UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

Computerized Tomography and the Radon Transform Through the Lens of Medical Imaging

Autor: Diego Viedma Gordillo

Director: Dr. F. Xavier Massaneda Clares
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 14 de gener de 2025

Abstract

This work explores the mathematical foundations and practical applications of Computerized Tomography (CT) within the context of medical imaging. By combining analytical concepts, we examine the process of generating cross-sectional images from X-ray data. The focus is placed on the properties of the Radon transform, including its relation to Fourier transforms, uniqueness theorems, and inversion formulas. Reconstruction algorithms, such as the filtered backprojection and the gridding method, are analyzed and computationally implemented, with performance evaluated using the Shepp-Logan phantom, a benchmark model for clinical image reconstruction. Additionally, we explore modern alternative geometries designed for enhanced efficiency. Beyond medical imaging, the broader implications of CT are discussed, illustrating how mathematical concepts drive transformative technological advancements.

2020 Mathematics Subject Classification. 44A12, 92C55, 65T50

Acknowledgments

I would like to extend my sincere and profound gratitude to my advisor, Dr. F. Xavier Massaneda Clares, for his continuous support throughout this project. During the weekly meetings we held over the past months, he has helped me shape, refine and deepen my understanding of this work, which would not have been possible without his dedication, patience and expertise.

Additionally, I would like to thank my parents and friends, specially Carla, for showing interest in my work from the very first day and for supporting me throughout these years. As doctors, my parents brought valuable insights into the medical aspect of this work, and it has been a wonderful opportunity to engage with them on a deeper level about their field.

Contents

1 Introduction

Computerized tomography (CT), also known as computed axial tomography (CAT), is a revolutionary imaging technique that combines advanced mathematical algorithms and modern computational power to produce highly detailed cross-sectional images of an object from its projections. Initially developed for medical purposes, CT has transformed the field of diagnostics by enabling non-invasive exploration of internal body structures. However, as we shall see, its applications extend far beyond medicine, finding use in areas such as geophysics, radar, astronomy, archaeology and industrial testing.

The term "tomography", derived from the Greek words $\tau o \mu o \sigma$ (*tomos*, meaning slice) and $\gamma \rho \alpha \phi \epsilon \iota \nu$ (*graphein*, meaning to record), denotes a method of imaging that reconstructs cross-sectional images of an object, particularly its interior, from its projections. Unlike traditional X-ray imaging, which creates a shadow-like image through the superposition of structures along the beam's path, CT scans reconstruct a numerical description of tissue density from multiple beams within a thin slice of the body. This allows for detailed examination of internal structures with remarkable clarity, all without the need for invasive procedures on the patient.

The process involves directing hundreds, or even thousands, of X-ray beams through a plane of interest at multiple angles. Each beam measures the attenuation of X-rays as they pass through the medium, with variations in intensity revealing the medium's absorption properties. These measurements, taken from a grid of directions, are processed using algorithms (such as those detailed in Section 5.2 and 5.3) to reconstruct the X-ray attenuation density across the slice. By analyzing the differences in absorption across different directions, CT enables the identification of regions with varying densities, which may correspond to distinct tissues, structures, or pathological features. In medical applications, this level of detail is critical. For instance, in head imaging, tissue densities typically range from 1 (water) to 1.05, with bones reaching values near 2. Even subtle variations in density, as small as 0.005, can indicate features of medical significance.

A simple physical model for CT can be described as follows: Let $f(x)$ denote the X-ray attenuation coefficient of the medium at a point x, quantifying the absorption of X-rays per unit distance. If a small segment of an X-ray beam traverses a distance dx, the relative loss in intensity is proportional to the attenuation coefficient at that point:

$$\frac{dI}{dx} = -f(x)I.$$

Noting that $\frac{d}{dx}(\ln I) = \frac{1}{I}\frac{dI}{dx}$ (here, instead, physicists and engineers often infamously treat the derivative as a fraction, a "funny" preparatory step for integration) and integrating, we get:

$$I = I_0 \exp\left(-\int_L f(x)\,dx\right). \tag{1.1}$$

where I_0 is the initial intensity of the beam, and I is its final intensity after passing through the object. This equation forms the foundation of CT: the scanning process measures the line integrals of $f(x)$, which represent the cumulative attenuation of X-rays along each line L and serve as the input data for the reconstruction process.

The mathematical framework behind this reconstruction is the *Radon transform*, named after the Austrian mathematician Johann Radon, who studied it extensively. The Radon

transform maps a function f defined on $\mathbb{R}^2$ to the set of its line integrals (hyperplanes, in the general n-dimensional case). The Radon transform $\mathcal{R}f$ is given by:

$$\mathcal{R}f(L) = \int_L f(x)\,dx = \ln \frac{I_0}{I}.$$

where in the last equality we have used (1.1). In 1917, Radon derived an explicit inversion formula for reconstructing f from $\mathcal{R}f$. While this formula provides a theoretical foundation for CT, it is not directly suitable for numerical computations.

In practice, the integrals can be measured only for a finite number of lines. Their arrangement, referred to as the *scanning geometry*, is determined by the design of the scanner. In Section 5.1, we discuss the *parallel beam geometry*, which was used in the first commercial scanner developed by G. Hounsfield and is the geometry we will focus on in this work. However, other geometries have been developed to improve scanning efficiency, as we shall explore in Section 5.5.

The mapping of attenuation data to density reconstruction is further complicated by several approximations. We must note that the mathematical models used in CT are idealizations of the complex relationships between the studied object and the measured data. For instance, in reality, X-ray beams are not monochromatic; they have an energy spectrum, and the function f depends not only on the spatial variable x but also on the energy E of the X-rays. Assuming $T(E)$ represents the energy spectrum of the X-ray source, equation (1.1) must be modified as:

$$I = I_0 \int T(E) \exp\left(-\int_L f(x, E)\,dx\right) dE.$$

Neglecting this polychromatic nature of the X-rays introduces artifacts, as lower-energy radiation is preferentially absorbed while the beam passes through the material, leaving higher-energy X-rays to dominate. This results in a "harder" X-ray beam, meaning the average energy of the beam increases as it traverses the object. Other factors contributing to discrepancies in numerical computations include the finite diameter of the X-ray source, inaccuracies in the detectors, numerical approximation errors, and other sources of uncertainty.

1.1 History of CT

We could simply attribute the beginnings of computed tomography (CT) to the groundbreaking work of the 1979 Nobel Prize in Medicine winners, Allan M. Cormack and Godfrey N. Hounsfield, placing their achievements on a timeline: from Cormack's theoretical developments in the late 1950s to Hounsfield's creation of the first practical device in the late 1960s. However, let us first explore a necessary discovery that laid its foundation: X-rays.

The discovery of X-rays was accidentally made on November 8, 1895, by the German physicist Wilhelm C. Röntgen, an event that marked the beginning of his extraordinary career. At the time, Röntgen was a professor of physics at the University of Würzburg. While working in his blacked-out laboratory, he was investigating the luminescent glow produced during electric discharges in an evacuated glass tube. By chance, Röntgen observed an unexpected phenomenon: a nearby screen coated with barium platinocyanide crystals began to glow. This fluorescent screen, routinely used to detect ultraviolet radiation, happened to be within range of what was then an unknown radiation emanating from the tube. To Röntgen's surprise, the glow persisted even when the tube was tightly

covered with cardboard, eliminating the possibility that visible or ultraviolet light was causing the effect.

Over the next few days, Röntgen conducted systematic experiments. By placing various objects between the tube and the screen, he discovered that the mysterious radiation could penetrate certain materials. During one of these experiments, he saw the shadowy outline of his own hand on the screen, revealing the bones within.

On December 28, 1895, he submitted a detailed report to the Würzburg Physical Medical Society. Attached to the report was one of the first X-ray images ever produced:

Figure 1: *Hand mit Ringen* (Hand with Rings): One of the first X-ray images, showing the left hand of Bertha Röntgen, Wilhelm's wife, extracted from https://wellcomecollection.org/works/wjc8ejn2.

Röntgen's discovery quickly captured the world's attention. By January 1896, news of X-rays had spread globally, and an "X-ray mania" swept through society. This newfound technology revolutionized medicine, allowing for non-invasive imaging of the human body. Over the next decades, researchers and inventors refined the technology, improving the quality and resolution of two-dimensional X-ray images.

Key contributors to the development of X-ray technology included Thomas Edison, who introduced significant improvements to X-ray tube designs, and Hermann von Helmholtz, who investigated the properties of X-rays and their interaction with materials. Yet, even as X-ray imaging became widespread, its limitations in capturing only two-dimensional projections of three-dimensional structures were apparent.

Attempts to address these limitations began almost immediately, advancing the journey toward modern CT. In 1896, E. Thompson proposed creating three-dimensional X-ray images using stereoscopic techniques, which involved taking two slightly displaced X-ray images and viewing them through a stereoscope to achieve depth perception. Later, in 1916, Karol Mayer created stratigraphic images by moving the X-ray tube while keeping

the film stationary, an approach resembling modern CT scanning. A similar method was independently patented by Carlo Baese in 1915, who described using simultaneous movements of the tube and film cassette to locate foreign objects within the body.

In 1917, the Austrian mathematician Johann Radon established the mathematical foundation for reconstructing functions from their projections, an idea that was originally conceived in 1826 by the young Norwegian Niels H. Abel in the context of theoretical physics. It is at this point that the path leading to the discovery and exploitation of X-rays met that leading to the development of computational techniques, which enabled the building of the computer and CT scans.

As we mentioned above, the people generally credited for the invention of computed tomography are Allan M. Cormack and Godfrey N. Hounsfield, both awarded the Nobel Prize for Medicine in 1979. Interestingly, they met for the first time during the Nobel Prize ceremony.

Figure 2: Co-creators of computed tomography posing with their respective image reconstruction apparatus: Allan M. Cormack (left) and Godfrey N. Hounsfield (right), extracted from [2].

Born in South Africa, Cormack was a physician who worked on the measurement of the X-ray absorption of various body tissues at the University of Cape Town. He later moved to Harvard University and, in 1956, began to work on the problem of image reconstruction of X-ray projections. He solved the problem theoretically and then, in 1963, experimentally validated his research using cutlets of pork and an apparatus that he built himself (shown in Figure 2).

Independently, the English engineer Hounsfield began his research on tomography in 1967. He approached the problem using the computers available at that time to perform the complicated calculations needed. In this way, the concept of computed tomography found its practical expression. His first laboratory scans were conducted on a human brain prepared in formalin and took as long as nine days. A reconstructed image of a preserved brain obtained using the first scanner used by Hounsfield is shown in the figure below.

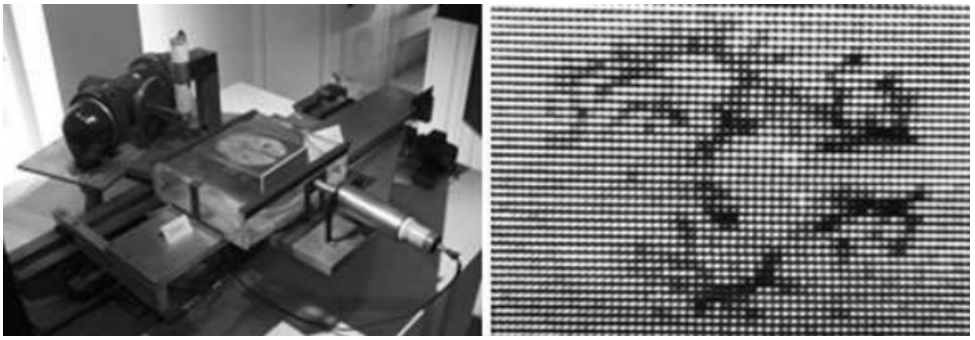

Figure 3: The laboratory scanner used by Hounsfield (left) and an image of a preserved brain, obtained using it in 1968 (right), extracted from [2].

Further experimental work was needed using living tissues and refining the design of the scanner. In 1971, with the participation of neurologists James Ambrose and Louis Kreel, the improved scanner EMI Mark I was installed at a hospital in Wimbledon (London). The first CT scan of a patient was carried out on a woman with a suspected brain tumor, which was clearly identifiable in the reconstructed image (see the darker ellipse area in Figure 4). The resolution of the image was 80×80 pixels, where each pixel represented an area 3×3 mm.

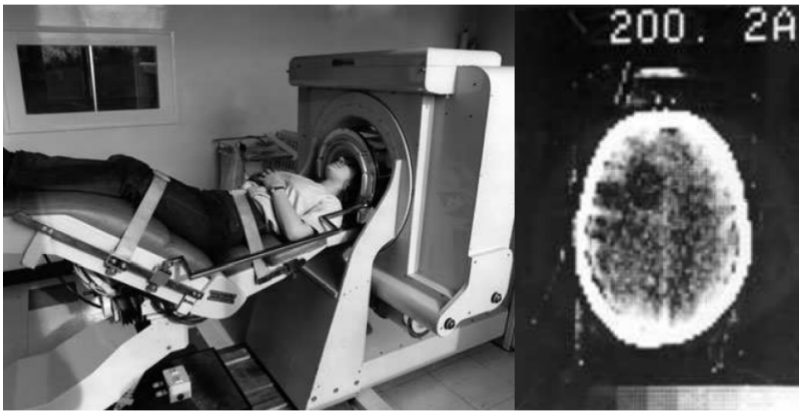

Figure 4: The EMI Mark I scanner (left) and the first CT scan of a living brain (right). Note the elliptical and darker area that corresponds to a tumor. Extracted from [2].

After the results were formally presented in 1972, the development of CT gathered momentum as numerous neurologists, radiologists, physicists, engineers and data processing specialists all started working on methods of obtaining and interpreting tomographic images.

By the end of 1973, the first commercial CT scanner was on the market: the EMI CT 1000, a development of the Mark I with a number of detectors increased to 30, allowing a resolution of 320×320 pixels. That year, six units were sold, each for the not inconsiderable sum of approximately £100,000. In the course of the following two years, driven by an avalanche of models from different competitors, the market for CT scanners reached a value of about £40,000,000.

Over the decades, scanners have evolved significantly, incorporating various hardware and

software advancements, as well as using different scanning geometries (see Section 5.5). The progress made is striking: contemporary CT scanners can scan in a few hundred milliseconds and reconstruct an image of 2048×2048 pixels.

1.2 Other applications

Standard CT occurs whenever the internal structure of an object is examined by exposing it to some kind of radiation which propagates along straight lines, the intensity loss being determined by (1.1). This can be classified as *transmission* CT, as the radiation traverses the body and is detected on the opposite side. There are many applications of this type besides standard CT. We mention only transmission electron microscopy, where an electron beam passes through a planar specimen under several incidence angles [4].

On the other hand, in *emission* CT, radiation is emitted inside the object and is detected emerging outside. Applications of this technique include nuclear reactor testing, detection of illicit nuclear materials, and clinical Single Photon Emission Computed Tomography (SPECT). Let us elaborate further on the latter example, given its significance in medical imaging.

In order to conduct the scan using SPECT, a short-lived radioactive isotope (usually fluorine-18) is injected in the blood or inhaled. As the radioisotope decays, it emits a positron which after travelling up to a few millimeters, encounters and annihilates with an electron, thus producing a pair of gamma photons. These photons are detected by a gamma camera collimated to detect only the photons coming in a given direction. The radiation intensity measured by the detector along the line L is given by

$$I = I_0 \int_L f(x) \exp\left(- \int_{L(x)} \mu(y)\, dy \right) dx. \tag{1.2}$$

where $L(x)$ is the section of L between x and the detector, f is the distribution of the radiopharmaceutical and μ is the attenuation coefficient of the studied tissue. If μ is negligible, then I is essentially the line integral of f and we end up with standard CT. However, in practice μ is not small. That means that we have to reconstruct f from weighted line integrals, the weight function being determined by the attenuation μ. The relevant integral transform is now a generalization of the Radon transform which we call *attenuated* Radon transform.

In SPECT, the goal is to compute f, not μ. Nevertheless, since μ enters the integral equation for f we have to determine it anyway, be it by additional measurements (e.g. a transmission scan) or by mathematical tools. For more information we refer the reader to Section II.6 in [8].

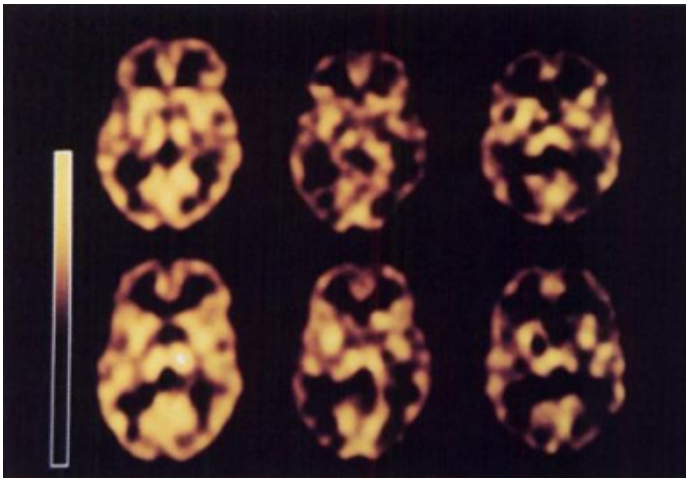

Figure 5: SPECT scans of a normal brain (left), a cocaine dependent brain (middle), and an AIDS dementia complex brain (right), extracted from [5].

A special case worth mentioning is Positron Emission Tomography (PET). In PET, the sources eject the particles pairwise in opposite directions and the radiation in opposite directions is measured in coincidence, i.e. only events with two positrons arriving at opposite detectors at the same time are counted. Then, (1.2) has to be replaced by

$$I = I_0 \int_L f(x) \exp\left(- \int_{L_+(x)} \mu(y)\, dy - \int_{L_-(x)} \mu(y)\, dy \right) dx.$$

where $L_+(x)$ and $L_-(x)$ are the two half-lines of L with endpoint x. Since the exponent adds up to the integral over L independently of x, we obtain

$$I = I_0 \exp\left(- \int_L \mu(y)\, dy \right) \int_L f(x) dx.$$

which does not lead to a new transform.

