UNIVERSITÀ DEGLI STUDI DI GENOVA

SCUOLA DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Corso di Laurea Magistrale in Matematica

Reconstruction of images from limited-angle computed tomography data

Candidata:

Veronica Raffetto

Relatore:

Prof. Giovanni Alberti

Correlatore:

Prof. Matteo Santacesaria

Anno accademico 2023/2024

Contents

Introduction

This thesis investigates algorithms for reconstructing images from limited-angle computed tomography data. In general, computed tomography is a widely used X-ray imaging technique, particularly in medical diagnostics. The method is based on measuring X-ray attenuation as the rays pass through the considered object from different angles, since this provides information about the internal density of the object. Indeed, the denser the material at a certain location, the more the X-rays are absorbed. As the object is exposed to X-rays from multiple directions, this attenuation data is collected and combined into a sinogram, which contains crucial information about the internal density distribution. However, the sinogram itself is not directly interpretable. Therefore, the goal is to reconstruct the internal structure of the object, essentially recovering its density, from the sinogram data.

The mathematical model for the full-data computed tomography reconstruction problem can be expressed as

$$\mathcal{R}f = y,$$

where $\mathcal{R}$ represents the Radon transform, whose main properties are presented in Section 2.2 (for an extensive treatment, see [29]), f denotes the density function of the object to be reconstructed and y corresponds to the measured data. When dealing with an inverse problem, it is crucial to determine whether the problem is well-posed in the sense of Hadamard. A problem is well-posed if it satisfies three conditions: the existence of a solution, the uniqueness of the solution, and the continuous dependence of the solution on the data. The third condition is particularly important, as it ensures that errors or noise in the measured data do not become excessively amplified, which is essential for achieving reliable and stable reconstructions.

It can be demonstrated that in the inverse problem of computed tomography, this stability condition is not satisfied, making the problem ill-posed. However, it is important to note that ill-posedness can vary in degree. Some problems exhibit mild instability, where data errors lead to controllable distortions, while others are severely ill-posed, with even minimal noise resulting in significant errors that are difficult to manage. Given a compact operator between Hilbert spaces, it is

possible to perform a singular value decomposition of the operator. In cases where the inverse problem is described by such an operator, the degree of ill-posedness can be inferred from the rate of decay of its singular values. Specifically, considering the Radon transform with domain $L^2(\mathbb{D})$, where $\mathbb{D} \subseteq \mathbb{R}^2$ is the unit disk, and codomain $L^2([0, 2\pi] \times [-1, 1])$, it can be shown that the Radon transform is a compact operator. Furthermore, it is demonstrated that its singular values decay at a polynomial rate, indicating that the problem is moderately ill-posed. Consequently, reliable reconstruction is still possible using appropriate regularization techniques, which balance fidelity to the measured data with the stability of the solution.

However, in some tomographic applications, such as digital breast tomosynthesis and dental tomography, full-angle data acquisition is not feasible, worsening the ill-posedness of the problem. In fact, it has been shown through the behavior of singular values that in the limited-angle case, the reconstruction problem becomes severely ill-conditioned. As a consequence, the problem becomes highly sensitive to noise and modeling errors, and classical regularization methods do not perform well. In particular, the missing angular ranges result in a substantial loss of information, making it challenging to accurately reconstruct pecific features of the image. These features correspond to singularities of f along certain curves, which are referred to as the *invisible boundaries*. It thus becomes crucial to understand how these boundaries relate to lines in the data set. This relationship is formally defined by the visibility principle, derived from microlocal analysis, as presented in Quinto's paper [30].

In certain medical cases, it may be sufficient to focus on slices of the reconstruction where the stable part of the boundaries captures clinically important tissue boundaries. However, developing methods that can more effectively recover the missing parts of the boundaries would significantly enhance reconstruction quality and open up new possibilities for limited-angle tomography. This challenge is particularly relevant in practical applications where acquiring full-angle data may be limited by physical, safety, or cost constraints. A promising alternative to classical regularization methods is the use of data-driven approaches, particularly deep learning techniques (see [1]). These methods do not rely on explicit data models but instead use large amounts of training data to learn implicit patterns. Recently, deep neural networks have gained popularity in solving inverse problems, often delivering excellent results. However, a significant drawback is that these networks often function as black boxes, offering little insight or control over the reconstruction process. This lack of transparency is particularly concerning in medical applications, where both reliability and interpretability are essential.

A possible solution to this problem is presented in [4], where the authors developed a hybrid reconstruction framework that combines model-based sparse regularization using shearlets with data-driven deep learning. In particular, shearlets are representation systems for multivariate data that allow for a decomposition of

images across multiple scales and directions, with the resulting shearlet coefficients capturing detailed information about edges.

The algorithm, based on the visibility principle, learns only the part that model-based methods likely cannot handle (the "invisible" part), while applying the theoretically controllable sparse regularization to the remaining parts (the "visible" part). Specifically, the algorithm follows three main steps. In the first step, an initial image reconstruction is performed using the ℓ^1 norm of the shearlet coefficients as a regularizer. This initial reconstruction yields good results for the visible portions of the image, but results in poor reconstruction of the invisible boundaries. Consequently, the shearlet coefficients aligned with the invisible edges are negligible and must be determined by other means. To address this, in the second step, neural networks are employed to learn these missing coefficients. Finally, the algorithm synthesizes the final image in the third step by combining the shearlet coefficients from the visible part, obtained in the first step, with those of the invisible part, reconstructed through neural networks in the second step. Therefore, this method strikes a balance between using neural networks to estimate the invisible components and ensuring the reliability of visible boundaries. The numerical results confirm that this combination is effective, as it yields high-quality reconstructions.

Ideally, the goal would be to achieve such good results using solely classical methods, without involving neural networks, as this could provide greater control over the reconstruction process. In this regard, prior work has already been done. Specifically, the authors in [12] propose a new variational regularization framework that integrates various regularizers, implemented through a combined curvelet-TV methodology.

In this thesis, we develop an algorithm inspired by the method in [4], where the use of neural networks is eliminated in the second step. Specifically rather than learning the invisible shearlet coefficients, the approach focuses on reconstructing them using classical regularization techniques. In particular, total variation of the image and the ℓ^1 norm for the invisible shearlet coefficients are incorporated as priors to infer information about the missing parts of the image. However, the numerical results obtained are not of the same quality as those achieved with the aid of neural networks, as no significant improvement is observed between the reconstruction from the first and second steps in the numerical tests. Additionally, a variant of the developed algorithm is considered, but the results from this alternative approach do not offer substantial improvements either.

Thus, this study highlights the challenges of identifying effective classical techniques, as they often do not achieve the same high-quality reconstruction results that neural networks can provide.

The thesis is organized into two parts: the first part introduces the theoretical concepts essential for understanding the problem, while the second part examines various algorithms to address it.

In Chapter 1, the fundamental concepts of distributions and Sobolev spaces are introduced. These tools are crucial, as they will be used throughout the thesis, particularly in the study of microlocal analysis, which is a rigorous theory that allows for determining the relationship between the singularities of a distribution and those of its Radon transform. This is particularly relevant in our study, since the edges of an image correspond to the singularities of the associated function.

Chapter 2 provides an introduction to computed tomography. While this is not the central focus of the thesis, it is important to offer an overview to properly understand the limited-angle tomography case. In this chapter, the mathematical model representing the inverse problem is analyzed, with particular emphasis on the Radon transform.

In Chapter 3, the focus shifts to the problem of limited-angle tomography, which, as previously mentioned, is severely ill-posed. The missing angular data prevents the accurate reconstruction of certain image features, particularly the boundaries where tangential directions are missing from the dataset. This phenomenon is formalized in the visibility principle introduced in Section 3.2, which derives from Quinto's Theorem 3.9, an important result in microlocal analysis. These insights form the basis of the method developed in [4]. Another key tool used in [4] are shearlets, which decompose the image while accounting for its anisotropic features. Shearlets are introduced in Chapter 4, along with an overview of the theory of frames and wavelets, which provide their theoretical foundation.

Following this theoretical groundwork, the final chapter of the thesis explores reconstruction algorithms specifically designed for limited-angle tomography. In particular, the method proposed in [4] is discussed, which leverages the decomposition strategy mentioned above. The final section presents the algorithm developed in this thesis, along with its variant, and the corresponding results from the numerical tests.

Chapter 1

Distributions and Sobolev Spaces

This chapter introduces key concepts of distributions and Sobolev spaces, necessary for understanding microlocal analysis, which plays an important role in examining the reconstruction problem in limited-angle tomography. For a more extensive treatment, readers can refer to the detailed discussions presented in the books [26], [7] and [8].

The traditional definition of a function refers to a mapping from the domain to the codomain. In this discussion, we examine functions defined over the d-dimensional Euclidean space $\mathbb{R}^d$, hence they can be regarded as maps that assign to each point of $\mathbb{R}^d$ a numerical value.

In certain mathematical contexts, it is advantageous to consider objects that share similarities with functions but may not adhere strictly to the standard definition of a function. For this reason, the theory of distributions has been developed to convey a generalization of functions.

This concept finds resonance within the theory concerning L^p spaces. Indeed, L^p spaces are essentially defined up to sets of measure zero, hence two functions that agree almost everywhere (i.e., they differ only on a set of measure zero) are considered equivalent in terms of their representation in L^p spaces. Therefore, the behavior of a function f as an element of the space is primarily determined by its integrability properties rather than its individual pointwise values.

We also know by the Riesz representation theorem that, given μ a σ-finite measure, $1 \leq p < +\infty$ and q such that $\frac{1}{q} + \frac{1}{p} = 1$, for any linear functional T in the dual space $(L^p)'(\mu)$, there exists a unique function $f \in L^q(\mu)$ such that $T(\phi) = \int f\phi \, d\mu$ for all $\phi \in L^p(\mu)$. Therefore, we can identify the space of the $L^q(\mu)$ functions with the dual space $(L^p)'(\mu)$.

This provides a similar idea to the one underlying the concept of distribution, as will be further explored in the chapter.

Throughout this chapter, we will denote by U an open set contained in $\mathbb{R}^d$. Moreover, given $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ a multi-index of length $|\alpha| = \alpha_1 + \cdots + \alpha_d$, we denote the partial derivative of ϕ of order α as $D^\alpha \phi = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \ldots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} \phi$.

1.1 The spaces $\mathcal{D}_K(U)$, $\mathcal{D}(U)$ and $\mathcal{D}'(U)$

We define $C_c^\infty(U)$ as the set of all the infinitely differentiable functions whose support is compact and contained in U. We want to define a topology on this space of functions.

In order to do this, we first consider the space of functions whose support is contained within a certain compact subset $K \subseteq U$, which is denoted by

$$\mathcal{D}_K(U) \coloneqq \{\phi \in C_c^\infty(U) : \operatorname{supp}(\phi) \subseteq K\}.$$

In this subspace the topology should be consistent with the natural notion of convergence, which is uniform convergence of functions and of their partial derivatives of any order. For each $j \in \mathbb{N}$, we define the norm $\|\cdot\|_{K,j}$ on $\mathcal{D}_K(U)$ by

$$\|\phi\|_{K,j} \coloneqq \sup_{\substack{x\in K \\ |\alpha|\leq j}} \{|D^\alpha \phi(x)|\}.$$

It is possible to prove that the family of norms $\{\|\cdot\|_{K,j}\}$ endows the space $\mathcal{D}_K(U)$ with local convexity and a basis for the induced topology, denoted by τ_K, is given by all the sets of the form

$$\left\{\phi \in \mathcal{D}_K(U) : \|\phi\|_{K,j_1} < \frac{1}{l_1}, \ldots, \|\phi\|_{K,j_k} < \frac{1}{l_k}\right\}$$

where $j_1, \ldots, j_k \in \mathbb{N}$ and $l_1, \ldots, l_k, k \in \mathbb{Z}_+$. Given $j \coloneqq \max\{j_1, \ldots, j_k\}$ and $l \coloneqq \max\{l_1, \ldots, l_k\}$, it follows that

$$\begin{aligned} V_{K,j,l} &\coloneqq \left\{\phi \in \mathcal{D}_K(U) : \|\phi\|_{K,j} < \frac{1}{l}\right\} \\ &\subseteq \left\{\phi \in D_K(U) : \|\phi\|_{K,j_1} < \frac{1}{l_1}, \ldots, \|\phi\|_{K,j_k} < \frac{1}{l_k}\right\}. \end{aligned}$$

Indeed, the family of sets $V_{K,j,l}$ forms a basis, where $j \in \mathbb{N}$ and $l \in \mathbb{Z}_+$. It can be proven that the topology τ_K is induced by a metric and that $(\mathcal{D}_K(U), \tau_K)$ is complete. Moreover, it can be verified that this topology is also induced by the family of seminorms $\{p_\alpha\}_{\alpha\in\mathbb{N}^d}$, where each p_α is defined as:

$$p_\alpha(\phi) = \sup_{x\in K}\{|D^\alpha\phi(x)|\} = \|D^\alpha\phi\|_\infty, \quad \phi \in \mathcal{D}_K(U). \tag{1.1}$$

Consequently, a sequence ϕ_n converges to ϕ in $\mathcal{D}_K(U)$ with respect to the topology τ_K if and only if $p_\alpha(\phi_n - \phi) \to 0$ as $n \to \infty$ for every multi-index α.

To define the topology on $C_c^\infty(U)$ the intuitive approach would be to consider the topology generated by the family of norms

$$\|\phi\|_j \coloneqq \sup_{\substack{x\in U \\ |\alpha|\leq j}} \{|D^\alpha \phi(x)|\}, \quad \phi \in C_c^\infty(U).$$

The problem with this topology is that it would not be complete. For instance, given $U = \mathbb{R}$ and $\phi \in C_c^\infty(\mathbb{R})$ with support in $[0,1]$, such that $\phi > 0$ in $(0,1)$, then the sequence

$$\phi_n(x) \coloneqq \phi(x-1) + \frac{1}{2}\phi(x-2) + \cdots + \frac{1}{n}\phi(x-n), \quad x \in \mathbb{R},$$

is a Cauchy sequence in the topology generated by the family of norms $\{\|\cdot\|_j\}_{j\in\mathbb{Z}_+}$. However, its limit does not have compact support, so it does not belong to $C_c^\infty(\mathbb{R})$.

