v(A) \in \mathbb{R}^{mn}$ of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as

$$v = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \qquad A = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}. \tag{4.6.17}$$

The following cute formula for Kronecker products is classical and can be found, for example, in (Horn and Johnson, 1991; Marcus and Minc, 1969).

Proposition 4.6.11. *For $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$ and $X \in \mathbb{R}^{n\times p}$ we have*

$$v(AXB) = (B^T \otimes A)v(X). \tag{4.6.18}$$

Before proving the proposition, let us first check that the dimensions coincide on both sides of (4.6.18). Since $AXB \in \mathbb{R}^{m\times q}$ the expression on the left hand side is a vector of size mq, the Kronecker product on the right hand side belongs to $\mathbb{R}^{qm\times pn}$ and $v(X)$ is an np–vector, so the right hand side indeed also gives a vector of size mq.

Proof: We write $X = \begin{pmatrix} x_1 & \dots & x_p \end{pmatrix}$ in terms of it colums, $B = \left(b_{jk} : \begin{array}{c} j = 1,\dots,p \\ k = 1,\dots,q \end{array} \right)$ and get for the ℓth column of the product that

$$\begin{aligned} (AXB)_\ell &= AXBe_\ell = AX\left(b_{j\ell} : j = 1,\dots,p\right) \\ &= A\left(\sum_{j=1}^{p} x_j b_{j\ell}\right) = \sum_{j=1}^{p} b_{j\ell}\, Ax_j = \begin{pmatrix} b_{1\ell}A & \dots & b_{p\ell}A \end{pmatrix} v(X) \\ &= \left((Be_\ell)^T \otimes A\right) v(X) \end{aligned}$$

and therefore

$$\begin{aligned} v(AXB) &= \begin{pmatrix} (AXB)_1 \\ \vdots \\ (AXB)_q \end{pmatrix} = \begin{pmatrix} \left((Be_1)^T \otimes A\right) v(X) \\ \vdots \\ \left((Be_p)^T \otimes A\right) v(X) \end{pmatrix} = \begin{pmatrix} \left((Be_1)^T \otimes A\right) \\ \vdots \\ \left((Be_p)^T \otimes A\right) \end{pmatrix} v(X) \\ &= \left(\begin{pmatrix} (Be_1)^T \\ \vdots \\ (Be_p)^T \end{pmatrix} \otimes A \right) v(X) = (B^T \otimes A)v(X) \end{aligned}$$

as claimed. □

The "Kronecker trick" (4.6.18) allows us to compute the product of a Kronecker product $A_1 \otimes \cdots \otimes A_s$, $A_j \in \mathbb{R}^{m_j \times n_j}$ and a vector $x \in \mathbb{R}^n$, $n = n_1 \cdots n_q$, which we interprete as $v(X)$ for a matrix $X \in \mathbb{R}^{n_2 \cdots n_q \times n_1}$. We obtain

$$\begin{aligned}(A_1 \otimes \cdots \otimes A_q)x &= (A_1 \otimes \cdots \otimes A_q)v(X) = (A_2 \otimes \cdots \otimes A_q)XA_1^T \\ &=: (A_2 \otimes \cdots \otimes A_q)X_1\end{aligned}$$

where

$$X_1 = XA_j^T = (A_j X^T)^T \in \mathbb{R}^{n_2 \cdots n_s \times n_1}.$$

is computed by determining the products

$$Ax^j, \qquad x^j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{i,n_1} \end{pmatrix}, \qquad j = 1, \ldots, n_2 \cdots n_s.$$

of A with the *rows* of X. If $A_j = V_n^1$ in the case of the multivariate DFT, then each of these products can be computed by means of the FFT which costs $O(n_1 \log n_1)$, but we have to compute it for $n_2 \cdots n_s$ columns, so the total cost is

$$\left(C\, n_1 \log n_1\right) n_2 \cdots n_s = Cn \log n_1.$$

Applying the same procedure to all n_1 columns of X_1 needs $C\left(C\, n_2 \log n_2\right) n_3 \cdots n_s$ operations, and since there are n_1 of these columns, the cost is

$$n\left(C\, n_2 \log n_2\right) n_3 \cdots n_s = Cn \log n_2.$$

Proceeding iteratively, it is easy to see that the total cost for the d-dimensional FFT for a signal of the form (4.6.3) has a cost of

$$C \sum_{j=1}^{d} n \log n_j = Cn \log(n_1 \cdots n_d) = Cn \log n,$$

with the same constant C as in the one-dimensional case.

In most applications, also in the ones we are considering here, the measurements are equidistant samples

$$c(\alpha) = (S_h f)(\alpha), \qquad \alpha \in \mathbb{Z}_n^d,$$

on an $n \times \cdots \times n$ grid. If we compute the DFT $\hat{c}$ of this signal c, then we compute, from the mathematical point of view, a discretization of the associated trigonometric polynomial

$$\hat{c}(\xi) = \sum_{\alpha \in \mathbb{Z}} c(\alpha)\, e^{i\alpha^T \xi} = \sum_{\alpha \in \mathbb{Z}} f(h\alpha)\, e^{i\alpha^T \xi}$$

on the grid $2\pi\mathbb{Z}_N/N$, that is,

$$(S_h f)^\wedge = \hat{c}_n(\alpha) = \sum_{\beta \in \mathbb{Z}_N} f(h\beta)\, e^{-2i\pi\beta^T \alpha/n}, \qquad \alpha \in \mathbb{Z}_n^d. \tag{4.6.19}$$

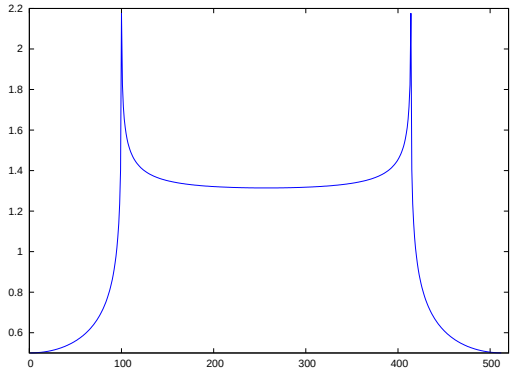

Figure 4.6.1: The modulus of the DFT of the sinc function. The substantial artifacts that we observe at the boundary we see nasty artifacts that cannot be explained by discretization alone.

Remark 4.6.12. (4.6.19) shows that there is no direct connection between the vector $(S_h f)^\wedge$ and what one usually wants to compute, namely the discretization of $\hat{f}$, the Fourier transform of the function f that was originally sampled.

It should come as no surprise that this leads to artifacts again.

Example 4.6.13. Let us consider the DFT of a sampled univariate sinc-function whose Fourier transform is[22] a characteristic function. We sample this function, in `octave` notation via

```
>> N = 512; c = sinc( 100*pi*(0:N-1)/N );
```

at locations that are no integers. The result of a DFT combined with the operation `fftshift`[23] is depicted in Fig. 4.6.1 and shows that this bandpass filter has a severe overshooting at the high frequencies.

To approximate the function f itself, we incorporate a so-called **quasi interpolant** which is defined, for a function $\phi \in L_{00}(\mathbb{R}^d)$ as

$$Q_{h,\phi} c := \phi * c\left(h^{-1}\cdot\right) = \sum_{\alpha\in\mathbb{Z}_n^d} c(\alpha)\phi\left(h^{-1}\cdot - \alpha\right) = \sum_{\alpha\in\mathbb{Z}_n^d} f\left(h\alpha\right)\ \phi\left(h^{-1}\cdot - \alpha\right); \quad (4.6.20)$$

if ϕ is even a cardinal function, that is, $\phi|_{\mathbb{Z}^d} = \delta$, then $Q_{h,\phi}(h\alpha) = S_h f(\alpha) = f(h\alpha)$, $\alpha \in \mathbb{Z}_n^d$, and the data is even interpolated without and "quasi". But in general, one at least hopes to obtain a reasonable approximation.

If we assume[24] that $Q_{h,\phi} c$ *approximates* the function f, meaning that $\left\|f - Q_{h,\phi} c\right\|_1$ is small, then the Fourier transform of $Q_{h,\phi} c$ is also a good approximation of the Fourier transform $\hat{f}$ of f and we can compute the latter from the samples as

$$\left(Q_{h,\phi} c\right)^\wedge(\xi) = \left(\sigma_{h^{-1}}\left(\phi * c\right)\right)^\wedge(\xi) = h\left(\phi * c\right)^\wedge(h\xi) = h\,\hat{\phi}(h\xi)\,\hat{c}(h\xi), \qquad \xi \in \mathbb{R}^d.$$

[22] Or at least should be.
[23] This is an `octave` function that shifts the zero frequency into the middle of the vector.
[24] Otherwise we definitely made a mistake with the quasi interpolant.

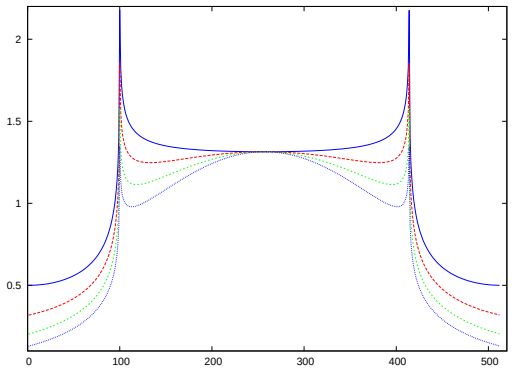

Figure 4.6.2: Filtering of the function from Fig. 4.6.1 with cardinal splines of order $0, 1, 2, 3$. The outliers become smaller with increasing order, however for the price of a bump at the boundary.

If we replace ξ by $h^{-1}\xi$, in this identity, and then discretize ξ on the *discrete torus* $\mathbb{T}_n := \frac{2\pi}{n}\mathbb{Z}_n^d$, we obtain

$$\hat{f}\left(\frac{2\alpha\pi}{nh}\right) \approx \left(Q_{h,\phi}c\right)^{\wedge}\left(\frac{2\alpha\pi}{nh}\right) = h^d\,\hat{\phi}\left(\frac{2\alpha\pi}{n}\right)\hat{c}(\alpha), \qquad \alpha \in \mathbb{Z}_n^d. \tag{4.6.21}$$

This simple formula eventually connects the discrete Fourier transform $\hat{c} = (S_h f)^{\wedge}$ of the samples with an *approximate* discretization of the Fourier transform $\hat{f}$ of f and also explains the practical relationship:

1. The **sampling rate** h determines which frequencies of f are really encoded in the DFT $\hat{c}$. The smaller h is, the smaller this frequency range becomes. This, of course, should come as no surprise.

2. The **frequency resolution** as the number of entries of the spectrum that are computer depends on the number n of the samples. The larger n is, the more precisely the spectrum is represented and the smaller is the distance between these frequencies. If n is small, on the other hand, more entries will be combined into one. Of course, the computational effort increases with n as well.

3. One cannot decouple these two parameters so easily. Normally the discrete data results from sampling over a substantial range or period of size nh, so that a high sampling rate will usually be related to a high frequency resolution as well.

4. If tensor product cardinal spline functions

$$\phi(x) = \prod_{j=1}^{d} \underbrace{\chi * \cdots * \chi}_{m}$$

of order m are chosen for the quasi interpolant, the correction filter $\hat{\phi}$ is simply a power of products of the sinc function; since we sample only the

first "hill" of this function which decreases faster with increasing order of the spline, a larger value of m results in a stronger damping of unwanted high frequency contributions, see Fig. 4.6.2.

5. If we simply drop or ignore ϕ in (4.6.20), then we replace $\hat{\phi}$ by $\chi_{[0,1]^d}$, which means that $\phi = \operatorname{sinc}$ and instead of $\hat{f}$ one discretizes the interpolatory reconstruction from the sampling theorem, Theorem 4.2.5. Although this sounds like a great idea, at least for band limited functions, the approximation quality is more on the poor side due to the slow decay of the sinc function.

4.7 Back to Fourier Reconstruction

As mentioned several times, the Radon transform is only known for finitely many discrete values in practical applications. In the most classical case this is done by sampling uniformly distributed directions v_j, $j = 0, \dots, p-1$ and offsets s_k, $k = -q, \dots, q$, more precisely, according to the following definition.