CT intersects with other fields outside medicine. For instance, both CT and quantum mechanics face the problem of reconstructing functions f from limited projection data, from which different f can be obtained (in Section 3.4, we address the issue of local uniqueness of the Radon transform). Article [6] shows how different functions f that can be obtained from the same limited projection data are close in certain metrics, particularly when convoluted with a Gaussian function (smoothed). This "closeness" improves as the number of data increases. Quantum states can be described by Wigner distributions, which can be used to reconstruct wavefunctions similarly as densities are reconstructed in CT using the Radon transform. The mathematical tools and stability estimates developed in CT can contribute to improving the accuracy and reliability of quantum state reconstructions (for more details, see [7]).

We can also find CT in archaeology. For instance, a CT scan of the mummy of Tutankhamen revealed that he wasn't murdered, as many previously speculated. Instead, the scan indicated that an infected broken leg may have been the cause of his death. In industry, CT techniques are used in nondestructive testing, where internal defects like cracks can be detected without damaging the material. The range of applications is even broader. Let us name just a few additional fields: geology, geophysics, seismology, astronomy, and radar.

2 Mathematical preliminaries

Let us consider the following model for T-periodic signals with finite energy; given $T > 0$, let the Hilbert space

$$L^2[0,T] = \left\{ f : [0,T] \to \mathbb{C} : \|f\|_2 = \int_0^T |f(t)|^2 \, dt < +\infty \right\},$$

with the inner product

$$\langle f, g \rangle = \frac{1}{T} \int_0^T f(t) \, \overline{g(t)} \, dt.$$

The factor $\frac{1}{T}$ is just a normalization which makes the formulas cleaner. To simplify the notation, we take $T = 2\pi$. The general case can be obtained just by rescaling.

Lema 2.1. *The system $\{e_n := e^{int}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $L^2[0,2\pi]$.*

As a result of this lemma, any function $f \in L^2[0,2\pi]$ can be represented in the following way

$$f(t) = \sum_{n\in\mathbb{Z}} \langle f, e_n \rangle e_n(t).$$

Definition 2.2. *Given $n \in \mathbb{Z}$, the n^{th} Fourier coefficient of a function $f \in L^2[0,2\pi]$ is*

$$\hat{f}(n) = \langle f, e_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} \, dt.$$

Observe that, in this language, the previous identity takes the form

$$f(t) = \sum_{n\in\mathbb{Z}} \hat{f}(n) e^{int}.$$

We can extend the definition of the Fourier transform on $\mathbb{R}$, in the space of integrable functions:

$$L^1(\mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{C} : \|f\|_1 = \int_{\mathbb{R}} |f(t)| \, dt < +\infty \right\}.$$

Definition 2.3. *Let $f \in L^1(\mathbb{R})$. The **Fourier transform** of f is:*

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(t) e^{-2\pi i t\xi} \, dt,$$

Notice that the Fourier transform is well-defined and bounded: $|\hat{f}(\xi)| \leq \int_{\mathbb{R}} |f(t)| \, dt < +\infty$.

Remark. We can get rid of the $f \in L^1(\mathbb{R})$ assumption and substitute it for $f \in L^2(\mathbb{R})$. Of course, this requires more than a few technicalities. We shall not discuss them here; we refer the reader to the reference [9].

The Fourier transform can be seen as the limit of the value of the Fourier coefficient of a T-periodic function as T tends to ∞, in the following sense. Assume that $f \in C^1(\mathbb{R})$ (not necessarily periodic). For any T, let f_T denote the T-periodic function that coincides with f on $(-T/2, T/2)$. Then, for $|t| < T/2$,

$$f(t) = f_T(t) = \sum_{n\in\mathbb{Z}} \hat{f}_T(n) e^{i\frac{2\pi}{T}nt}.$$

Hence

$$f(t) = \lim_{T\to\infty} \sum_{n\in\mathbb{Z}} \left(\frac{1}{T} \int_{-T/2}^{T/2} f(s) e^{-i\frac{2\pi}{T}ns}\, ds \right) e^{i\frac{2\pi}{T}nt}.$$

Let us try to identify this limit, at least at a formal level. Let $\xi_n = \frac{n}{T}$, $n \in \mathbb{Z}$, and consider the partition of $\mathbb{R}$ given by these nodes. In these terms, the sum above is

$$\sum_{n\in\mathbb{Z}} \left(\int_{-T/2}^{T/2} f(s) e^{-i2\pi\xi_n s}\, ds \right) e^{i2\pi\xi_n t} (\xi_{n+1} - \xi_n).$$

Letting $T \to \infty$ in the integral, this turns into

$$\sum_{n\in\mathbb{Z}} \hat{f}(\xi_n) e^{2\pi i\xi_n t} (\xi_{n+1} - \xi_n),$$

which is a Riemann sum of the integral

$$\int_{\mathbb{R}} \hat{f}(\xi) e^{2\pi i\xi t}\, d\xi.$$

Thus, formally, the "inverse" formula of the Fourier transform

$$f(t) = \int_{\mathbb{R}} \hat{f}(\xi) e^{2\pi i\xi t}\, d\xi$$

works with the definition of $\hat{f}$ just given.

Let us state without proof the following properties of the Fourier transform.

Theorem 2.4. *Let $f \in L^1(\mathbb{R})$. Then:*

a) $\hat{f}$ is uniformly continuous and $|\hat{f}(\xi)| \le \|f\|_1$.

b) If $f, f' \in L^1(\mathbb{R})$, then

$$\widehat{f'}(\xi) = 2\pi i\xi \hat{f}(\xi) \quad \xi \in \mathbb{R}.$$

c) If $tf(t) \in L^1(\mathbb{R})$ then $\hat{f}$ is differentiable and

$$(\hat{f})'(\xi) = (-2\pi i t f)^{\wedge}(\xi) \quad \xi \in \mathbb{R}.$$

*d) **Translations:** Let $\tau_s f(t) = f(t-s)$. Then*

$$(\tau_s f)(\xi) = \hat{f}(\xi) e^{-2\pi i s\xi}, \quad \xi \in \mathbb{R}.$$

*e) **Riemann-Lebesgue lemma:** $\lim_{|\xi|\to\infty} \hat{f}(\xi) = 0$.*

*f) **Multiplication formula:** if $f, g \in L^1(\mathbb{R})$*

$$\int_{\mathbb{R}} f(t)\hat{g}(t)\, dt = \int_{\mathbb{R}} \hat{f}(t) g(t)\, dt.$$

*g) **Convolution theorem:** if $f, g \in L^1(\mathbb{R})$*

$$\widehat{f * g} = \hat{f} \cdot \hat{g}$$

Remark. Note that Property (b) can be applied iteratively; if $f, f', \ldots, f^{(k)} \in L^1(\mathbb{R})$ we have

$$\widehat{f^{(k)}}(\xi) = (2\pi i\xi)^k \hat{f}(\xi).$$

It is precisely this property, which converts something difficult (derivatives) into something easy (products), what triggered the development of Fourier analysis.

Although in most practical applications we work with functions whose input space is the plane, let us consider the case of an arbitrary dimension n, i.e. $f : \mathbb{R}^n \to \mathbb{C}$. As we mentioned before, the function f might represent the attenuation coefficient in a cross-section of a human brain. Therefore, it is reasonable to assume that f is smooth and vanishes outside the area that the brain occupies. Let's put it formally.

Definition 2.5. *The support of a function $f : \mathbb{R}^n \to \mathbb{C}$, denoted supp$(f)$, is defined as:*

$$supp(f) = \overline{\{x \in \mathbb{R}^n \mid f(x) \neq 0\}},$$

*where $\overline{\{\cdot\}}$ denotes the closure of the set. Let $f \in C^\infty(\mathbb{R}^n)$. We say f is **compactly supported**, $f \in C_c^\infty(\mathbb{R}^n)$, if supp$(f)$ is bounded.*

In other words, supp(f) is the smallest closed set outside of which the function f is identically zero.

Functions with compact support have a particularly good Fourier transform.

Proposition 2.6. *Let $f \in C_c^\infty(\mathbb{R}^n)$, then its Fourier transform extends to an entire function in $\mathbb{C}^n$, and can therefore be expanded as a series:*

$$\hat{f}(\xi) = \sum_{k=0}^{\infty} p_k(\xi),$$

where p_k is a homogeneous polynomial of degree k.

Proof. For simplicity, we will consider the one-dimensional case. Without loosing any generality, we can assume that there exists $T \in \mathbb{R}$ such that $f(x) = 0$, for all $x \notin [-T, T]$. Then:

$$\hat{f}(\xi) = \int_{-T}^{T} f(x) e^{-2\pi i x \cdot \xi}\, dx.$$

Now taking absolute values we get:

$$\begin{aligned} |\hat{f}(\xi)| &\leq \int_{-T}^{T} |f(x)||e^{-2\pi i x \cdot Re(\xi)}||e^{2\pi x \cdot Im(\xi)}|\, dx = \\ &= \int_{-T}^{T} |f(x)| e^{2\pi |x \cdot Im(\xi)|}\, dx \leq e^{2\pi T |Im(\xi)|} \left(\int_{-T}^{T} |f(x)|\, dx \right) := C e^{2\pi |Im(\xi)|} \end{aligned}$$

Now that we have seen that the integral converges for all $\xi \in \mathbb{C}$, we apply Morera's theorem in order to prove that $\hat{f}$ is holomorphic, and therefore can be expanded as a series of homogeneous polynomials.

Since $f \in C_c^\infty(\mathbb{R}^n)$, we've got that $\hat{f}$ is continuous. Now let's see that its integral along any triangle $\mathcal{T}$ vanishes. In order to do that, given that the exponential function is entire, we apply Cauchy's theorem:

$$\int_{\partial\mathcal{T}} \hat{f}(\xi)\, d\xi = \int_{-T}^{T} f(x) \underbrace{\left(\int_{\partial\mathcal{T}} e^{-2\pi i x\cdot\xi}\, d\xi \right)}_{0} dx = 0$$

Note that we've permuted the integrals because $\hat{f}$ converges uniformly on compact sets (such as $\partial\mathcal{T}$), as we have seen. □

As we have just seen, if $f \in C_c^\infty(\mathbb{R}^n)$, then $\hat{f} \notin C_c^\infty(\mathbb{R}^n)$. This is quite inconvenient when trying to use the Fourier formulas for these functions. Luckily, the winner of the Fields Medal in 1950, Laurent Schwartz, found a way to circumvent this obstacle by defining a class of regular functions, bigger than $C_c^\infty(\mathbb{R}^n)$, for which the Fourier transform is stable.

Definition 2.7. *The **Schwartz class**, denoted by S, consists of the functions $f \in C^\infty(\mathbb{R})$ such that for all $m, k \in \mathbb{N}$*

$$P_{m,k}(f) := \sup_{j\le k} \sup_{t\in\mathbb{R}} (1+|t|)^m |f^{(j)}(t)| < +\infty.$$

Notice that $C_c^\infty(\mathbb{R}) \subset S$ and that there are functions which are in S but not in $C_c^\infty(\mathbb{R})$, such as $f(t) = e^{-t^2}$. Functions in S (and all their derivatives) have very fast decay, but not necessarily compact support.

We can extend the definition of the Schwartz class to $\mathbb{R}^n$ using the following multi-index notation. Let $f \in C^\infty(\mathbb{R}^n)$, $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, $|\alpha| := \alpha_1 + \cdots + \alpha_n$ and

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

Then $f \in S(\mathbb{R}^n)$ if for all $m, k \in \mathbb{N}$

$$P_{m,k}(f) := \sup_{|\alpha|\le k} \sup_{t\in\mathbb{R}^n} (1+||x||)^m |D^\alpha f(x)| < +\infty.$$

Let us introduce a notation that will be used in the proof of the following proposition and later on as well. The notation $A \lesssim B$ indicates that there exists a constant $C > 0$ such that $A \le CB$. This implies that A is bounded by a constant multiple of B, though the precise value of C may not be explicitly specified.

Proposition 2.8. *If $f \in S$, then $\hat{f} \in S$.*

Proof. As seen in Theorem 2.4 (c), for $j, m \ge 0$

$$(\hat{f})^{(j)}(\xi) = \left[(-2\pi i t)^j f\right]^\wedge(\xi).$$

Denote $\psi(t) = (-2\pi i t)^j f(t)$, so that $(\hat{f})^{(j)}(\xi) = \hat{\psi}(\xi)$, and recall that, by iterating Theorem 2.4 (b),

$$\widehat{\psi^{(m)}}(\xi) = (2\pi i\xi)^m \hat{\psi}(\xi).$$

Therefore, by Theorem 2.4 (a),

$$(1+|\xi|)^m \left|(\hat{f})^{(j)}(\xi)\right| = (1+|\xi|)^m \left|\hat{\psi}(\xi)\right| \lesssim \left|\widehat{\psi^{(m)}}(\xi)\right| = \left| \left(\frac{\partial^m}{\partial t^m} \left((-2\pi i t)^j f\right) \right)^\wedge(\xi) \right|$$
$$\le \left\| \frac{\partial^m}{\partial t^m} \left((-2\pi i t)^j f\right) \right\|_1.$$

This L^1 norm is finite, because by hypothesis the seminorms $P_{m,k}(f) < +\infty$ for all $m, k \in \mathbb{N}$. □

This allows us to use the Fourier formulas (analogous from those in Theorem 2.4) without any trouble.

We finish this chapter by stating without proof a couple of useful theorems for checking the continuity and differentiability of functions defined by integrals depending on parameters (such as, spoiler alert, the Radon transform).

Theorem 2.9. (Continuity Under the Integral Sign) *Let I and J be two intervals in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and let $f : I \times J \to \mathbb{R}$ such that:*

a) For each $y \in J$, the function from I to $\mathbb{R}$ defined by $x \mapsto f(x, y)$ is measurable.

b) There exists $g \in L^1(I)$ such that for each $y \in J$, it holds that $|f(x, y)| \leq g(x)$ almost everywhere in I.

c) The function $f(x, y)$ is continuous in the variable y almost everywhere in $x \in I$.

Then $f(., y) \in L^1(I)$ for each y and its integral is a continuous function in y, that is,

$$\lim_{y \to t} \int_I f(x, y)\, dx = \int_I \lim_{y \to t} f(x, y)\, dx = \int_I f(x, t)\, dx.$$

Theorem 2.10. (Differentiation Under the Integral Sign) *Let I and J be open intervals in $\mathbb{R}^p$ and $\mathbb{R}$, respectively, and let $f : I \times J \to \mathbb{R}$ such that*

a) For each $y \in J$, the function $f(., y)$ is measurable, and for a fixed $a \in J$, $f(., a) \in L^1(I)$.

b) The partial derivative $\frac{\partial f}{\partial y}(x, y)$ exists for each $(x, y) \in I \times J$.

c) There exists $g \in L^1(I)$ such that

$$\left| \frac{\partial f}{\partial y}(x, y) \right| \leq g(x) \quad \text{for } (x, y) \in I \times J.$$

Then, for each $y \in J$, $f(., y) \in L^1(I)$, the function $F(y) = \int_I f(x, y)\, dx$ is differentiable at all its points, and

$$F'(y) = \int_I \frac{\partial f}{\partial y}(x, y)\, dx.$$

3 Mathematical fundamentals

In this section, we delve into the core theoretical tools of CT and the properties that ensure stability and regularity in their application. We begin with the definition of the Radon transform, including its key properties and its connection with the Fourier transform. Next, we explore the conditions under which functions can be uniquely reconstructed from their projections. Finally, we introduce the backprojection operator, laying the groundwork for understanding the inversion formulas discussed in the subsequent section. The primary reference for this section is [1], with additional details drawn from [8].

3.1 Definition and Preliminaries

Let $f : \mathbb{R}^n \to \mathbb{C},\ n \geq 2$, be a function in the Schwartz space $S(\mathbb{R}^n)$. The Radon transform of such function, $\mathcal{R}f$, is a powerful tool that helps us reconstruct f. $\mathcal{R}f$ is defined on the set of all hyperplanes in $\mathbb{R}^n$ (lines, in our 2-dimensional case). These are identified by two parameters: a signed distance from the origin $s \in \mathbb{R}$ and a normal unit vector $\theta \in \mathbb{S}^{n-1}$, which is perpendicular to the hyperplane. Hence we will write a hyperplane as θ_s.

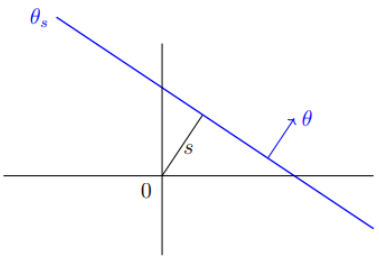

Figure 6: A hyperplane on $\mathbb{R}^2$, a line, determined by a signed distance from the origin s and a normal unit vector θ.

Let's now consider the hyperplane perpendicular to θ passing through the origin, $\theta^\perp$. We note that for every point x in the hyperplane θ_s, there exists a unique point $y \in \theta^\perp$ such that $x = s\theta + y$. Therefore, $\langle x, \theta \rangle = s$, and we can identify θ_s as:

$$\theta_s = \{x \in \mathbb{R}^n : \langle x, \theta \rangle = s\} = s\theta + \theta^\perp$$

Let f be integrable along all hyperplanes. Now, the Radon transform $\mathcal{R}f$ is simply the integral of the function f over a hyperplane:

$$\mathcal{R}f(\theta, s) = \int_{\theta_s} f(x)\, dx = \int_{\theta^\perp} f(s\theta + y)\, dy.$$

One could think that all we have to do afterwards is do some reverse-engineering to figure out what f, the internal structure, looks like. Easy, right? Well, almost. Let's not get ahead of ourselves.

3.2 Basic Properties of the Radon Transform

Before jumping to the inversion formulas for obtaining f, we should make sure that we can always figure out the original function f from its Radon transform $\mathcal{R}f$. But we can,

right? In order to answer this question and, in general, to better understand what we are doing, we will see a few properties of the Radon transform.