To construct a topology on $C_c^\infty(U)$ that ensures its completeness, the following definition is useful.

Definition 1.1. *A subset V of a vector space X is balanced if $tx \in V$ for all $x \in V$ and $t \in [-1,1]$.*

Let $\mathcal{B}_0$ be the collection of all convex and balanced sets $V \subset C_c^\infty(U)$ such that $V \cap \mathcal{D}_K(U) \in \tau_K$ for every compact set $K \subset U$. The following theorems hold. Their proofs are omitted.

Theorem 1.2. *The family*

$$\mathcal{B} \coloneqq \{\phi + V : \phi \in C_c^\infty(U), V \in \mathcal{B}_0\}$$

is a basis for a locally convex Haussdorff topology τ on $C_c^\infty(U)$ that turns $C_c^\infty(U)$ into a topological vector space.

The space $C_c^\infty(U)$ equipped with the topology τ is denoted as $\mathcal{D}(U)$ and its elements are called *test functions*. We set $\mathcal{D} = \mathcal{D}(\mathbb{R}^d)$.

Theorem 1.3. *For every compact set $K \subseteq U$ the topology τ_K coincides with the relative topology of $\mathcal{D}_K(U)$ as a subset of $\mathcal{D}(U)$.*

Theorem 1.4. *The space $\mathcal{D}(U)$ is complete. Moreover, a sequence $\{\phi_n\} \subseteq \mathcal{D}(U)$ converges to $\phi \in \mathcal{D}(U)$ with respect to τ if and only if*

(i) there exists a compact set $K \subseteq U$ such that the support of every ϕ_n and of ϕ is contained in K,

(ii) $\lim_{n\to\infty} D^\alpha \phi = D^\alpha \phi$ uniformly on K for every multi-index α.

While it can be shown that $\mathcal{D}(U)$ is not metrizable, it is still possible to prove that linear functionals defined on $\mathcal{D}(U)$ are continuous if and only if they are sequentially continuous. Specifically, the following result holds (the proof is omitted).

Theorem 1.5. *Let $T\colon \mathcal{D}(U) \to \mathbb{R}$ be a linear functional. Then the following properties are equivalent:*

(i) T *is continuous.*

(ii) *If* $\{\phi_n\} \subseteq \mathcal{D}(U)$ *converges to* $\phi \in \mathcal{D}(U)$ *with respect to* τ*, then* $\lim_{n\to\infty} T(\phi_n) = T(\phi)$*.*

(iii) *The restriction of* T *to* $\mathcal{D}_K(U)$ *is continuous for every compact set* $K \subseteq U$*.*

(iv) *For every compact set* $K \subseteq U$ *there exist an integer* $j \in \mathbb{N}$ *and a costant* $C_k > 0$ *such that* $|T(\phi)| \leq C_K \|\phi\|_{K,j}$ *for all* $\phi \in \mathcal{D}_K(U)$*.*

Definition 1.6. *A distribution on* U *is a continuous linear functional on* $\mathcal{D}(U)$*.*

The space of all distributions on U, denoted as $\mathcal{D}'(U)$, is the dual space of $\mathcal{D}(U)$ equipped with the weak* topology. Under this topology, a sequence $\{T_n\} \subseteq \mathcal{D}'(U)$ converges to $T \in \mathcal{D}'(U)$ if $T_n(\phi) \to T(\phi)$ for every $\phi \in \mathcal{D}(U)$. We define $\mathcal{D}' = \mathcal{D}'(\mathbb{R}^d)$ and in general, we will denote $\langle T, \phi\rangle$ instead of $T(\phi)$ to represent the action of the distribution T on the test function $\phi \in \mathcal{D}(U)$.

We now examine some examples of distributions.
Let $f \in L^1_{loc}(U)$, the functional

$$\langle T_f, \phi\rangle \coloneqq \int_U \phi(x) f(x)\, dx, \quad \phi \in \mathcal{D}(U) \tag{1.2}$$

is a distribution. Indeed, it is linear on $\mathcal{D}(U)$ and its continuity can be easily proved using Theorem 1.5, (iv). Moreover, if two functions define the same distribution, then they are equal almost everywhere.

This example illustrates why distributions can be regarded as generalized functions, as mentioned in the introduction. Indeed, we can associate a distribution T_f with the locally integrable function f. Hence, we denote T_f also by f, identifying L^1_{loc} with a subspace of $\mathcal{D}'(U)$. A distribution that can be represented by a $L^1_{loc}(U)$ function is called a *regular distribution*. It is important to note that not all linear functionals on $\mathcal{D}(U)$ can be expressed in this manner, leading to the notion of "generalized functions".

An example of this is provided by the Dirac delta. Given $x_0 \in U$, the functional $\langle \delta_{x_0}, \phi\rangle \coloneqq \phi(x_0)$, with $\phi \in \mathcal{D}(U)$, defines a distribution known as the *Dirac delta with mass at* x_0. Indeed, the functional defined is linear on $\mathcal{D}(U)$ and its continuity follows from the inequality $|\phi(x_0)| \leq \|\phi\|_\infty$. However, it can be proven that this distribution is not regular.

Proposition 1.7. *Given* $x_0 \in U$*, there is no function* $f \in L^1_{loc}(U)$ *verifying* $\phi(x_0) = \int_U \phi(x) f(x)\, dx$ *for all* $\phi \in \mathcal{D}(U)$*.*

Proof. Without loss of generality we can consider $x_0 = 0$ and so we suppose that $0 \in U$. Since U is open, there exist $r > 0$ such that the ball centered in zero and of radius r, denoted by $B(0, r)$, is contained in U. Assume by contradiction

that exists $f \in L^1_{\text{loc}}(U)$ satisfying $\phi(0) = \int_U \phi(x) f(x)\,dx$ for all $\phi \in \mathcal{D}(U)$. Let $\phi \in \mathcal{D}(U)$ such that its support is in $B(0,r)$, $0 \leq \phi(x) \leq 1$ for all $x \in U$ and $\phi(0) = 1$.

Given $n \in \mathbb{Z}_+$ and $x \in U$, we set $\phi_n(x) = \phi(nx)$, so that ϕ_n is supported in $B(0, \frac{r}{n})$ and $\phi_n(0) = 1$. Then we have for any $n \in \mathbb{Z}_+$:

$$1 = \phi_n(0) = \int_U f(x)\phi_n(x)\,dx \leq \int_{B(0,\frac{r}{n})} |f(x)|.$$

Since f is integrable on $B(0,r)$, we have that, by the Lebesgue's dominated convergence theorem,

$$\int_{B(0,\frac{r}{n})} |f(x)| \to 0 \quad \text{for } n \to +\infty.$$

This gives a contradiction. □

1.2 Operations on distributions

In light of the observations in the previous section, it is natural to try to extend certain operations defined on classical functions to distributions. A general procedure exists for this purpose, which will now be explored in depth.

Suppose that U and V are open sets in $\mathbb{R}^d$ and A is a linear map from some subspace X of $L^1_{loc}(U)$ into $L^1_{loc}(V)$. We would like to extend A as a map from $\mathcal{D}'(U)$ to $\mathcal{D}'(V)$. Suppose that there is another linear map A' from $\mathcal{D}(V)$ into $\mathcal{D}(U)$ such that

$$\int (Af)\phi = \int f(A'\phi), \quad f \in X, \quad \phi \in \mathcal{D}(V).$$

Suppose also that A' is continuous. Then A can be extended to a map from $\mathcal{D}'(U)$ to $\mathcal{D}'(V)$, still denoted by A, by

$$\langle AT, \phi \rangle = \langle T, A'\phi \rangle, \quad T \in \mathcal{D}'(U), \quad \phi \in \mathcal{D}(V).$$

Thanks to the continuous map A', the map A extended to distributions is continuous with respect to the weak* topology on distributions.

A fundamental operation on functions is differentiation, which is well-defined for smooth functions. Therefore, in this case the linear map A is defined on $X = C^{|\alpha|}(U)$ and $Af = D^\alpha f$ for all $f \in C^{|\alpha|}(U)$. We observe that, given a function $f \in C^1(U)$ and $\phi \in \mathcal{D}(U)$, the integration by parts formula yields:

$$\int_U (\partial_{x_i} f)(x)\phi(x)\,dx = -\int_U f(x)(\partial_{x_i}\phi)(x)\,dx \quad \forall i = 1, \ldots, n,$$

where $x = (x_1, \ldots, x_n)$. There are no boundary terms, since ϕ has compact support in U and thus vanishes near ∂U.

If $f \in C^{|\alpha|}(U)$, by applying the formula (1.2) $|\alpha|$ times we obtain

$$\int_U (D^\alpha f)(x)\phi(x)\,dx = (-1)^{|\alpha|}\int_U f(x)(D^\alpha \phi)(x)\,dx, \quad \phi \in \mathcal{D}(U).$$

Therefore, referring to the procedure explained above, we set $A' = (-1)^{|\alpha|}A|_{\mathcal{D}(U)}$, which is continuous. Starting by this, we can provide a more general definition of derivative, valid for distributions in general:

Definition 1.8. *Let $T \in \mathcal{D}'(U)$ and α a multi-index. We define the α- derivative of T as*

$$\langle D^\alpha T, \phi\rangle \coloneqq (-1)^{|\alpha|}\langle T, D^\alpha \phi\rangle, \quad \phi \in \mathcal{D}(U).$$

It can be verified that the α- derivative of T is still a distribution.

Example 1.1: Consider the Heaviside function $H\colon \mathbb{R} \to \mathbb{R}$ defined as follows:

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases}$$

The Heaviside function belongs to $L^1_{\text{loc}}(\mathbb{R})$, allowing us to consider the associated distribution T_H as described in formula (1.2).

Given a test function $\phi \in \mathcal{D}$, we can compute the action of the distributional derivative DT_H as follows:

$$\begin{aligned} \langle DT_H, \phi\rangle &= -\int_{\mathbb{R}} H(x)\phi'(x)\,dx \\ &= -\int_0^{+\infty} \phi'(x)\,dx \\ &= -\lim_{a\to+\infty}(\phi(a) - \phi(0)) \\ &= \phi(0). \end{aligned}$$

From the definition of the Dirac delta distribution, we conclude that

$$\langle DT_H, \phi\rangle = \langle \delta, \phi\rangle,$$

where δ is the Dirac delta with mass at 0. Therefore, the distributional derivative of the Heaviside function is given by

$$DT_H = \delta.$$

It is evident that this procedure allows us to define derivatives for any locally integrable function f, even when it does not possess classical differentiability, by considering the associated regular distribution T_f. Furthermore, if the distributional derivative of this distribution is also a regular distribution, we can define the concept of weak derivative.

Definition 1.9. *Suppose $f, g \in L^1_{loc}(U)$ and α a multi-index. We say that g is the α^{th}-weak partial derivative of f if $T_g = D^\alpha T_f$, that is*

$$\int_U g(x)\phi(x)\,dx = (-1)^{|\alpha|}\int_U f(x)(D^\alpha\phi)(x)\,dx, \quad \phi \in \mathcal{D}(U). \tag{1.3}$$

In this case, g is denoted by $D^\alpha f$.

The weak derivative is uniquely defined up to a set of measure zero, because two functions belonging to $L^1_{loc}(U)$ that define the same regular distribution are equal almost everywhere. Since the derivative defined in the classical manner satisfies (1.3), it follows that if a function is differentiable, then the weak derivative coincides with the classical one.

It is important to recognize that not every locally integrable function has a weak derivative. The Heaviside function from Example 1.1 illustrates this, as its distributional derivative is the Dirac delta function, which is not a regular distribution.

Example 1.2: Given the function $f(x) = |x|$ defined on $\mathbb{R}$, it can be shown that this function is weakly differentiable, and its weak derivative is given by the sign function $\operatorname{sgn}(x)$.

In fact, by integration by parts,

$$-\int_{\mathbb{R}} f(x)\phi'(x)\,dx = -\int_{-\infty}^{0}\phi(x)\,dx + \int_0^{+\infty}\phi(x)\,dx = \int_{\mathbb{R}}\operatorname{sgn}(x)\phi(x)\,dx,$$

where

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

Therefore, we have that $Df = \operatorname{sgn}$.

Another instance of extending operations to distributions is the multiplication by smooth functions. In particular, given $\psi \in C^\infty(U)$, we define $Af = \psi f$ with $f \in L^1_{loc}$ and consider $A' = A|_{\mathcal{D}(U)}$. We can define the product $\psi T \in \mathcal{D}'(U)$ for $T \in \mathcal{D}'(U)$ by

$$\langle \psi T, \phi\rangle = \langle T, \psi\phi\rangle, \quad \phi \in \mathcal{D}(U).$$

A crucial operation that can be extended to distributions is the Fourier transformation. To achieve this, the idea is following similar principles as those utilized for differentiation. Indeed, we recall that for $f, g \in L^1$ we have

$$\int \hat{f}(y)g(y)\,dy = \int f(y)\hat{g}(y)\,dy,$$

where $\hat{f}$ represents the Fourier transform of the function f, that is

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^d} f(x)e^{-i\xi\cdot x}\,dx, \quad \xi \in \mathbb{R}^d,$$

where $\xi \cdot x$ is the scalar product. Therefore, an intuitive approach to extend the Fourier transform of a distribution T is by defining:

$$\langle \hat{T}, \phi \rangle = \langle T, \hat{\phi} \rangle.$$

However, this is not a good definition in general because the Fourier transform of a function belonging to $\mathcal{D}$ generally does not belong to $\mathcal{D}$. Consequently, the right hand side of the equation is not well-defined. However there exists a specific class of functions for which this property holds true, which is the Schwartz class $\mathcal{S}$. Therefore, it becomes necessary to define tempered distributions, which will be addressed in the upcoming section. In it, we first introduce compactly supported distributions, a specific instance of tempered distributions for which an alternative definition of the Fourier transform is possible.