Definition 4.7.1 (Parallel scanning). In **parallel scanning** with resolution $\rho > 0$, the Radon transform Rf is measured for the parameters

$$v_j \in \mathbb{S}^{d-1} \qquad j = 0, \dots, p-1, \tag{4.7.1}$$

$$s_k = \frac{\rho k}{q}, \qquad k = -q, \dots, q, \tag{4.7.2}$$

that is, we know the $p(2q+1)$ values

$$y_{jk} = y\left(v_j, s_k\right) = Rf\left(v_j, s_k\right), \qquad j = 0, \dots, p-1, \quad k = -q, \dots, q.$$

Remark 4.7.2. For $d = 2$, the v_j can be explicitly written as

$$v_j = \begin{pmatrix} \cos\varphi_j \\ \sin\varphi_j \end{pmatrix}, \qquad \varphi_j = \frac{\pi j}{p}, \qquad j = 0, \dots, p-1. \tag{4.7.3}$$

Moreover, the Radon transform and the X-ray transform coincide in this case. The geometry is depicted in Fig. 4.7.1.

Remark 4.7.3. The resolution of the offsets in (4.7.2) reflects the fundamental assumption that the object to be measured in of finite size and contained in in the ball $B_\rho = B_\rho(0)$ of radius ρ around the origin, in other words,

$$f(x) = 0, \qquad \|x\|_2 > \rho. \tag{4.7.4}$$

In accordance with the Shannon sampling theorem, Theorem 4.2.5, on the other hand, we would also like the function f to be bandlimited. This, however, contradicts the finite support requested in (4.7.4), as a consequence of the Heisenberg uncertainty principle. Therefore, there will be no exact reconstructions[25] any more, only approximation; but this is just numerical reality.

[25] Which in general would be impossible for finite data anyway.

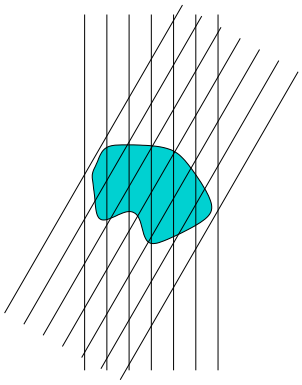

Figure 4.7.1: Parallel scan geometry: bundles of parallel lines intersect the object at different angles.

For each $j \in \mathbb{Z}_p$ and the associated v_j, we therefore can compute a discrete approximation [26]

$$\big(R(v_j,\cdot)\big)^\wedge(\xi) \approx \frac{q}{\rho}\sum_{k=-q}^{q} e^{-is_k\xi} Rf(v_j,s_k) = \frac{q}{\rho}\sum_{k=-q}^{q} e^{-ik\frac{\rho}{q}\xi} Rf(v_j,s_k). \tag{4.7.5}$$

The way of approximating an integral by a finite number of point evaluations is not only the idea behind the Riemann sum in the definition of the Riemann integrarl, it is also a fundamental concept in Numerical Analysis, cf. (Gautschi, 1997; Isaacson and Keller, 1966).

Definition 4.7.4. A **quadrature formula** for an integral $I(f) := \int f(x)w(x)dx$ is a sum of the form

$$Q(f) = \sum_{j=0}^{M} w_j f(x_j).$$

The quadrature formula based on equidistant sampling that we used in (4.7.5) is known as the **rectangular rule**.

Sampling (4.7.5) at $\xi_\ell = \frac{2\pi q}{\rho(2q+1)}\ell$, we get for[27] $\ell \in \mathbb{Z}_{2q+1}$ that

$$\begin{aligned}\big(R(v_j,\cdot)\big)^\wedge(\xi_\ell) &\approx \big(R(v_j,\cdot)\big)^\wedge_{2q+1}(\ell) = \big(R(v_j,\cdot)\big)^\wedge\left(\frac{2\pi q}{\rho(2q+1)}\ell\right)\\ &= \frac{q}{\rho}\sum_{k\in\mathbb{Z}_{2q+1}} e^{-2\pi ik\ell/(2q+1)} Rf(v_j,s_k) = \hat{c}_j(\ell),\end{aligned} \tag{4.7.6}$$

where

$$c_j = \left(\frac{q}{\rho}\, Rf(v_j,s_k) : k \in \mathbb{Z}_{2q+1}\right).$$

[26] The difference between two points s_k is $\frac{\rho}{q}$, hence the approximation for the integral is an expression of the form

$$\int f(x)\,dx = \frac{1}{h}\sum_{j=0}^{M} f(hj),$$

where Mh is the length of the region of integration.

[27] Here we use $\mathbb{Z}_{2q+1}$ with the representer set $\{-q,\dots,q\}$.

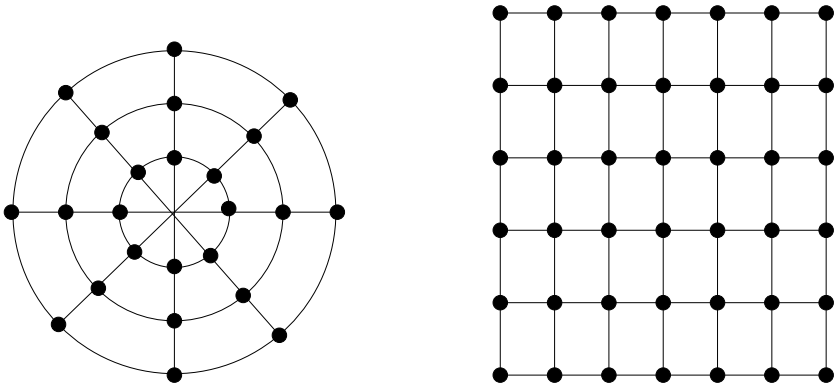

Figure 4.7.2: Two different domains for the Fourier transform. It can be determined from the Radon transform on the polar grid *(left)* but has to be transferred to the cartesian grid *(right)*.

Therefore, (4.7.6) can be computed by an FFT and everything that has been said about the computational and numerical aspects of the FFT above can be applied here.

By (4.6.1) we thus approximate

$$\hat{c}_j(k) \approx \hat{f}\left(v_j \xi_k\right), \qquad j \in \mathbb{Z}_p, k \in \mathbb{Z}_{2q+1},$$

and therefore we know $\hat{f}$ on the *polar grid*. Here is the problem with the Fourier approach: we can determine the Fourier transform on the polar grid, but in order to apply an (inverse) FFT, the values have to be available on the rectangular grid, see Fig. 4.7.2.

The transformation of the Fourier transform is performed by **interpolation**: let $Q = [a_1, b_1] \times \cdots \times [a_d, b_d]$ be a d-dimensional cube such that

$$v_j \xi_k \in Q, \qquad j \in \mathbb{Z}_p, k \in \mathbb{Z}_{2q+1},$$

and find a function $\varphi : Q \to \mathbb{C}$ such that

$$\varphi(v_j \xi_k) = \hat{c}_j(k), \qquad j \in \mathbb{Z}_p, k \in \mathbb{Z}_{2q+1}, \tag{4.7.7}$$

from which one can determine f on a grid of the form $\mathbb{Z}_n^d$ by

$$f|_{\mathbb{Z}_n^d} = \left(\varphi(h\alpha + \alpha_0) : \alpha \in \mathbb{Z}_n^d\right)^\vee,$$

again by means of an FFT. Of course, the quality of the reconstruction depends significantly on the interpolation process; details can be seen in (Natterer, 1986).

Remark 4.7.5 (Interpolation and Fourier method).

1. The simplest method mentioned in (Natterer, 1986) is a *nearest neighbor interpolation*: for any grid point $\xi_\alpha := \alpha_0 + h\alpha$ use the value $\hat{c}_j(k)$ such that

$$\left|\xi_\alpha - v_j \xi_k\right| = \min_{j',k'} \left|\xi_\alpha - v_{j'} \xi_{k'}\right|.$$

2. Since the points are more dense in the interior of the circle, the deviation between the next point in the polar and the cartesian grid becomes larger away from the origin. This is yet another reason to apply low pass filters.

4.8 Reconstruction by Filtered Backprojection

Reconstruction by means of the filtered backprojection involves filtering, i.e., a convolution, as well as the dual Radon transform which can also be expressed in terms of a Fourier transform as shown in (4.3.6). To write the right hand side of the reconstruction formula (4.3.1) in this way, we let $y = y(v, s)$ be the measurement function with Fourier transform $\hat{y}_v(\sigma)$, $\sigma \in \mathbb{R}$. Next, we pick a radial filter with $\hat{g}(\sigma) = |\sigma|^{d-1}\psi(\sigma)$ and thus

$$(g * y)^\wedge(\xi) = |\xi|_2^{d-1} \hat{y}_{\xi/|\xi|_2}(|\xi|_2)\ \hat{\psi}(|\xi|_2),$$

which leads to

$$\left(R^*(g * y)\right)^\wedge(\xi) = \hat{y}_{\xi/|\xi|_2}(|\xi|_2)\ \psi(|\xi|_2). \tag{4.8.1}$$

Hence, the Fourier method to filtered backprojection compares (4.8.1) with

$$\left((R^*g) * f\right)^\wedge(\xi) = (R^*g)^\wedge(\xi)\ \hat{f}(\xi) = |\xi|_2^{1-d}\psi(|\xi|_2)\ \hat{f}(\xi).$$

Now $\hat{\psi}$ cancels in this representation and it seems that the backprojection filter plays no role. That the factors cancels is actually due to a somewhat cyclical reasoning since it is this cancellation that makes the filtered backprojection work in Theorem 3.2.8; and practically it plays a role again due to the incompatibility of grids that we encountered in the last section.

Now we assume that $f \in C_{00}^\infty(\Omega)$ is compactly supported and consider a more direct discretization for the reconstruction from the Radon transform by means of

$$G * f = R^*(g * y) \tag{4.8.2}$$

where G by construction is "almost 'δ" and therefore we have to provide a discretization for the convolution integral

$$(g * y)(v, s) = \int_{\mathbb{R}^d} g(s - t)\ y(v, s)\ ds = \int_\Omega g(s - t)\ y(v, s)\ ds, \tag{4.8.3}$$

since we assumed that $y(v, \cdot)$ vanishes outside Ω, where we can assume that $\Omega = B_\rho(0)$ for some $\rho > 0$. For bandlimited functions, the integral can even be computed exactly from (infinitely many) samples, which is yet another sampling theorem in the tradition of Shannon, see (Natterer and Wübbeling, 2001, Theorem 4.1).

Lemma 4.8.1. *If $f \in L_1(\mathbb{R}^d)$ is T-bandlimited and $0 < h < \frac{\pi}{T}$, then*

$$\int_{\mathbb{R}^d} f(x)\, dx = h^d \sum_{\alpha \in \mathbb{Z}^d} f(h\alpha). \tag{4.8.4}$$

Proof: The proof relies on the Poisson Summation Formula, Theorem 2.2.5; we set $g = h^d f(h\cdot)$, so that (2.2.5) and (2.1.7) yields

$$\begin{aligned} h^d \sum_{\alpha \in \mathbb{Z}^d} f(h\alpha) &= \sum_{\alpha \in \mathbb{Z}^d} g(\alpha) = \sum_{\alpha \in \mathbb{Z}^d} \widehat{g}(2\alpha\pi) = \sum_{\alpha \in \mathbb{Z}^d} \left(h^d f(\cdot/h)\right)^\wedge (2\alpha\pi) \\ &= \sum_{\alpha \in \mathbb{Z}} \widehat{f}\left(\frac{2\alpha\pi}{h}\right) = \widehat{f}(0) + \sum_{\alpha \neq 0} \widehat{f}\left(\frac{2\alpha\pi}{h}\right). \end{aligned} \tag{4.8.5}$$

Since for $\alpha \neq 0$,

$$\left|\frac{2\alpha\pi}{h}\right|_2 \geq |2\alpha T|_2 = 2\underbrace{|\alpha|_2}_{>1} > T$$

the sum on the right hand side of (4.8.5) vanishes and

$$h^d \sum_{\alpha\in\mathbb{Z}^d} f(h\alpha) = \widehat{f}(0) = \int_{\mathbb{R}^d} f(x)\,dx$$

remains. □

As already mentioned, a function cannot be compactly supported and bandlimited at the same time, so that Lemma (4.8.1) is nice to have but cannot be satisfied. Having to give up one of them, we relax the Fourier condition.