At first glance, from its definition, we observe that $\mathcal{R}f$ is an even function defined on $Z := \mathbb{S}^{n-1} \times \mathbb{R}$, i.e. $\mathcal{R}f(-\theta, -s) = \mathcal{R}f(\theta, s)$, for all $(\theta, s) \in Z$.

Another basic property of the Radon Transform is that it commutes with orthogonal transformations, meaning that it is invariant under rotations and reflections:

$$\mathcal{R}f(U\theta, s) = \mathcal{R}(f \circ U)(\theta, s), \quad \forall U \in O(n)$$

where $O(n) = \{g : \mathbb{R}^n \to \mathbb{R}^n,\ \|g(x)\| = \|x\| \text{ for all } x \in \mathbb{R}^n\}$. Note that, equivalently, $O(n)$ is the group of endomorphisms which associated matrix is orthogonal (its inverse coincides with its transpose).

The following results about the regularity of $\mathcal{R}f$ are crucial in medical image reconstruction. To follow the proofs, we need to recall (or learn) how to integrate in polar coordinates on hyperplanes:

Lema 3.1. *If $\theta \in \mathbb{S}^{n-1}$ and f is integrable over $\theta^\perp$, then the following holds:*

$$\int_{\theta^\perp} f(y)\, dy = \int_0^\infty r^{n-2} \left(\int_{\theta^\perp \cap \mathbb{S}^{n-1}} f(r\omega)\, d\omega \right) dr,$$

where r is the radial coordinate and ω is a point on the unit sphere $\mathbb{S}^{n-1}$.

Let us give a first result on the regularity of the Radon transform.

Proposition 3.2. *If $f \in C(\mathbb{R}^n)$ satisfies the growth condition*

$$\sup_{x \in \mathbb{R}^n} |x|^a |f(x)| < \infty,$$

for some $a > n - 1$, then $\mathcal{R}f \in C(Z)$ and

$$\sup_{(\theta,s) \in Z} |s|^{a-(n-1)} |\mathcal{R}f(\theta, s)| < \infty.$$

In particular, if $f \in C(\mathbb{R}^n)$ is rapidly decreasing on $\mathbb{R}^n$, i.e.,

$$\sup_{x \in \mathbb{R}^n} |x|^k |f(x)| < \infty, \quad \forall\, k \in \mathbb{N},$$

then so is $\mathcal{R}f$ on Z, i.e.,

$$\sup_{(\theta,s) \in Z} |s|^k |\mathcal{R}f(\theta, s)| < \infty, \quad \forall\, k \in \mathbb{N}.$$

Proof. The growth condition of f can be changed replacing $|x|^a$ by $(1 + |x|^2)^{a/2}$. Thus, $|f(x)| \lesssim (1 + |x|^2)^{-a/2}$. Integrating in polar coordinates on $\theta^\perp$, i.e. using Lema 3.1, we get:

$$\begin{aligned} |\mathcal{R}f(\theta, s)| &\leq \int_{\theta^\perp} |f(s\theta + y)|\, dy \lesssim \int_{\theta^\perp} \frac{dy}{(1 + s^2 + |y|^2)^{a/2}} \\ &\lesssim \int_0^\infty \frac{r^{n-2} dr}{(1 + s^2 + r^2)^{a/2}} \sim \int_0^s \frac{r^{n-2} dr}{(1 + s^2)^{a/2}} + \int_s^\infty \frac{r^{n-2} dr}{(1 + r^2)^{a/2}}. \end{aligned}$$

Note that in the last estimate we've divided the range of the integral in order to get an integral where $r < s$ and another where $s < r$, and therefore estimate the denominator: $1 + s^2 + r^2 \sim 1 + s^2$ and $1 + s^2 + r^2 \sim 1 + r^2$, respectively. The first integral can be computed directly:

$$\int_0^s \frac{r^{n-2}dr}{(1+s^2)^{a/2}} = \frac{1}{(1+s^2)^{a/2}} \frac{s^{n-1}}{n-1} \sim \frac{1}{s^{a-(n-1)}}.$$

Similarly, for the second integral and taking into account that, by hypothesis $a > n-1$, we get:

$$\int_s^\infty \frac{r^{n-2}dr}{(1+r^2)^{a/2}} \sim \int_s^\infty r^{n-2-a}dr = \frac{1}{a-(n-1)} \frac{1}{s^{a-(n-1)}} \sim \frac{1}{s^{a-(n-1)}}.$$

Hence,

$$\sup_{(\theta,s)\in Z} |s|^{a-(n-1)} |\mathcal{R}f(\theta,s)| < \infty.$$

Now, let us prove that $\mathcal{R}f$ is continuous on Z. By the $O(n)$-invariance of the Radon transform, it suffices to show that $\mathcal{R}f$ is continuous on $N \times \mathbb{R}$, where

$$N = \{\theta \in \mathbb{S}^{n-1} : \theta_n > 0\}$$

is the Northern hemisphere of $\mathbb{S}^{n-1}$. Let us check the continuity of $\mathcal{R}f$ at the North Pole, $\theta^* = (0, \dots, 0, 1)$. Consider the standard parametrization, since $\langle \theta, \theta \rangle = 1$,

$$\theta_n = \left(1 - \sum_{j=1}^{n-1} \theta_j^2\right)^{1/2}.$$

Therefore, we can parametrize a neighborhood of θ^* by the coordinates $(\theta_1, \dots, \theta_{n-1})$. Moreover, for all $y \in \theta^\perp$, we have $\langle \theta, y \rangle = 0$ if and only if $\sum_{j=1}^{n-1} \theta_j y_j + \theta_n y_n = 0$. Then $y_n = -\frac{1}{\theta_n} \sum_{j=1}^{n-1} \theta_j y_j$. Hence $\theta^\perp \cong \mathbb{R}^{n-1}$ and:

$$\mathcal{R}f(\theta, s) = \int_{\mathbb{R}^{n-1}} f(s\theta + y)\, dy,$$

where $dy := dy_1 \cdots dy_{n-1}$ and $y := (y_1, \cdots, y_n)$. Since the growth condition of f implies that its module is regulated by an integrable function:

$$|f(s\theta + y)| \lesssim \frac{1}{(1 + s^2 + y_1^2 + \cdots + y_n^2)^{a/2}} \lesssim \frac{1}{(1 + y_1^2 + \cdots + y_{n-1}^2)^{a/2}}.$$

Then, by the Theorem of Continuity Under the Integral Sign (Theorem 2.9), we conclude that $\mathcal{R}f$ is continuous on $N \times \mathbb{R}$ □

A result along the same lines, involving the derivatives of the Radon transform is the following.

Proposition 3.3. *Assume $f \in C^1(\mathbb{R}^n)$ satisfies the growth conditions*

$$\sup_{x\in\mathbb{R}^n} |x|^a |f(x)| < \infty \quad \textit{and} \quad \sup_{x\in\mathbb{R}^n} |x|^b |D_j f(x)| < \infty \quad (j = 1, \dots, n)$$

for some exponents $a > n-1$ and $b > n$. Consider $\tilde{R}f$, the extension of $\mathcal{R}f$ to $(\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}$ with homogeneity of degree -1, i.e. $\tilde{R}f(\theta, s) := \frac{1}{|\theta|}\mathcal{R}f\left(\frac{\theta}{|\theta|}, \frac{s}{|\theta|}\right)$. Then $\tilde{R}f$ is C^1 on $(\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}$ and, for $\theta \in \mathbb{R}^n \setminus \{0\}$ and $s \in \mathbb{R}$:

$$\frac{\partial \tilde{R}f}{\partial s}(\theta, s) = \frac{1}{|\theta|^2} \sum_{k=1}^{n} \theta_k R(D_k f)\left(\frac{\theta}{|\theta|}, \frac{s}{|\theta|}\right),$$

and:

$$\frac{\partial \tilde{R}f}{\partial \theta_j}(\theta, s) = -\frac{\partial \tilde{R}(x_j f)}{\partial s}(\theta, s)$$

for $\theta \in \mathbb{R}^n \setminus \{0\}$, $s \in \mathbb{R}$, and $j = 1, \ldots, n$.

Proof. In order to simplify the solution, we recall that, by the $O(n)$-invariance of the Radon transform, it is enough to prove it for the Northern hemisphere of $\mathbb{S}^{n-1}$, i.e. for $\theta_n > 0$. Also recall that then:

$$\mathcal{R}f(\theta, s) = \int_{\mathbb{R}^{n-1}} f(s\theta + y)\, dy, \qquad \theta_n = \left(1 - \sum_{k=1}^{n-1} \theta_j^2\right)^{1/2} \quad \text{and} \quad y_n = -\frac{1}{\theta_n} \sum_{k=1}^{n-1} \theta_j y_j.$$

Note that, for $1 \leq j < n$,

$$\frac{\partial \theta_n}{\partial \theta_j} = -\frac{\theta_j}{\theta_n} \quad \text{and} \quad \frac{\partial y_n}{\partial \theta_j} = \frac{\theta_j}{\theta_n^2} y_n - \frac{y_j}{\theta_n}.$$

Then, since $f(s\theta + y) = f(s\theta_1 + y_1, \ldots, s\theta_n + y_n)$:

$$\begin{aligned} \frac{\partial}{\partial \theta_j}(f(s\theta + y)) &= sD_j f(s\theta + y) + D_n f(s\theta + y) \frac{\partial}{\partial \theta_j}(s\theta_n + y_n) \\ &= sD_j f(s\theta + y) + \left(\frac{\theta_j}{\theta_n^2} y_n - \frac{s\theta_j + y_j}{\theta_n}\right) D_n f(s\theta + y). \end{aligned}$$

Moreover, we can quickly see that:

$$\frac{\partial}{\partial s}(f(s\theta + y)) = \sum_{k=1}^{n} \theta_k D_k f(s\theta + y).$$

For the North Pole of $\mathbb{S}^{n-1}$, $\theta^* = (0, \ldots, 0, 1)$, we can substitute its coordinates and get $f(s\theta^* + y) = f(y_1, \ldots, y_{n-1}, s + y_n)$. Then, using the hypothesis on the growth of f and its derivatives, we can see that the partial derivatives of f are bounded by an integrable function. On the one hand,

$$\left|\frac{\partial}{\partial s}(f(s\theta^* + y))\right| \leq \sum_{k=1}^{n} |\theta_k| |D_k f(s\theta^* + y)| \leq \frac{1}{(1 + s^2 + y_1^2 + \cdots + y_{n-1}^2)^{b/2}}$$

and

$$|sD_j f(s\theta^* + y)| \leq \frac{|s|}{(1 + s^2 + y_1^2 + \cdots + y_{n-1}^2)^{b/2}}.$$

On the other hand,

$$|y_n D_n f(s\theta^* + y)| \leq \frac{|y_n|}{(1 + s^2 + y_1^2 + \cdots + y_{n-1}^2)^{b/2}} \leq \frac{\left(\sum_{k=1}^{n-1} |y_k|^2\right)^{1/2}}{(1 + s^2 + y_1^2 + \cdots + y_{n-1}^2)^{b/2}}$$
$$\leq \frac{1 + |s|}{(1 + y_1^2 + \cdots + y_{n-1}^2)^{(b-1)/2}}.$$

Altogether:

$$\left|\frac{\partial}{\partial s}(f(s\theta^* + y))\right| + \sum_{j=1}^{n-1} \left|\frac{\partial}{\partial \theta_j}(f(s\theta^* + y))\right| \lesssim \frac{1 + |s|}{(1 + y_1^2 + \cdots + y_{n-1}^2)^{(b-1)/2}}.$$

By the Theorem of Differentiation Under the Integral Sign (Theorem 2.10), we conclude that $\mathcal{R}f \in C^1(N \times \mathbb{R})$.

Let us obtain the formulas of the derivatives of $\mathcal{R}f$ for the North Pole. Since $f(s\theta^* + y) = f(y_1, \ldots, y_{n-1}, s + y_n)$, we have:

$$\frac{\partial \mathcal{R}f}{\partial s}(\theta^*, s) = \int_{\mathbb{R}^{n-1}} D_n f(s\theta^* + y)\, dy = R(D_n f)(\theta^*, s).$$

Note that:

$$\frac{\partial R(x_j f)}{\partial s}(\theta^*, s) = \sum_{k=1}^{n} \theta_k^* \mathcal{R}(D_n(x_j f))(\theta^*, t) = \mathcal{R}(D_n(x_j f))(\theta^*, t)$$
$$= \mathcal{R}(x_j D_n f)(\theta^*, t) = \int_{\mathbb{R}^{n-1}} y_j D_n f(y, t)\, dy.$$

Therefore:

$$\frac{\partial \mathcal{R}f}{\partial \theta_j}(\theta^*, s) = s \underbrace{\int_{\mathbb{R}^{n-1}} D_j f(s\theta^* + y)\, dy}_{0} - \int_{\mathbb{R}^{n-1}} y_j D_n f(s\theta^* + y)\, dy = -\frac{\partial R(x_j f)}{\partial s}(\theta^*, s),$$

where the first term of the latter derivative vanishes by the fundamental theorem of calculus and the rapidly decreasing nature of f.

The analogous results for $\tilde{\mathcal{R}}f$ can be obtained taking into account its homogeneity. Finally, to extend the formulas to an arbitrary point, we can use once again the $O(n)$-invariance of the Radon transform and operate with arbitrary rotations and reflections. These details can be found in [1]. □

Corollary 3.4. *The Radon transform maps $S(\mathbb{R}^n)$ into $S(Z) := S(\mathbb{R}^{n+1})|_Z$.*

This means that $\mathcal{R}f$ behaves nicely, i.e. that the transformed functions (which live in Z, the space of hyperplanes) are still Schwartz functions. This ensures stability in numerical computations.

3.3 Radon transform and Fourier transform

The Radon transform is deeply connected to another math tool that turns wavy data into something more understandable, the Fourier transform we have discussed in the previous chapter. For the functions defined on Z, we take the Fourier transform and the convolution with respect to the variable in $\mathbb{R}$ (the second variable).

Definition 3.5. *Let $g, h \in S(Z)$, then for any $\theta \in \mathbb{S}^{n-1}$ and $s, \sigma \in \mathbb{R}$, the Fourier transform of g is:*

$$\hat{g}(\theta, \sigma) = \int_{\mathbb{R}} g(\theta, s) e^{-2\pi i \sigma s} ds,$$

and the convolution of g and h is:

$$(g * h)(\theta, s) = \int_{\mathbb{R}} g(\theta, s-u) h(\theta, u) du.$$

Now we are ready to state and prove the main connection between the Radon and the Fourier transforms: the *Fourier projection-slice theorem*. Essentially, it says that the one-dimensional Fourier transform of $\mathcal{R}f$, the *projection*, coincides with the restriction of the Fourier transform of f to the *slice* θ.

Theorem 3.6. (Projection-slice theorem). *For any $f \in S(\mathbb{R}^n), \theta \in \mathbb{S}^{n-1}, \sigma \in \mathbb{R}$,*

$$\widehat{\mathcal{R}f}(\theta, \sigma) = \hat{f}(\sigma\theta).$$

Proof. By definition,

$$\hat{f}(\sigma\theta) = \int_{\mathbb{R}^n} f(t) e^{-2\pi i \langle t, \sigma\theta \rangle} dt,$$

where $\langle \, , \, \rangle$ denotes the usual dot product. Given a unit vector θ, by foliating $\mathbb{R}^n$ with hyperplanes orthogonal to θ, we can rewrite the points t in $\mathbb{R}^n$ as $t = s\theta + y$, with $s \in \mathbb{R}$ and $y \in \theta^{\perp}$. Since $\langle \theta, \theta^{\perp} \rangle = 0$ and $\langle \theta, \theta \rangle = 1$, then $\langle t, \sigma\theta \rangle = s\sigma$. Then, by Fubini's theorem,

$$\hat{f}(\sigma\theta) = \int_{\mathbb{R}} \int_{\theta^{\perp}} f(s\theta + y) e^{-2\pi i \sigma s} dy ds = \int_{\mathbb{R}} e^{-2\pi i \sigma s} \left(\int_{\theta^{\perp}} f(s\theta + y) dy \right) ds = \widehat{\mathcal{R}f}(\theta, \sigma).$$

□

Corollary 3.7. *The map $\mathcal{R} : S(\mathbb{R}^n) \to S(Z)$ is injective.*

Proof. If $\mathcal{R}f \equiv 0$, then $\hat{f}(\sigma\theta) = \widehat{\mathcal{R}f}(\theta, \sigma) = 0$ for all $\theta \in \mathbb{S}^{n-1}$ and $\sigma \in \mathbb{R}$. Hence $\hat{f} \equiv 0$, and by Fourier's uniqueness theorem, $f \equiv 0$. □

This answers the question from above. Thankfully, yes, if you know $\mathcal{R}f$ you can uniquely recover f. So, in theory, no two brains share the same projections unless they are identical.

3.4 Local Uniqueness

In most practical applications we won't know $\mathcal{R}f$ for all lines in every direction. Usually we only have the data on a proper subset of Z. Do these data determine the brain scan, i.e. the restriction of f to a certain subset of $\mathbb{R}^n$? We present two results that deal with this "local" uniqueness. They basically tell us that if we see nothing, it's because there is nothing indeed.

Theorem 3.8. (Hole theorem) *Let $f \in S(\mathbb{R}^n)$ and let $K \subset \mathbb{R}^n$ be a compact and convex set (the "hole") such that $\mathcal{R}f(\theta, s) = 0$ for each θ_s not meeting K. Then $f \equiv 0$ on $\mathbb{R}^n \setminus K$.*

Proof. We begin with some reductions to simplify the proof. Since K is a compact convex set, K is the intersection of all closed balls which contain it, we may assume that K is a closed ball. By translation, we may also suppose that it is centered at the origin, i.e. that $K = B(0, R)$, for some $R > 0$.

Now we only have to show that $\mathcal{R}(x_j f)(\theta, s) = 0$, for $j = 1, \ldots, n$, $\theta \in \mathbb{S}^{n-1}$ and $s > R$. Indeed, should we have $\mathcal{R}(x_j f)(\theta, s) = 0$, then we could apply the same to $g = x_k f \in S(\mathbb{R}^n)$ and we'd have $\mathcal{R}(x_j x_k f)(\theta, s) = 0$. Therefore, iterating this process, and since integrating is a linear operation, we would get $\mathcal{R}(pf)(\theta, s) = 0$ for all polynomials.