1.3 Compactly supported and tempered distributions

Definition 1.10. *Let $T \in \mathcal{D}'(U)$. If $U' \subseteq U$ is open, then we write that $T = 0$ in U' if $\langle T, \phi \rangle = 0$ for all $\phi \in \mathcal{D}(U')$. The support of T is the complement of the union of all open subsets $U' \subseteq U$ in which $T = 0$.*

The space of all distributions on U whose support is a compact subset of U is denoted by $\mathcal{E}'(U)$. As usual, we set $\mathcal{E}' = \mathcal{E}'(\mathbb{R}^d)$.

It can be proven that $\mathcal{E}'(U)$ is the dual space of $C^\infty(U)$ equipped with the C^∞ topology, that is the topology of uniform convergence of functions, together with all their derivatives, on compact subsets of U. To prove this it is useful the following result (the proof is omitted).

Proposition 1.11. *Suppose X and Y are vector spaces with topologies defined respectively by the families $\{p_\alpha\}_{\alpha\in A}$ and $\{q_\beta\}_{\beta\in B}$ of seminorms, and $T\colon X \to Y$ is a linear map. Then T is continuous if and only if for each $\beta \in B$ there exist $\alpha_1, \ldots, \alpha_k \in A$ and $C > 0$ such that $q_\beta(Tx) \leq C \sum_{j=1}^k p_{\alpha_j}(x)$.*

The topology on $C^\infty(U)$ can be defined by a countable family of seminorms as follows. Let $\{V_m\}_{m\in\mathbb{Z}_+}$ be an increasing sequence of precompact open subsets of U, whose union is U. Then, for each $m \in \mathbb{Z}_+$ and each multi-index $\alpha \in \mathbb{N}^d$, we can define the seminorm

$$\|f\|_{[m,\alpha]} = \sup_{x\in\bar{V}_m} |D^\alpha f(x)|.$$

Clearly, $D^\alpha f_j \to D^\alpha f$ uniformly on compact sets as $j \to \infty$ for all α if and only if $\|f_j - f\|_{[m,\alpha]} \to 0$ for all m and α. It is worth noting that varying choices of sets V_m lead to equivalent families of seminorms. The following result holds.

Proposition 1.12. *$\mathcal{D}(U)$ is dense in $C^\infty(U)$.*

Proof. Let $\{V_m\}_{m\in\mathbb{Z}_+}$ be as above. For each m, by the Urysohn lemma, we can pick $\psi_m \in \mathcal{D}(U)$ with $\psi_m = 1$ on $\bar{V}_m$. If $\phi \in C^\infty(U)$, clearly $\psi_m\phi \in \mathcal{D}(U)$ and $\|\psi_m\phi - \phi\|_{[m_0,\alpha]} = 0$ if $m > m_0$, hence $\psi_m\phi \to \phi$ in the C^∞ topology. □

Theorem 1.13. *$\mathcal{E}'(U)$ is the dual space of $C^\infty(U)$. More precisely if $F \in \mathcal{E}'(U)$, then F extends uniquely to a continuous linear functional on $C^\infty(U)$ and if G is a continuous linear functional on $C^\infty(U)$, then $G|_{\mathcal{D}(U)} \in \mathcal{E}'(U)$.*

Proof. If $F \in \mathcal{E}'(U)$, since $\operatorname{supp}(F)$ is compact for hypothesis, for the Urysohn lemma, we can consider $\psi \in \mathcal{D}(U)$ with $\psi = 1$ on $\operatorname{supp}(F)$. Consequently, we define the linear functional G on $C^\infty(U)$ by $\langle G, \phi\rangle = \langle F, \psi\phi\rangle$. This is well-defined because $\psi\phi \in \mathcal{D}(U)$ and the result does not rely on the particular choice of ψ, as what matters is the value the function in question assumes over the support of F. Moreover, for construction $G|_{\mathcal{D}(U)} = F$ and G is a functional whose action on the element of $C^\infty(U)$ is determined by F's actions on functions in $\mathcal{D}(U)$ with support contained in $\operatorname{supp}(\psi)$, which is compact. By restricting F to this subspace, corresponding to $\mathcal{D}_{\operatorname{supp}(\psi)}(U)$, continuity is preserved. We recall that the topology on this space can also be regarded as induced by the family of seminorms (1.1), hence, by Proposition 1.11, there exist $N \in \mathbb{Z}_+$ and $C > 0$ such that $|\langle G, \phi\rangle| \le C \sum_{|\alpha|\le N} \|D^\alpha(\psi\phi)\|_\infty$ for all $\phi \in C^\infty(U)$. By the definition of ψ and by the product rule, there exist $C' > 0$ so that

$$|\langle G, \phi\rangle| \le C' \sum_{|\alpha|\le N} \sup_{x\in\operatorname{supp}\psi} |D^\alpha\phi(x)| \quad \forall\phi \in C^\infty(U).$$

If we choose m large enough so that $\operatorname{supp}(\psi) \subseteq V_m$, this implies that

$$|\langle G, \phi\rangle| \le C' \sum_{|\alpha|\le N} \|\phi\|_{[m,\alpha]} \quad \forall\phi \in C^\infty(U),$$

hence G is continuous on $C^\infty(U)$. Moreover G is the unique continuous extension of F by Proposition 1.12.

Now suppose that G is a continuous linear functional on $C^\infty(U)$.

By Proposition 1.11, G is continuous on $\mathcal{D}(U)$ if there exist C, m, N such that $|\langle G, \phi\rangle| \leq C \sum_{|\alpha|\leq N} \|\phi\|_{[m,\alpha]}$ for all $\phi \in C^\infty(U)$. Since $\|\phi\|_{[m,\alpha]} \leq \|D^\alpha \phi\|_\infty$, this implies that G is continuous on $\mathcal{D}_K(U)$ for each compact $K \subseteq U$. Hence, by Theorem 1.5 (*iii*), $G|_{\mathcal{D}(U)} \in \mathcal{D}'(U)$. Moreover, if $\text{supp}(\phi) \cap \bar{V}_m = \emptyset$, then $\langle G, \phi\rangle = 0$. Hence, $\text{supp}(G) \subseteq \bar{V}_m$ and $G|_{\mathcal{D}(U)} \in \mathcal{E}'(U)$.

□

To define tempered distributions, we first introduce the Schwartz space $\mathcal{S}$, given by

$$\mathcal{S} = \{f \in C^\infty(\mathbb{R}^d) : p_{N,\alpha}(f) < +\infty \quad \forall N \in \mathbb{N}, \alpha \in \mathbb{Z}_+^d\},$$

where $p_{N,\alpha}(f) = \sup_{x\in\mathbb{R}^d}(1 + |x|)^N |D^\alpha f(x)|$. This family of seminorms induces a topology on $\mathcal{S}$. With this topology established, we can define the dual space $\mathcal{S}'$, where we will extend the Fourier transform operation.

Definition 1.14. *A tempered distribution is a continuous linear functional on $\mathcal{S}$. The space of tempered distributions is denoted by $\mathcal{S}'$ and it is equipped with the weak* topology.*

It can be proven that the space $\mathcal{D}$ is dense in $\mathcal{S}$. If $F \in \mathcal{S}'$, then $F|_{\mathcal{D}}$ is clearly a distribution, since convergence in $\mathcal{D}$ implies convergence in $\mathcal{S}$ and $F|_{\mathcal{D}}$ determines F uniquely since $\mathcal{D}$ is dense in $\mathcal{S}$. Hence $\mathcal{S}'$ can be identified with the set of distributions that extends continuously from $\mathcal{D}$ to $\mathcal{S}$.

Herein are provided some examples of tempered distributions.

If $f \in L^1_{loc}(\mathbb{R}^d)$ and $\int (1 + |x|)^{-N} |f(x)|\, dx < +\infty$ for some $N \in \mathbb{N}$, then it can be identified with a tempered distribution, since

$$\left| \int f(x)\phi(x)\, dx \right| \leq C \cdot p_{N,0}(\phi) \quad \forall \phi \in \mathcal{S}.$$

It follows that L^p functions, with $1 \leq p \leq +\infty$, can also be viewed as tempered distributions.

By Theorem 1.13, $\mathcal{E}'$ can be regarded as the dual space of $C^\infty(\mathbb{R}^d)$. Therefore, we can restrict $T \in \mathcal{E}'$ to the Schwartz space and $T|_{\mathcal{S}} \in \mathcal{S}'$. It follows that every compactly supported distribution can be regarded as a tempered distribution.

As deduced previously, it is possible to extend the Fourier transformation to a tempered distribution F through the following definition:

$$\langle \hat{F}, \phi\rangle := \langle F, \hat{\phi}\rangle \quad \forall \phi \in \mathcal{S}. \tag{1.4}$$

Given that the Fourier transform of a Schwartz function remains in the same space, the definition is properly formulated, and $\hat{F}$ is a tempered distribution. Moreover, by construction, when $F \in L^1$ or $F \in L^2$, which we have previously shown can be identified as tempered distributions, the provided definition aligns with the classical Fourier transformation of the functions.

The basic properties of the Fourier transform continue to hold in this setting. In particular:

$$\begin{aligned} D^\alpha \hat{F} &= [\widehat{(-i\cdot)^\alpha F}] \\ \widehat{D^\alpha F} &= (i\cdot)^\alpha \hat{F}. \end{aligned} \tag{1.5}$$

If $F \in \mathcal{E}'$, there is an alternative way to define $\hat{F}$. Indeed, $\langle F, \phi \rangle$ makes sense for any $\phi \in C^\infty$ and if we take $\phi(x) = e^{-2\pi i \xi x}$, we obtain a function of ξ that has a strong claim to be called $\hat{F}$. In fact, the following result holds:

Proposition 1.15. *If $F \in \mathcal{E}'$, then $\hat{F}$ is a slowly increasing C^∞ function, and it is given by $\hat{F}(\xi) = \langle F, E_{-\xi} \rangle$ where $E_\xi(x) = e^{2\pi i \xi \cdot x}$.*

1.4 Sobolev Spaces

This section provides an introduction to Sobolev spaces, which encompass functions possessing certain, but not overly strong, smoothness properties. Defining such functions necessitates the concept of weak derivatives, introduced in Chapter 1. We recall also that if $f \in L^p(\mathbb{R}^d)$, with $1 \leq p \leq +\infty$, it can be regarded as a tempered distribution.

Definition 1.16. *Fix $1 \leq p \leq +\infty$, and let $k \in \mathbb{Z}_+$. The Sobolev space $W^{k,p}(U)$ consists of all the functions $u \in L^p(U)$ such that for each multiindex α with $|\alpha| \leq k$, $D^\alpha u$ exists in the weak sense and belongs to $L^p(U)$. Therefore:*

$$W^{k,p}(U) \coloneqq \{u \in L^p(U) : D^\alpha u \in L^p(U), \forall \alpha \in \mathbb{N}^d, |\alpha| \leq k\}.$$

Definition 1.17. *If $u \in W^{k,p}(U)$ we define its norm to be*

$$\|u\|_{W^{k,p}(U)} = \begin{cases} (\sum_{|\alpha| \leq k} \int_U |D^\alpha u|^p \, dx)^{\frac{1}{p}} & 1 \leq p \leq +\infty \\ \sum_{|\alpha| \leq k} \operatorname{ess\,sup}_U |D^\alpha u| & p = +\infty. \end{cases}$$

We will also write $\| \cdot \|_{k,p}$ in place of $\| \cdot \|_{W^{k,p}(U)}$.

The validity of the norms just defined will now be assessed.

Clearly $\|\lambda u\|_{k,p} = |\lambda| \|u\|_{k,p}$ for all $\lambda \in \mathbb{R}$, and $\|u\|_{k,p} = 0$ if and only if $u = 0$ almost everywhere. Moreover, if $u, v \in W^{k,p}(U)$ and $1 \leq p < +\infty$, by the Minkowski inequality:

$$\begin{aligned} \|u+v\|_{k,p} &= (\sum_{|\alpha| \leq k} \|D^\alpha (u+v)\|_p^p)^{\frac{1}{p}} \\ &\leq (\sum_{|\alpha| \leq k} (\|D^\alpha u\|_p + \|D^\alpha v\|_p)^p)^{\frac{1}{p}} \\ &\leq (\sum_{|\alpha| \leq k} \|D^\alpha u\|_p^p)^{\frac{1}{p}} + (\sum_{|\alpha| \leq k} \|D^\alpha v\|_p^p)^{\frac{1}{p}} \\ &= \|u\|_{k,p} + \|v\|_{k,p}. \end{aligned}$$

If $p = +\infty$, the triangle inequality holds immediately.

Definition 1.18. *Given $1 \leq p \leq +\infty$, a sequence $\{u_m\}_{m=1}^{\infty} \subseteq W^{k,p}(U)$, and $u \in W^{k,p}(U)$, we say that u_m converges to u in $W^{k,p}(U)$ if $\lim_{m\to+\infty} \|u_m - u\|_{k,p} = 0$.*

The Sobolev spaces have a good mathematical structure:

Theorem 1.19. *For each $k \in \mathbb{Z}_+$ and $1 \leq p \leq +\infty$, the Sobolev space $W^{k,p}(U)$ is a Banach space.*

Proof. $(W^{k,p}(U), \|\cdot\|_{k,p})$ is a normed space. It remains to show that $W^{k,p}(U)$ is complete. Let $\{u_m\}_{m=1}^{\infty}$ a Cauchy sequence in $W^{k,p}(U)$. Then for each $|\alpha| \leq k$, $\{D^\alpha u_m\}_{m=1}^{\infty}$ is a Cauchy sequence in $L^p(U)$. Since, $L^p(U)$ is complete, there exist functions $u_\alpha \in L^p(U)$ such that $D^\alpha u_m \to u_\alpha$ in $L^p(U)$ for each $|\alpha| \leq k$. Particularly, $u_m \to u_{(0,\dots,0)} =: u$ in $L^p(U)$.