Definition 4.8.2. A function $f \in L_1(\mathbb{R}^d)$ is called **essentially bandlimited** with bandwidth T or *essentially T-bandlimited* for short, if there exists some[28] $\varepsilon > 0$ such that

$$\int_{|\xi|\geq T} \left|\hat{f}(\xi)\right| d\xi < \varepsilon. \tag{4.8.6}$$

One can show that for essentially bandlimited functions the Shannon Sampling Theorem holds up to an error of $C\varepsilon$ if $h < \frac{\pi}{T}$ where the constant C can be given explicitly and independently of f. Something similar holds for (4.8.4) since $\hat{f}$ is a uniformly continuous function.

To sketch the main idea in the bivariate case, let assume that the s-sampling is fine enough for the Quadrature in Lemma 4.8.1 to be (almost) exact, then we can approximate the convolution integral by

$$(g * y)(v,s) \approx \frac{\rho}{q} \sum_{k=-q}^{q} g\left(s - \frac{\rho\,k}{q}\right) y\left(v, \frac{\rho\,k}{q}\right). \tag{4.8.7}$$

For the numerical treatment of the backprojection

$$R^*(g * y)(x) = \frac{1}{2\pi} \int_{\mathbb{S}^2} (g * y)\left(v, v^T x\right) dv, \qquad x \in \mathbb{R}^2, \tag{4.8.8}$$

we have to make sure that for any $x \in \mathbb{R}^2$ the function

$$h(v) = (g * y)\left(v, v^T x\right), \qquad v \in \mathbb{S}^2 \simeq \mathbb{T}$$

is essentially bandlimited, that is, that its Fourier coefficients[29]

$$\widehat{h}(k) = \int_{\mathbb{T}} e^{-ik\theta} (g * y)\left(v_\theta, v_\theta^T x\right), \qquad v_\theta = \begin{pmatrix}\cos\theta \\ \sin\theta\end{pmatrix}$$

[28] This is part of the definition which therefore defines T and an associated ε that should of course be small.

[29] Since $\mathbb{S}^1 \simeq \mathbb{T}$, we can interpret this in terms of 2π-periodic functions.

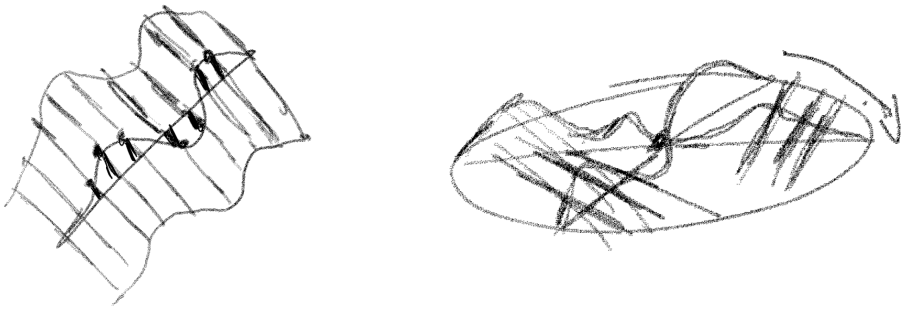

Figure 4.8.1: The nature of a ridge function *(left)* and the idea of the reconstruction as an overlay of rotated ridge functions *(right)*.

are neglegible for large $|k|$ which allows us to apply Lemma 4.8.1 to (4.8.8) giving the final simple discretization

$$f(x) \approx (G * f)(x) = \frac{\pi\rho}{pq} \sum_{j=0}^{p-1} \sum_{k=-q}^{q} g\left(v_j^T x - \frac{\rho\, k}{q}\right) y_{jk} \tag{4.8.9}$$

of the reconstruction.

Remark 4.8.3. The formula (4.8.9) has a nice geometric interpretation. For each j, the function

$$\frac{\pi\rho}{pq} \sum_{k=-q}^{q} g\left(v_j^T x - \frac{\rho\, k}{q}\right) y_{jk} \tag{4.8.10}$$

is a so-called **ridge function**, i.e., a function of the form

$$f(x) = g(v^T x), \qquad f : \mathbb{R}^d \to \mathbb{R}, \qquad g : \mathbb{R} \to \mathbb{R}. \tag{4.8.11}$$

Such functions are constant along $v^\perp$, see Fig. 4.8.1, which is easily seen by writing $x = \lambda v + y$, $y \in v^\perp$ and then noting that

$$f(x) = g\left(v^T(\lambda v + y)\right) = g\left(\lambda |v|_2^2\right).$$

These ridge functions depend on the data y_{jk}, $k = -q, \ldots, q$, i.e., on the measurements of the jth parallel beam, and thus can already be computed after this measurement is finished without waiting for the other measurements. As also shown in Fig. 4.8.1, these ridge functions are rotated over the reconstruction domain[30] and summed up. Making use of such geometric observations not only gives insight into reconstruction, but also allows for fast implementations of the algorithm.

To finally check for the validity of the discretization, let us fix $x \in \mathbb{R}^2$ and look

[30] Here it makes good sense to choose it circular.

at h. Since we restricted the support of $\widehat{g}$ to $[-\omega, \omega]$, see (4.4.2), we have

$$\begin{aligned} h(v) = (g * y)\left(v, v^T x\right) &= \left((g * y)^\wedge\right)^\vee \left(v, v^T x\right) = \left(\widehat{g}\,\widehat{y}\right)^\vee \left(v, v^T x\right) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{g}(\sigma)\, \widehat{y}\,(v, \sigma)\; e^{i\sigma v^T x}\, d\sigma = \frac{1}{2\pi} \int_{-\omega}^{\omega} \widehat{g}(\sigma)\; \underbrace{\widehat{y}\,(v, \sigma)}_{=(Rf)^\wedge(v,\sigma)}\; e^{i\sigma v^T x}\, d\sigma \\ &= \frac{1}{2\pi} \int_{-\omega}^{\omega} \widehat{g}(\sigma)\, \widehat{f}\,(\sigma v)\; e^{i\sigma v^T x}\, d\sigma = \frac{1}{2\pi} \int_{-\omega}^{\omega} \widehat{g}(\sigma) \int_{\mathbb{R}^2} f(t)\, e^{-i\sigma v^T t}\, dt\, e^{i\sigma v^T x}\, d\sigma \\ &= \frac{1}{2\pi} \int_{|t|_2 \le \rho} f(t) \int_{-\omega}^{\omega} \widehat{g}(\sigma)\, e^{i\sigma v^T (x-t)}\, d\sigma\, dt \end{aligned}$$

and therefore

$$\widehat{h}(k) = \frac{1}{2\pi} \int_{|t|_2 \le \rho} f(t) \int_{-\omega}^{\omega} \widehat{g}(\sigma) \underbrace{\int_{\mathbb{T}} e^{i\sigma v_\theta^T (x-t) - ik\theta}\, d\theta}_{\sim J_k(\sigma \|x-t\|_2)}\, d\sigma\, dt, \qquad k \in \mathbb{Z}. \tag{4.8.12}$$

Now the asymptotics of the Bessel functions can be used to verify that indeed the function is essentially bandlimited.

Remark 4.8.4.

1. There exist quantitative estimates for all parts of these reconstruction methods; they are sometimes tedious but important and valuable since they not only show which quality of reconstruction can be guaranteed but often also highligh which parameters are critical. We will not consider them here as they would be a lecture of its own, but only refer to (Natterer, 1986; Natterer and Wübbeling, 2001).

2. In the same way, backprojection methods usually have to be crafted for any scan geometry separately. Again, see (Kak and Slaney, 2001) or (Natterer and Wübbeling, 2001). We will consider a more universal technique, called Algebraic Reconstruction Technique or ART, in the next chapter.

3. For example, in fan beam scans where the projections are not computed along parallel rays but along rays emerging fan-like from the source, one makes use of the fact that in 2D the Radon transform and the X-ray transform are in fact the same. But the computations are quite a bit more complicated.

Algebraic Reconstruction Techniques

5

Art is never finished, only abandoned.

(Leonardo da Vinci)

In the last chapter, we consider a totally different type of reconstruction algorithm, called *Algebraic Reconstruction Technique* (*ART*). It consists of writing the reconstruction problem as a linear system

$$Ax = b \qquad \Leftrightarrow \qquad a_j^T x = b_j$$

where each component of x corresponds to a pixel or voxel of the solution and each equation $a_j^T x = b_j$ to one projection datum. The approach is highly flexible, but

1. it is difficult to make statements about the existence of solutions, in particular since $Ax = b$ can be overdetermined[1] or underdertermined[2],
2. the vector x can and will be very large.

Especially the second point will limit us to special and rather simple iterative methods since keeping the whole matrix A or even the vector x in memory can be quite nontrivial.

5.1 The Setup

To discretize the tomography problem, we put a pixel or voxel grid over the area of measurement and assume that the data is piecewise constant on this grid, see Fig. 5.1.1. Any ray that we shoot through the domain, say ray number j, intersects some grid elements $k_1(j), \ldots, k_{n(j)}(j)$, where both the number and the index of the element depend on j. Now we use the assumption that the value there is constant, then the contribution of the intersection of the ray with this element is the length of the intersection, say $a_{k_\ell(j)}$, and the whole discrete line integral is

$$\sum_{\ell=1}^{n(j)} a_{k_\ell(j)}\, x_{k_\ell(j)}$$

which should be b_j. If we write this as $a_j^T x = b_j$, $j = 1, \ldots, M$, where M denotes the number of projections, then we found a linear system $Ax = b$ to be solved.

[1] There will usually be *no* solution.

[2] We have to decide between a lot of solutions.

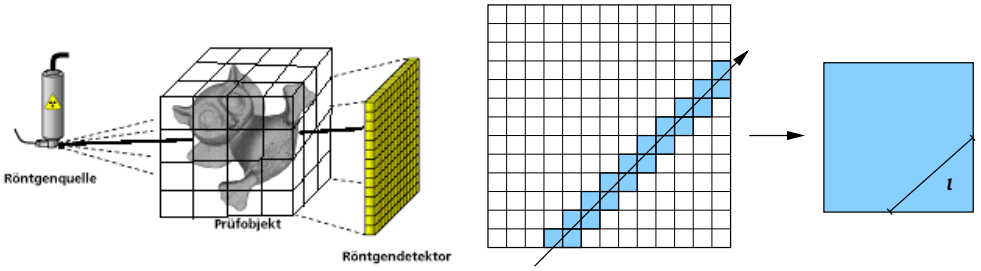

Figure 5.1.1: Main idea of the ART discretization. We put a grid into the domain of interest and then reconstruct values on this grid *(left)*. Then rays are traced through the grid an the length of the intersection of the ray with a grid element defines the contribution of this element to the discretized line integral *(right)*.

Definition 5.1.1. The matrix A of the above discretization is called the **system matrix** and depends only on the scan geometry.

This approach has two obvious advantages:

1. it works with arbitrary scan geometries since any projection is simply one linear equation and the origin and the direction of the beam can be encoded in the system matrix,

2. the resolution of the measurements and the resolution of the reconstruction can be decoupled.

As nice as this advantages are, they do not come for free. First, of all, the relationship between resolutions affect the solvability of the problem:

1. if the resolution of the reconstruction is too high, it can happen that some voxels are not visited by any ray and the system is definitely underdetermined as the value of these voxels does not affect the right hand side.

2. If the resolution is too small, the system may become overdetermined and unsolvable; just consider the extremal case of a single grid element and a non-constant function.

The main problem, however, is the size of the system. Modern flat panel detectors used in industrial tomography have a resolution of 1000×1000 pixel, so a single image already gives 10^8 equations and $200 - 1000$ of these images are not a rarity, which means that we may easily have $10^{11} - 10^{12}$ equations. If we now try to reconstruct 1000^3 voxels[3], then also have 10^{12} variables.