Then, by Weierstrass approximation theorem, we could take a sequence of polynomials p_n that conveniently converge uniformly on $\overline{f}$ and have

$$\mathcal{R}(|f|^2) = \mathcal{R}(\overline{f} f) = \lim_n \mathcal{R}(p_n f) = 0.$$

Therefore, since $|f|^2 \geq 0$, we would conclude that $f \equiv 0$ on $\mathbb{R}^n \setminus K$.

Let's see that $\mathcal{R}(x_j f) = 0$. We'll start with the easy coordinate, x_n. We pick $s > R$ and, as before, assume that θ is the North Pole $\theta^* = (0, \ldots, 0, 1)$. Then

$$\mathcal{R}(x_n f)(\theta^*, s) = \int_{\theta_s^*} x_n f(x)\, dx = s \int_{\theta_s^*} f(x)\, dx = s\mathcal{R}f(\theta^*, s) = 0,$$

so we only need to check that $\mathcal{R}(x_j f)(\theta^*, s) = 0$ for $1 \leq j < n$, i.e.,

$$\int_{\mathbb{R}^{n-1}} y_j f(y, s)\, dy = 0 \quad (1 \leq j < n).$$

Let's fix $1 \leq j < n$. Since $s > R$, we have $\mathcal{R}f(\theta, t) = 0$ for every $\theta \in \mathbb{S}^{n-1}$ and $t > s$. Then, we differentiate and, using Proposition 3.3 and $\theta^\perp \cong \mathbb{R}^{n-1}$, obtain:

$$0 = \frac{\partial \mathcal{R}f}{\partial \theta_j}(\theta^*, t) = -\frac{\partial R(x_j f)}{\partial s}(\theta^*, s) = -\int_{\mathbb{R}^{n-1}} y_j D_n f(y, t)\, dy.$$

Now, an integration on t along the interval $(s, +\infty)$ produces the required result:

$$0 = \int_s^\infty \int_{\mathbb{R}^{n-1}} y_j D_n f(y, t)\, dy\, dt = \int_{\mathbb{R}^{n-1}} y_j \left(\int_s^\infty D_n f(y, t)\, dt \right) dy = \int_{\mathbb{R}^{n-1}} y_j f(y, s)\, dy$$

□

The Hole theorem tells us that if $\mathcal{R}f$ vanishes outside a bounded interval for all directions, then f also vanishes outside that interval. The following result assumes kind of the opposite: that $\mathcal{R}f$ vanishes only on a certain set of directions, but for all $s \in \mathbb{R}$.

Definition 3.9. *We say that $A \subset \mathbb{S}^{n-1}$ is a uniqueness set for the homogeneous polynomials in $\mathbb{R}^n$ if the vanishing of a homogeneous polynomial p on A implies $p \equiv 0$.*

Theorem 3.10. *Let $A \subset \mathbb{S}^{n-1}$ be a uniqueness set for the homogeneous polynomials in $\mathbb{R}^n$. If $\mathcal{R}f(\theta, s) = 0$ for all $\theta \in A$ and all $s \in \mathbb{R}$, then $f \equiv 0$.*

Proof. As we have seen in the preliminaries, since $f \in C_c(\mathbb{R}^n)$, its Fourier transform extends to an entire function in $\mathbb{C}^n$ and can therefore be expanded as a series

$$\hat{f}(\xi) = \sum_{k=0}^{\infty} p_k(\xi),$$

where p_k is a homogeneous polynomial of degree k.

By the Fourier projection slice theorem (Theorem 3.6) and the homogeneity of p_k of degree k, for $\theta \in A$ and $s \in \mathbb{R}$,

$$0 = \widehat{\mathcal{R}f}(\theta, s) = \hat{f}(s\theta) = \sum_{k=0}^{\infty} p_k(s\theta) = \sum_{k=0}^{\infty} s^k p_k(\theta).$$

Viewing this as a power series in s, we deduce that $p_k(\theta) = 0$ for all $\theta \in A$, which by hypothesis implies $p_k \equiv 0$ for all $k \geq 0$. Thus, $\hat{f} \equiv 0$ and therefore $f \equiv 0$. □

3.5 Backprojection operator

Let's introduce an operator that plays a crucial role in the inversion formulas as a substitute of the Fourier transform. The *backprojection* of $g \in S(Z)$ is the function $\mathcal{R}^{\#}g \in S(\mathbb{R}^n)$ defined by

$$\mathcal{R}^{\#}g(t) = \int_{\mathbb{S}^{n-1}} g(\theta, \langle t, \theta \rangle) d\theta, \quad t \in \mathbb{R}^n.$$

As we could expect, in our case $g = \mathcal{R}f$. Therefore, the operation $\mathcal{R}f(\theta, \langle t, \theta \rangle)$ is the integral of f on the hyperplane passing through $t \in \mathbb{R}^n$ and orthogonal to $\theta \in \mathbb{S}^{n-1}$. In that sense, $\mathcal{R}^{\#}g$ is the average of the integrals of f along the hyperplanes passing through t.

The following proposition summarizes some properties of the backprojection.

Proposition 3.11.

1. *If $g \in C(Z)$, then $\mathcal{R}^{\#}g \in C(\mathbb{R}^n)$.*
2. *If $g \in C^1(Z)$, then $\mathcal{R}^{\#}g \in C^1(\mathbb{R}^n)$ and*

$$\frac{\partial}{\partial t_j}(\mathcal{R}^{\#}g) = \mathcal{R}^{\#}\left(\theta_j \frac{\partial g}{\partial s}\right).$$

Analogously, if $g \in C^\infty(Z)$, then $\mathcal{R}^{\#}g \in C^\infty(\mathbb{R}^n)$ and for any multi-index α,

$$\frac{\partial^{|\alpha|}}{\partial t_1^{\alpha_1} \cdots \partial t_n^{\alpha_n}}(\mathcal{R}^{\#}g) = \mathcal{R}^{\#}\left(\theta^{\alpha} \frac{\partial^{|\alpha|} g}{\partial s^{|\alpha|}}\right).$$

3. *The backprojection and the Laplacian commute. Let* $\Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial t_j^2}$ *be the Laplacian in* $\mathbb{R}^n$ *and let* $\Box = \frac{\partial^2}{\partial s^2}$ *be the Laplacian in* $\mathbb{R}$*. For* $g \in C^2(Z)$*,*

$$\mathcal{R}^{\#}(\Box g) = \Delta(\mathcal{R}^{\#} g).$$

4. *The operator* $\mathcal{R}^{\#}$ *is the formal adjoint of* $\mathcal{R}$*, in the sense that for* $f \in S(\mathbb{R}^n)$ *and* $g \in S(Z)$*,*

$$\int_{\mathbb{S}^{n-1}} \int_{\mathbb{R}} \mathcal{R}f(\theta, s) g(\theta, s) ds d\theta = \int_{\mathbb{R}^n} f(t) \mathcal{R}^{\#} g(t) dt.$$

In particular,

$$f * (\mathcal{R}^{\#} g) = \mathcal{R}^{\#}(\mathcal{R}f * g).$$

Proof. Properties 1 and 2 follow directly from the definition. Property 3 is a consequence of Property 2 and the fact that $\|\theta\|^2 = \sum_{j=1}^{n} \theta_j^2 = 1$, so

$$\Delta(\mathcal{R}^{\#} g)(t) = \sum_{j=1}^{n} \mathcal{R}^{\#} \left(\theta_j^2 \frac{\partial^2 g}{\partial s^2} \right) = \mathcal{R}^{\#} \left(\frac{\partial^2 g}{\partial s^2} \right).$$

Property 4 follows from Fubini's theorem. Writing $t = s\theta + y$, with $y \in \theta^{\perp}$,

$$\begin{aligned} \int_{\mathbb{R}^n} f(t) \mathcal{R}^{\#} g(t)\, dt &= \int_{\mathbb{S}^{n-1}} \int_{\mathbb{R}^n} f(t) g(\theta, \langle t, \theta \rangle)\, dt\, d\theta \\ &= \int_{\mathbb{S}^{n-1}} \int_{\mathbb{R}} \int_{\theta^{\perp}} f(s\theta + y) g(\theta, s)\, dy\, ds\, d\theta \\ &= \int_{\mathbb{S}^{n-1}} \int_{\mathbb{R}} \left(\int_{\theta^{\perp}} f(s\theta + y)\, dy \right) g(\theta, s)\, ds\, d\theta. \end{aligned}$$

Thus, $f * (\mathcal{R}^{\#} g) = \mathcal{R}^{\#}(\mathcal{R}f * g)$. □

4 Inversion formulas

We are ready to introduce the inversion formulas based on the Radon transform and the backprojection operator. They are the basis of reconstruction algorithms used in medical imaging. For instance, the first formula we will see is the foundation of the *gridding algorithm*, which will be discussed later. This explicit formula expresses f in terms of the Fourier transform of $\mathcal{R}f$, which is not numerically ideal (as we will also see later) but it's a good start.

Theorem 4.1. (Fourier inversion formula). *Let $f \in S(\mathbb{R}^n)$. Then*

$$f(t) = \frac{1}{2}\int_{\mathbb{S}^{n-1}} \left(\int_{\mathbb{R}} |\sigma|^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma \right) d\theta.$$

Proof. First, we can use the inverse of the Fourier transform and switch to polar coordinates $\xi = \sigma\theta \in \mathbb{R}^n$, $\sigma \geq 0$, $\theta \in \mathbb{S}^{n-1}$:

$$f(t) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i \langle t,\xi \rangle} d\xi = \int_0^\infty \int_{\mathbb{S}^{n-1}} \hat{f}(\sigma\theta) e^{2\pi i \sigma \langle t,\theta \rangle}\, d\theta\, \sigma^{n-1} d\sigma.$$

Now, using the Fourier projection-slice theorem (Theorem 3.6), we get:

$$f(t) = \int_{\mathbb{S}^{n-1}} \left(\int_0^\infty \sigma^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma \right) d\theta.$$

For simplicity, let's denote:

$$F(\theta,t) = \int_0^\infty \sigma^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma.$$

By the parity of $\widehat{\mathcal{R}f}$ and changing the integration range,

$$\begin{aligned} F(-\theta,t) &= \int_0^\infty \sigma^{n-1} e^{2\pi i \sigma \langle t,-\theta \rangle} \widehat{\mathcal{R}f}(-\theta,\sigma)\, d\sigma = \int_0^\infty \sigma^{n-1} e^{2\pi i (-\sigma) \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,-\sigma)\, d\sigma \\ &= \int_{-\infty}^0 |\sigma|^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma. \end{aligned}$$

Thus, since $\mathbb{S}^{n-1}$ is symmetric, we can observe that:

$$\int_{\mathbb{S}^{n-1}} F(-\theta,t) d\theta = \int_{\mathbb{S}^{n-1}} F(\theta,t) d\theta.$$

Therefore,

$$f(t) = \int_{\mathbb{S}^{n-1}} F(\theta,t)\, d\theta = \frac{1}{2}\int_{\mathbb{S}^{n-1}} (F(\theta,t) + F(-\theta,t))\, d\theta.$$

Finally, we conclude that:

$$\begin{aligned} f(t) &= \frac{1}{2}\int_{\mathbb{S}^{n-1}} \left(\int_0^\infty \sigma^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma + \int_{-\infty}^0 |\sigma|^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma \right) d\theta \\ &= \frac{1}{2}\int_{\mathbb{S}^{n-1}} \left(\int_{-\infty}^\infty |\sigma|^{n-1} e^{2\pi i \sigma \langle t,\theta \rangle} \widehat{\mathcal{R}f}(\theta,\sigma)\, d\sigma \right) d\theta \end{aligned}$$

as desired. □

As we have mentioned before, we can get rid of the Fourier transform using the backprojection. The following formula is only valid for odd dimensions. Following its proof, we will later see the even (and more technical) case.

Theorem 4.2. (Backprojection inversion formula for n odd) *Let n be odd, let $f \in S(\mathbb{R}^n)$ and denote $g = \mathcal{R}f$. Then*

$$f(t) = \frac{1}{2(2\pi i)^{n-1}} \mathcal{R}^{\#} \left(\frac{\partial^{n-1} g}{\partial s^{n-1}} \right)(t), \quad t \in \mathbb{R}^n.$$

In particular, for $n = 3$,

$$f(t) = -\frac{1}{8\pi^2} \mathcal{R}^{\#} \left(\frac{\partial^2 g}{\partial s^2} \right)(t) = -\frac{1}{8\pi^2} \mathcal{R}^{\#}(\Box g)(t) = -\frac{1}{8\pi^2} \Delta(\mathcal{R}^{\#} g)(t), \quad t \in \mathbb{R}^3.$$

Proof. On the one hand, by definition of the backprojection:

$$\frac{1}{2(2\pi i)^{n-1}} \mathcal{R}^{\#} \left(\frac{\partial^{n-1} g}{\partial s^{n-1}} \right)(t) = \frac{1}{2(2\pi i)^{n-1}} \int_{\mathbb{S}^{n-1}} \frac{\partial^{n-1} g}{\partial s^{n-1}}(\theta, \langle t, \theta \rangle)\, d\theta.$$

On the other hand, by the Fourier inversion formula:

$$f(t) = \frac{1}{2} \int_{\mathbb{S}^{n-1}} \left(\int_{\mathbb{R}} |\sigma|^{n-1} e^{2\pi i \sigma \langle t, \theta \rangle} \hat{g}(\theta, \sigma)\, d\sigma \right) d\theta.$$

Therefore, we only need to show that:

$$\int_{\mathbb{R}} |\sigma|^{n-1} e^{2\pi i \sigma \langle t, \theta \rangle} \hat{g}(\theta, \sigma)\, d\sigma = \frac{1}{(2\pi i)^{n-1}} \frac{\partial^{n-1} g}{\partial s^{n-1}}(\theta, \langle t, \theta \rangle).$$

Since n is odd, we have $|\sigma|^{n-1} = \sigma^{n-1}$. Note that, by Corollary 3.4, $g(\theta, \sigma) \in S(Z)$ and therefore we can use the iterative version of Theorem 2.4 (b):

$$\begin{aligned} \int_{\mathbb{R}} \sigma^{n-1} e^{2\pi i \sigma \langle t, \theta \rangle} \hat{g}(\theta, \sigma)\, d\sigma &= \frac{1}{(2\pi i)^{n-1}} \int_{\mathbb{R}} (2\pi i \sigma)^{n-1} \hat{g}(\theta, \sigma) e^{2\pi i \sigma \langle t, \theta \rangle}\, d\sigma \\ &= \frac{1}{(2\pi i)^{n-1}} \int_{\mathbb{R}} \widehat{\frac{\partial^{n-1} g}{\partial s^{n-1}}}(\theta, \sigma) e^{2\pi i \sigma \langle t, \theta \rangle}\, d\sigma. \end{aligned}$$

Finally, using the inverse of the Fourier transform:

$$\int_{\mathbb{R}} \sigma^{n-1} e^{2\pi i \sigma \langle t, \theta \rangle} \hat{g}(\theta, \sigma)\, d\sigma = \frac{1}{(2\pi i)^{n-1}} \frac{\partial^{n-1} g}{\partial s^{n-1}}(\theta, \langle t, \theta \rangle).$$

□

When n is even, we indeed have:

$$|\sigma|^{n-1} = \operatorname{sgn}(\sigma) \cdot \sigma^{n-1} = \begin{cases} \sigma^{n-1} & \text{if } \sigma > 0 \\ 0 & \text{if } \sigma = 0 \\ -\sigma^{n-1} & \text{if } \sigma < 0. \end{cases}$$

Hence, we need to identify the function g such that $\hat{g}(\sigma) = \operatorname{sgn}(\sigma) \hat{f}(\sigma)$. This requires us to naturally introduce the Hilbert transform, named after David Hilbert (1862-1943).

Definition 4.3. *The Hilbert transform of $f \in S(\mathbb{R})$ is the function*

$$Hf(x) = \frac{i}{\pi}\, p.v. \int_{-\infty}^{\infty} \frac{f(x-s)}{s}\, ds := \frac{i}{\pi} \lim_{\epsilon \to 0} \int_{|s|>\epsilon} \frac{f(x-s)}{s}\, ds.$$

There is a substantial body of deep theory behind this transform, which lies beyond the scope of this work. Here, we focus only on the following property, which is precisely what we need for the case n even above.