We now show that $u \in W^{k,p}(U)$ and $D^\alpha u = u_\alpha$ with $|\alpha| \leq k$. Given $\phi \in C_c^\infty(U)$, we have that:

$$\begin{aligned}\int_U u D^\alpha \phi \, dx &= \lim_{m\to+\infty} \int_U u_m D^\alpha \phi \, dx \\ &= \lim_{m\to+\infty} (-1)^{|\alpha|} \int_U D^\alpha u_m \phi \, dx \\ &= (-1)^{|\alpha|} \int_U u_\alpha \phi \, dx.\end{aligned}$$

Thus, $D^\alpha u_m \to D^\alpha u$ in $L^p(U)$ for all $|\alpha| \leq k$. Therefore, $u_m \to u$ in $W^{k,p}(U)$. □

It is important to note that $W^{k,2}(U)$ is not only a Banach space but also a Hilbert space, equipped with the scalar product:

$$\langle u, v \rangle := \sum_{|\alpha|\leq k} \int_U D^\alpha u \overline{D^\alpha v} \, dx \qquad \forall u, v \in W^{k,2}(U).$$

Therefore, we will denote $W^{k,2}(U)$ by $H^k(U)$.

It is possible to provide an alternative characterization of the spaces $H^k(\mathbb{R}^d)$ by employing the Fourier transform.

Theorem 1.20. *Let $k \in \mathbb{Z}_+$.*

(i) A function $u \in L^2(\mathbb{R}^d)$ belongs to $H^k(\mathbb{R}^d)$ if and only if $(1+|\cdot|^2)^{\frac{k}{2}} \hat{u} \in L^2(\mathbb{R}^d)$.

(ii) There exist two positive constants, C and C', such that

$$\frac{1}{C'} \|u\|_{H^k(\mathbb{R}^d)} \leq \|(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u}\|_2 \leq C \|u\|_{H^k(\mathbb{R}^d)}$$

for each $u \in H^k(\mathbb{R}^d)$.

Proof. Assume first $u \in H^k(\mathbb{R}^d)$. Then for each multiindex $|\alpha| \leq k$, we have $D^\alpha u \in L^2(\mathbb{R}^d)$. From the formula (1.5) and the fact that functions in $L^2(\mathbb{R}^d)$ can be identified as tempered distributions, we have $\widehat{D^\alpha u}(\xi) = (i\xi)^\alpha \hat{u}(\xi)$. Moreover, since $\|f\|_2 = \|\hat{f}\|_2$ for all $f \in L^2(\mathbb{R}^d)$ by Plancherel theorem, $\widehat{D^\alpha u} \in L^2(\mathbb{R}^d)$ for each $|\alpha| \leq k$. In particular, $\|D^\alpha u\|^2 = \|\widehat{D^\alpha u}\|^2 = \int_{\mathbb{R}^d} \prod_{i=1}^d |\xi_i|^{2\alpha_i} |\hat{u}(\xi)|^2 \, d\xi$, where $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $\sum_{i=1}^d \alpha_i \leq k$. Through mathematical calculations, it can be shown that there exists a positive constant C such that

$$\int_{\mathbb{R}^d} (1 + |\xi|^2)^k |\hat{u}(\xi)|^2 \, d\xi \leq C \sum_{|\alpha| \leq k} \int_{\mathbb{R}^d} |D^\alpha u(x)|^2 \, dx = C \|u\|_{H^k(\mathbb{R}^d)}.$$

Hence, $(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u} \in L^2(\mathbb{R}^d)$ and $\|(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u}\|_2 \leq C \|u\|_{H^k(\mathbb{R}^d)}$.

Suppose $(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u} \in L^2(\mathbb{R}^d)$ and $|\alpha| \leq k$. Then there exists a positive constant $\tilde{C}$ for which

$$\|(i\cdot)^\alpha \hat{u}\|_2^2 \leq \int_{\mathbb{R}^d} |\xi|^{2|\alpha|} |\hat{u}(\xi)|^2 \, d\xi \leq \tilde{C} \|(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u}\|_2^2. \tag{1.6}$$

From the previous observation, $\|D^\alpha u\|_2 = \|(i\cdot)^\alpha \hat{u}\|_2$. Therefore, $u \in H^k(\mathbb{R}^d)$ as required. Moreover, from (1.6) there exists a positive constant C' such that

$$\|u\|_{H^k(\mathbb{R}^d)} \leq C' \|(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u}\|_2^2.$$

□

Note that requiring $(1 + |\cdot|^2)^{\frac{k}{2}} \hat{u} \in L^2(\mathbb{R}^d)$ equivalent to requiring that

$$\int_{\mathbb{R}^d} \hat{u}(\xi)^2 (1 + |\xi|^2)^k \, d\xi < +\infty.$$

We will denote this condition by saying $\hat{u} \in L^2(\mathbb{R}^d, (1 + |\xi|^2)^k)$.

Thanks to this characterization, it is also possible to define fractional Sobolev spaces.

Definition 1.21. *Assume $0 < s < +\infty$ and $u \in L^2(\mathbb{R}^d)$. Then $u \in H^s(\mathbb{R}^d)$ if $\hat{u} \in L^2(\mathbb{R}^d, (1 + |\xi|^2)^s)$. For noninteger s, we set $\|u\|_{H^s(\mathbb{R}^d)} := \|(1 + |\cdot|^2)^{\frac{s}{2}} \hat{u}\|_2$.*

Chapter 2

Computed tomography

2.1 Introduction

Computed tomography (CT) is a fundamental technique in medical imaging, aimed at determining the density function of an unknown object. This is achieved by measuring and analyzing the attenuation of X-rays as they pass through the object along numerous lines. The following section outlines the underlying physics and mathematical framework.

Consider a line ℓ along which X-rays propagate. Let $I(x)$ be the intensity (i.e., the number of photons) of the X-rays at the point $x \in \ell$. The attenuation coefficient at x, denoted $f(x)$, correlates directly with the density of the object at that point for monochromatic light, and they can be considered equivalent by appropriate scaling.

It is verified that the rate of decrease in intensity along the path is proportional to the intensity itself, with $-f(x)$ the proportionality constant:

$$\frac{dI(x)}{dx} = -f(x)I(x). \tag{2.1}$$

This equation is consistent with the fact that denser materials at x (i.e., larger $f(x)$) cause increased attenuation of the X-ray beam, resulting in a steeper decline in intensity. Equation (2.1) is a differential equation, solvable by separation of variables. If I_0 represents the intensity at the X-ray source, located at $x_0 \in \ell$, and I_1 the intensity at the detector, located at $x_1 \in \ell$, integration of (2.1) yields:

$$\ln\left(\frac{I_0}{I_1}\right) = \int_{x_0}^{x_1} f(x)\,dx = \int_{x\in\ell} f(x)\,dx. \tag{2.2}$$

The integral on the right side of the equation defines an operator known as the Radon transform:

$$\mathcal{R}(f)(\ell) = \int_{x\in\ell} f(x)\,dx,$$

where dx is the arc length measure on ℓ. The basic properties of this transform will be analysed in the following section.

From equation (2.2), it follows that the problem of reconstructing an object f from CT data translates into the following inverse problem:

$$y = \mathcal{R}f + \eta, \tag{2.3}$$

where y represents the measurements and η denotes the noise.

To better understand the problem of image reconstruction, it is essential to study the underlying mathematical model (2.3). To this end, the following section will outline the key properties of the Radon transform and related results. For further details, we refer to [29].

2.2 The Radon transform

As introduced earlier, the Radon transform is crucial in computed tomography. Indeed, the goal in this context is to reconstruct the internal structure of an object from its Radon transform, which represents the collection of line integrals corresponding to different angles and positions.

The Radon transform of a function f is defined as the integral of f over all hyperplanes π in $\mathbb{R}^n$, provided the integral exists. Specifically, the Radon transform is given by $\mathcal{R}f(\pi) = \int_\pi f\, dS$, where dS denotes the Euclidean surface measure on the hyperplane π. For simplicity, and because it is of primary interest in this context, we will focus on the two-dimensional case. In this scenario, the Radon transform is often referred to as the X-ray transform because of its application in tomography.

Definition 2.1. *Assuming that the integral exists, the Radon transform of a function f in $\mathbb{R}^2$ is $\mathcal{R}f(\ell) = \int_\ell f\, ds$, where ℓ is a given line in $\mathbb{R}^2$ and ds is the unit length measure on ℓ.*

To obtain a more explicit mathematical expression for the Radon transform, we can parameterize the line ℓ by its distance from the origin, $s \in \mathbb{R}$, and its direction, identified by the unit vector ω perpendicular to the line. Specifically, this line ℓ in $\mathbb{R}^2$ can be defined as

$$L(\omega, s) = \{x \in \mathbb{R}^2 \mid x \cdot \omega = s\},$$

where $x \cdot \omega$ denotes the scalar product.

It is beneficial to express ω as a function of the angle $\theta \in [0, 2\pi)$. Specifically, we define $\omega(\theta) = (\cos(\theta), \sin(\theta))$. This parameterizes the unit sphere in $\mathbb{R}^2$, denoted by $\mathbb{S}^1$. Under this parameterization, we denote

$$L(\theta, s) = \{x \in \mathbb{R}^2 \mid x_1 \cos(\theta) + x_2 \sin(\theta) = s\},$$

where $x = (x_1, x_2)$.

The following definition holds.

Definition 2.2. *The Radon transform $\mathcal{R}$ of a function f can be expressed as a function on $[0, 2\pi) \times \mathbb{R}$ given by:*

$$\mathcal{R}f(\theta, s) = \int_{x \in L(\theta,s)} f(x)\, dx = \int_{\mathbb{R}} f(s\omega(\theta) + t\omega(\theta)^{\perp})\, dt.$$

Here, $\omega(\theta)^{\perp}$ is the unit vector obtained by rotating $\omega(\theta)$ counterclockwise by $\frac{\pi}{2}$ radians.

With this parameterization, the following symmetry property holds: if $\theta_1, \theta_2 \in [0, 2\pi)$ and $|\theta_1 - \theta_2| = \pi$, then $\mathcal{R}f(\theta_1, s) = \mathcal{R}f(\theta_2, -s)$ for any $s \in \mathbb{R}$. For this reason, it would also be possible to parameterize the lines considering, for instance, $\theta \in [0, \pi)$ or $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$, as this would still cover all possible lines since $s \in \mathbb{R}$. It is important to note that the Radon transform is a linear operator. Specifically, given α, β real constants, and f, g functions within the domain of the Radon transform, $\mathcal{R}(\alpha f + \beta g) = \alpha\mathcal{R}f + \beta\mathcal{R}g$.

At this point, the domain over which the Radon transform is considered has not yet been specified. A natural domain of the Radon transform is the set of the functions whose restriction to every line $L(\theta, s)$ is an absolutely integrable function. For simplicity, we will consider the Radon transform on the domain of functions that are absolutely integrable over $\mathbb{R}^2$, which is the space $L^1(\mathbb{R}^2)$.
The Radon transform is well-defined on this domain, since by Fubini's Theorem

$$\int_{\mathbb{R}} |\mathcal{R}f(\theta, s)|\, ds = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(s\omega(\theta) + t\omega(\theta)^{\perp})\, dt \right| ds \leq \int_{\mathbb{R}^2} |f(x)|\, dx < +\infty. \quad (2.4)$$

Therefore, for each fixed $\theta \in [0, 2\pi)$, $\mathcal{R}f(\theta, s)$ is well-defined and finite for almost every $s \in \mathbb{R}$.

We denote $L^1([0, 2\pi) \times \mathbb{R})$ as the space of measurable functions such that the absolute integral on $[0, 2\pi) \times \mathbb{R}$ is finite. Hence, $h \in L^1([0, 2\pi) \times \mathbb{R})$ if $\|h\|_{L^1([0,2\pi)\times\mathbb{R})} = \int_0^{2\pi} \int_{\mathbb{R}} |h(\theta, s)|\, ds\, d\theta < +\infty$.

Theorem 2.3. *The Radon transform is a continuous operator from $L^1(\mathbb{R}^2)$ into $L^1([0, 2\pi) \times \mathbb{R})$. In particular, for $f \in L^1(\mathbb{R}^2)$, $\|Rf\|_{L^1([0,2\pi)\times\mathbb{R})} \leq 2\pi\|f\|_{L^1(\mathbb{R}^2)}$.*

Proof. By (2.4), we have

$$\|\mathcal{R}f\|_{L^1([0,2\pi)\times\mathbb{R})} = \int_0^{2\pi} \int_{\mathbb{R}} |\mathcal{R}f(\theta, s)|\, ds\, d\theta \leq 2\pi\|f\|_{L^1(\mathbb{R}^2)},$$

as claimed. □

As mentioned earlier, our goal is to reconstruct an object from its Radon transform. The key questions are whether an inversion method exists and if it provides accurate results. We will demonstrate that such an inversion method does indeed exist, but it must be treated with caution. To gain a better understanding,

it is useful to explore the concept of backprojection. Intuitively, the value $f(x)$ contributes to $\mathcal{R}f(\theta, s)$ whenever x lies on the line $L(\theta, s)$, namely when $x \cdot \omega(\theta) = s$. In the absence of additional information, we assume that $f(x)$ contributes equally in all such cases. Therefore, to estimate the value of $f(x)$, we can average the values of $\mathcal{R}f$ along all lines that pass through x.

Definition 2.4. *The backprojection of $f \in L^1(\mathbb{R}^2)$ is defined as:*

$$\mathcal{B}f(x) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{R}f(\theta, x \cdot \omega(\theta))\, d\theta,$$

where $x \in \mathbb{R}^2$.

However, $\mathcal{B}f$ is not equal to f, meaning this is not an inversion formula. For instance, through calculations, we find that in the case of the indicator function of the unit disk

$$\mathcal{B}f(x) = \frac{1}{\pi} \int_0^{2\pi} \left(1 - |x \cdot \omega(\theta)|^2\right)^{\frac{1}{2}} \chi_{[-1,1]}(x \cdot \omega(\theta))\, d\theta,$$

where $\chi_{[-1,1]}$ denotes the characteristic function of the interval $[-1, 1]$. Therefore, the backprojection resembles the unit disk and appears as a blurred version of f. This suggests that some additional operations might lead to the true inversion formula. A crucial theorem in this context is the Fourier slice theorem, which provide the relationship between the Radon and Fourier transforms.

Let $\theta \in [0, 2\pi)$. We introduce the following notation

$$R_\theta f(s) = Rf(\theta, s) \quad \forall s \in \mathbb{R}.$$

This notation is useful for considering the Fourier transform of the Radon transform of a function with respect to the second variable s, as needed for the Fourier Slice Theorem.