Remark 5.1.2. It first occurs that the such linear system of the size $10^{12} \times 10^{12}$ is out of scope for almost any type of computer. However, the matrices are **sparse**, that is, the number of entries in each row of the matrix is small relative to the size of the matrix. If, for example, the grid is an $N \times N \times N$ one, then we have N^3 variables, but any ray we shoot only has $O(N)$ entries that can be determined on the fly while tracing the line.

[3] This is no longer considered to be really large, the good objects have something like $6k \times 10k \times 10k$.

The observation of Remark 5.1.2 suggests that we should use an *iterative* solver for the linear system that uses rows of the matrix A. Such an algorithm is the classical Kaczmarz algorithm from (Kaczmarz, 1937), which we will consider next.

5.2 Naive Kaczmarz

Definition 5.2.1. We write a matrix $A \in \mathbb{R}^{M\times N}$ as

$$A = \left(a_j^T : j = 1, \dots, M\right) = \begin{pmatrix} a_1^T \\ \vdots \\ a_M^T \end{pmatrix}, \qquad a_j \in \mathbb{R}^N, \tag{5.2.1}$$

and therefore the linear system $Ax = b$ as

$$a_j^T x = b_j, \qquad j = 1, \dots, M. \tag{5.2.2}$$

Note that (5.2.2) means that

$$0 = a_j^T x - b_j \qquad \Leftrightarrow \qquad x \in H(a_j, b_j),$$

so that a condition of the linear system is satisfied iff x lies on the respective hyperplane. If it does not, it sounds like a reasonable idea to improve this condition by projecting x on this hyperplane.

Definition 5.2.2. The **update step** of the Kaczmarz algorithm consist of

$$x \to x + \frac{1}{|a_j|_2^2} a_j \left(b_j - a_j^T x\right). \tag{5.2.3}$$

Since the update step (5.2.3) is undefined at best if $a_j = 0$, we clearly exclude this situation. The meaning of (5.2.3) is clear if we note that

$$a_j^T \left(x + \frac{1}{|a_j|_2^2} a_j \left(b_j - a_j^T x\right)\right) = a_j^T x + \underbrace{\frac{a_j^T a_j}{|a_j|_2^2}}_{=1} \left(b_j - a_j^T x\right) = b_j.$$

This means that the update step "corrects" x in such a way that the jth equation is satisfied. Moreover, x is modified by multiple of a_j, the normal vector to the hyperplane $H(a_j, b_j)$ which gives the geometric interpretation of (5.2.3) as the projection of x onto $H(a_j, b_j)$.

Definition 5.2.3. The **Kaczmarz algorithm** computes the sequence

$$x^{(n)} = x^{(n-1)} + \delta_{j(n)}^{(n)} a_{j(n)} := x^{(n-1)} + \frac{b_{j(n)} - a_{j(n)}^T x^{(n-1)}}{\left|a_{j(n)}\right|_2^2} a_{j(n)}, \qquad n \in \mathbb{N}, \tag{5.2.4}$$

starting with some initial value $x^{(0)}$ and some choice $n \mapsto j(n)$ of rows. For the particular choice $j(kN + \ell) = \ell$, $k \in \mathbb{N}$, $\ell \in \mathbb{Z}_N$, we speak of the **cyclic Kaczmarz algorithm**, see Fig. 5.2.1.

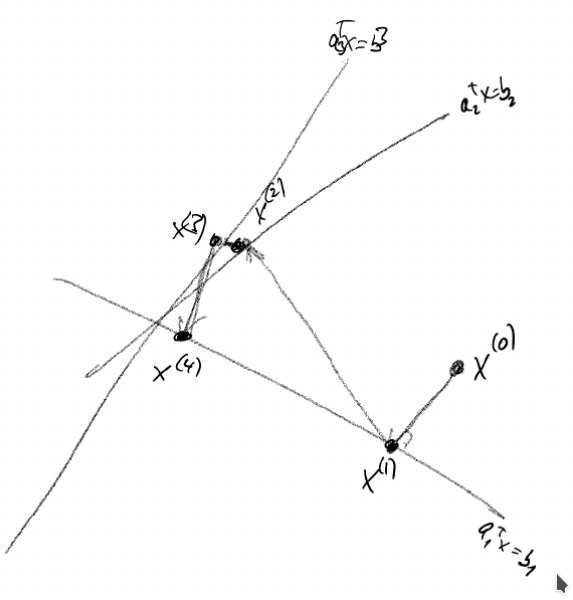

Figure 5.2.1: One cycle in the cyclic Kaczmarz algorithm for three equations in 2d.

Definition 5.2.4. By

$$A^T\mathbb{R}^M := \left\{ \sum_{j=1}^{M} \alpha_j a_j : \alpha_j \in \mathbb{R} \right\} \subseteq \mathbb{R}^N \tag{5.2.5}$$

we denote the **span** of the rows of A.

And this simple algorithm really works. This is indeed surprising since, in contrast to other iterative methods like the *Jacobi method* or *Gauss-Seidel iteration* the Kaczmarz method needs no requirements on the underlying matrix like being diagonally dominant or symmetric and positive definite[4].

Theorem 5.2.5. *If the system $Ax = b$ has a solution, then the cyclic Kaczmarz algorithm converges to a solution x^* and if the starting vector $x^{(0)} \in A^T\mathbb{R}^M$, then the solution is of minimal norm.*

Proof: Let x be a solution of $Ax - b$, then, using the abbreviation $j = j(n)$, we have

[4] And usually the system matrices of tomography have neither of these properties.

for $n \in \mathbb{N}$,

$$\begin{aligned}
\left|x^{(n)} - x\right|_2^2 &= \left|x^{(n-1)} + \frac{1}{\left|a_j\right|_2^2} a_j \left(b_j - a_j^T x^{(n-1)}\right) - x\right|_2^2 \\
&= \left|x^{(n-1)} - x\right|_2^2 + 2\frac{\left(x^{(n-1)} - x\right)^T a_j}{\left|a_j\right|_2^2}\left(b_j - a_j^T x^{(n-1)}\right) + \frac{\left(b_j - a_j^T x^{(n-1)}\right)^2}{\left|a_j\right|_2^4}|a_j|_2^2 \\
&= \left|x^{(n-1)} - x\right|_2^2 - 2\frac{\left(b_j - a_j^T x^{(n-1)}\right)^2}{\left|a_j\right|_2^2} + \frac{\left(b_j - a_j^T x^{(n-1)}\right)^2}{\left|a_j\right|_2^2} \\
&= \left|x^{(n-1)} - x\right|_2^2 - \frac{1}{\left|a_j\right|_2^2}\frac{\left|a_j\left(b_j - a_j^T x^{(n-1)}\right)\right|_2^2}{|a_j|_2^2} = \left|x^{(n-1)} - x\right|_2^2 - \frac{\left|x^{(n)} - x^{(n-1)}\right|_2^2}{|a_j|_2^2}.
\end{aligned}$$

Therefore,

$$\left|x^{(n)} - x\right|_2^2 = \left|x^{(n-1)} - x\right|_2^2 - \frac{\left|x^{(n)} - x^{(n-1)}\right|_2^2}{|a_{j(n)}|_2^2} \tag{5.2.6}$$

yields that the sequence $|x^{(n)} - x|_2$ is monotonically decreasing, strictly if $a_j^T x^{(n-1)} \neq b_j$, and therefore ist must converge for *any* $x^{(0)}$ since it is bounded from below by 0. This implies that

$$\lim_{n\to\infty} \frac{\left|x^{(n)} - x^{(n-1)}\right|_2^2}{|a_{j(n)}|_2^4} = 0 = \lim_{n\to\infty} \left|x^{(n)} - x^{(n-1)}\right|_2, \tag{5.2.7}$$

and since $|x^{(n)}|_2$ is bounded, there exists a subsequence $k(n)$ such that $x^{k(n)} \to x^*$ for some $x^* \in \mathbb{R}^N$.

Since

$$\left|x^{k(n)+\ell} - x^*\right|_2 \leq \underbrace{\sum_{j=0}^{\ell} \left|x^{k(n)+j+1} - x^{k(n)+j}\right|_2}_{\to 0} + \underbrace{\left|x^{k(n)} - x^*\right|_2}_{\to 0} \to 0,$$

it also follows that $x^{k(n)+\ell} \to x^*$ for any $\ell \in \mathbb{N}$. Now we make use of the cyclic version of the algorithm and consider the sequence $k(n)$ modulo N; since there are only finitely many possibilities, the sequence $k(n)$ must contain infinitely many values of the form $k(n) = N\,k'(n) + m$ for at least one $m \in \mathbb{Z}_N$. For all of them, we have that

$$b_m = a_m^T x^{N\,k'(n)+m} \to a_m^T x^*,$$

hence $(Ax^*)_m = b_m$. Passing to the sequence $k(n)+\ell$, we get $(Ax^*)_{m+\ell} = b_{m+\ell}$ where the summation in the index takes place in $\mathbb{Z}_N$. Hence $Ax^* = b$.

Now let x' be the limit of another convergent subsequence and consider the sequence

$$\left|x^{(n)} - x^*\right|_2 - \left|x^{(n)} - x'\right|_2 \tag{5.2.8}$$

which converges to a unique limit due to (5.2.6) since both sequences converge individually. But for the subsequence converging to x^* the limit of (5.2.8) is $-|x^* - x'|_2$ while for the one converging to x' it is $|x' - x^*|_2 = |x^* - x'|_2$, which gives $0 = |x^* - x'|_2$, and the limit is independent of the subsequence and satsfies $Ax^* = b$. Since the sequence is bounded and all convergent subsequences have the same limit, it finally follows that

$$\lim_{n\to\infty} x^{(n)} = x^*, \qquad Ax^* = b. \tag{5.2.9}$$

It remains to prove the minimality of the norm. To that end we note that $x^{(0)} \in A^T\mathbb{R}^M$ implies that $x^{(n)} \in A^T\mathbb{R}^M$, $n \in \mathbb{N}$, since the update step (5.2.3) only adds multiples of rows of A. Hence, $x^* = A^T y^*$ for some $y^* \in \mathbb{R}^M$ and for *any* solution x of $Ax = b$ we have that

$$Ax = b = Ax^* = AA^T y^* \qquad \Rightarrow \qquad 0 = A\left(x - A^T y^*\right)$$

so that y^* is the *least squares* solution of $A^T y = x$ or the projection of x on $A^T\mathbb{R}^M$ and hence of smaller or the same norm. □

Remark 5.2.6. Since trivially $0 \in A^T\mathbb{R}^M$ for *any* matrix $A \in \mathbb{R}^{M\times N}$, the Kaczmarz algorithm always converges to a minimal norm solution if $x^{(0)} = 0$.

Remark 5.2.7. From the proof we can extract two important ingredients that make the Kaczmarz algorithm work.

1. The equation $j(n)$ has to be chosen in such a way that the update step really modifies $x^{(n)}$ which requires that

   $$b_{j(n)} - a_{j(n)}^T x^{(n-1)} \neq 0.$$

   If we could find no such $j(n)$, then $Ax^{(n-1)} = b$ and the problem has been completely solved already.

2. The squared improvement towards a solution x is

   $$\gamma_j := \frac{\left(b_j - a_j^T x^{(n-1)}\right)^2}{|a_j|_2^2} = \left(\frac{b_j}{|a_j|_2} - \frac{a_j^T x^{(n-1)}}{|a_j|_2}\right)^2. \tag{5.2.10}$$

   which is the gain of the scaled system whose rows are unit vectors. In order to find best directions, these numbers have to be checked. It is an advantage in terms of performance to precompute or at least cache these norms or to compute the normalized versions.

3. It is impossible that $j(n+1) = j(n)$, so the projection changes is every step.

4. To ensure that the solution satisfied at the equations, one has to guarantee that all $j = 1, \ldots, M$ appear infinitely many times in the sequence $j(n)$. Even if the sequence $k(n)$ may only have one accumulation point, the shifted sequences $k(n) + \ell$, $\ell \in \mathbb{Z}_M$, will take care of the other equations.