Proposition 4.4. *Let $f \in S(\mathbb{R})$. Then:*

$$\widehat{Hf}(\sigma) = \operatorname{sgn}(\sigma)\, \hat{f}(\sigma).$$

Proof. By definition

$$p.v. \int_{-\infty}^{\infty} \frac{f(t-s)}{s} ds = \lim_{\epsilon \to 0} \int_{|s|>\epsilon} \frac{f(t-s)}{s} ds.$$

Let us omit this notation and simply write

$$\int_{-\infty}^{\infty} \frac{f(t-s)}{s} ds.$$

Let $\varphi \in S(\mathbb{R})$; then, using the translation property of the Fourier transform, d) in Theorem 2.4,

$$\begin{aligned}
\langle \widehat{Hf}, \varphi \rangle = \langle Hf, \hat{\varphi} \rangle &= \frac{i}{\pi} \int_{\mathbb{R}} \left(\int_{-\infty}^{\infty} \frac{f(t-s)}{s} ds \right) \left(\int_{\mathbb{R}} \varphi(u) e^{-2\pi i u t} du \right) dt \\
&= \frac{i}{\pi} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \frac{1}{s} \varphi(u) \left(\int_{\mathbb{R}} f(t-s) e^{-2\pi i u t} dt \right) ds du \\
&= \frac{i}{\pi} \int_{\mathbb{R}} \int_{-\infty}^{\infty} \frac{1}{s} \varphi(u) \hat{f}(u) e^{-2\pi i u s} ds du \\
&= \frac{i}{\pi} \int_{\mathbb{R}} \hat{f}(u) \varphi(u) \left(\int_{-\infty}^{\infty} \frac{e^{-2\pi i u s}}{s} ds \right) du.
\end{aligned}$$

Thus we shall be done as soon as we prove that

$$p.v. \int_{-\infty}^{\infty} \frac{e^{-2\pi i u s}}{s}\, ds = \frac{\pi}{i} \operatorname{sgn}(u).$$

We have

$$\begin{aligned}
p.v. \int_{-\infty}^{\infty} \frac{e^{-2\pi i u s}}{s}\, ds &= \int_0^{\infty} \frac{e^{-2\pi i u s} - e^{2\pi i u s}}{s}\, ds = -2i \int_0^{\infty} \frac{\sin(2\pi u s)}{s}\, ds \\
&= -4\pi i u \int_0^{\infty} \frac{\sin(2\pi u s)}{2\pi u s}\, ds.
\end{aligned}$$

Since the cardinal sine function is even

$$\int_0^{\infty} \frac{\sin(2\pi u s)}{2\pi u s}\, ds = \frac{1}{2} \int_{\mathbb{R}} \frac{\sin(2\pi u s)}{2\pi u s}\, ds,$$

and since, for $u > 0$,

$$\widehat{\frac{1}{2u}\chi_{[-u,u]}}(s) = \frac{1}{2u}\int_{-u}^{u} e^{-2\pi ist}\,dt = \frac{1}{2u}\left[\frac{e^{-2\pi ist}}{-2\pi is}\right]_{t=-u}^{t=u} = \frac{1}{2u}\frac{e^{-2\pi isu} - e^{2\pi isu}}{-2\pi is} = \frac{\sin(2\pi us)}{2\pi us},$$

the Fourier inversion formula gives

$$\int_{\mathbb{R}} \frac{\sin(2\pi us)}{2\pi us}\,ds = \int_{\mathbb{R}} \frac{1}{2u}\chi_{[-u,u]}(s)e^{2\pi i0s}\,ds = \frac{1}{2u}\chi_{[-u,u]}(0) = \frac{1}{2u}.$$

When $u < 0$, following the argument above to deduce the analogous result

$$\int_{\mathbb{R}} \frac{\sin(2\pi us)}{2\pi us}\,ds = -\frac{1}{2u}.$$

All combined

$$\int_0^{\infty} \frac{\sin(2\pi us)}{2\pi us}\,ds = \mathrm{sgn}(u)\frac{1}{4u},$$

hence

$$p.v.\int_{-\infty}^{\infty} \frac{e^{-2\pi ius}}{s}\,ds = (-4\pi iu)\mathrm{sgn}(u)\frac{1}{4u} = \frac{\pi}{i}\mathrm{sgn}(u),$$

as desired. □

Therefore, going back to the proof of Theorem 4.2, in the case of even n we have:

$$\int_{\mathbb{R}} |\sigma|^{n-1}e^{2\pi i\sigma\langle t,\theta\rangle}\,\hat{g}(\theta,\sigma)\,d\sigma = \frac{1}{(2\pi i)^{n-1}}\frac{\partial^{n-1}Hg}{\partial s^{n-1}}(\theta,\langle t,\theta\rangle).$$

Then, since $g \in S(Z)$, we can use differentiation under integral sign (Theorem 2.10) and the definition of the derivative to see that:

$$\frac{\partial^{n-1}Hg}{\partial s^{n-1}} = H\frac{\partial^{n-1}g}{\partial s^{n-1}},$$

Putting it altogether, we obtain the following result.

Theorem 4.5. (Backprojection inversion formula for n even) *Let n be even, let $f \in S(\mathbb{R}^n)$ and denote $g = Rf$. Then*

$$f(t) = \frac{1}{2(2\pi i)^{n-1}}\mathcal{R}^{\#}\left[H\frac{\partial^{n-1}g}{\partial s^{n-1}}\right](t), \quad t \in \mathbb{R}^n.$$

In particular, for $n = 2$,

$$f(t) = \frac{1}{4\pi i}\mathcal{R}^{\#}\left(H\frac{\partial g}{\partial s}\right)(t), \quad t \in \mathbb{R}^2$$

This is Radon's original inversion formula (1917), though expressed here in contemporary notation.

Remark. Note that, given $\theta \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}^n$, the derivative $\frac{\partial^{n-1}g}{\partial s^{n-1}}(\theta,\langle t,\theta\rangle)$ is determined by the values of $g(\theta,x)$ for $x \in \mathbb{R}$ near $\langle t,\theta\rangle$. Recall also that the hyperplanes containing $t \in \mathbb{R}^n$ are those of the form θs, with $s = \langle t,\theta\rangle$. Therefore, to compute

$$\int_{\mathbb{S}^{n-1}} \frac{\partial^{n-1} g}{\partial s^{n-1}}(\theta, \langle t, \theta \rangle)\, d\theta, \quad t \in \mathbb{R}^n$$

we only need the values of the integrals of f along all the hyperplanes through a neighborhood of t. In this sense, the inversion of the Radon transform in odd dimensions is a local problem.

However, the same conclusion is not possible when n is even, because the Hilbert transform is not local. Indeed, the value $Hf(t)$ depends on the value of f everywhere, not just in a neighborhood of t.

5 Image reconstruction

Having established the theoretical foundations of CT, we now delve into its practical aspects. We begin this section by defining the function space used in applications and discussing the sampling process. We then examine two primary reconstruction algorithms: the filtered backprojection and the gridding method, emphasizing their derivation and practical implementation. To illustrate these techniques in action, we present a specific case study, the reconstruction of the Shepp-Logan phantom. Finally, we conclude with a discussion on alternative geometries, which aim to improve the efficiency and resolution of image reconstruction.

5.1 Sampling

The starting point in the implementation of the inversion formulas in the practical reconstruction of images from X-ray data is that only a finite number of values of $\mathcal{R}f$ are available in a real study. This raises the question: how many values are needed in order to generate a clinically useful image?

Before answering this question we need to clarify the function space we will consider. We assume that we are only interested in the details (or frequencies) up to a certain scale. This can be modeled by considering only bandlimited functions, that is, functions f with Fourier transform $\hat{f}$ supported in a ball $B(0,\Omega)$, $\Omega > 0$ (the smallest such Ω is called the bandwidth of f). According to our definition of the Fourier transform, the finest details that such functions can deal with are of scale $1/\Omega$.

The issue is that bandlimited functions cannot be in the Schwartz class. Indeed, as we have seen in Proposition 2.8, if $f \in S(\mathbb{R}^n)$ then $\hat{f} \in S(\mathbb{R}^n)$, but not necessarily in the subclass $C_c^\infty(\mathbb{R}^n)$. Hence $\hat{f}$ doesn't necessarily vanish outside a ball $B(0,\Omega)$.

However, in the real world it is impossible to verify whether a function is bandlimited or compactly supported. Indeed, we would have to take measurements at arbitrarily high frequencies or at arbitrarily remote places, i.e. experiments that can never be carried out.

It is possible, nevertheless, to have functions essentially band-limited in the following sense:

Definition 5.1. *A function $f \in L^2$ is essentially Ω-bandlimited at level ϵ if*

$$\int_{B(0,\Omega)} |\hat{f}(\xi)|^2 \, d\xi \geq (1-\epsilon)\|f\|_2^2.$$

Similarly, a function $f \in L^2$ is essentially R-supported at level ϵ if

$$\int_{B(0,R)} |f(x)|^2 \, dx \geq (1-\epsilon)\|f\|_2^2.$$

We can interpret $\|f\|_2^2$ as the energy of the signal. In this sense, essentially Ω-bandlimited functions have its Fourier transform mostly contained within the ball $B(0,\Omega)$, but there can be a small amount of energy outside it, quantified by ϵ. Analogously, essentially R-supported functions contain most of its energy, up to $(1-\epsilon)$, inside the ball $B(0,R)$.

From now on, we will analyze functions that are essentially Ω-bandlimited and, for simplicity, supported in $B(0,1)$. The following "folklore" theorem from signal theory provides an

answer to the question posed above: the dimension of the space of functions Ω-bandlimited with support in $B(0,R)$ can be approximated by:

$$|B(0,R)||B(0,\Omega)| = v_n^2 \, (R\Omega)^n \,,$$

where $v_n = \pi^{n/2}\Gamma(1+n/2)$ is the volume of the unit ball in $\mathbb{R}^n$. A proof for the 1-dimensional case of this result can be found in the appendix of [11].

This gives a lower bound on the number of samples that determine a function in this subspace. In our 2-dimensional case, this bound is simply $(\pi\Omega)^2$.

We assume that the samples are obtained using *parallel standard geometry*, that is, with a certain number p of equidistributed directions, and taking $2q+1$ equispaced samples in each of them. Hence, the data take the form $\mathcal{R}f(\theta_j, s_l) = g(\theta_j, s_l)$, where:

$$\theta_j = e^{i\varphi_j}, \quad \varphi_j = j\pi/p, \quad j = 0, \ldots, p-1, \quad \Delta\varphi = \pi/p$$

and

$$s_l = l/q, \quad l = -q, \ldots, q, \quad \Delta s = 1/q.$$

The specific values of p and q essentially depend on the design of the machine itself and on the sizes of the objects the machine is designed to scan. The following figure shows the lines that a scan would take for $p = 10$ and $q = 5$:

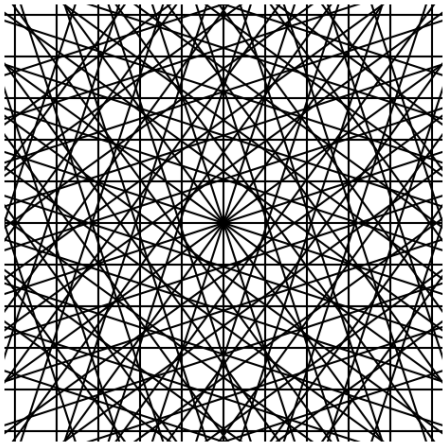

Figure 7: Sample lines obtained using parallel beam geometry with $p = 10$ and $q = 5$.

We are now set to accomplish our goal of reconstructing an attenuation-coefficient function f using a discrete set of samples of its Radon transform. We will present two algorithms: the *filtered backprojection* and the *gridding method* (Fourier method). Let us work our way through each algorithm step by step.

5.2 Filtered backprojection

The filtered backprojection is based on the following relation (see 4 in Proposition 3.11):

$$(\mathcal{R}^{\#}v) * f = \mathcal{R}^{\#}(v * Rf),$$

where $f \in S(\mathbb{R}^n)$, $v \in S(Z)$, and $\mathcal{R}^{\#}$ is the backprojection operator.

For $g = Rf$ and setting $V = \mathcal{R}^{\#}v$, this identity becomes:

$$(V * f)(x) = \mathcal{R}^{\#}(v * g)(x) = \int_{S^{n-1}} (v * g)(\theta, x \cdot \theta)\, d\theta. \tag{5.1}$$

This explains the name of the algorithm: the data g are first *filtered* by v (producing $v*g$), and then the *backprojection* $\mathcal{R}^{\#}$ is applied. As we will see, since the number of samples is finite, the integrals from the convolution and the backprojection must be discretized.

The filter v is applied to mitigate the impact of a constant challenge in signal analysis: the presence of noise. This refers to any effect that distorts or contaminates the signal, which can come from background radiation, interfering stray signals, errors in the measurement process, or other factors related to the random nature of radioactivity.

Besides dealing with noise, filtering is essential because, due to the nature of the backprojection operation, reconstructing the data directly without it would introduce artifacts in the reconstructed images.

This filter is determined by the so-called *point-spread function* $V = \mathcal{R}^{\#}v$. The choice of V is crucial not only for this reason, but also because the left-hand side of (5.1) can approximate $f(x)$ if V approximates the Dirac delta function δ_0.

Convolution is a computationally intensive task, so it is better to avoid it when possible. Since the convolution in the spatial domain is equivalent to a multiplication in the frequency domain (see g) in Theorem 2.4), the filtering by V in the spatial domain can be performed by a simple multiplication by $\hat{V}$ in the frequency domain. That is no trouble as, in the usual cases, V can be computed explicitly from $\hat{V}$.

Since f has (essential) bandwidth Ω and high frequencies are mostly noise, we look for V such that

$$\widehat{(V * f)}(\zeta) \simeq \begin{cases} \hat{f}(\zeta), & \text{if } |\zeta| \leq \Omega, \\ 0, & \text{if } |\zeta| > \Omega. \end{cases}$$

The relationship between $\hat{V}$ and $\hat{v}$ is explicit through the following distributional identity (Theorem 1.4 in Section II from [8]): if $g \in S(Z)$ is even, then

$$\widehat{(\mathcal{R}^{\#}g)}(\zeta) = 2|\zeta|^{1-n}\hat{g}(\zeta/|\zeta|, |\zeta|).$$

Since we can choose the filter, the condition of the theorem is conveniently satisfied. Indeed, considering radial symmetric functions $V(x) = V(|x|)$, v does not depend on θ and it is an even function of s. In this particular situation, the identity above gives:

$$\hat{V}(\zeta) = \widehat{(\mathcal{R}^{\#}v)}(\zeta) = 2|\zeta|^{1-n}\hat{v}(|\zeta|). \tag{5.2}$$

In order to accurately reconstruct functions f with essential bandwidth Ω, we can take, for instance, $\hat{V}(\zeta) = \mathcal{X}_{B(0,\Omega)}(\zeta)$ (called the Ram-Lak filter, as we will soon see), where $\mathcal{X}_{B(0,\Omega)}$ is the characteristic function of the ball $B(0, \Omega)$.

More generally, let us consider a filter $\hat{\varphi}(\sigma)$ close to 1 when $|\sigma| \leq 1$ and with $\hat{\varphi}(\sigma) = 0$ for $|\sigma| > 1$ such that:

$$\hat{V}_{\Omega}(\zeta) = \hat{\varphi}\left(\frac{|\zeta|}{\Omega}\right).$$

According to (5.2), the corresponding function v_Ω (such that $\mathcal{R}^\# v_\Omega = V_\Omega$) is determined by the identity

$$\hat{v}_\Omega(\sigma) = \frac{1}{2}|\sigma|^{n-1}\hat{\varphi}\left(\frac{|\sigma|}{\Omega}\right). \tag{5.3}$$

Since multiplication by $|\sigma|^{n-1}$ in the Fourier domain corresponds (except for a constant) to the operation $\frac{\partial^{n-1}}{\partial s^{n-1}}$, for odd n, and $H\frac{\partial^{n-1}}{\partial s^{n-1}}$, for even n, the previous identity can be viewed as the Fourier counterpart of the Backprojection inversion formulas (see Theorems 4.2 and 4.5).

Many different filters have been proposed. It seems, however, that there is no justification for any specific choice other than the experimental results. In other words, the choice of the filter is still more an art than a science. Next, for dimension $n = 2$, i.e. $\mathbb{R}^2$, we show three common filters.

(a) **Ram-Lak filter.** Introduced by Ramachandran and Lakshminarayanan in 1971. As we have mentioned before, it is associated with the standard low-pass filter $\hat{\varphi}(\sigma) = \mathcal{X}_{[0,1]}(\sigma)$. Here, (5.3) gives

$$\hat{v}_\Omega(\sigma) = \frac{1}{2}|\sigma|\mathcal{X}_{[0,1]}\left(\frac{|\sigma|}{\Omega}\right),$$

and therefore

$$v_\Omega(s) = \int_{\mathbb{R}} \hat{v}_\Omega(\sigma)e^{2\pi i\sigma s}\,d\sigma = \frac{1}{2}\int_{-\Omega}^{\Omega} |\sigma|e^{2\pi i\sigma s}\,d\sigma.$$

Splitting the integral for $\sigma > 0$ and $\sigma < 0$ and integrating by parts, we obtain:

$$\begin{aligned}
\int_{-\Omega}^{\Omega} |\sigma|e^{2\pi i\sigma s}\,d\sigma &= \int_{-\Omega}^{0} -\sigma e^{2\pi i\sigma s}\,d\sigma + \int_0^{\Omega} \sigma e^{2\pi i\sigma s}\,d\sigma = \int_0^{\Omega} \sigma\left(e^{-2\pi i\sigma s} + e^{2\pi i\sigma s}\right)\,d\sigma \\
&= 2\int_0^{\Omega} \sigma\cos(2\pi\sigma s)\,d\sigma = 2\left[\frac{\sigma\sin(2\pi\sigma s)}{2\pi s}\right]_0^{\Omega} - 2\int_0^{\Omega}\frac{\sin(2\pi\sigma s)}{2\pi s}\,d\sigma \\
&= 2\Omega^2\frac{\sin(2\pi\Omega s)}{2\pi\Omega s} + 2\frac{\cos(2\pi\Omega s) - 1}{(2\pi s)^2} \\
&= 2\Omega^2\left(\operatorname{sinc}(2\pi\Omega s) - \frac{1}{2}(\operatorname{sinc}(\pi\Omega s))^2\right)
\end{aligned}$$

where $\operatorname{sinc}(x) = \frac{\sin(x)}{x}$ is the cardinal sine, and finally:

$$v_\Omega(s) = \Omega^2 u(2\pi\Omega s),$$

where $u(s) = \operatorname{sinc}(s) - \frac{1}{2}\left(\operatorname{sinc}\left(\frac{s}{2}\right)\right)^2$.

(b) **Cosine filter.** Here $\hat{\varphi}(\sigma) = \cos\left(\frac{\sigma\pi}{2}\right)\mathcal{X}_{[0,1]}$ and the corresponding filter is

$$v_\Omega(s) = \frac{\Omega^2}{2}\left(u\left(2\pi\Omega s + \frac{\pi}{2}\right) + u\left(2\pi\Omega s - \frac{\pi}{2}\right)\right),$$

where u is the same as in (a).