Theorem 2.5 (General Projection Slice Theorem)**.** *Let $f \in L^1(\mathbb{R}^2)$ and $\theta \in [0, 2\pi)$. Let $h \in L^\infty(\mathbb{R})$. Then*

$$\int_{\mathbb{R}^2} f(x)h(x \cdot \omega(\theta))\, dx = \int_{\mathbb{R}} \mathcal{R}f(\theta, s)h(s)\, ds.$$

Proof. Let $\theta \in [0, 2\pi)$. First, note that the function $x \mapsto f(x)h(x \cdot \omega(\theta))$ is in $L^1(\mathbb{R}^2)$, since h is bounded. For the same reason, the function $(s, t) \mapsto f(s\omega(\theta) + t\omega(\theta)^\perp)h(s)$ is in $L^1(\mathbb{R}^2)$. Therefore, by Fubini's theorem,

$$\int_{x \in \mathbb{R}^2} f(x)h(x \cdot \omega(\theta))\, dx = \int_{\mathbb{R}} \int_{\mathbb{R}} f(s\omega(\theta) + t\omega(\theta)^\perp)h(s)\, dt\, ds = \int_{\mathbb{R}} \mathcal{R}f(\theta, s)h(s)\, ds.$$

□

The Fourier slice theorem is a special case of the general projection slice theorem, which corresponds to choosing $h(s) = \frac{e^{-is\xi}}{2\pi}$, with $\xi \in \mathbb{R}$.

Corollary 2.6 (Fourier Slice Theorem)**.** *Let $f \in L^1(\mathbb{R}^2)$ and $(\theta, \xi) \in [0, 2\pi) \times \mathbb{R}$. Then,*

$$\widehat{\mathcal{R}_\theta f}(\xi) = (2\pi)^{\frac{1}{2}} \hat{f}(\xi\omega(\theta)).$$

This corollary establishes a connection between the one-dimensional Fourier transform of the Radon transform of f, with the parameter θ fixed, and the two-dimensional Fourier transform of the function f. Moreover, it shows that $\mathcal{R}$ is injective on the domain $L^1(\mathbb{R}^2)$. Indeed, if $\mathcal{R}f \equiv 0$, then $\hat{f} \equiv 0$ and, by the injectivity of the Fourier transform on $L^1(\mathbb{R}^2)$, it follows that $f \equiv 0$.

Theorem 2.7 (Radon inversion formula)**.** *If $f \in L^1(\mathbb{R}^2)$ and $\hat{f} \in L^1(\mathbb{R}^2)$, then*

$$f(x) = \frac{1}{2(2\pi)^{\frac{3}{2}}} \int_0^{2\pi} \int_{\mathbb{R}} \widehat{\mathcal{R}_\theta f}(r)\, |r|\, e^{irx\cdot\omega(\theta)}\, dr\, d\theta$$

Proof. As f and $\hat{f}$ are absolutely integrable, it is known from Fourier analysis theory that

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi.$$

By re-expressing the Fourier inversion formula using polar coordinates we have:

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} \hat{f}(r\omega(\theta)) e^{irx\cdot\omega(\theta)} r\, dr\, d\theta \end{aligned}$$

It can be observed that

$$\begin{aligned} \int_0^{2\pi} \int_{-\infty}^{0} \hat{f}(r\omega(\theta)) e^{irx\cdot\omega(\theta)} |r|\, dr\, d\theta &= \int_0^{2\pi} \int_0^{+\infty} \hat{f}(-r\omega(\theta)) e^{-irx\cdot\omega(\theta)} |r|\, dr\, d\theta \\ &= \int_0^{2\pi} \int_0^{+\infty} \hat{f}(r\omega(\theta)) e^{irx\cdot\omega(\theta)} |r|\, dr\, d\theta, \end{aligned}$$

where the last equality is obtained by making appropriate changes of variables in θ. Since this equality holds, it follows that $f(x) = \frac{1}{4\pi} \int_0^{2\pi} \int_{\mathbb{R}} \hat{f}(r\omega(\theta)) e^{irx\cdot\omega(\theta)} |r|\, dr\, d\theta$. For the Fourier slice theorem, we have

$$\begin{aligned} f(x) &= \frac{1}{4\pi} \int_0^{2\pi} \int_{\mathbb{R}} \hat{f}(r\omega(\theta)) e^{irx\cdot\omega(\theta)} |r|\, dr\, d\theta \\ &= \frac{1}{2(2\pi)^{\frac{3}{2}}} \int_0^{2\pi} \int_{\mathbb{R}} \widehat{\mathcal{R}_\theta f}(r)\, |r|\, e^{irx\cdot\omega(\theta)}\, dr\, d\theta, \end{aligned}$$

as desired. □

The inversion formula for the Radon transform remains valid under weaker conditions than those specified in Theorem 2.7. In imaging applications, the data are typically piecewise continuous and vanish outside a bounded set, which means they belong to $L^1(\mathbb{R}^2)$. However, it is well-known that this does not guarantee that $\hat{f}$ is absolutely integrable. As a result, the Fourier inversion formula requires careful interpretation in this case. It can be noted that such data are also square integrable, and, by the properties of the Fourier transform on $L^2(\mathbb{R}^2)$, it follows that

$$f = \lim_{\rho\to\infty} \frac{1}{2(2\pi)^{\frac{3}{2}}} \int_0^{2\pi} \int_{-\rho}^{\rho} \widehat{\mathcal{R}_\theta f}(r)\, |r|\, e^{irx\cdot\omega(\theta)}\, dr\, d\theta.$$

The inversion formula can be understood in relation to backprojection. Specifically, the radial integral in the formula is viewed as a filtering operation applied to the Radon transform, acting only on the affine parameter. The angular integral that follows is then interpreted as the backprojection of the filtered Radon transform. Consequently, the Radon inversion formula is often referred to as the filtered backprojection formula. The fact that the inverse problem is ill-posed is intrinsically linked to the filter that must be applied to reconstruct f. Specifically, it can be shown that the unbounded nature of $|r|$ leads to $\mathcal{R}^{-1}$ being unbounded. This aspect will be further explored in the following section.

The operation of backprojection has an alternative mathematical interpretation. We recall the following definition.

Definition 2.8. *If $(X, \langle\cdot,\cdot\rangle_X)$ and $(Y, \langle\cdot,\cdot\rangle_Y)$ are inner product spaces, given $A\colon X \to Y$ a linear map, the adjoint of A, denoted by $A^*\colon Y \to X$, is defined by the relation*

$$\langle Ax, y\rangle_Y = \langle x, A^*y\rangle_X, \quad \forall x \in X, y \in Y.$$

We now consider the spaces $L^2(\mathbb{R}^2)$ and $L^2([0,2\pi)\times\mathbb{R})$.

Particularly, $L^2([0,2\pi)\times\mathbb{R})$ consists of locally integrable functions h for which the square-integral,

$$\|h\|^2_{L^2([0,2\pi)\times\mathbb{R})} = \int_0^{2\pi}\int_{\mathbb{R}} |h(\theta,s)|^2\, ds\, d\theta \tag{2.5}$$

is finite. For simplicity, assume that f is a function of bounded support on $\mathbb{R}^2$ and h is a function of bounded support on $[0,2\pi)\times\mathbb{R}$.

Using the L^2-inner product for functions on $\mathbb{R}^2$ and the inner product for functions on $[0,2\pi)\times\mathbb{R}$ compatible with the L^2-norm defined in (2.5), we have

$$\begin{aligned}\langle \mathcal{R}f, h\rangle_{L^2([0,2\pi)\times\mathbb{R})} &= \int_0^{2\pi}\int_{\mathbb{R}} \mathcal{R}f(\theta,s)h(\theta,s)\, ds\, d\theta \\ &= \int_0^{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f(s\omega(\theta) + t\omega(\theta)^\perp)h(\theta,s)\, ds\, dt\, d\omega.\end{aligned}$$

Let $x = s\omega(\theta) + t\omega(\theta)^{\perp}$, so $s = x \cdot \omega(\theta)$. By interchanging the θ- and x-integrals, we obtain

$$\langle \mathcal{R}f, h\rangle_{L^2([0,2\pi)\times\mathbb{R})} = \int_{\mathbb{R}^2}\int_0^{2\pi} f(x)h(\theta, x\cdot\omega(\theta))\,d\theta\,dx = \langle f, \mathcal{R}^*h\rangle_{L^2(\mathbb{R}^2)}, \qquad (2.6)$$

where

$$\mathcal{R}^*h(x) = \int_0^{2\pi} h(\theta, x\cdot\omega(\theta))\,d\theta.$$

This is only a formal adjoint because it can be shown that the Radon transform does not extend to define a continuous map from $L^2(\mathbb{R}^2)$ to $L^2([0,2\pi)\times\mathbb{R})$. From the definition of $\mathcal{R}^*$, it follows that $\mathcal{R}^*\mathcal{R}f = 2\pi\mathcal{B}f$. Hence, $\mathcal{R}^* \neq \mathcal{R}^{-1}$.

The domain of the Radon transform can be extended to the space of tempered distributions. Specifically, since the Schwartz space $\mathcal{S}(\mathbb{R}^2)$ is a subspace of $L^2(\mathbb{R}^2)\cap L^1(\mathbb{R}^2)$, equation (2.6) remains valid if we consider $h \in \mathcal{S}(\mathbb{R}^2)$. Furthermore, it can be shown that the Radon transform is a continuous operator from $\mathcal{S}(\mathbb{R}^2)$ to $\mathcal{S}([0,2\pi)\times\mathbb{R})$. Here, $\mathcal{S}([0,2\pi)\times\mathbb{R})$ is the space of functions that can be extended to be smooth and $2\pi-$periodic in θ, and that decay (along with all derivatives in the second variable s) faster than any power of $\frac{1}{|s|}$ uniformly in θ. Starting from equation (2.6), it is possible to define $\mathcal{R}\colon \mathcal{S}'(\mathbb{R}^2) \to \mathcal{S}'([0,2\pi)\times\mathbb{R})$ as the operator such that

$$\langle \mathcal{R}f, h\rangle = \langle f, \mathcal{R}^*h\rangle \quad \forall f \in \mathcal{S}'(\mathbb{R}^2),\ h \in \mathcal{S}([0,2\pi)\times\mathbb{R}).$$

It has been demonstrated that this operator is well-defined and continuous with respect to the weak* topology on tempered distributions. Moreover, it can be verified that both the Fourier slice theorem and the Radon inversion formula also hold for tempered distributions. This definition is useful since, in Section 3.2, which focuses on microlocal analysis, the Radon transform of tempered distributions will be considered.

2.3 Ill posedeness of the inverse problem

Given an inverse problem, the first thing to consider when addressing it is whether it is well-posed in the sense of Hadamard, meaning it satisfies the following three conditions:

(a) Existence of a solution;

(b) Uniqueness of the solution;

(c) Continuous dependence of the solution on the data.

The inversion formula derived in Theorem 2.7 guarantees both the existence of a solution within the image of $\mathcal{R}$ and its uniqueness. We now proceed to examine the third condition.

We recall that the Radon inversion formula is given by

$$f(x) = \frac{1}{2(2\pi)^{\frac{3}{2}}} \int_0^{2\pi} \int_{\mathbb{R}} \widehat{\mathcal{R}_\theta f}(r)\, |r|\, e^{irx\cdot\omega(\theta)}\, dr\, d\theta.$$

As already observed, up to a constant, the function is reconstructed by applying backprojection to the Radon transform filtered in Fourier with the function $g(r) = |r|$. It can be observed that this operation is similar to differentiation. Indeed, if we had r instead of $|r|$, then, since $\partial_t \hat{h}(\xi) = (i\xi)\hat{h}(\xi)$, the formula would give

$$\frac{1}{2(2\pi)i} \int_0^{2\pi} \partial_s \mathcal{R} f(\theta, x\cdot\omega(\theta)) d\theta. \tag{2.7}$$

The first thing to observe is that if f is real-valued, then this function is purely imaginary. This indicates that the presence of the absolute value applied to r is significant and makes a substantial difference. Moreover, (2.7) represents, up to a constant factor, the backprojection of the radial derivative of the Radon transform, which suggests that the filter essentially makes objects more singular. This is consistent with the fact that the backprojection formula produces a blurring effect, smoothing out edges, as demonstrated by the example of the backprojection of the indicator function of the unit disk discussed earlier. In order to obtain a reconstruction, the edges must be somehow recovered, and this is achieved through the filtering process. However, this poses a challenge, as the filtering process also amplifies the errors in the data, which are inevitable in practical applications.

For this reason, the Radon inverse problem is ill-posed. However, it is still possible to achieve good reconstructions with appropriate techniques, as the problem is not severely ill-posed. In fact, there are different degrees of ill-posedness, which can be studied by analyzing the decay rate of the singular values obtained through the decomposition of the operator describing the problem, when such a decomposition exists.

Since the objects of interest are typically images, which can generally be viewed as compactly supported functions, the problem can be reduced to studying the Radon transform on $L^2(\mathbb{D})$, where $\mathbb{D}$ is the unit disk, with codomain $L^2([0, 2\pi) \times [-1, 1])$. Under these conditions, it is shown that the operator is compact, which allows for a singular value decomposition. Thus, there exists a singular system $\{\sigma_k, u_k, v_k\}$ associated with $\mathcal{R}$, such that

$$\mathcal{R} f = \sum_{k=1}^{+\infty} \sigma_k \langle f, u_k \rangle v_k,$$

where $\{u_k\}$ and $\{v_k\}$ are orthonormal systems in the domain and codomain of $\mathcal{R}$, respectively, and $\{\sigma_k\}$ are singular values of $\mathcal{R}$, which are strictly positive,

non-increasing, and decay to zero as $k \to +\infty$. In terms of the singular values, reconstruction via the inverse Radon transform takes the form

$$f = \sum_{k=0}^{+\infty} \frac{1}{\sigma_k} \langle y, v_k \rangle u_k, \tag{2.8}$$

where $y = \mathcal{R}f$. Therefore, when $y = \mathcal{R}f + \eta$, where η represents noise, the components of y corresponding to errors are multiplied by $\frac{1}{\sigma_k}$, and since the sequence given by σ_k decay to zero, these terms grow unbounded as k increases, leading to amplification of noise in the reconstruction. However, in this case, the decay of the singular values σ_k has been shown to be of a polynomial type, which moderates the severity of the ill-posedness and allows for controlled regularization techniques to improve reconstruction quality.