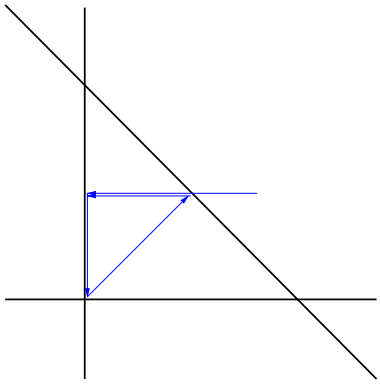

Figure 5.2.2: An overdetermined system of three equations in 2 variables that does not have a solution. Then it can happen that the sequence of projections cycles between the lines and thus never converges.

5. The existence of a solution is fundamental for the convergence of the algorithm. This is not only a technical issue of the proof. Fig. 5.2.2 shows a simple example where the algorithm will not converge in the case of an overdetermined system without a solution.

Example 5.2.8. The easiest way to mess up the algorithm would be to take $j(n) = j$ for some fixed j. We then have that $x^{(n)} = x^{(1)}$ which obviously converges, but can only guarantee that $(Ax)_1 = b_1$.

Remark 5.2.9 (Computational effort of the Kaczmarz algorithm). Let us have a brief look at the number of computations occuring in the Kaczmarz algorithm:

1. Any evaluation of the stepwidth

$$\delta_j := \frac{b_j - a_j^T x^{(n-1)}}{|a_j|_2^2}$$

takes $2N$ operations for the inner product $a_j^T x^{(n-1)}$ and maybe the same for computing[5] $|a_j|_2^2$. Together with one subtraction and the division the complete effort is $4N + 3$ or $2N + 3$ depending on whether $|a_j|_2^2$ is precomputed.

2. Precomputaion of $|a_j|_2^2$ can be avoided by scaling the system which could actually be done "on the fly" when accessing a row for the first time.

3. The update step $x^{(n)} = x^{(n-1)} + \delta_j a_j$ also consist of $2N$ operations.

4. In summary, a full cycle of the Kaczmarz algorithm costs $4MN$ or $6MN$ operations[6].

[5] So precomputing or caching could accelerate the algorithm by a factor of 2.

[6] Keep in mind, however, that we do not consider the effort of generating the row a_j or handling the storage.

The fact that a full cycle of the Kaczmarz algorithm can be really costly is well-known and it appears to be necessary since otherwise we cannot guarantee that we really get convergence in all components. Therefore, there are different selection strategies for $j(n)$, the simplest being a random choice; in fact, this strategy even works surprisingly well, but cannot ensure convergence. Another strategy is more successful but not for free.

Definition 5.2.10. The **pivot** $j^* \in \{1, \dots, M\}$

$$\gamma_{j^*} = \max_{j=1,\dots,M} \gamma_j = \max_{j=1,\dots,M} \frac{\left(b_j - a_j^T x^{(n-1)}\right)^2}{\left|a_j\right|_2^2} \tag{5.2.11}$$

indexes the equation such that the (quadratic) **gain**

$$\left|x^{(n-1)} - x\right|_2^2 - \left|x^{(n)} - x\right|_2^2$$

is maximized. The **pivoted Kaczmarz** algorithm uses $j(n) = j^*$.

The pivoted Kaczmarz algorithm searches M values γ_j, so one step costs $2MN$ or $4MN$ operations which is just $\frac{1}{2}$ or $\frac{2}{3}$ of the effort for the full cycle. However, the search operations can be performed in parallel which could speed up the process significantly.

Theorem 5.2.11. *If the system $Ax = b$ has a solution, then the pivoted Kaczmarz algorithm converges to a solution x^*.*

Proof: Instead of $Ax = b$ we consider the *scaled system*

$$A'x =: \begin{pmatrix} \frac{1}{|a_1|_2^2} & & \\ & \ddots & \\ & & \frac{1}{|a_M|_2^2} \end{pmatrix} Ax = \begin{pmatrix} \frac{1}{|a_1|_2^2} & & \\ & \ddots & \\ & & \frac{1}{|a_M|_2^2} \end{pmatrix} b =: b',$$

and note that because of (5.2.10), pivoting chooses the same sequence $j(n)$ for $A'x = b'$. By pivoting, the sequence $|x^{(n)} - x|_2^2$ decays in a strictly monotonic way, hence

$$0 = \lim_{n\to\infty} \left|b'_{j^*(n)} - a_{j^*(n)}^T x^{(n-1)}\right| = \lim_{n\to\infty} \max_{j=1,\dots,M} \left|(A'x^{(n-1)} - b')_j\right| = \lim_{n\to\infty} |A'x^{(n)} - b|_\infty,$$

hence at least a subsequence of $x^{(n)}$ converges to a solution. That all subsequences converge to the same limit follows by the same argument based on (5.2.8) as in the proof of Theorem 5.2.5. And clearly $Ax = b$ if and only if $A'x = b'$. □

Since parallelization capabilities are usually limited in practice, one could also incorporate *batch pivoting* by decomposing $\{1, \dots, M\} = J_1 \cup \dots \cup J_n$ and cycle over the n pivots

$$\gamma_{j_k^*} = \max_{j \in J_k} \gamma_j, \qquad k = 1, \dots, n.$$

Exercise 5.2.1 Prove that batch pivoting converges. ◊

5.3 SOR Methods for the Kaczmarz Algorithm

The Kaczmarz algorithm can be related to relaxation methods which extends to an algorithm with a defined behavior even if the system $Ax = b$ has *no* solution. In presenting this method, we follow the exposition from (Natterer, 1986). Here we decompose the system matrix A into blocks of the form

$$A = \big(A_j : j = 1, \dots, m\big) = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}, \qquad A_j \in \mathbb{R}^{M_j \times N}, \qquad \sum_{j=1}^{m} M_j = M. \tag{5.3.1}$$

In the preceding section we considered the special case $M_j = 1$ and $m = M$. Moreover, we assume that $A_j \mathbb{R}^N = \mathbb{R}^{M_j}$ which means that $M_j \leq N$ and that the matrices A_j all are of full rank. This can also be expressed as the linear map $A : \mathbb{R}^N \to \mathbb{R}^{M_j}$ being **surjective** or **onto**. The linear system to be solved now takes the form

$$A_j x = b_j, \qquad j = 1, \dots, m. \tag{5.3.2}$$

Let $P_j : \mathbb{R}^N \to \mathbb{R}^N$ denote the orthogonal projections onto the affine subspace $A_j x = b_j$, i.e.,

$$A_j P_j x = b_j, \qquad P_j x = x + A_j^T y, \quad y \in \mathbb{R}^{M_j}. \tag{5.3.3}$$

The second condition is due to orthogonality[7]: $x - P_j x \in (\ker A)^\perp = \operatorname{span} A^T$. This implies that

$$b_j = A_j \left(x + A_j^T y\right) = A_j x + \underbrace{A_j A_j^T}_{\in \mathbb{R}^{M_j \times M_j}} y \qquad \Leftrightarrow \qquad y = \left(A_j A_j^T\right)^{-1} (b_j - A_j x);$$

our assumption that A_j has full rank yields that $A_j A_j^T$ is a symmetric and strictly positive matrix which has an inverse. This leads to the modified update step

$$x \mapsto P_j x = x + A_j^T \left(A_j A_j^T\right)^{-1} (b_j - A_j x), \tag{5.3.4}$$

which generalizes (5.2.3).

Exercise 5.3.1 Show that (5.3.4) and (5.2.3) coincide for $M_j = 1$. ◊

Remark 5.3.1. The formula (5.3.4) is *not* a way to really compute the update as the computation of matrix inverses is inefficient and numerically unstable. The same holds true for the explicit computation of $A_j A_j^T$ which leads to a quadratically conditioned system of equations. What one is really doing is first compute the *residual* $r = b_j - A_j x$, then solve the system $A_j A_j^T y = r$ based on a QR factorization of A_j only and finally compute $A_j^T y$. Note that, like the norms $|a_j|_2$ in the previous section, the factorizations of the A_j can be precomputed to speed up the computations.

[7] And the classical result $\mathbb{R}^n = \ker A \oplus \operatorname{span} A^T$ whenever $A \in \mathbb{R}^{m \times n}$ has full rank.

Now, we use a slightly different approach and also restrict ourselves to the cyclic version of the algorithm by combining m successive steps into a single one.

Definition 5.3.2. For a given **relaxation parameter** $\omega \in (0,2)$, we define the **relaxation** of the projection as

$$P_{j,\omega} := (1-\omega)I + \omega P_j. \tag{5.3.5}$$

If $\omega < 1$ one speaks of **underrelaxation**, for $\omega > 1$ of **overrelaxation**. Moreover, we define

$$P_\omega = P_{m,\omega} \cdots P_{1,\omega}. \tag{5.3.6}$$

Relaxation is in some way compensating (for $\omega < 1$) or overdoing (for $\omega > 1$) the improvement provided by the projection. For $\omega = 1$ it reduces to the standard algorithm. Moreover, the definition (5.3.6) simply combines all m relaxed projections into a single projection step, so we consider a cyclic version of the Kaczmarz algorithm. The convergence indeed looks familiar to us.

Theorem 5.3.3. *If the problem $Ax = b$ has a solution and $x^{(0)} \in A^T\mathbb{R}$, then, for any $0 < \omega < 2$, the sequence*

$$x^{(n+1)} = P_\omega x^{(n)} \tag{5.3.7}$$

of the cyclic relaxed Kaczmarz algorithm converges to a solution x^* of minimal norm, i.e., $Ax^* = b$ and $Ax = b$ implies $|x|_2 \geq |x^*|_2$.

Assume that $x^* \in \mathbb{R}^N$ is a solution of $Ax = b$, hence of $A_j x^* = b_j$, then the iteration (5.3.4) takes the form

$$P_j x = x + A_j^T \left(A_j A_j^T\right)^{-1} A_j (x^* - x) =: x + Q_j(x^* - x), \tag{5.3.8}$$

where[8]

$$Q_j := A_j^T \left(A_j A_j^T\right)^{-1} A_j, \qquad j = 1, \ldots, m,$$

now defines a *linear* projection with

$$Q^2 = Q_j^T Q_j = A_j^T (A_j A_j^T)^{-T} A_j A_j^T (A_j A_j^T)^{-1} A_j = Q_j.$$

We will continued to work with the linear projections Q_j and define, in the same way as in (5.3.5) and (5.3.6) the matrices $Q_{j,\omega}$ and Q_ω.

Definition 5.3.4. The **operator norm** of a matrix $A \in \mathbb{R}^{p\times q}$ is defined as

$$\|A\| := \|A\|_2 := \max_{x\neq 0} \frac{|Ax|_2}{|x|_2} = \max_{|x|_2=1} |Ax|_2. \tag{5.3.9}$$

[8] Do not make the mistake to compute

$$A_j^T \left(A_j A_j^T\right)^{-1} A_j = A_j^T A_j^{-T} A_j^{-1} A_j = I,$$

which is only admissible if A_j is square and nonsingular, as this never happens in the situations we consider.

Since[9]

$$\|Q\| = \|Q^2\| \le \|Q\|^2,$$

it follows that[10]

$$\|Q_j\| \le 1, \qquad j = 1, \dots, m, \tag{5.3.10}$$

Since

$$\begin{aligned} |Q_{j,\omega}x|_2^2 &= x^T Q_{j,\omega}^T Q_{j,\omega} x = x^T \left((1-\omega)I + \omega Q_j\right)^2 x \\ &= (1-\omega)^2 |x|_2^2 + 2\omega(1-\omega) \underbrace{x^T Q_j x}_{=x^T Q_j^2 x} + \omega^2 \underbrace{x^T Q_j^2 x}_{=|Q_j x|_2^2} \\ &= |x|_2^2 - \left(2\omega - \omega^2\right) + \left(2\omega(1-\omega) + \omega^2\right) |Q_j x|_2^2 = |x|_2^2 + \omega(2-\omega)\left(|Q_j x|_2^2 - |x|_2^2\right), \end{aligned}$$

i.e.,

$$|Q_{j,\omega}x|_2^2 - |x|_2^2 = \omega(2-\omega)\left(|Q_j x|_2^2 - |x|_2^2\right). \tag{5.3.11}$$

Since $\|Q_j\| \le 1$, for $|x|_2 \le 1$, hence $|Q_j x| \le 1$, (5.3.11) yields $|Q_{j,\omega}x|_2^2 - |x|_2^2 \le 0$, and therefore

$$\left\|Q_{j,\omega}\right\| \le 1, \qquad 0 \le \omega \le 2, \tag{5.3.12}$$

which also explains the restrictions on ω. The contractivity of the linear projections can now be used for the convergence proof of Theorem 5.3.3.