(c) **Shepp-Logan filter.** Now $\hat{\varphi}(\sigma) = \operatorname{sinc}\left(\frac{\sigma\pi}{2}\right)\mathcal{X}_{[0,1]}$ and the filter is

$$v_\Omega(s) = \frac{2\Omega^2}{\pi}t(2\pi\Omega s), \quad \text{where } t(s) = \begin{cases} \frac{\pi/2 - s\sin s}{(\pi/2)^2 - s^2}, & \text{if } s \neq \pm\pi/2, \\ 1/\pi, & \text{if } s = \pm\pi/2. \end{cases}$$

The following figure helps us to visualize the three filters:

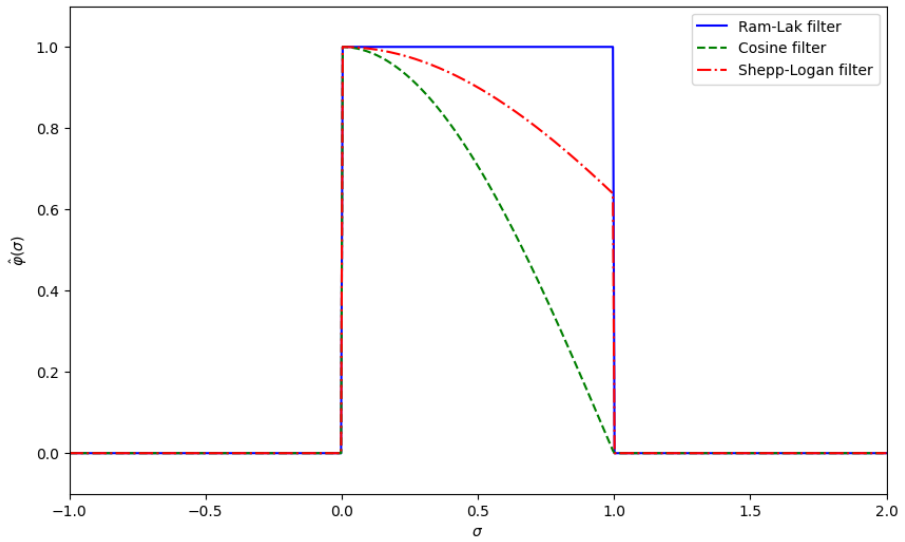

Figure 8: Three common filters for $\mathbb{R}^2$.

Once we have chosen our filter, we are ready to evaluate the right side of (5.1). As we have mentioned before, the integrals must be discretized. First, the convolution:

$$(v_\Omega * g)(\theta, s) = \int_{\mathbb{R}} v_\Omega(s-t)g(\theta,t)\,dt = \int_{-1}^{1} v_\Omega(s-t)g(\theta,t)\,dt.$$

We will use the following result (Theorem 1.1 in Section III from [8]):

Theorem 5.2. (Shannon's theorem).

(a) Let $f_1, f_2 \in L^2(\mathbb{R})$ have bandwidth Ω and let $0 < \Delta x \leq \frac{1}{2\Omega}$. Then

$$\int_{\mathbb{R}} f_1(x)f_2(x)\,dx = \Delta x \sum_{l\in\mathbb{Z}} f_1(l\Delta x)f_2(l\Delta x).$$

(b) Let $f \in L^2(\mathbb{R})$ have bandwidth Ω and let $0 < \Delta x \leq \frac{1}{2\Omega}$. Then

$$\int_{\mathbb{R}} f(x)\,dx = \Delta x \sum_{l\in\mathbb{Z}} f(l\Delta x).$$

Now, according to (5.2), $v_\Omega = v$ has bandwidth Ω. On the other hand, by the projection-slice theorem (Theorem 3.6), g, as a function of s, is essentially bandlimited. Thus, except for a negligible error (since g is only essentially bandlimited), Shannon's theorem (a) can be applied to $f_1(t) = v_\Omega(s-t)$, $f_2(t) = g(\theta,t)$, and the grid $(\Delta s)\mathbb{Z}$, with $\Delta s \leq 1/(2\Omega)$. This yields:

$$(v_\Omega * g)(\theta, s) = \Delta s \sum_{l=-q}^{q} v_\Omega(s-s_l)g(\theta,s_l). \tag{5.4}$$

Next step consists of discretizing the backprojection:

$$(V * f)(x) = \int_0^{2\pi} (v * g)(\theta, x \cdot \theta)\, d\phi, \quad \text{where } \theta = e^{i\phi}.$$

A computation shows that the π-periodic function $h(\phi) = (v * g)(\theta, x \cdot \theta)$ has essential bandwidth $4\pi\Omega$, in the sense that

$$\hat{h}(k) = \frac{1}{2\pi} \int_0^{2\pi} (v * g)(\theta, x \cdot \theta) e^{-ik\phi}\, d\phi$$

is negligible for $|k| > 4\pi\Omega$ (see Section V.I in [8]). Thus, we can apply Shannon's theorem (b), at the cost of only a negligible error: if $\Delta\phi \leq 1/(2\Omega)$, and we obtain the approximation:

$$(V * f)(x) = \int_0^{2\pi} (v * g)(\theta, x \cdot \theta)\, d\phi = \frac{\pi}{p} \sum_{j=0}^{2p-1} (v * g)(\theta_j, x \cdot \theta_j) = \frac{2\pi}{p} \sum_{j=0}^{p-1} (v * g)(\theta_j, x \cdot \theta_j),$$

where the last identity follows by π-periodicity of h.

This, together with (5.4), and always taking $\max\{\Delta\phi, \Delta s\} \leq 1/(2\Omega)$, yields:

$$(V * f)(x) = \frac{2\pi}{p} \Delta s \sum_{j=0}^{p-1} \sum_{l=-q}^{q} v_\Omega(x \cdot \theta_j - s_l) g(\theta_j, s_l). \tag{5.5}$$

The algorithm, as given by (5.5), is computationally too demanding. It requires $O(pq)$ operations for each $f(x)$, and since f has (essential) bandwidth Ω, it is necessary to compute $f(x)$ in a grid with stepsize $1/(2\Omega)$. This gives a total number of operations of order $O(\Omega^2 pq) \simeq O(\Omega^4)$.

We can reduce this complexity by doing a linear interpolation. Let us see what weight η (which is the parameter of the segment between two known values) should be used. For a given direction θ_j, we have $2q + 1$ points that are equidistributed with a spacing of $\Delta s = 1/q$. Therefore, a point $x \in \mathbb{R}^2$ lies on the sample line if there exists an integer k such that $k\Delta s = x \cdot \theta_j$.

Now, $\eta \in [0, 1]$, and:

$$\left[\frac{x \cdot \theta_j}{\Delta s}\right] \leq \frac{x \cdot \theta_j}{\Delta s} \leq \left[\frac{x \cdot \theta_j}{\Delta s}\right] + 1,$$

Thus we will pick $\eta = \eta(j, x) = \frac{x \cdot \theta_j}{\Delta s} - \left[\frac{x \cdot \theta_j}{\Delta s}\right]$.

Since $v_\Omega * g$ has bandwidth Ω, it is determined by the data $(v_\Omega * g)(\theta_j, s_l)$, which can be computed with $O(pq^2)$ operations. Then, the values $(v_\Omega * g)(\theta_j, x \cdot \theta_j)$ required to compute $V * f$ are obtained from the previous ones by linear interpolation. This reduces the number of operations to $O(\Omega^3)$.

Final algorithm:

Step 1. For every direction θ_j, $j = 1, \ldots, p$, we take the discrete convolution:

$$h_{j,k} = \Delta s \sum_{l=-q}^{q} v_\Omega(s_k - s_l) g_{j,l}, \quad k = -q, \ldots, q.$$

Step 2. For each x, we compute the discrete backprojection using a linear interpolation of the values obtained in *Step 1*:

$$f_{FB}(x) = \frac{2\pi}{p} \sum_{j=0}^{p-1} \left[(1-\eta) h_{j,k} + \eta h_{j,k+1} \right],$$

where $k = k(j,x) = \left[\frac{x \cdot \theta_j}{\Delta s} \right]$, $\eta = \eta(j,x) = \frac{x \cdot \theta_j}{\Delta s} - \left[\frac{x \cdot \theta_j}{\Delta s} \right]$, and $[a]$ denotes the integer part of a.

5.3 Fourier methods

A second family of reconstruction methods seeks a numerical implementation of the relationship given by the projection-slice theorem (Theorem 3.6). To better understand the previously mentioned *gridding method*, let us first examine the algorithm obtained by directly applying the theorem. Using it, we can compute $\hat{f}$ on a grid in polar coordinates with the data $g_{j,l} = g(\theta_j, s_l)$. However, Shannon's formula requires knowing $f(x)$ on a Cartesian grid. This leads to approximating the Cartesian grid with a polar grid, which introduces errors that make the algorithm unreliable. Therefore, the direct approach is not used in applications.

In the second instance, we will see how the *gridding method* solves this issue by multiplying f by an appropriate weight W, so that the effective computation of $\widehat{(W \cdot f)}$ on a Cartesian grid avoids the errors of the direct method.

5.3.1 Direct Fourier method

Given that f is (essentially) band-limited to Ω, Shannon's Theorem states that it is determined, except for a negligible error, by the values on a grid $f_m := f(m/2\Omega)$, $m \in \mathbb{Z}^2$. Since f is supported in $B(0,1)$, in fact, it suffices to consider $|m| \leq 2\Omega$.

Using the fact that $\{e^{i\pi k \cdot x}\}_{k \in \mathbb{Z}^2}$ is an orthogonal basis for $L^2[-1,1]$, and taking into account that $\hat{f}(k)$ is negligible if $|k| > \Omega$, we obtain the following expression:

$$f_m = f\left(\frac{m}{2\Omega}\right) = \sum_{k \in \mathbb{Z}^2} \hat{f}(k) e^{i\pi k \cdot \frac{m}{2\Omega}} = \sum_{|k| \leq \Omega} \hat{f}(k) e^{i\frac{\pi}{2\Omega} k \cdot m}. \tag{5.6}$$

Thus, we need to approximate the coefficients $\hat{f}(k)$, with $k \in \mathbb{Z}^2$ and $|k| \leq \Omega$. With the data and the projection-slice theorem, we can numerically compute:

$$\hat{f}(N\theta_j) = \hat{g}(\theta_j, N), \tag{5.7}$$

where $N = -Q, \ldots, Q$ and Q is such that $N\theta_j \subset [-\Omega, \Omega]$. We can obtain the values $\hat{g}(\theta_j, N)$ from the data $g_{j,l} := g(\theta_j, s_l)$ with a Riemann sum of step size $1/q$:

$$\hat{g}(\theta_j, N) = \frac{1}{q} \sum_{l=-q}^{q} e^{-2\pi i N l/q} g_{j,l} \quad (j = 0, \ldots, p-1; N = -Q, \ldots, Q).$$

Once we have these values, for each $k \in \mathbb{Z}^2$, $|k| \leq \Omega$, we can take N and θ_j such that $N\theta_j$ is closest to k, so that the value of (5.7) gives an approximation of $\hat{f}(k)$. This is the so-called *nearest neighbor approximation*, which is exactly what causes the problems

in this algorithm. Finally, with the values $\hat{f}(k)$, we can proceed to (5.6) and approximate f_m.

The main theoretical advantage of this method is its computational efficiency. The discrete Fourier transforms in (5.6) and (5.7), if computed with an FFT (Fast Fourier Transform), require a number of operations of the order $\Omega^2 \log \Omega$, which is smaller than the Ω^3 order obtained with the backprojection.

Unfortunately, as we have pointed out, the approximation of the Cartesian grid by the polar grid results in sparse images and serious distortions in practice, which makes the method not very reliable. Moreover, these distortions do not decrease by taking more data: in the polar grid (θ_j, N), the step in the second variable is constant, regardless of the p and q of the data, while almost all numerical approximation methods require a reduction of the step size to decrease the error.

5.3.2 Gridding method

Unlike the previous method, this algorithm is used in many practical applications (including medical imaging) since it's the most accurate Fourier reconstruction method. Its main feature is the use of a weight function W, close to 1 in the reconstruction region, $B(0,1)$, vanishing when $|x| \geq a > 1$, and with Fourier transform concentrated near 0. Then, the product $W \cdot f$ is close to f in the reconstruction region, and we can therefore approximate f as soon as we can approximate the Fourier transform:

$$\begin{aligned}
\widehat{(W \cdot f)}(\zeta) = (\widehat{W} * \hat{f})(\zeta) &= \int_{\mathbb{R}^2} \widehat{W}(\zeta - \eta)\hat{f}(\eta)\, d\eta \\
&= \int_0^\infty \int_{S^1} \widehat{W}(\zeta - \sigma\theta)\hat{f}(\sigma\theta)\, \sigma\, d\theta\, d\sigma \\
&= \int_0^\infty \int_{S^1} \widehat{W}(\zeta - \sigma\theta)\hat{g}(\theta, \sigma)\, \sigma\, d\theta\, d\sigma, \qquad (5.8)
\end{aligned}$$

where, as we already know, the last identity follows from the projection-slice theorem (Theorem 3.6). The name of the method comes form the left-hand side of equation (5.8), which is evaluated on a Cartesian grid. With an inverse 2-dimensional FFT, we can obtain $W \cdot f$, and finally, by dividing by W, recover f.

Since $\widehat{(W \cdot f)}$ has bandwidth 1 (f is supported in $B(0,1)$), in order to apply Shannon's theorem, we need to consider a lattice with stepsize $1/2$. Thus, for $k \in \mathbb{Z}^2$, we want to evaluate

$$\widehat{(W \cdot f)}(k/2) = \int_0^\infty \int_{S^1} \widehat{W}(k/2 - \sigma\theta)\hat{g}(\theta, \sigma)\, \sigma\, d\theta\, d\sigma. \qquad (5.9)$$

The success of this method resides in an appropriate discretization of these integrals. The values θ_j in this discretization are determined by the data $g_{j,l} = g(\theta_j, s_l)$, since the Fourier transform $\hat{g}$ does not affect the first variable.

As a first approach, we could try to apply the famous numerical integration method of the trapezoidal rule, with stepsizes $\Delta\varphi = \pi/p$ in the angle and $\Delta\sigma$ in the radial variable, which would yield the approximation

$$\widehat{(W \cdot f)}(k/2) = \sum_{l=0}^{\infty} \sum_{j=0}^{p-1} \widehat{W}\big(k/2 - l\Delta\sigma\theta_j\big)\hat{g}(\theta_j, l\Delta\sigma)(l\Delta\sigma)\frac{\pi}{p}\Delta\sigma. \qquad (5.10)$$

However, in the derivation of this formula, we ignore the fact that the integral is defined only on $[0, \infty)$, and not on the whole $\mathbb{R}$. This results in the annihilation of all the information stored in the cell represented by $l = 0$ (the information given by $\hat{g}(\theta_j, 0)$). This produces artifacts and must be corrected conducting a more delicate approach.

Let us define

$$G(\sigma) = \int_{S^1} \widehat{W}(\zeta - \sigma\theta)\hat{g}(\theta, \sigma)\, d\theta.$$

Then (5.9) becomes:

$$\widehat{(W \cdot f)}(k/2) = \int_0^\infty \sigma G(\sigma)\, d\sigma.$$

We will apply a general formula to the function $g(\sigma) := \sigma G(\sigma)$ in order to obtain the discretization of (5.9).

First, let us denote $\tilde{f} = \mathcal{F}^{-1} f$, the inverse Fourier transform of f, and start with the classical *Poisson's summation formula*, for $f \in S(\mathbb{R})$:

$$\sum_{l \in \mathbb{Z}} f(l\Delta\sigma) = \frac{1}{\Delta\sigma} \sum_{l \in \mathbb{Z}} \tilde{f}\left(\frac{l}{\Delta\sigma}\right).$$

In the second sum, separate the term corresponding to $l = 0$, which is

$$\frac{1}{\Delta\sigma}\tilde{f}(0) = \frac{1}{\Delta\sigma}\int_{\mathbb{R}} f(\sigma)\, d\sigma,$$

so that

$$\int_{\mathbb{R}} f(\sigma)\, d\sigma = \Delta\sigma \sum_{l \in \mathbb{Z}} f(l\Delta\sigma) - \sum_{l \in \mathbb{Z}\setminus\{0\}} \tilde{f}\left(\frac{l}{\Delta\sigma}\right). \tag{5.11}$$

This general formula will be used for $f(\sigma) = g(\sigma)\operatorname{sgn}(\sigma)$, with g such that $g(0) = 0$. As pointed out before in Proposition 4.4, the Hilbert transform of f satisfies the relationship $\widehat{Hf}(\zeta) = \operatorname{sgn}(\zeta)\hat{f}(\zeta)$. Therefore,

$$\widehat{H\tilde{g}}(\zeta) = \operatorname{sgn}(\zeta)\widehat{\tilde{g}}(\zeta) = \widehat{\tilde{f}}(\zeta).$$

Hence, $\tilde{f} = H\tilde{g}$. Then (5.11) yields:

$$\int_{\mathbb{R}} g(\sigma)\operatorname{sgn}(\sigma)\, d\sigma = \Delta\sigma \sum_{l \in \mathbb{Z}} g(l\Delta\sigma)\operatorname{sgn}(l\Delta\sigma) + r,$$

where

$$\begin{aligned} r &= -\sum_{l \in \mathbb{Z}\setminus\{0\}} (H\tilde{g})\left(\frac{l}{\Delta\sigma}\right) = -\frac{i}{\pi} \sum_{l \in \mathbb{Z}\setminus\{0\}} \int_{\mathbb{R}} \frac{\tilde{g}(s)}{l/\Delta\sigma - s}\, ds \\ &= -\frac{i}{\pi} \sum_{l=1}^{\infty} \left(\int_{\mathbb{R}} \frac{\tilde{g}(s)}{l/\Delta\sigma - s}\, ds + \int_{\mathbb{R}} \frac{\tilde{g}(s)}{-l/\Delta\sigma - s}\, ds \right) \\ &= -\frac{i}{\pi} \sum_{l=1}^{\infty} \int_{\mathbb{R}} \frac{2s\tilde{g}(s)}{(l/\Delta\sigma)^2 - s^2}\, ds \end{aligned}$$

Finally, using that $s\tilde{g}(s) = -\frac{1}{2\pi i}\tilde{g}'(s)$, we obtain:

$$r = \frac{1}{\pi^2}\sum_{l=1}^{\infty}\int_{\mathbb{R}}\frac{\tilde{g}'(s)}{(l/\Delta\sigma)^2 - s^2}\,ds.$$