It is important to note that the division by the singular values in (2.8) is analogous to the amplification of higher frequencies by the factor $|r|$ in the filtered back-projection.

A potential solution to the issue of noise magnification is to apply a filter within the filtered back-projection process that attenuates high frequencies. Mathematically, this is equivalent to introducing damping factors ψ_k in the singular value decomposition formula (2.8). These factors, for large k, can counteract the effect of σ_k^{-1}. Various filters can be employed for this purpose, resulting in different forms of the filtered back-projection (FBP) algorithm. However, these methods are beyond the primary focus of this thesis and will not be explored in detail.

Moreover, the analysis carried out so far applies to the ideal, infinite-dimensional case. When approaching the problem from a practical perspective, additional complications inevitably arise. In particular, it is not feasible to perform measurements along an infinite number of lines, resulting in only a limited amount of information being available. Despite this limitation, it can be shown that the existence and uniqueness of solutions can still be guaranteed. Moreover, it is necessary to discretize the entire problem, which adds further complexity to the reconstruction process. The ill-posedness of the continuous problem carries over to the discrete case, and in addition to measurement errors, there are also errors due to sampling and the discretization of the problem. Despite these challenges, the implementation of reconstruction techniques developed in the continuous case, particularly the filtered back-projection algorithm, yields good reconstruction results.

Chapter 3

Limited-angle computed tomography

3.1 Introduction

Many tomographic problems involve limited data because it is often impractical to collect measurements from all directions around the object, or the focus is on only a small region, requiring data only from that part. As a result, different types of limited data problems arise. This thesis specifically addresses the case of limited-angle tomography, where the projection angles θ are restricted, though all values of s are allowed.

To formalize the inverse problem in the limited-angle case, we consider the restricted Radon transform. Specifically, let $U \subset [0, 2\pi)$ be an open non-empty subinterval representing the limited set of available projection angles. The restricted Radon transform is then defined as

$$\mathcal{R}_U f = \mathcal{R}f|_{U\times\mathbb{R}}.$$

Given the symmetry properties of the Radon transform and the fact that the parameter s continues to vary over all of $\mathbb{R}$, we can assume without loss of generality that the interval U satisfies the condition

$$U = U + \pi \mod 2\pi.$$

The objective is to recover an approximation of f from the noisy measurements of $\mathcal{R}_U f$, which involves solving the following inverse problem:

$$y = \mathcal{R}_U f + \eta,$$

where η represents the measurement errors. This section presents the main results related to this inverse problem. For a more detailed treatment, see, for instance, [29] and [17].

It is a classical result that $\mathcal{R}_U$ is injective on compactly supported distributions.

Theorem 3.1. *Let $f \in \mathcal{E}'$ and let $U \subset [0, 2\pi)$ as above. If $\mathcal{R}f(\theta, s) = 0$ for $\theta \in U$ and all $s \in \mathbb{R}$, then $f = 0$.*

Proof. Let $f \in \mathcal{E}'(\mathbb{R}^2)$. Since the Radon transform can be extended to tempered distributions, as shown in Section 2.2, and compactly supported distributions can be regarded as tempered distributions, as shown in Section 1.3, it follows that $\mathcal{R}f$ makes sense for $f \in \mathcal{E}'(\mathbb{R}^2)$. Assume $\mathcal{R}f(\theta, s) = 0$ for $\theta \in U$ and all $s \in \mathbb{R}$. By the Fourier slice theorem, which is true for $\mathcal{E}'(\mathbb{R}^2)$,

$$\hat{f}(\xi\omega(\theta)) = \frac{1}{(2\pi)^{\frac{1}{2}}}\widehat{\mathcal{R}_\theta f}(\xi) = 0,$$

for $\theta \in U$ and $\xi \in \mathbb{R}$. Hence, $\hat{f}$ is zero on the open cone

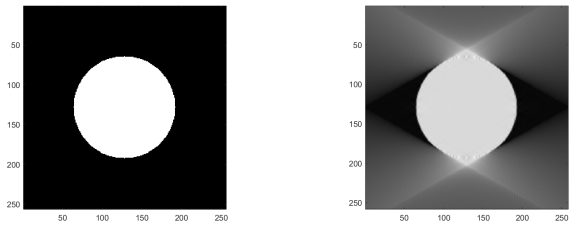
$$V = \{\xi\omega(\theta) \,:\, \xi \neq 0,\, \theta \in U\}.$$

Since f is a compactly supported distribution, by the Paley Wiener Schwartz theorem (see, e.g., [35]), $\hat{f}$ is analytic. Therefore, $\hat{f}$ must be zero everywhere since it is zero on the open set V. This shows $f = 0$. □

Therefore f is uniquely determined. However, the inverse problem of recovering f from $\mathcal{R}_U f$ is severely ill-posed, as evidenced by the exponential decay of its singular values (see [29]). This decay implies that the problem is highly sensitive to noise, making stable reconstruction challenging. This is connected to the fact that there is a complete lack of information for certain angular ranges, leading to gaps in the data.

As a consequence, in the limited-angle case, classical algorithms such as the filtered backprojection algorithm, which is applied by extending $\mathcal{R}_U f$ to zero, do not perform well. The following example illustrates this issue.

Example 3.1: Consider reconstructing a unit circle from noisy measurements using the filtered back-projection algorithm. In this example, no measurements are taken in the directions within the interval $[\frac{\pi}{3}, \frac{2\pi}{3}] \cup [\frac{4\pi}{3}, \frac{5\pi}{3}]$. The result is shown in Figure 3.1. From the figure, it is evident that the filtered back-projection algorithm in limited-angle tomography reconstructs only certain features of the object, while introducing additional artifacts, which are features in the reconstruction that do not correspond to actual singularities of the object. In particular, the poorly reconstructed regions correspond to the parts of the object whose boundaries are tangent to lines within the missing angular wedge. A partial explanation of this can be derived by examining the sinogram from which the reconstruction is derived. Specifically, the mathematical representation of the image is provided



(a) (b)

Figure 3.1: (a) Object to be reconstructed. (b) FBP reconstruction.

by the characteristic function of the unit disk in $\mathbb{R}^2$. It is easy to verify that the sinogram is given by

$$\mathcal{R}f(\theta, s) = \begin{cases} 2(1-s^2)^{\frac{1}{2}} & \text{if} \quad |s| \leq 1 \\ 0 & \text{if} \quad |s| > 1. \end{cases}$$

The function $\mathcal{R}f$ is smooth except at $s = \pm 1$, corresponding to the lines $L(\theta, \pm 1)$, which are tangent to the boundary of the disk (Figure 3.2). Thus, $\mathcal{R}f$ remains smooth as long as the integration line does not touch the boundary tangentially, even if it crosses interior singularities. Moreover, the irregularity in $\mathcal{R}f$ appears as a jump in the derivative rather than a discontinuity, due to the smoothing effect of $\mathcal{R}$.

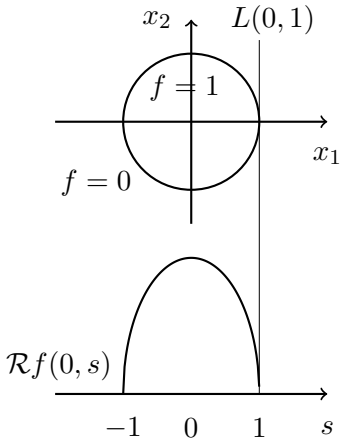


Figure 3.2: Graphical representation of the Radon transform obtained in Example 3.1 considering lines corresponding to $\theta = 0$.

This suggests that data obtained from lines tangent to the edges of the object carry crucial information about the boundaries themselves. Hence, critical information about boundaries that are tangent to lines in the missing angular range is lost, suggesting that recovering those features in the reconstruction may be challenging. This observation aligns with the result shown in Figure 3.1.

Developing a rigorous mathematical theory to formalize the observation derived from the example and thus establishing a general criterion for identifying

features that are difficult to reconstruct is essential, as it provides deeper insights into the reconstruction process and a clearer understanding of the problem. Since the structure and shapes of the image are determined by the singularities of the corresponding mathematical function, a detailed analysis of these is essential. To clarify the discussion, the following definitions are provided:

- visible singularities correspond to the boundaries or features of the object that can be accurately recovered from the data;
- invisible singularities are those that remain blurred or undetectable in the reconstruction;

The next section discusses microlocal analysis, which enables the classification of visible and invisible singularities based on the X-ray data collected. For a more in-depth treatment, see [30] and [31].

3.2 Microlocal analysis

Microlocal analysis provides a framework for identifying which singularities can be reliably reconstructed from limited data and which cannot be stably recovered, regardless of the quality of the algorithm. To accomplish this, the concept of singularities will be refined using the wavefront set. In this section, we will use the notation $\mathcal{F}f$ to denote the Fourier transform of a distribution f.

To understand the concept of wavefront set it is important to give some definitions and results. Given U an open set in $\mathbb{R}^2$, a distribution $f \in \mathcal{D}'$ is said to be C^∞ smooth on $U \subseteq \mathbb{R}^2$ if there is a smooth function $\psi \in C^\infty(U)$ such that

$$\langle f, \phi \rangle = \langle \psi, \phi \rangle = \int \phi(x)\psi(x)\, dx$$

for all test functions ϕ supported on U. This means that f is equal to a C^∞ function almost everywhere in U.

Definition 3.2. *Let $f \in \mathcal{D}'$. The singular support of f, denoted by* $\operatorname{ssupp}(f)$*, is the complement of the largest open set on which f is C^∞ smooth.*

Therefore, $x \notin \operatorname{ssupp}(f)$ if and only if there is a neighborhood U of x such that the restriction of f to U is a smooth function.

For instance, let f be the characteristic function of the closed unit disk $\bar{\mathbb{D}} \subset \mathbb{R}^2$, that is

$$f(x) = \begin{cases} 1 & \text{if } x \in \bar{\mathbb{D}} \\ 0 & \text{if } x \notin \bar{\mathbb{D}}. \end{cases}$$

Consider a point $x_0 \in \mathbb{D}$. Around such a point, we can find a small neighborhood that is fully contained inside $\mathbb{D}$. In this neighborhood, f is identically

equal to 1, which implies that f is smooth at x_0. Similarly, for any point x_0 outside the closed disk, f is constantly zero in a neighborhood around x_0, making it smooth there as well. However, at points on the boundary of the disk, f is not smooth, since f jumps discontinuously from 1 to 0. Therefore, the singular support $\operatorname{ssupp}(f)$ is the boundary of the disk.

Smoothness of a distribution $f \in \mathcal{E}'$ is related to the rapid decay of its Fourier transform. Recall the following definition and theorem.

Definition 3.3. *A function $g\colon \mathbb{R}^2 \to \mathbb{C}$ is rapidly decaying at infinity if for every $N \geq 0$, there is a C_N such that*

$$|g(x)| \leq C_N(1 + \|x\|)^{-N}$$

for all $x \in \mathbb{R}^2$.

Theorem 3.4. *A distribution $f \in \mathcal{E}'$ is C^∞ smooth if and only if its Fourier transform is rapidly decaying.*

This statement is well-founded because, for compactly supported distributions, the Fourier transform can be interpreted as a function, as outlined in Proposition 1.15. Consequently, it is appropriate to analyze the decay of its Fourier transform.

Theorem 3.4 relates global smoothness of f to rapid decrease of its Fourier transform. However, singularities are at specific points, so it is important to understand the behaviour of f locally near each of them. To do this at a particular point $x_0 \in \mathbb{R}^2$, we can multiply f by a smooth cut-off function ϕ that satisfies $\phi(x_0) \neq 0$. By considering the Fourier transform of the localized product $f\phi$ we can investigate whether this transform exhibits rapid decay in every direction, which in turn provides insights into the local smoothness of f around x_0.

Theorem 3.5. *A distribution $f \in \mathcal{E}'$ is C^∞ smooth at x_0 if and only if there exists a smooth cut-off function ϕ such that the localized Fourier transform $\mathcal{F}(\phi f)$ is rapidly decreasing.*

Note that this theorem implies that if the localized Fourier transform of f does not decay rapidly in a certain direction, then f is not smooth at x_0. By focusing on the specific directions in which this rapid decay fails, we obtain a more precise characterization of singularities, leading to the definition of the wavefront set.

Definition 3.6. *Let $f \in \mathcal{D}'$, $x_0 \in \mathbb{R}^2$ and $\xi_0 \in \mathbb{R}^2 \setminus \{0\}$. We say that $(x_0, \xi_0) \in \operatorname{WF}(f)$, the wavefront set of f, if for each cut-off function at x_0, $\phi \in \mathcal{D}$ with $\phi(x_0) \neq 0$, $\mathcal{F}(f\phi)$ does not decrease rapidly in any open conic neighborhood of the ray $\{t\xi_0 : t > 0\}$.*

The wavefront set of a distribution f is a closed subset of $\mathbb{R}^2 \times (\mathbb{R}^2 \setminus \{0\})$ and it is conic (if $(x, \xi) \in \operatorname{WF}(f)$ then so is $(x, a\xi)$ for any $a > 0$). Note that

it is a refinement of the singular support, in the sense that the singular support of a distribution is the set of points $x \in \mathbb{R}^2$, such that $(x, \xi) \in \mathrm{WF}(f)$ for some nonzero $\xi \in \mathbb{R}^2$.

For instance, consider the characteristic function of the unit disk. The singular support of this function is the boundary of the disk, as previously noted.The wavefront set, however, provides a more detailed picture: it consists of pairs (x, ξ) where x lies on the disk's boundary, and ξ is the direction normal to the boundary at x. A detailed calculation of the localized Fourier transform shows that only these normal directions contribute to the wavefront set $\mathrm{WF}(f)$. This aligns with the intuition, as these normal directions correspond to the sharpest transition, where f jumps most abruptly from zero to one, making them the "worst" directions for the discontinuity.

More generally, we can consider the characteristic function of any domain $\Omega \subset \mathbb{R}^2$ limited by smooth boundary S, denoted by χ_Ω. The following result can be proven.