This, however, needs some preliminary observations. The first says that for sequences where Q_ω is not contractive, the limit of the iteration is the identity.

Lemma 5.3.5. *If $0 < \omega < 2$ and $x^{(n)}$ is a sequence such that*

$$|x^{(n)}|_2 \le 1, \qquad \lim_{n\to\infty} \left|Q_\omega x^{(n)}\right|_2 = 1, \tag{5.3.13}$$

then

$$\lim_{n\to\infty} \left(I - Q_\omega\right) x^{(n)} = 0. \tag{5.3.14}$$

Proof: We use induction on the number m of factors of Q_ω. For[11] $m = 1$, we have by (5.3.11) that

$$\begin{aligned} \left|\left(I - Q_\omega\right) x^{(n)}\right|_2^2 &= \left|\left(I - (1-\omega)I - \omega Q_1\right) x^{(n)}\right|_2^2 \\ &= \omega^2 \left|(I - Q_1)x^{(n)}\right|_2^2 = \omega^2 \left(\left|x^{(n)}\right|_2^2 - \left|Q_1 x^{(n)}\right|_2^2\right) = \frac{\omega}{2-\omega}\left(\left|x^{(n)}\right|_2^2 - \left|Q_{1,\omega}x^{(n)}\right|_2^2\right) \\ &= \frac{\omega}{2-\omega}\left(\left|x^{(n)}\right|_2^2 - \left|Q_\omega x^{(n)}\right|_2^2\right), \end{aligned}$$

hence, if $\left|Q_\omega x^{(n)}\right|_2^2 \le \left|x^{(n)}\right|_2^2 \le 1$ coverges to 1 this implies (5.3.14).

[9] The operator norm is submultiplicative, i.e., $\|AB\| \le \|A\|\|B\|$; try to verify this by yourself.
[10] Projections are weakly contractive.
[11] This is the SOR algorithm for a single matrix without any cylcles.

For the induction step $m-1 \to m$, we write

$$Q_\omega = Q_{m,\omega} Q_{m-1,\omega} \cdots Q_{1,\omega} =: Q_{m,\omega} Q'_\omega$$

and note that

$$I - Q_\omega = I - Q'_\omega + \left(Q'_\omega - Q_\omega\right) = \left(I - Q'_\omega\right) - \left(I - Q_{m,\omega}\right) Q'_\omega. \tag{5.3.15}$$

Since

$$\left|Q_\omega x^{(n)}\right|_2 = \left|Q_{m,\omega} Q'_\omega x^{(n)}\right|_2 \leq \underbrace{\|Q_{m,\omega}\|}_{\leq 1} \left|Q'_\omega x^{(n-1)}\right|_2 \leq \left|x^{(n-1)}\right|_2, \tag{5.3.16}$$

the assumptions (5.3.13) ensure that $\left|Q'_\omega x^{(n-1)}\right|_2 \to 1$ and thus $(I - Q'_\omega)x^{(n)} \to 0$ by the induction hypothesis. Moreover, setting $y^{(n)} = Q'_\omega x^{(n)}$, (5.3.16) yields that

$$1 \leftarrow \left|Q_\omega x^{(n)}\right|_2 \leq \left|Q_{m,\omega} y^{(n)}\right|_2 \leq \left|y^{(n)}\right|_2 \leq \left|x^{(n-1)}\right|_2,$$

and therefore, by the case $m = 1$, $\left(I - Q_{m,\omega}\right) y^{(n)} \to 0$. Substituing this into (5.3.15) advances the induction hypothesis and proves the claim. □

Definition 5.3.6. The **kernel** $\ker A$ of a matrix A is defined as the subspace

$$\ker A := \{x \in \mathbb{R}^n : Ax = 0\}. \tag{5.3.17}$$

Lemma 5.3.7. *For $0 < \omega < 2$, the sequence Q_ω^k, $k \in \mathbb{N}_0$, converges to the orthogonal projection matrix T onto* $\ker(I - Q_\omega)$.

Proof: Being a projection means that $(I - T)(I - Q_\omega) = (I - Q_\omega)$ and mapping to the kernel means $(I - Q_\omega)T = 0$. The first of these identities is equivalent to $T(I - Q_\omega) = 0$, hence the two identities are in turn equivalent to

$$TQ_\omega = T = Q_\omega T. \tag{5.3.18}$$

Hence

$$T = TQ_\omega^k = Q_\omega T Q_\omega^{k-1} = \cdots = Q_\omega^k T, \qquad k \in \mathbb{N}_0,$$

and since for any $x \in \mathbb{R}^n$ the sequence

$$\left|Q_\omega^k x\right|_2 = \underbrace{\|Q_\omega\|}_{\leq 1} \left|Q_\omega^{k-1} x\right|_2 \leq \left|Q_\omega^{k-1} x\right|_2$$

is decreasing and bounded from below by 0, it must converge to some limit

$$q = \lim_{k\to\infty} \left|Q_\omega^k x\right|_2.$$

If $q = 0$ then

$$Tx = \lim_{k\to\infty} TQ_\omega^k x = 0 = \lim_{k\to\infty} Q_\omega^k x$$

as claimed. If $q > 0$, on the other hand, we consider the bounded sequence

$$g_k := \frac{Q_\omega^k x}{\left|Q_\omega^k x\right|_2}, \qquad k \in \mathbb{N},$$

with $|g_k|_2 = 1$ and

$$|Q_\omega g_k|_2 = \frac{\left|Q_\omega^{k+1} x\right|_2}{\left|Q_\omega^k x\right|_2} \to \frac{q}{q} = 1, \qquad k \to \infty.$$

Hence, by the preceding Lemma 5.3.5,

$$0 = \lim_{k\to\infty} (I - Q_\omega) g_k = \lim_{k\to\infty} (I - Q_\omega) Q_\omega^k x = \lim_{k\to\infty} Q_\omega^k (I - Q_\omega) x.$$

Hence $Q_\omega^k y \to 0$ for any y of the form $(I - Q_\omega)x$ while for $x \in \ker(I - Q_\omega)$, i.e., $Q_\omega x = x$, we trivially have that $Q_\omega^k = I$. □

Lemma 5.3.8. *For* $0 < \omega < 2$,

$$\ker (I - Q_\omega) = \bigcap_{j=1}^{m} \left(I - Q_j\right). \tag{5.3.19}$$

Proof: If $Q_{j,\omega} x = x$, $j = 1, \ldots, m$, then clearly $Q_\omega x = x$, and the inclusion $\supseteq$ is settled. Suppose now that $Q_\omega x = x$, then

$$|x|_2 = |Q_\omega x|_2 = \left(\prod_{j=2}^{m} \|Q_{j,\omega}\| \right) |Q_{1,\omega} x|_2 \le |Q_{1,\omega} x|_2 \le |x|_2,$$

hence $|Q_{1,\omega} x|_2^2 - |x|_2^2 = 0$ from which (5.3.11) allows us to conclude that $|Q_1 x|_2^2 - |x|_2^2 = 0$, i.e., $x \in \ker(I - Q_1)$ or $x = Q_1 x = Q_{1,\omega} x$. But then, as above,

$$|x|_2 = |Q_\omega x|_2 = \left|Q_{m,\omega} \cdots Q_{2,\omega} Q_{1,\omega} x\right|_2 = \left|Q_{m,\omega} \cdots Q_{2,\omega} x\right|_2,$$

and thus $x \in \ker(I - Q_2)$. Iterating the argument, we thus get $x \in \ker(I - Q_j)$, $j = 1, \ldots, m$, which is the inclusion $\subseteq$. □

Now, we can complete the convergence proof for the Kaczmarz algorithm.

Proof of Theorem 5.3.3: Let x^* be an arbitrary solution of $Ax = b$, then

$$P_j x - x^* = Q_j (x - x^*), \qquad P_{j,\omega} x - x^* = Q_{j,\omega} (x - x^*), \qquad P_\omega x - x^* = Q_\omega (x - x^*).$$

Hence,

$$P_\omega^k x = x^* + Q_\omega^k (x - x^*), \tag{5.3.20}$$

and, by Lemma 5.3.7,

$$x^{(n)} = P_\omega^n x^{(0)} = x^* + Q_\omega^n \left(x^{(0)} - x^*\right) \to x^* + T\left(x^{(0)} - x^*\right) = (I - T)x^* + Tx^{(0)},$$

where T projects onto

$$\ker A = \bigcap_{j=1}^{m} \ker A_j,$$

and if $x^{(0)} = A^T y$ for some y, then $Tx^{(0)} = 0$ and the algorithm converges to the minimal norm projection $(I - T)x^*$. □

Remark 5.3.9. The convergence rate of the Kaczmarz algorithm is *geometric*, or, as one would say in Numerical Analysis, *linear*. From (5.3.20) we know that

$$\left|x^{(n)} - x^*\right|_2 = \left|Q_\omega^n \left(x^{(0)} - x^*\right)\right|_2$$

From Lemma 5.3.5 we know that the only possible eigenvalue λ of Q_ω with $|\lambda| = 1$ is $\lambda = 1$ and that all other eigenvalues λ satisfy $|\lambda| < 1$. Let u be the respective normalized eigenvector, $Q_\omega u = u$, $|u|_2 = 1$, and write

$$x^{(0)} - x^* = \alpha u + v, \qquad v \in u^\perp,$$

then $|Q_\omega v|_2 \leq \lambda' |v|_2$, where $\lambda' < 1$ is the second largest eigenvalue of Q_ω, and therefore

$$\left|x^{(n)} - x^*\right|_2 = \left|Q_\omega^n (\alpha u + v)\right|_2 = \left|\alpha u + Q_\omega^n v\right|_2 \geq \alpha \underbrace{|u|_2}_{=1} - \left|Q_\omega^n v\right|_2,$$

hence

$$\alpha \leq \left|x^{(n)} - x^*\right|_2 + (\lambda')^n |v|_2 \to 0,$$

i.e., $\alpha = 0$ and

$$\left|x^{(n)} - x^*\right|_2 \leq (\lambda')^n |v|_2,$$

which is linear convergence.

Remark 5.3.10. Although the convergence takes place independently of ω, over-relaxation usually accelerated the convergence, often by chosing $\omega \sim 1.5$.