If g has bandwidth A and $\Delta\sigma \ll 1/A$, the factor s^2 in the denominator is negligible compared to $(l/\Delta\sigma)^2$, and we have approximately:

$$r \simeq \frac{1}{\pi^2}(\Delta\sigma)^2\sum_{l=1}^{\infty}\frac{1}{l^2}\int_{-A}^{A}\tilde{g}'(s)\,ds = \frac{1}{6}(\Delta\sigma)^2 g'(0),$$

where we have used the standard result $\sum_{l=1}^{\infty} 1/l^2 = \pi^2/6$. Hence, we obtain:

$$\int_{\mathbb{R}} g(\sigma)\,\mathrm{sgn}(\sigma)\,d\sigma = \Delta\sigma\sum_{l\in\mathbb{Z}} g(l\Delta\sigma)\,\mathrm{sgn}(l\Delta\sigma) + \frac{1}{6}(\Delta\sigma)^2 g'(0).$$

Using this formula for $g(\sigma) = \sigma G(\sigma)$. Since $g(0) = 0$, $g'(0) = 1\cdot G(0) + 0\cdot G'(0) = G(0)$ and G is even:

$$\int_0^{\infty}\sigma G(\sigma)\,d\sigma = \Delta\sigma\sum_{l=1}^{\infty} l\Delta\sigma G(l\Delta\sigma) + \frac{\Delta\sigma^2}{12}G(0).$$

Note that the sum is on positive integers. This can be written in a unified form as:

$$\int_0^{\infty}\sigma G(\sigma)\,d\sigma = \Delta\sigma\sum_{l=0}^{\infty}\sigma_l G(l\Delta\sigma), \quad \text{where } \sigma_l = \begin{cases} l\Delta\sigma, & \text{if } l > 0,\\ \frac{\Delta\sigma}{12}, & \text{if } l = 0.\end{cases}$$

In our case, we have:

$$\int_0^{\infty}\int_{\mathbb{S}^1}\widehat{W}(\zeta - \sigma\theta)\hat{g}(\theta,\sigma)\sigma\,d\theta\,d\sigma = \Delta\sigma\sum_{l=0}^{\infty}\sigma_l\int_{\mathbb{S}^1}\widehat{W}(\zeta - l\Delta\sigma\theta)\hat{g}(\theta, l\Delta\sigma)\,d\theta.$$

Finally, this yields the discretization of (5.8) using the Riemann sum with step size π/p for the angular integral:

$$\begin{aligned}\widehat{(W\cdot f)}(k/2) &= \Delta\sigma\sum_{l=0}^{\infty}\sigma_l\int_{S^1}\widehat{W}\big(k/2 - l\Delta\sigma\theta\big)\hat{g}(\theta, l\Delta\sigma)\,d\theta\\ &= \Delta\sigma\sum_{l=0}^{\infty}\sigma_l\sum_{j=0}^{2p-1}\widehat{W}\big(k/2 - l\Delta\sigma\theta_j\big)\hat{g}(\theta_j, l\Delta\sigma)\frac{\pi}{p}.\end{aligned}$$

Notice that this discretization differs from that in (5.10) only in the term corresponding to $l = 0$, which corrects the error derived from applying the trapezoidal rule on a function defined in $[0,\infty)$.

Before showing the final algorithm for the gridding method, we need to find a good stepsize for $\Delta\sigma$. In order to do this, we need to determine the bandwidth of the integrand in (5.8) as a function of σ. Let us see that the function $\sigma \mapsto \sigma\widehat{W}(\zeta - \sigma\theta)\hat{g}(\sigma\theta)$ has bandwidth $a + 1$.

We already know that $\hat{g}(\theta,\sigma)$ has bandwidth 1 and W is supported in $[-a,a]$, $a>1$. Now we need to prove that $\sigma \mapsto \widehat{W}(\zeta-\sigma\theta)$ has bandwidth a. By the distributional identity

$$\delta(x\cdot\theta-s)=\int_{\mathbb{R}}\hat{\delta}(\sigma)e^{2\pi i(x\cdot\theta-s)\sigma}d\sigma=\int_{\mathbb{R}}e^{2\pi i(x\cdot\theta-s)\sigma}d\sigma,$$

we see that, for $|s|>a$,

$$\begin{aligned}\int_{\mathbb{R}}\widehat{W}(\zeta-\sigma\theta)e^{-2\pi is\sigma}d\sigma &= \int_{\mathbb{R}}\left(\int_{|x|\le a}W(x)e^{-2\pi ix\cdot(\zeta-\sigma\theta)}\,dx\right)e^{-2\pi is\sigma}\,d\sigma\\ &= \int_{|x|\le a}W(x)e^{-2\pi ix\cdot\zeta}\left(\int_{\mathbb{R}}e^{2\pi i(x\cdot\theta-s)\sigma}d\sigma\right)dx\\ &= \int_{|x|\le a}W(x)e^{-2\pi ix\cdot\zeta}\delta(x\cdot\theta-s)dx=0.\end{aligned}$$

Therefore, we should choose a stepsize $\Delta\sigma \ll 1/(a+1)$.

Final algorithm

Step 1: Discretization of $\hat{g}(\theta_j, N\Delta\sigma)$, with $\Delta\sigma \ll 1/(a+1)$ and $N=-Q,\dots,Q$, where Q is such that $N\Delta\sigma\subset[-\Omega,\Omega]$ (i.e., $N\le\Omega/\Delta\sigma$):

$$\hat{g}_{j,N}=\frac{1}{q}\sum_{l=-q}^{q}e^{-2\pi iN\Delta\sigma l/q}g_{j,l}\quad(j=0,\dots,p-1;N=-Q,\dots,Q).$$

Step 2: For every $k\in\mathbb{Z}^2$, $|k|\le q$, compute the discretization of $(\widehat{W\cdot f})(k/2)$:

$$z_k=\frac{\pi}{p}\Delta\sigma\sum_{N=0}^{Q}\sum_{j=0}^{p-1}\sigma_N\widehat{W}(k/2-N\Delta\sigma\theta_j)\hat{g}_{j,N}.$$

Since $\widehat{W}$ decays rapidly, only a few terms are relevant in this sum (those corresponding to the values with $|k/2-N\Delta\sigma\theta_j|\le C$, where C is a constant depending on the choice of the weight W).

Step 3: Compute an approximation of $f(m/2\Omega)$, $m\in\mathbb{Z}^2$, $|m|\le 2\Omega$, by taking the inverse Fourier transform of $(\widehat{W\cdot f})$ and dividing by W:

$$f_m=\frac{1}{W\left(\frac{m}{2\Omega}\right)}\sum_{|k|\le\Omega}e^{2\pi ik\cdot\frac{m}{2\Omega}}z_k.$$

As often happens in this area, the choice of W is justified by empirical reasons, rather than by theoretical studies.

5.4 Reconstruction of the Shepp-Logan Phantom

The most widely used technique for comparing algorithms has been to compare the reconstructions when applied to data taken from *phantoms*: this means taking data from an object (physical or mathematical) of known structure instead of a human subject. This is useful because we know what the true object is. Additionally, for mathematical phantoms,

there is no measurement error, so any errors in the reconstruction are due to the algorithm, which is appropriate for comparison.

In 1974, Larry Shepp and Benjamin F. Logan created what later would become a standard test image, the Shepp-Logan phantom. They simulated a phantom of a cross-section of a human brain with different tumors and a blood clot in order to test the Hounsfield (who shared the 1979 Nobel prize in Medicine with Allan Cormack) algorithm: one of the earliest practical implementations of CT image reconstruction that paved the way for advancements like the filtered backprojection.

Here, we will use this standard test image to test the implementation of the two algorithms previously presented: the *filtered backprojection* and the *gridding method*. In particular, for the *filtered backprojection*, we will compare the results of applying the three presented filters: the *Ram-Lak* filter, the *cosine* filter and, of course, the *Shepp-Logan* filter. Before that, however, let us describe the phantom.

5.4.1 The Shepp-Logan Phantom

The figure below shows the original representation of the Shepp-Logan phantom from the article from 1974 [10] (left) and a computer generated representation (right).

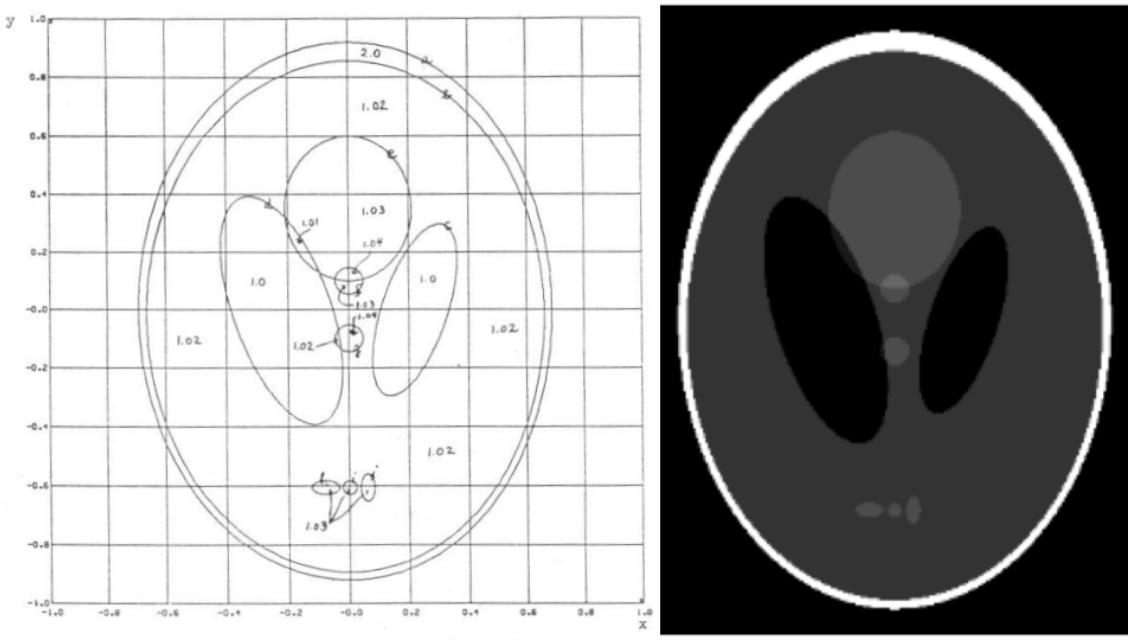

Figure 9: Sketch of the Shepp-Logan phantom (left), extracted from [10], and its computer-generated representation (right). The larger, nearly concentric ellipses represent the cross-section of a human skull containing gray matter, 2 ventricles, 5 tumors and a blood clot.

The density, or the attenuation coefficient f, is represented in a gray scale on the computer generated image: the darker areas correspond to lower densities and the lighter areas correspond to higher densities. The original figure represents the density with numbers, which are associated to letters that correspond to different substances found in the brain.

Besides using ellipses to create the phantom, Shepp and Logan attempted to be consistent with known facts about the human head. In that sense, the skull is thicker at the forehead

(the top of the figures) and is about twice as dense as the interior brain tissue. The ventricles (1.0 at d and c) are filled with spinal fluid, which is basically water, hence they are the least dense tissue. The gray matter (1.02) fills the interior of the head except for tumors (1.03 at h, i, j, g, and f) and the blood clot (1.03 at e). Note that the intersections between different substances have their mean density.

Finally, we want to comment that Shepp-Logan model supposed a water bag to surround the head to reduce the range of radiation intensities registered, which has the effect (after subtraction) of reducing each density by unity. Hence the skull corresponds to $f = 1$ and the water areas correspond to $f = 0$.

5.4.2 Implementing the filtered backprojection

In order to apply the filtered backprojection algorithm, we first need to choose the parameters of the data obtained by parallel standard geometry. In particular, the value of p (the number of angles) and q (which determines $2q+1$, the number of lines for each direction).

On the one hand, since, for a given sample line, rotating the angle by 180° doesn't change the sample line, we can take $p = 180$ to have a set of $2q + 1$ samples for each degree. On the other hand, since the computer generated image has a size of 256×256 pixels, we can take $q = 128$ so that every pixel is covered by a line.

One could have the idea to increase the values of the parameters in order to get a higher resolution image. However, as Shepp and Logan mentioned in [10], "*it appears that further increasing n and m* (analogous to our p and q), *once large enough, does not produce large changes in the reconstruction inside the skull*".

For these values and the three different filters (see Figure 8), we obtained the following results:

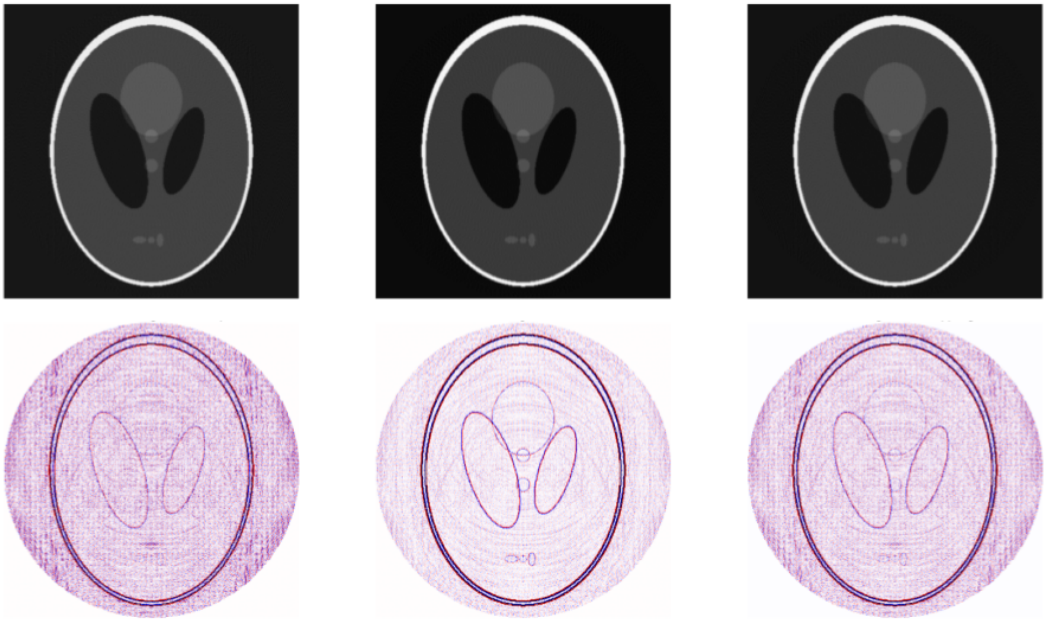

Figure 10: Reconstructions of the Shepp-Logan phantom using the *Ram-Lak* filter (left), the *cosine* filter (middle), and the *Shepp-Logan* filter (right). The difference between the reconstructed and the original image is plotted on the bottom figures. The images were obtained using the code in Appendix.

At first glance, all three reconstructed images (the top figures) might look identical. However, we can distinguish them by plotting the difference between the reconstructed and the original image (the bottom figures). The blue color indicates areas where f in the reconstructed image is lower, up to 0.1, than the original one. Analogously, the red pixels correspond to higher densities, up to 0.1, in the reconstructed image.

We observe that among the three filters, the *Ram-Lak* filter is the one that differs the most from the original in the non-edge areas, while the *cosine* filter provides an image that is the closest to the original in those regions. However, at the edges, the *cosine* filter shows the largest deviation from the original, whereas the *Ram-Lak* filter is the most consistent in preserving edge details.

In particular, in the bottom left figure, we can't distinguish all the tumors' edges. This was expected, as edges are associated with abrupt changes in intensity, corresponding to high-frequency components of the image. These components are better preserved by the *Ram-Lak* filter but at the cost of increased noise.

On the other hand, the *cosine* filter suppress high frequencies, resulting in less pronounced edges but also less noise. The *Shepp-Logan* filter represents a middle ground.

5.4.3 Implementing the gridding method

Let us reconstruct the Shepp-Logan algorithm using the gridding method. Since we want to compare it with the filtered backprojection, let us take the same filter for both methods. For the sake of simplicity, we will only present the results of applying the *Ram-Lak* filter, as the three filters produce nearly indistinguishable images unless further computations are performed, which we have already carried out.

Additionally, for the sake of comparison, we take the same values for the sampling parameters: $p = 180$ and $q = 128$, which correspond to taking 256 parallel samples for every degree. We map the sampling points onto a grid in the Fourier domain and then perform an inverse 2-dimensional FFT. The reconstructed image is shown in Figure 11 (Left).

The result is evidently unsatisfactory, as significant artifacts emerge. Similarly as the grid pattern that emerged in the scan conducted in 1968 by Hounsfield on a preserved brain (see Figure 3), we observe an artifact superimposed on the expected structure of the Shepp-Logan phantom, which is barely distinguishable. Moreover, the entire image exhibits an increase in density, particularly along the edges, where not only $f \neq 0$, but an additional shape appears that is not part of the original phantom.

In an attempt to get a better reconstruction, we can oversample the image taking the double of samples per degree ($q = 256$). As we can see in Figure 11 (Right), most of the previous artifacts don't appear anymore. In particular, the grid-like pattern has been corrected, as well as the edges, which now have density $f \approx 0$.

However, the overall reconstructed image still shows an increase in density compared to the original. This is particularly problematic, as the little ellipses, which represent tumors (see Figure 9), are hardly distinguishable from the gray matter! On the other hand, the blood clot is clearly identified, meaning a hypothetical doctor examining this case could prevent a stroke, but might miss some tumors.