Proposition 3.7. *Let $\Omega \subset \mathbb{R}^2$ be a region with smooth boundary S and let χ_Ω be the characteristic function of Ω. Then $\mathrm{WF}(\chi_\Omega) = \{(x, \xi) : x \in S, \xi \text{ normal to } S\}$.*

We have established that regularity is related to the decay rate of the Fourier transform. This relationship can also be described in terms of Sobolev spaces.

Definition 3.8. *A distribution is in the Sobolev space H^s microlocally near (x_0, ξ_0) if there exists a cut-off function $\phi \in \mathcal{D}$ with $\phi(x_0) \neq 0$ and function u which is homogeneous of degree zero (i.e., $u(\lambda\xi) = u(\xi)$ for all $\lambda > 0$ and $\xi \in \mathbb{R}^2$), smooth on $\mathbb{R}^2 \setminus \{0\}$ and with $u(\xi_0) \neq 0$, such that $u\mathcal{F}(\phi f) \in L^2(\mathbb{R}^2, (1 + |\xi|^2)^s)$.*

First, the distribution f is localized near x_0 by multiplying it by ϕ and then taking the Fourier transform. Next, microlocalization near ξ_0 is achieved by considering $u\mathcal{F}f$ and determining whether it belongs to $\mathcal{F}(H^s(\mathbb{R}^2))$. It follows from the definition that, if $(x_0, \xi_0) \notin \mathrm{WF}(f)$, then for all s, f is H^s near (x_0, ξ_0).

There is a very important theorem that connects the wavefront set of a distribution with that of its Radon transform. To understand this theorem, it is useful to view the wavefront set as a subset of the cotangent space $T^*(\mathbb{R}^2)$. We will consider the wavefront set of functions defined on $Y = [0, 2\pi] \times \mathbb{R}$. To address this, we extend functions h defined on Y periodically in θ and use localization functions ϕ with support within one period of θ. This allows us to treat ϕh as a function on $\mathbb{R}^2$, making it possible to compute the two-dimensional Fourier transform using these coordinates.

We define $d\theta$ and ds as the standard basis of $T^*([0, 2\pi] \times \mathbb{R})$, where $d\theta$ is the dual covector to $\frac{\partial}{\partial\theta}$, and ds is the dual covector to $\frac{\partial}{\partial s}$. The wavefront set is extended to distributions on $[0, 2\pi] \times \mathbb{R}$ using these local coordinates, and it forms a subset of $T^*([0, 2\pi] \times \mathbb{R})$.

One can think of $(x; \mathbf{r}\, dx)$ as representing the vector $(x; \mathbf{r})$, where x is a point in the plane and $\mathbf{r}$ is a tangent vector at x. Similarly, $(\theta, s; a\, d\theta + b\, ds)$ can be viewed as the vector $(\theta, s; a, b)$.

The fundamental theorem relating the wavefront set of a function to its Radon transform is as follows.

Theorem 3.9. *Let f be a distribution with compact support, $f \in \mathcal{E}'$. Let $x_0 \in L(\theta_0, s_0)$, $\bar{\theta}_0 = \omega(\theta_0)$, $\eta_0 = ds - (x_0 \cdot \theta_0^{\perp})\, d\theta$, and $a \neq 0$. The correspondence between $\mathrm{WF}(f)$ and $\mathrm{WF}(\mathcal{R}f)$ is*

$$(x_0; a\bar{\theta}_0\, dx) \in \mathrm{WF}(f) \text{ if and only if } (\theta_0, s_0; a\eta_0) \in \mathrm{WF}(Rf). \tag{3.1}$$

Given $(\theta_0, s_0; a\eta_0)$, $(\theta_0, s_0; a\bar{\theta}_0\, dx)$ is uniquely determined by (3.1). Moreover, f is in H^s microlocally near $(x_0; a\bar{\theta}_0\, dx)$ if and only if $\mathcal{R}f$ is in $H^{s+\frac{1}{2}}$ microlocally near $(\theta_0, s_0; a\eta_0)$.

The proof of this theorem relies on the Fourier Slice Theorem, which links the Fourier and Radon transforms globally. The main challenge in the proof, detailed in [30], lies in adapting this global relationship to a local context.

Theorem 3.9 provides an exact correspondence between singularities of f and those of Rf. In fact, the Radon transform $\mathcal{R}f$ for (θ, s) arbitrarily close to (θ_0, s_0) detects singularities of f whose directions in the wavefront set are perpendicular to the line $L(\theta_0, s_0)$, but not those in other directions. Moreover, the singularities of $\mathcal{R}f$ that are detected are of Sobolev order $\frac{1}{2}$ smoother than the corresponding singularities of f.

Building on these important results, we can connect them to the goal of image reconstruction in limited-angle computed tomography. In this context, the data $\mathcal{R}f(\theta, s)$ are known for all $s \in \mathbb{R}$ and $\theta \in U$ with $U \subset [0, 2\pi)$ and $U = U + \pi \mod 2\pi$. Therefore, the only singularities of f that affect the smoothness of $\mathcal{R}f$ are those corresponding to $(x, \omega(\theta))$ with $\theta \in U$. All other singularities, although not entirely absent, appear smoothed in the acquired data, because they are not conormal to any measured line. Moreover, we observe that the density function we aim to reconstruct can generally be considered piecewise smooth. For such functions, Theorem 3.7 provides a precise characterization of the wavefront set. Together with Theorem 3.9, this leads to the following visibility principle:

(i) If a boundary is tangent to a line in a limited data set, then that boundary should be easy to reconstruct from that limited data. Such boundaries are the visible boundaries.

(ii) If a boundary is not tangent to any line in a limited data set, then that boundary should be difficult to reconstruct from the limited data. Such boundaries are the invisible boundaries.

While this principle is based on the results regarding how singularities of f affect $\mathcal{R}f$ on a limited data set, it remains only partially justified. What is missing is the connection to the actual reconstruction process. That is provided for filtered back projection type reconstruction methods in [10].

Moreover, this principle is reflected in practical applications, as illustrated in Example 3.1, where the poorly reconstructed areas correspond to the boundaries that are tangent to the lines that are not acquired, as previously noted.

It is important to note that the visibility principle explains only part of the issue. A good algorithm should be able to reconstruct singularities more clearly if they are visible in the data set. However, other factors, such as noisy data or a bad algorithm, can have a significant impact on the reconstruction. In any case, the result does not predict how an algorithm will reconstruct invisible singularities. However, utilizing the classification of singularities into visible and invisible categories can be beneficial for developing an efficient reconstruction algorithm in limited-angle computed tomography. In fact, it is logical to adopt distinct approaches for visible and invisible boundaries, as the latter require more careful handling. One way to distinguish between the different boundaries is to utilize the image decomposition obtained through shearlets, which will be introduced in the next chapter.

Chapter 4

Shearlets

Applied harmonic analysis plays a crucial role in image reconstruction, since it provides the mathematical tools to analyze and represent images efficiently. A key aspect of many images is the sparsity of their essential information, meaning that critical features can be efficiently captured using only a few terms from an appropriate dictionary. Shearlets ([25]), in particular, excel at capturing directional information and anisotropic features like edges, which are crucial for reconstructing images. Furthermore, shearlets are capable of resolving the wavefront set of a distribution, and, given Theorem 3.9, this makes them highly relevant in the context of limited-angle tomography. Hence, analyzing an image through its shearlet decomposition provides a natural foundation for developing effective reconstruction algorithms from limited-angle computed tomography data. Before introducing shearlets, it is useful to briefly outline the theory of frames.

For a more detailed treatment, see, for instance, [13] and [5].

4.1 Frame Theory

When designing representation systems of functions, it is sometimes advantageous or necessary to go beyond the setting of orthonormal bases and consider redundant systems. In the case of orthonormal bases, the associated coefficient sequence is uniquely determined. However, with a redundant yet stable system, it is possible to choose sequences that are significantly sparser.

Definition 4.1. *A sequence $\{e_j : j \in J\}$ in a separable Hilbert space $\mathcal{H}$ is called a frame if there exist positive constants A, $B > 0$ such that for all $f \in \mathcal{H}$*

$$A\|f\|_{\mathcal{H}}^2 \leq \sum_{j\in J} |\langle f, e_j\rangle|^2 \leq B\|f\|_{\mathcal{H}}^2, \tag{4.1}$$

where $\|\cdot\|_{\mathcal{H}}$ denotes the norm in $\mathcal{H}$. Any two constants A, B satisfying (4.1) are called frame bounds.

The frame constants A and B are called *lower* and *upper frame bounds*, respectively. The supremum over all A and the infimum over all B such that the frame inequalities hold are the optimal frame bounds. If A and B can be chosen with $A = B$, then the frame is called $A-tight$. In particular, if $A = B = 1$ is possible, then $\{e_j : j \in J\}$ is a *Parseval frame*. An orthonormal basis is a specific type of Parseval frame. More broadly, frames generalize orthonormal bases by allowing for redundancy in their representation of vectors. Specifically, while orthonormal bases consist of vectors that are both orthogonal and linearly independent, frame elements do not need to satisfy these conditions.

Given a frame, data can be not only analysed but also synthesised using the frame coefficients $(\langle f, e_j\rangle)_{j\in J}$. To understand frames and reconstruction methods better, it is useful to study first some important associated operators.

Definition 4.2. *Given a separable Hilbert space $\mathcal{H}$, for any subset $\{e_j : j \in J\} \subseteq \mathcal{H}$, the coefficient operator or analysis operator C is given by*

$$Cf = (\langle f, e_j\rangle)_{j\in J} \quad \forall f \in \mathcal{H}.$$

The synthesis operator or reconstruction operator D is defined for a finite sequence $c = (c_j)_{j\in J}$ by

$$Dc = \sum_{j\in J} c_j e_j \in \mathcal{H},$$

and the frame operator S is defined on $\mathcal{H}$ by

$$Sf = \sum_{j\in J} \langle f, e_j\rangle e_j \quad \forall f \in \mathcal{H}. \tag{4.2}$$

Proposition 4.3. *Suppose that $\{e_j : j \in J\}$ is a frame for $\mathcal{H}$.*

(i) C is a bounded operator from $\mathcal{H}$ into $\ell^2(J)$ with closed range.

(ii) The operators C and D are adjoint to each other; that is, $D = C^*$. Consequently, D extends to a bounded operator from $\ell^2(J)$ into $\mathcal{H}$ and satisfies

$$\left\| \sum_{j\in J} c_j e_j \right\|_{\mathcal{H}} \leq B^{\frac{1}{2}} \|c\|_2 \quad \text{for } j \in J,$$

where $\|\cdot\|_2$ denotes the ℓ^2 norm.

(iii) The frame operator $S = C^*C = DD^*$ maps $\mathcal{H}$ onto $\mathcal{H}$ and is a positive invertible operator satisfying

$$AI_{\mathcal{H}} \leq S \leq BI_{\mathcal{H}} \quad \text{and} \quad B^{-1}I_{\mathcal{H}} \leq S^{-1} \leq A^{-1}I_{\mathcal{H}}.$$

In particular, $\{e_j : j \in J\}$ is a tight frame if and only if $S = AI_{\mathcal{H}}$.

(iv) The optimal frame bounds are $B_{opt} = \|S\|_{op}$ *and* $A_{opt} = \|S^{-1}\|_{op}^{-1}$*, where* $\|\cdot\|_{op}$ *denotes the operator norm of* S*.*

Proof. (i) The statement is a direct consequence of the frame inequalities (4.1).

(ii) Let $c = (c_j)_{j\in J}$ be a finite sequence. Then

$$\langle c, Cf\rangle = \sum_{j\in J} \overline{\langle f, e_j\rangle} c_j = \langle \sum_{j\in J} c_j e_j, f\rangle = \langle Dc, f\rangle, \quad \forall f \in \mathcal{H}.$$

Therefore, $D = C^*$. Moreover, since C is bounded on $\mathcal{H}$ and has operator norm $\|C\|_{\text{op}} \leq B^{\frac{1}{2}}$, it follows that $D = C^*$ is also bounded with the same operator norm.

(iii) The frame operator is $S = C^*C = DD^*$. It follows that S is self-adjoint and positive. Moreover, since

$$\langle Sf, f\rangle = \sum_{j\in J} |\langle f, e_j\rangle|^2,$$

it holds the operator inequality $AI_{\mathcal{H}} \leq S \leq BI_{\mathcal{H}}$. Hence S is invertible since $A > 0$. It follows also that $AS^{-1} \leq SS^{-1} \leq BS^{-1}$.

(iv) The optimal upper bound B_{opt} follows from the frame inequalities (4.1) and the fact that the operator norm of positive operator is given by $\|S\|_{\text{op}} = \sup\{\langle Sf, f\rangle : \|f\|_{\mathcal{H}} \leq 1\}$. The argument for A_{opt} is similar.

□

Statement (ii) shows that $\sum_{j\in J} c_j e_j$ is well defined for arbitrary sequences in $\ell^2(J)$, even though the frame vectors e_j are not orthogonal in general. The following corollary holds (the proof is omitted).

Corollary 4.4. *Let* $\{e_j : j \in J\}$ *be a frame for* $\mathcal{H}$*. If* $f = \sum_{j\in J} c_j e_j$ *for some* $c \in \ell^2(J)$*, then for every* $\epsilon > 0$*, there exists a finite subset* $F_0 = F_0(\epsilon) \subseteq J$ *such that*

$$\left\| f - \sum_{j\in F} c_j e_j \right\|_{\mathcal{H}} < \epsilon \text{ for all finite subsets } F \supseteq F_0.$$

We say that the series $\sum_{j\in J} c_j e_j$ *converges unconditionally to* $f \in \mathcal{H}$*.*

Unconditional convergence is a very important notion of convergence for non-orthogonal series over general and unstructured index set.

Another consequence of Proposition 4.1 is a reconstruction formula for f from the frame coefficients $\langle f, e_j\rangle$.