5.4 More on SOR

Now, we will consider aspects of the SOR Kaczmarz that are of a more theoretical nature. To that end, we assume that the algorithm was started with a proper initialization $x^{(0)} = A^T y^{(0)}$, which leads to $A^T y^{(n)} := x^{(n)} \in A^T \mathbb{R}^M$. With the block representation

$$y = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \quad u_k \in \mathbb{R}^{M_k}, \qquad \text{hence} \qquad x = A^T y = \begin{pmatrix} A_1^T & \dots & A_m^T \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = \sum_{k=1}^{n} A_k^T u_k$$

we can rewrite the recurrence (5.3.4) for the first m steps in two ways as

$$\begin{aligned} x^{(j)} &= x^{(0)} + \sum_{k=1}^{j} A_k^T u_k \\ &= P_{j,\omega} x^{(j-1)} = (1-\omega) x^{(j-1)} + \omega \left(x^{(j-1)} + A_j^T (A_j A_j^T)^{-1} \left(b - A_j x^{(j-1)} \right) \right) \\ &= \underbrace{x^{(0)} + \sum_{k=1}^{j-1} A_k u_k}_{= x^{(j-1)}} + \omega\, A_j^T (A_j A_j^T)^{-1} \left(b - x^{(0)} \right) - \omega\, A_j^T (A_j A_j^T)^{-1} A_j \sum_{k=1}^{j-1} A_k u_k, \end{aligned}$$

i.e.,

$$A_j^T u_j = \omega\, A_j^T (A_j A_j^T)^{-1} \left(b - A_j x^{(0)} - A_j \sum_{k=1}^{j-1} A_k^T u_k \right).$$

Multplying both sides from the left with A_j we thus get

$$A_j A_j^T u_j = \omega \underbrace{A_j A_j^T (A_j A_j^T)^{-1}}_{=I} \left(b - A_j x^{(0)} - A_j \sum_{k=1}^{j-1} A_k^T u_k \right),$$

hence

$$u_j = \omega (A_j A_j^T)^{-1} \left(b - A_j x^{(0)} - A_j \sum_{k=1}^{j-1} A_k^T u_k \right) \tag{5.4.1}$$

or

$$\underbrace{\begin{pmatrix} A_1 A_1^T & & \\ & \ddots & \\ & & A_m A_m^T \end{pmatrix}}_{=:D} y = \omega \left(b - A x^{(0)} \right) - \omega \underbrace{\begin{pmatrix} 0 & & & \\ A_2 A_1^T & 0 & & \\ \vdots & \ddots & \ddots & \\ A_m A_1^T & \dots & A_m A_{m-1}^T & 0 \end{pmatrix}}_{=:L} y. \tag{5.4.2}$$

Remark 5.4.1. (5.4.1) and (5.4.2) are the relaxation of the (iterative) **Jacobi method** for the solution of the *symmetric* linear system

$$By = d, \qquad B := AA^T = \begin{pmatrix} A_1 A_1^T & \dots & A_1 A_m^T \\ \vdots & \ddots & \vdots \\ A_m A_1^T & \dots & A_m A_m^T \end{pmatrix}, \quad d := b - A x^{(0)}.$$

The Jacobi method consist of decomposing $B = L + D + L^T$ where L is a lower triangular matrix and D a nonsingular block diagonal one and then iterating

$$y^{(n)} = D^{-1} \left(b - L y^{(n-1)} \right),$$

cf. (Golub and van Loan, 1996; Sauer, 2013). In fact, (5.4.2) is the block version of the algorithm where the iteration is again multiplied by the relaxation parameter $\omega \in (0, 2)$, where $\omega = 1$ is once more the standard case.

Solving in (5.4.2) for u, we find that

$$y = \omega \, (D + \omega L)^{-1} \left(b - A x^{(0)} \right)$$

which is well defined since $D+L$ is a block upper triangular matrix with nonsingular diagonal blocks. This allows us to rewrite the cyclic Kaczmarz algorithm as

$$\begin{aligned} x^{(n)} &= P_\omega x^{(n-1)} = x^{(n-1)} + A^T y^{(n)} = x^{(n-1)} + \omega A^T (D + \omega L)^{-1} \left(b - A x^{(n-1)} \right) \\ &= \left(I - \omega A^T (D + \omega L)^{-1} A \right) x^{(n-1)} + \omega A^T (D + \omega L)^{-1} b \qquad (5.4.3) \\ &=: B_\omega x^{(n-1)} + b_\omega. \qquad (5.4.4) \end{aligned}$$

The goal is now to relate the iteration (5.4.4) to the **successive overrelaxation** or **SOR** method for a *symmetric* system of the form

$$b = AA^T y = (L + D + L^T)y. \tag{5.4.5}$$

The modified *Gauss-Seidel iteration* then reads as[12]

$$y^{(n)} = (1-\omega)y^{(n-1)} + \omega D^{-1}\left(b - Ly^{(n)} - L^T y^{(n-1)}\right), \tag{5.4.6}$$

hence

$$\begin{aligned}(D+\omega L)y^{(n)} &= \omega b + Dy^{(n-1)} - \omega \underbrace{(D+L^T)}_{=AA^T-L} y^{(n-1)} \\ &= \omega\left(b - AA^T y^{(n-1)}\right) + (D+\omega L)y^{(n-1)},\end{aligned}$$

and therefore

$$y^{(n)} = \left(I - \omega(D+\omega L)^{-1}AA^T\right)y^{(n-1)} + \omega(D+\omega L)^{-1}b =: C_\omega y^{(n-1)} + c_\omega, \tag{5.4.7}$$

which already looks very much like (5.4.3). In fact, $A^T C_\omega = B_\omega A^T$ and $A^T c_\omega = b_\omega$. Hence, if we set $x^{(n-1)} = A^T y^{(n-1)}$ in (5.4.4), then

$$x^{(n)} = B_\omega A^T y^{(n-1)} + b_\omega = A^T\left(C_\omega y^{(n-1)} + c_\omega\right) = A^T y^{(n)},$$

hence the iteration (5.4.7) is the restriction of (5.4.4) on $A^T\mathbb{R}^M$.

Remark 5.4.2. It is important to note that in all this derivation we *never* inverted A or AA^T, we only used D^{-1} which is allowed since this is the block diagonal matrix with diagonal elements $A_j A_j^T$ and *those* were assumed to be nonsingular.

There are strong relations between the two iterations.

Lemma 5.4.3 (Properties of the iteration).

1. *The iteration* (5.4.3) *leaves* $A^T\mathbb{R}^M$ *invariant.*
2. *The eigenspace of* B_ω *with respect to the eigenvalue* 1 *is* $\ker A$.
3. *All eigenvalues* λ *with* $\lambda \neq 1$ *of* B_ω *and* C_ω *coincide.*

Proof: 1) follows since

$$B_\omega A^T y + b_\omega = A^T\left(C_\omega y + c_\omega\right) \in A^T\mathbb{R}^M,$$

for 2) we note that

$$x = B_\omega x = x - \omega A^T(D+\omega L)^{-1}Ax$$

[12] This can be perfomed efficiently in such a way that any $x_j^{(n)}$ depends only on $x_1^{(n)}, \ldots, x_{j-1}^{(n)}$, i.e., in a reasonably iterative way.

if and only if

$$A^T(D+\omega L)^{-1}Ax = 0 \tag{5.4.8}$$

which holds for $Ax = 0$. Conversely, (5.4.8) implies that

$$(D+\omega L)^{-1}Ax = z, \qquad z \in \ker A^T, \text{ i.e., } A^T z = 0,$$

hence,

$$z^T(D+\omega L)z = z^T Ax = \underbrace{(A^T z)^T}_{=0} x = 0 = \underbrace{z^T(D+\omega L)^T z}_{=\left(z^T(D+\omega L)z\right)^T} = z^T\left(D+\omega L^T\right)z,$$

and therefore

$$0 = z^T\left(2D+\omega L+\omega L^T\right)z = z^T\left((2-\omega)D+\omega AA^T\right)z;$$

since the matrix inside is symmetric and strictly positive definite, this implies $z = 0$ and also $Ax = (D+\omega L)z = 0$. Hence, (5.4.8) holds if and only if $Ax = 0$ which can be rephrased as

$$\ker A^T \cap (D+\omega L)^{-1} A\,\mathbb{R}^N = \{0\}. \tag{5.4.9}$$

For 3), we assume that $\lambda \neq 1$ is an eigenvalue of B_ω, i.e., there exists $x \neq 0$ such that

$$\lambda x = B_\omega x = \left(I - \omega A^T(D+\omega L)^{-1}A\right)x, \quad \text{i.e.,} \quad (1-\lambda)x = \omega A^T(D+\omega L)^{-1}Ax,$$

which implies that $x \in A^T\mathbb{R}^M \perp \ker A$. We write $x = A^T y$, and $x = 0$ together with (5.4.9) yields that $y \in (D+\omega L)^{-1} A\,\mathbb{R}^N$, hence

$$x = A^T\,(D+\omega L)^{-1}\,Az$$

for some $z \neq 0$. Consequently,

$$\begin{aligned} 0 &= B_\omega x - \lambda x = (B_\omega - \lambda I)\,A^T\,(D+\omega L)^{-1}\,Az = \left(B_\omega A^T - \lambda A^T\right)(D+\omega L)^{-1}\,Az \\ &= A^T\,(C_\omega - \lambda I)\,(D+\omega L)^{-1}\,Az. \end{aligned}$$

Since $C_\omega\,(D+\omega L)^{-1}\,A\mathbb{R}^N \subseteq (D+\omega L)^{-1}\,A\mathbb{R}^N$, again (5.4.9) yields that

$$(C_\omega - \lambda I)\,(D+\omega L)^{-1}\,Az, \tag{5.4.10}$$

and λ is an eigenvalue of C_ω with eigenvector $(D+\omega L)^{-1}\,Az$. Conversely, let λ be an eigenvalue of C_ω with eigenvector y, then

$$(B_\omega - \lambda I)\,A^T y = B_\omega A^T y - \lambda A^T y = A^T C_\omega y - \lambda A^T y = A^T\,(C_\omega y - \lambda y) = 0$$

shows that λ is also an eigenvalue of B_ω. □

The lesson from the lemma is the fact that the Kaczmarz iteration (5.4.4) produces a sequence of elements of $A^T\mathbb{R}^M$ and this subspace contains *no* eigenvectors for the eigenvalue 1, those are in the orthogonal complement $\ker A$ - this follows form the key observation (5.4.9). If we now start the Kaczmarz iteration with $x^{(0)} = A^T y^{(0)}$, then we produce a sequence of elements of this subspace and now properties of *the restriction* of the matrix B_ω to this subspace become relevant.

Definition 5.4.4. The **spectral radius**[13] $\rho(A)$ of a square matrix $A \in \mathbb{R}^{n\times n}$ is defined as

$$\rho(A) = \max\{|\lambda| : \det(A - \lambda I) = 0,\ \lambda \in \mathbb{C}\}, \tag{5.4.11}$$

that is, as the largest modulus of eigenvalues.

The proof of the following lemma is a slight variation of the fact that the Gauss-Seidel method converges for symmetric positive definite matrices, cf. (Golub and van Loan, 1996; Sauer, 2013).

Lemma 5.4.5. *The restriction B'_ω of B_ω on $A^T\mathbb{R}^M$ satisfies $\rho(B'_\omega) < 1$.*

Proof: By Lemma 5.4.3, 3), we can also consider the eigenvalues $\neq 1$ of C_ω, i.e.,

$$\begin{aligned}(1-\lambda)y &= \omega(D+\omega L)^{-1}AA^Ty = (D+\omega L)^{-1}\left(\omega D + \omega L + \omega L^T\right)y\\ &= (D+\omega L)^{-1}\left(D + \omega L + (\omega-1)D + \omega L^T\right)y\\ &= y + (D+\omega L)^{-1}\left((\omega-1)D + \omega L^T\right)y,\end{aligned}$$

or

$$\lambda(D+\omega L)y = \left((1-\omega)D - \omega L^T\right)y. \tag{5.4.12}$$

We normalize y such that[14] $|y|_D^2 := y^HDy = 1$ and multiply (5.4.12) from the left by y^H yielding

$$\lambda(1 + \omega\underbrace{y^HLy}_{=:a}) = (1-\omega) - \omega\underbrace{y^HL^Ty}_{=\overline{a}},$$

that is, $\lambda(1+\omega a) = 1 - \omega - \omega\overline{a}$. Since

$$0 \le y^TAA^Ty = y^T(D+L+L^T)y = 1 + a + \overline{a} = 1 + 2\alpha, \qquad a = \alpha + i\beta,$$

the real part α of a satisfies $\alpha \ge -\frac{1}{2}$. Hence,

$$\begin{aligned}|\lambda|^2 &= \frac{((1-\omega)-\omega\overline{a})\,((1-\omega)-\omega a)}{(1+\omega a)(1+\omega\overline{a})}\\ &= \frac{(1-\omega)^2 - 2\omega(1-\omega)\alpha + \omega^2(\alpha^2+\beta^2)}{1 + 2\omega\alpha + \omega^2\alpha^2 + \beta^2} = \frac{((1-\omega)-\omega\alpha)^2 + \omega^2\beta^2}{(1+\omega\alpha)^2 + \omega^2\beta^2},\end{aligned} \tag{5.4.13}$$

and since

$$((1-\omega)-\omega\alpha)^2 - (1+\omega\alpha)^2 = \omega(\omega-2) - 2\omega(2-\omega)\alpha = -\underbrace{\omega(2-\omega)}_{>0}\underbrace{(2\alpha+1)}_{\ge 0},$$

the numerator in (5.4.13) is smaller than the denominator and we have that $|\lambda| < 1$ whenever $\alpha > -\frac{1}{2}$, which occurs iff $y \in A^T\mathbb{R}^M$ by means of Lemma 5.4.3. □

[13] Even if the spectral radius is defined by complex eigenvalues, it can be computed in real terms as $\rho(A) = \lim \|A^n\|^{1/n}$ with an arbitrary matrix norm.