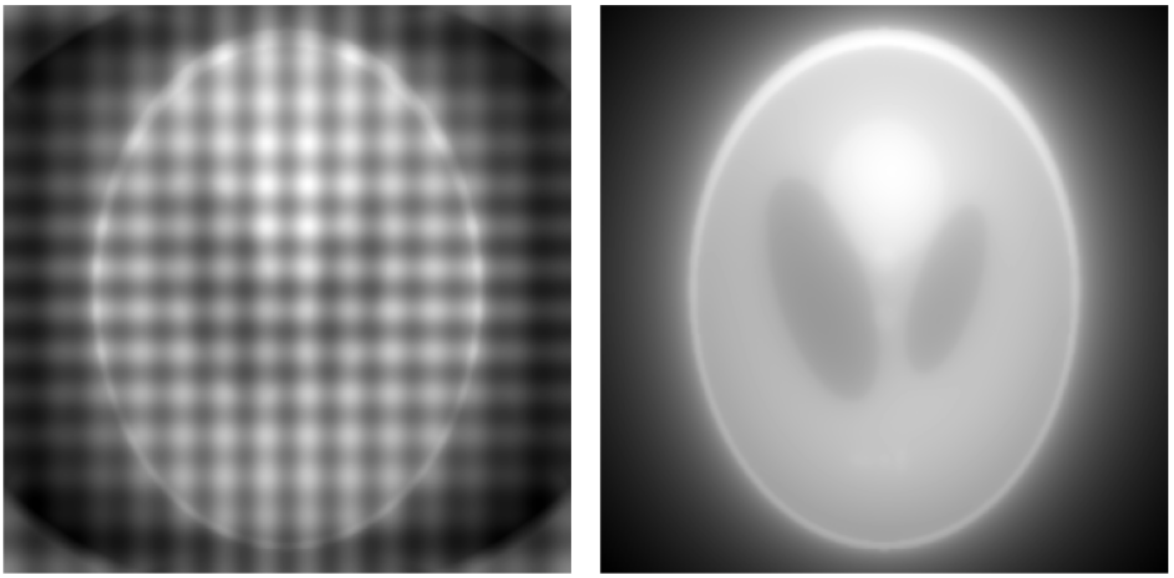

Figure 11: Comparison of Shepp-Logan phantom reconstructions using the gridding method with the Ram-Lak filter applied as the weighting function W. Left: Reconstruction with $p = 180$ and $q = 128$. Right: Reconstruction with $p = 180$ and $q = 256$. The images were obtained using the code in Appendix.

As our simple example indicates, the filtered backprojection has been for decades the golden standard in CT scan and other applications. It has at least two advantages over interpolation schemes in the frequency domain. First, the reconstruction procedure can start as soon as the first projection is measured, which speeds up the process and reduces the amount of data stored at any time. Second, from a numerical perspective, interpolation in the spatial domain is usually simpler (linear interpolation is often sufficient) compared to the more sophisticated interpolation methods required in the Fourier domain. Therefore, the reconstructed images are not only more accurate, but also faster obtained. However, gridding-based methods are also used in some situations.

5.5 Alternative geometries

5.5.1 Fan beam geometry

The parallel beam geometry has been the basis of our analysis throughout. The setup corresponding to this geometry involves a scanning machine that sends out a set of parallel X-ray beams at each selected angle θ_j and records the corresponding values of the Radon transform. Such a machine would need a strip of distinct transmitters spaced at an appropriate sample spacing and capable of rotating as a single unit during the scanning process.

In practice, however, it is easier (and cheaper) to design a machine with a single X-ray beam transmitter that emits a fan of beams. An arc of detectors on the opposite side measures the values of the Radon transform along the lines corresponding to the beams in the fan.

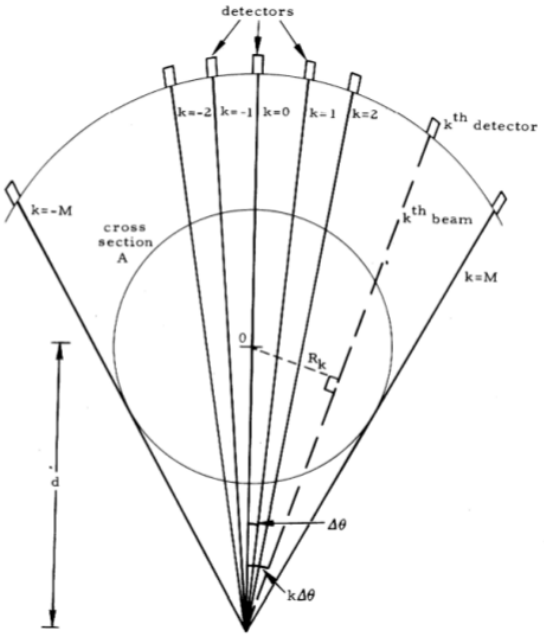

Figure 12: Fan beam scanning geometry, extracted from [12].

There are a total of $2M+1$ beams separated by equal angular intervals $\Delta\theta$. The equipment is rotated around the cross-section at equal angular intervals $\Delta\beta = \pi/N$, where N is the number of angular positions covering half a rotation.

To ensure a complete sampling of the cross-section, N rotations are required for the central beam. Additionally, since the off-centered beams also need to cover all directions, the system needs to account for an additional $2M$ rotations to sample all beam directions. Hence, the total number of rotations is $N+2M$. Then, the set of data for the j^{th} rotation is given by

$$\sum_{k=-M}^{M} \mathcal{R}f(j\Delta\beta - k\Delta\theta, R_k), \quad j = 0, \ldots, N+2M-1 \tag{5.12}$$

where $R_k = |d\sin(k\Delta\theta)|$ is the distance of the k^{th} beam from the origin O, with d being the distance of the source from the origin (or approximately the radius of the scanner). If the scanner is set up so that $\Delta\beta = \Delta\theta$, then (5.12) becomes

$$\sum_{h=-M}^{M} \mathcal{R}f(i\Delta\theta, R_h), \quad i = -k, -k+1, \ldots, N+2M-k-1.$$

Thus, in order to obtain N sets of rearranged projection data as if they were generated from the parallel-beam scan, a minimum number of $N+2M$ rotations for the fan-beam scan will be required.

Indeed, observe that if two beams in the fan make an angle of $k\Delta\theta$ with each other when they are emitted, then, when the transmitter itself has been rotated by the same angle $k\Delta\theta$, and the two beams are emitted again, one of the new beams will be parallel to one of the beams from the previous transmission. The following figure illustrates this:

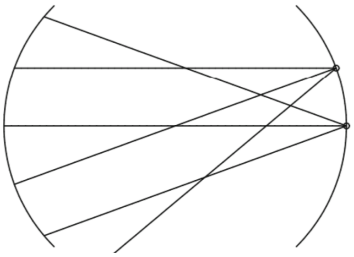

Figure 13: Different fans yield parallel beams, extracted from [3].

In other words, once the scanning process has been completed, it is possible to reorganize the fan beam data into an equivalent collection of parallel beam data. Then, the image reconstruction algorithms previously detailed can be applied to this reorganized data to produce an image.

5.5.2 3D geometries

The parallel and fan beam geometries are inherently two-dimensional. To obtain a 3D CT scan using these geometries, the patient table must be translated after each 2D scan is completed. Hence, when using the early designs of CT scanners, the time to acquire the images is quite long.

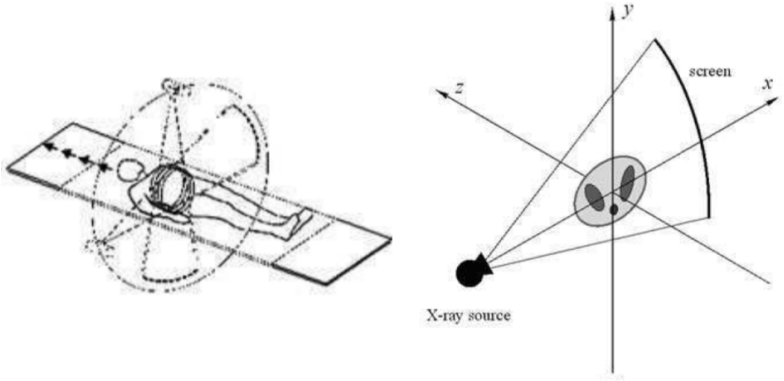

Figure 14: 3D view of a fan beam scanner, extracted from [2].

Besides that, the first scanners faced other limitations. The limited power of the X-ray tube and its constrained rotation speed around the test object required a compromise between several projection parameters: scan resolution, noise levels, and the ability to adapt radiation power to the material being examined.

To overcome these challenges, in the beginning of the 1990s, the *spiral tomography* was developed and since then its use has been increasingly widespread. By using a system that moved in a spiral path, it is possible to collect data for more than one slice simultaneously, hence increasing the scanning resolution and reducing the time needed for the whole examination, as well as the time for each individual scan. This benefits both patients, who are less exposed to radiation, and healthcare providers, who can perform more scans per day.

We can identify three consecutive stages in the development of spiral tomography:

1. *single-slice computed tomography* (SSCT),
2. *multi-slice computed tomography* (MSCT),
3. *cone-beam computed tomography* (CBCT).

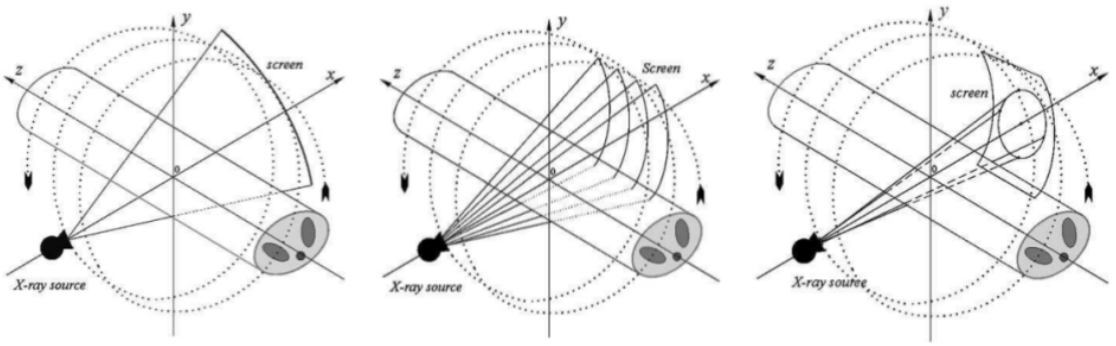

Figure 15: 3D view of SSCT (left), MSCT (center) and CBCT (right) scanners, extracted from [2].

The very first design of spiral tomography device, the SSCT scanner, used an array of X-ray detectors arranged in a single row. As with fan-beam scanners, these detectors were placed on an arc-shaped screen in a plane perpendicular to the axis of the patient, as shown in Figure 15 (left). The main distinguishing feature of the new design, however, was the combination of the rotary motion of the projection system with the continuous forward movement of the patient during the scan. This innovation eliminated the time interval between each individual scan, significantly reducing the time taken for the entire CT examination.

While the SSCT scanners had only one row of detectors in the array, adding additional rows of detectors to the array created the MSCT scanner. By using them, it is possible to acquire projections simultaneously for the subsequent reconstruction of up to four slices.

Despite the procedure for acquiring the projections in MSCT is the same as in SSCT for the most part, MSCT scanners have improved specifications. There is an eightfold increase in the rate of acquisition of the reconstructed images (four times the number of reconstructed slices in half the time). Additionally, the scanning resolution is increased, the level of image noise is reduced, and the power of X-ray tube is more effectively used.

An important factor in the design of the MSCT spiral scanner is the assumption that the individual fan-beams composing the cone-shaped radiation beam are parallel. This assumption initially posed a challenge to increasing the number of rows in the detector array. However, the development of CBCT marked a departure from this constraint, enabling a significant expansion in the width of the detector array.

This breakthrough led to a dramatic increase in scanning speed, making it possible to image organs in physiological motion, such as the heart. Additionally, the reduced spacing between detector rows allowed for a substantial improvement in scan resolution along the z-axis.

Another major advantage of the new projection geometry was the increased effective solid

angle of the X-rays. This improvement greatly enhanced the scanner's efficiency in extracting information from the available radiation energy, minimizing energy loss in the X-ray tube. As a result, the tube current could be increased, which significantly reduced noise levels in the reconstructed images.

With these advancements, spiral scanners have become the prevailing standard in the medical imaging market. For a more comprehensive explanation of these geometries and their associated reconstruction algorithms, we refer the reader to [2].

Conclusion

The basic tools used in CT, the Radon and the Fourier transforms and the backprojection operator, were presented in this work; as well as their properties, which ensure regularity in their use. Using these tools, we constructed inversion formulas, from which we derived reconstruction algorithms. Their accuracy and limitations were shown by reconstructing the Shepp-Logan phantom. Additionally, we explored alternative geometries to highlight the evolving nature of CT.

While this work has addressed many aspects of CT, some theoretical and practical elements of image reconstruction remain open for further exploration. For a deeper mathematical analysis of the Radon transform, we refer the reader to [8]. On the other hand, for those interested in the technical aspects of X-ray CT scanners, such as the physics of X-ray sources and detectors, we recommend consulting [2]. Additionally, readers who wish to explore other imaging techniques, such as Magnetic Resonance Imaging (MRI), may find [3] a good starting point.

Appendix

The following code was used to generate the reconstructed images of the Shepp-Logan phantom using the filtered backprojection algorithm (with the Ram-Lak, cosine, and Shepp-Logan filters) and the gridding method, corresponding to Figures 10 and 11, respectively. The code is also available at this link: https://colab.research.google.com/drive/1naCS6cQI3CH4W02i8vjRei3NBlSgk55l?usp=sharing.

```
import numpy as np
import matplotlib.pyplot as plt
from skimage.transform import radon, iradon
from skimage.data import shepp_logan_phantom

# Plotting the Shepp-Logan phantom
image = shepp_logan_phantom()
plt.figure(figsize=(6, 6))
plt.title("Original Image")
plt.imshow(image, cmap='gray')
plt.colorbar()
plt.show()

# Generating Projections
angles = np.linspace(0., 180., 180, endpoint=False)
sinogram = radon(image, theta=angles)

# Reconstructing using three filters
filters = ['ramp', 'cosine', 'shepp-logan']
reconstructed_images = {}

for filter_type in filters:
    reconstructed_images[filter_type] = iradon(sinogram, theta=
        angles, filter_name=filter_type)

# Plotting reconstructed images and their differences
fig, axs = plt.subplots(2, 3, figsize=(18, 10))

for i, (filter_type, reconstructed_image) in enumerate(
    reconstructed_images.items()):
    axs[0, i].imshow(reconstructed_image, cmap='gray')
    axs[0, i].set_title(f"Reconstructed ({filter_type.capitalize()
        })")
    axs[0, i].axis('off')

    diff = reconstructed_image - image
    axs[1, i].imshow(diff, cmap='seismic', vmin=-0.1, vmax=0.1)
    axs[1, i].set_title(f"Difference (Original vs {filter_type.
        capitalize()})")
    axs[1, i].axis('off')

plt.tight_layout()
plt.show()
```

```python
import numpy as np
import matplotlib.pyplot as plt
from pynufft import NUFFT
from skimage.data import shepp_logan_phantom
from skimage.transform import resize

# Plotting the Shepp-Logan phantom
image = shepp_logan_phantom()
size = 256
image = resize(image, (size, size))

# Generating sampling points
p = 180
q = 128

angles = np.linspace(0, np.pi, p, endpoint=False)
s = np.linspace(-1, 1, 2 * q + 1)

om = np.zeros((p * (2 * q + 1), 2))
for j, angle in enumerate(angles):
    for l, sl in enumerate(s):
        kx = sl * np.cos(angle)
        ky = sl * np.sin(angle)
        om[j * (2 * q + 1) + l] = [kx * np.pi, ky * np.pi]

#Initializing the Non-Uniform Fast Fourier Transform (NUFFT)
nufft_obj = NUFFT()
Kd = (size, size)
Jd = (6, 6)
Nd = (size, size)
nufft_obj.plan(om, Nd, Kd, Jd)

# Performing forward NUFFT to get data
k_space_data = nufft_obj.forward(image)

# Performing adjoint NUFFT to reconstruct the image
reconstructed_image = nufft_obj.adjoint(k_space_data)

# Plotting the reconstructed image
plt.imshow(np.abs(reconstructed_image), cmap='gray')
plt.axis('off')
plt.show()
```

Bibliography

[1] Cascante, C., Massaneda, X., Ortega, J., and Pascuas, D. "Computerized Tomography and the Radon Transform". In: *Boletín de la Sociedad Española de Matemática Aplicada* 41 (2007), pp. 41–74.

[2] Cierniak, R. *X-Ray Computed Tomography in Biomedical Engineering*. Springer, 2011. ISBN: 978-0-85729-027-4.

[3] Feeman, T. G. *The Mathematics of Medical Imaging*. Springer, 2015. ISBN: 978-3-319-22665-1.

[4] Hegerl, R. and Hoppe, W. "Three-Dimensional Structure Determination by Electron Microscopy (Nonperiodic Specimens)". In: *Topics in Current Physics* 13 (1980), pp. 127–185.

[5] Holman, B. L. e. a. "A comparison of brain perfusion SPECT in cocaine abuse and AIDS dementia complex". In: *Journal of Nuclear Medicine* 33.7 (1992), pp. 1312–1315.

[6] Khalfin, L. A. and Klebanov, L. B. "A Solution of the Computer Tomography Paradox and Estimating the Distances Between the Densities of Measures with the Same Marginals". In: *The Annals of Probability* 22.4 (1994), pp. 2235–2241.

[7] Klebanov, L. B. and Rachev, S. T. "Computer Tomography and Quantum Mechanics". In: *Advances in Applied Probability* 29.3 (1997), pp. 595–606.

[8] Natterer, F. *The Mathematics of Computerized Tomography*. Springer, 1986. ISBN: 978-3-519-02103-2.

[9] Pereyra, M. C. and Ward, L. *Harmonic Analysis. From Fourier to Wavelets*. American Mathematical Society, 2012. ISBN: 978-0-8218-7566-7.

[10] Shepp, L. and Logan, B. "The Fourier Reconstruction of a Head Section". In: *Institute of Electrical and Electronics Engineers (IEEE)* 21.3 (1974), pp. 21–43.

[11] Slepian, D. "On Bandwidth". In: *Society for Industrial and Applied Mathematics (SIAM)* 64.3 (1976), pp. 292–300.

[12] Wang, L. "Cross-section reconstruction with a fan-beam scanning geometry". In: *Institute of Electrical and Electronics Engineers (IEEE)* 26.3 (1977), pp. 264–268.