[14] The eigenvalue as well as the eigenvector can be complex, even if all our matrices were real so far.

Remark 5.4.6. (5.4.13) is the core argument in favor of SOR. The goal is to reduce the value $|\lambda|$ for all eigenvalues of modulus < 1 by an appropriate ω. Obviously, the relationship is far from trivial, but is is easy to see that the value of $|\lambda(\omega)|^2$ of (5.4.13) is 1 if $\omega = 0, 2$ while for $\omega = 1$ it is $\frac{\alpha^2+\beta^2}{(1+\alpha)^2+\beta^2} < 1$, as long as $\alpha > -\frac{1}{2}$. Therefore, the function must assume its minimum for some ω, but to determine this value, we must know the eigenstructure of C_ω which is impossible to determine in practice. Nevertheless, there are results and many heuristics, usually $\omega \sim 1.5$ seems to be a good choice.

Lemma 5.4.7 (**Banach Fixpoint Theorem**, adapted). *If $\rho(B'_\omega) < 1$ and $x^{(0)} \in A^T\mathbb{R}^M$, then the iteration* (5.4.4) *converges to a unique **fixpoint** x^* with*

$$(I - B_\omega)x^* = b_\omega. \tag{5.4.14}$$

Proof: By Lemma 5.4.3, we know that $x^{(0)} \in A^T\mathbb{R}^M$ implies that[15] $x^{(n)} \in A^T\mathbb{R}^M$, hence $x^{(n)} = B'_\omega x^{(n-1)} + b_\omega$ for any $n \in \mathbb{N}$. For convergence, we note that, with $\rho := \rho(B'_\omega) = \|B'_\omega\|$, that

$$\begin{aligned}\left|x^{(n+1)} - x^{(n)}\right|_2 &= \left|B_\omega\left(x^{(n)} - x^{(n-1)}\right)\right|_2 = \left|B'_\omega\left(x^{(n)} - x^{(n-1)}\right)\right|_2 \\ &\leq \|B'_\omega\|\left|x^{(n)} - x^{(n-1)}\right|_2 = \rho\left|x^{(n)} - x^{(n-1)}\right|_2 \leq \cdots \leq \rho^n\left|x^{(1)} - x^{(0)}\right|_2,\end{aligned}$$

hence

$$\left|x^{(n+k)} - x^{(n)}\right|_2 \leq \sum_{j=0}^{k-1}\left|x^{(n+j+1)} - x^{(n+j)}\right|_2 \leq \sum_{j=0}^{k-1}\rho^{n+j}\left|x^{(1)} - x^{(0)}\right|_2 \leq \rho^n\frac{\left|x^{(1)} - x^{(0)}\right|_2}{1-\rho},$$

which shows that $x^{(n)}$ is a Cauchy sequence that converges to a limit $x^* \in A^T\mathbb{R}^M$ which satsfies $x^* = B_\omega x^* + b_\omega$. Since for any two fixpoints $x^*, x^\dagger \in A^T\mathbb{R}^M$,

$$|x^* - x^\dagger|_2 = \left|B_\omega x^* + b_\omega - B_\omega x^\dagger - b_\omega\right| = \left|B'_\omega\left(x^* - x^\dagger\right)\right|_2 \leq \rho|x^* - x^\dagger|_2,$$

it follows that $|x^* - x^\dagger|_2 = 0$ and thus uniqueness. □

Theorem 5.4.8. *For any $0 < \omega < 2$ and any $x^{(0)} \in A^T\mathbb{R}^M$, especially for $x^{(0)} = 0$, the SOR Kaczmarz iteration* (5.4.4) *converges to a unique limit x^* such that*

$$A^T(D + \omega L)^{-1}(Ax^* - b) = 0 \tag{5.4.15}$$

that solves If the system $Ax = b$ has a solution, then x^* is the solution of minimal norm, if the problem has no solution, then $x^* = x^\dagger + O(\omega)$ where

$$x^\dagger = \operatorname*{argmin}_{x \in A^T\mathbb{R}^M} |Ax - b|^2_{D^{-1}} := (Ax - b)^T D^{-1}(Ax - b)$$

[15] This is a condition in the Banach fix point theorem which is easily overlooked: The iteration has to map a certain set or space to *itself* and be a contraction there.

Proof: Convergence and (5.4.15) follow directly from Lemma 5.4.7. If $b = Ax'$ for some x', and $x^* = A^T y^*$, then (5.4.9) implies that $Ax' = AA^T y^*$ and this is the least square solution from $A^T\mathbb{R}^m$. For the last statement, we set $\omega = 0$ in (5.4.15) and consider the scaling

$$A^T D^{-1}(Ax - b) = (D^{-1/2}A)^T \left((D^{-1/2}A)x - D^{-1/2}b\right).$$

The rest is standard least-squares theory. □

Remark 5.4.9. The convergence issue is more tricky that in appears. In particular, the recurrence (5.4.7) for the $y^{(n)}$ only converges if $c_\omega \in (D + \omega L)^{-1}A\mathbb{R}^n$ which is equivalent to $AA^T y = b$ having a solution. Otherwise, it could happen that $c_\omega = c'_\omega + w$, $w \neq 0$, $A^T w = 0$, and this w would be added to $y^{(n)}$ in any iteration, giving $y^{(n)} = z^{(n)} + nw$, where $z^{(n)}$ is a convergent series while the $y^{(n)}$ obviously diverge. This nuisance however, is not affecting

$$x^{(n)} = A^T y^{(n)} = A^T \left(z^{(n)} + nw\right) = A^T z^{(n)}.$$

This is what eventually makes the $x^{(n)}$ converge *unconditionally* provided the initial value is chosen properly.

5.5 Inverse Problems and Regularization

References 6

A Literatur

Abramowitz, M. and Stegun, I. A., editors (1972). *Handbook of mathematical functions.* Dover. 10th printing.

Benson, D. J. (2007). *Music. A Mathematical Offering.* Cambridge University Press.

Courant, R. and Hilbert, D. (1931). *Methoden der mathematischen Physik I.* Berlin.

Fisher, S. D. (1990). *Complex Variables.* Wadsworth & Brooks. Dover Reprint 1999.

Forster, O. (1984). *Analysis 3. Integralrechung im $\mathbb{R}^n$ mit Anwendungen.* Vieweg, 3. edition.

Gasquet, C. and Witomski, P. (1998). *Fourier Analysis and Applications. Filtering, Numerical Computation, Wavelets*, volume 30 of *Texts in Applied Mathematics.* Springer.

Gautschi, W. (1997). *Numerical Analysis. An Introduction.* Birkhäuser.

Gelbaum, B. R. and Olmstedt, J. M. H. (1964). *Counterexamples in Analysis.* Holden–Day. Dover reprint 2003.

Golub, G. and van Loan, C. F. (1996). *Matrix Computations.* The Johns Hopkins University Press, 3rd edition.

Grüningen, D. C. v. (1993). *Digitale Signalverarbeitung.* VDE Verlag, AT Verlag.

Hamming, R. W. (1989). *Digital Filters.* Prentice–Hall. Republished by Dover Publications, 1998.

Helmholtz, H. (1885). *On the Sensations of Tone.* Longmans & Co. Translated by A. J. Ellis, Dover reprint 1954.

Heuser, H. (1984). *Lehrbuch der Analysis. Teil 1.* B. G. Teubner, 3. edition.

Hille, E. (1982). *Analytic Function Theory.* Chelsea Publishing Company, 2nd edition.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press.

Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis.* Cambridge University Press.

Isaacson, E. and Keller, H. B. (1966). *Analysis of Numerical Methods.* John Wiley & Sons.

Jackson, D. (1941). *Fourier Series and Orthogonal Polynomials.* Number 6 in The Carus Mathematical Monographs. Mathematical Association of America. Dover reprint 2004.

Kaczmarz, S. (1937). Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Intern. Acad. Polonaise Sci. Lett., Cl. Sci. Math. Nat. A*, 35:355–357.

Kak, A. C. and Slaney, M. (2001). *Principles of Computerized Tomographic Imaging.* SIAM.

Kammeyer, K. D. and Kroschel, K. (1998). *Digitale Signalverarbeitung.* Teubner Studienbücher Elektrotechnik. B. G. Teubner, Stuttgart.

Katznelson, Y. (1976). *An Introduction to Harmonic Analysis.* Dover Books on advanced Mathematics. Dover Publications, 2. edition.

Kotelnikov (1933). On the carrying capacity of the "ether" and wire in telecommunications. In *First All Union Conference of Communications I, zd. Red. Upr. Svyazi RKKA, Moscov.* In Russian.

Kreyszig, E. (1978). *Introductionary Functional Analysis with Applications.* John Wiley & Sons.

Loomis, L. H. (1953). *Introduction to Abstract Harmonic Analysis.* Van Nostrand. Dover reprint, 2011.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing.* Academic Press, 2. edition.

Mallat, S. (2009). *A Wavelet Tour of Signal Processing: The Sparse Way.* Academic Press, 3rd edition.

Marcus, M. and Minc, H. (1969). *A Survey of Matrix Theory and Matrix Inequalities.* Prindle, Weber & Schmidt. Paperback reprint, Dover Publications, 1992.

Natterer, F. (1986). *The Mathematics of Computerized Tomography.* John Wiley & Sons.

Natterer, F. and Wübbeling, F. (2001). *Mathematical Methods in Image Reconstruction.* SIAM.

Olafsson, G. and Quinto, E. T., editors (2006). *The Radon Transform, Inverse Problems, and Tomography*, volume 65 of *Proceedings of Symposia in Applied Mathematics.* AMS.

Riesz, F. and Sz.-Nagy, B. (1955). *Functional Analysis.* Frederick Ungar, New York. Dover reprint 1990.

Sauer, T. (2013). Einführung in die Numerische Mathematik. Vorlesungsskript, Universität Passau.

Sauer, T. (2014). Analysis 1. Vorlesungsskript, Universität Passau.

Sauer, T. (2017). Constructive Approximation. Lecture notes, University of Passau.

Schüßler, H. W. (1992). *Digitale Signalverarbeitung.* Springer, 3. edition.

Shannon, C. E. (1949). Communications in the presence of noise. *Proc. of the IRE*, 37:10–21.

Steger, A. (2001). *Diskrete Strukturen 1. Kombinatorik – Graphentheorie – Algebra.* Springer.

Szegö, G. (1939). *Orthogonal Polynomials.* AMS, 5th edition.

Taylor, A. E. and Lay, D. C. (1980). *Introduction to Functional Analysis.* John Wiley & Sons, 2nd edition.

Tuy, H. K. (1983). An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.*, 43:546–552.

Van Loan, C. F. (2000). The ubiquitous Kronecker product. *J. Comput. Appl. Math.*, 123:85–100.

Van Loan, C. F. (2009). The Kronecker product. A product of times. `http://www.siam.org/meetings/la09/talks/vanloan.pdf`.

Whittaker, J. (1935). *Interpolatory function theory*, volume 33 of *Cambridge Tracts in Math. and Math. Physics.*

Williamson, J. H. (1962). *Lebesgue Integration.* Holt, Rinehart and Winston. Dover reprint 2014.

Yosida, K. (1965). *Functional Analysis.* Grundlehren der mathematischen Wissenschaften. Springer–Verlag.

Index 6