Corollary 4.5. *If $\{e_j : j \in J\}$ is a frame with frame bounds $A, B > 0$, then $\{S^{-1}e_j : j \in J\}$ is a frame with frame bounds B^{-1}, $A^{-1} > 0$, known as the dual frame. Every $f \in \mathcal{H}$ has non-orthogonal expansions:*

$$f = \sum_{j \in J} \langle f, S^{-1}e_j \rangle e_j \tag{4.3}$$

and

$$f = \sum_{j \in J} \langle f, e_j \rangle S^{-1}e_j, \tag{4.4}$$

where both sums converge unconditionally in $\mathcal{H}$.

Equation (4.3) provides a non-orthogonal expansion of f in terms of the frame vectors e_j, where the coefficients are the inner products of f with the dual frame vectors. In contrast, equation (4.4) represents the reconstruction of f from its frame coefficients, using the dual frame vectors as the expansion functions. For orthonormal bases and tight frames, the two expansion formulas are the same. In contrast to orthonormal bases, for general frames, the coefficients in the expansion provided by equation (4.3) are in general not unique. It can be verified that $(\langle f, S^{-1}e_j \rangle)_{j \in J}$ is the smallest in ℓ^2 norm of all expansion coefficients sequences. If the coefficients in (4.3) are unique, the frame is called a *Riesz basis* of $\mathcal{H}$.

For general frames, it might be difficult to use (4.3) or (4.4), since inverting S might be numerically unfeasible, in particular if the frame rate $\frac{B}{A}$ is large.

Note that the reconstruction of f can be calculated through the pseudo inverse. Property (4.1) guarantees that C is an injective and bounded operator. However, there is no guarantee that C is surjective. We define the pseudoinverse of C,

$$C^\dagger \colon \ell^2(J) \to \mathcal{H},$$

as the left inverse of C, such that

$$C^\dagger c = 0, \quad c \in \mathrm{Im}(C^\perp).$$

It can be proved that the pseudo inverse satisfies

$$C^\dagger = (C^*C)^{-1}C^*,$$

and

$$f = C^\dagger C f = \sum_{j \in J} \langle f, e_j \rangle S^{-1}e_j.$$

By using frames, we can achieve sparse representation of signals, where the degree of sparsity is measured by the decay rate of the error in the best N-term

approximation, which is obtained by selecting the indices of the N largest amplitude coefficients to represent the signal. In general, given S the frame operator (4.2) associated to a frame $\{e_i\}$, one can define the best $N-$term approximation of $f \in \mathcal{H}$ to be

$$f_N = \sum_{i \in I} \langle f, e_i \rangle S^{-1} e_i, \tag{4.5}$$

where I is the index-set of cardinality N associated with the $N-$largest coefficients $|\langle f, e_i \rangle|$. This is a non-linear approximation, in the sense that if f_N is an N-term approximation to $f \in \mathcal{H}$ with indices I_N and g_N is an N-term approximation to $g \in \mathcal{H}$ with indices J_N, then $f_N + g_N$ is only an N-term approximation to $f + g$ if $I_N = J_N$. On the contrary, if $J = \mathbb{N}$ and we consider the classical linear approximation $f_N = \sum_{i=1}^{N} \langle f, e_i \rangle S^{-1} e_i$, then the signal is projected over N vectors of the basis chosen a priori, which corresponds to a projection onto a vector subspace. Therefore, in this case, linearity holds.

4.2 Wavelet Theory

Shearlets originate from the concept of wavelets, making it important to first understand wavelet theory in order to comprehend shearlets. A wavelet system is a framework that allows us to decompose a signal into different scales and positions, particularly well-suited at processing one-dimensional signals with a limited number of pointwise singularities. In this section, we will introduce the fundamental aspects of wavelet theory. To explore the topic further, we refer to [27] and [22].

Definition 4.6. *Let $\psi \in L^2(\mathbb{R})$. For $f \in L^2(\mathbb{R})$, the continuous wavelet transform with respect to ψ is defined as*

$$\mathcal{W}_\psi f(u, a) := a^{-\frac{1}{2}} \int_{\mathbb{R}} f(t) \bar{\psi} \left(\frac{t - u}{a} \right) dt, \quad (u, a) \in \mathbb{R} \times \mathbb{R}_+.$$

If we define $\psi_{u,a}(t) = a^{-\frac{1}{2}} \psi \left(\frac{t-u}{a} \right)$, then the wavelet transform can be expressed as $\mathcal{W}_\psi f(u, a) = \langle f, \psi_{u,a} \rangle$. Moreover, by defining $\tilde{\psi}_a(t) = \bar{\psi} \left(-\frac{t}{a} \right)$, the continuous wavelet transform ca be reformulated as a convolution product

$$\mathcal{W}_\psi f(u, a) = f * \tilde{\psi}_a(u).$$

If the essential support of ψ is contained in $E \subseteq \mathbb{R}$ and E is centered at the origin, then $\operatorname{supp} \psi_{u,a} \subseteq u + aE$ is a neighborhood of u with size proportional to a. As a result, $\mathcal{W}_\psi f$ encodes local information about f at the point u, with the scale a indicating the resolution at which local details are captured. Based on these observations, it is natural to expect that, under certain conditions, the signal f can be reconstructed from its wavelet transform. Let us first see how to characterize the suitable functions ψ.

Definition 4.7. *A function $\psi \in L^2(\mathbb{R})$ such that $\|\psi\|_2 = 1$ is said to be a mother wavelet if*

$$C_\psi := \int_0^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi}\, d\xi < +\infty. \tag{4.6}$$

Given a mother wavelet ψ, we consider the family of functions $\psi_{u,a}$ and from the condition $\|\psi\|_2 = 1$, it follows that $\|\psi_{u,a}\|_2 = 1$ for all $u \in \mathbb{R}$, $a \in \mathbb{R}_+$.

Note that, according to Definition 4.7, a necessary condition for ψ to be a mother wavelet is that $\int_{\mathbb{R}^2} \psi(t)\, dt = 0$. Consequently, it follows that

$$\int_{\mathbb{R}^2} \psi_{u,a}(t)\, dt = 0,$$

for any $u \in \mathbb{R}$, $a \in \mathbb{R}_+$. This implies that the wavelet transform $\mathcal{W}_\psi f(u,a)$ measures the variation of f within a neighborhood of u of size a. Specifically, if f is smooth, its variation at u is minimal, leading to a small wavelet transform at u. Conversely, a singularity in f at u corresponds to a significant change in f, resulting in large values of $\mathcal{W}_\psi f(u,a)$. Therefore, the wavelet transform serves as a tool for measuring the local smoothness of functions.

Given ψ that satisfies condition (4.6), it is possible to reconstruct a signal $f \in L^2(\mathbb{R}^2)$ from its wavelet transform.

Theorem 4.8. *Let $\psi \in L^2(\mathbb{R})$ be a mother wavelet. Every function $f \in L^2(\mathbb{R})$ satisfies the following reconstruction formula*

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} \mathcal{W}_\psi f(u,a) \psi_{u,a}(t)\, du\, \frac{da}{a^2}. \tag{4.7}$$

Moreover,

$$\int_{\mathbb{R}} |f(t)|^2\, dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} |\mathcal{W}_\psi f(u,a)|^2\, du\, \frac{da}{a^2}.$$

Note that the definition of the wavelet transform can be extended to distributions. Specifically, if $f \in \mathcal{E}'$ and ψ is a smooth mother wavelet, the wavelet transform of the distribution f is defined as

$$\mathcal{W}_\psi f(u,a) = \langle f, \psi_{u,a} \rangle,$$

where, in this context, the notation $\langle \cdot, \cdot \rangle$ represents the action of the distribution f on a test function.

If ψ is smooth and f is a function or distribution that is regular everywhere except for a point singularity at u_0, then it can be verified that $\mathcal{W}f(u,a)$ has a rapid asymptotic decay as $a \to 0$ for all values of u, unless $u = u_0$. In this sense the continuous wavelet transform of f detect the location of the singularity through its

asymptotic decay at fine scales. Consequently, the continuous wavelet transform can be used to characterize the singular support of a function or distribution.

Given the continuous wavelet transform and its reconstruction formula, a discretized version can be derived. In particular, for a regular sampling of the wavelet transform we discretize the dilation parameter by taking a subgroup $\{\beta^j : j \in \mathbb{Z}\}$, with $\beta \in \mathbb{R}_+$. Given a scale β^j, we discretize the translation parameter by considering a subgroup $\alpha\beta^j\mathbb{Z}$, with $\alpha \in \mathbb{R}_+$. From the samples $\mathcal{W}_\psi f(\alpha\beta^j k, \beta^j)$ with $j, k \in \mathbb{Z}$, we obtain a wavelet system. Usually, it is considered a dyadic structure by choosing $\beta = 2$ and $\alpha = 1$. In particular, it is possible to find a mother wavelet ψ such that the resulting wavelet system forms an orthonormal basis for $L^2(\mathbb{R})$, which is a specific type of frame.

Definition 4.9. *An orthonormal wavelet is a function $\psi \in L^2(\mathbb{R})$, such that $\|\psi\|_2 = 1$ and the family of the translated and dilated functions*

$$\{\psi_{j,n} := 2^{\frac{j}{2}}\psi(2^j \cdot -n) : j, n \in \mathbb{Z}\}$$

is an orthonormal basis of $L^2(\mathbb{R})$.

If ψ is an orthonormal wavelet and $f \in L^2(\mathbb{R})$, then

$$f = \sum_{j,n\in\mathbb{Z}} \langle f, \psi_{j,n}\rangle \psi_{j,n}.$$

Note that, in frame theory, this corresponds to synthesizing a signal f using a Parseval frame. Indeed, in this case, we have

$$\sum_{j,n\in\mathbb{Z}} |\langle f, \psi_{j,n}\rangle|^2 = \|f\|_2^2.$$

As in the continuous case, the wavelet coefficients $\langle f, \psi_{j,n}\rangle$ carry local information of f. In particular, if ψ is supported near the origin, then $\psi_{j,n}$ is supported in a neighborhood of $\frac{n}{2^j}$ with the support size of order $O(2^{-j})$.

As mentioned above, wavelets are particularly effective for analyzing one-dimensional signals with isolated singularities. However, when they are extended to higher dimensions, their efficiency diminishes. Intuitively, multidimensional wavelets are supported on cubes that through the dilation parameter can only be changed in size and not shape. This isotropic nature of wavelets, while simple, lacks the directional sensitivity needed to effectively analyze multidimensional structures. As a result, they struggle to efficiently capture singularities that occur along curves rather than at points. In particular, the wavelet transform is unable to provide additional information about the geometry of the set of singularities of a function or distribution in terms of resolving the wavefront set. The same phenomenon can be illustrated using discrete wavelet transform. Therefore, wavelets

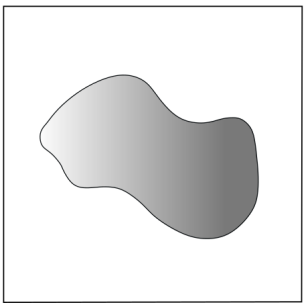


Figure 4.1: Example of a cartoon-like image represented using a grey scale map.

are not well-suited for dealing with images, which are typically multidimensional piecewise regular functions. The essential results, which highlight these issues, are summarized here; for further details, see [27] and [6].

A class of functions of particular interest in imaging is the class of the so-called *cartoon-like images*. This class provides a simplified model of natural images, which emphasizes anisotropic features.

Definition 4.10. *The class of cartoon-like images is defined as the set of functions* $f\colon \mathbb{R}^2 \to \mathbb{C}$ *of the form*

$$f = f_0 + f_1 \chi_B,$$

where $B \subset [0,1]^2$ *is a set with* ∂B *being a closed* C^2*-curve with bounded curvature, and* $f_i \in C^2(\mathbb{R}^2)$ *are functions with supp* $f_i \subset [0,1]^2$ *and* $\| \sum_{|\alpha| \leq 2} D^\alpha f_i \|_\infty \leq 1$ *for each* $i = 0, 1$.

An example of such a function is illustrated in Figure 4.1.

It has been shown in [6] that for any cartoon-like image f and for any positive integer N, there exists a triangulation of $[0,1]^2$ into N triangles such that the piecewise linear interpolation f_N satisfies

$$\|f - f_N\|_2^2 \lesssim N^{-2}, \quad \text{as } N \to \infty. \tag{4.8}$$

This result is optimal, meaning that this bound is the best possible. However, the approximation rate obtained using wavelet approximations f_N given by formula (4.5) is far from optimal, since it is $\mathrm{O}(N^{-1})$ (see [27]). While the result (4.8) provides an ideal error estimate, its practical usefulness can be limited. The key insight is that to achieve such optimal error decay, the supports of the functions in the dictionary must be adapted to align with the singularities of the function being analyzed. In this way the representations capture not only location but also directional information. The potential for improvement offered by this approach is further illustrated by Figure 4.2.

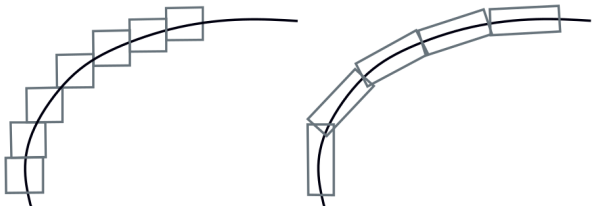


Figure 4.2: Approximation of a curve by isotropic basis elements and anisotropic basis elements.

4.3 Continuous Shearlet Systems

To overcome the limitations of wavelet systems in higher dimensions, shearlet systems have been developed as a more effective solution. The basic idea behind this is that the capability of elongating and orienting the supports of the functions in the dictionary along the singularity set of the function we are analyzing is fundamental in order to achieve the optimal error decay estimate. For a detailed treatment, see [6] and [22].

Before formally defining the shearlet system, it is important to introduce the key concepts behind its construction. Our observations from the previous section indicate that to achieve optimally sparse approximations of signals with anisotropic singularities, such as cartoon-like images, the analyzing elements must vary across multiple scales, orientations, and locations. This requires the use of a scaling operator to generate elements at different scales, an orthogonal operator to adjust their orientations, and a translation operator to position these elements across the 2D plane. We use a family of dilation operators D_{A_a} on $L^2(\mathbb{R}^2)$ defined as

$$D_{A_a}\psi(x_1, x_2) := |\det A_a|^{-\frac{1}{2}}\psi\left(A_a^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = a^{-\frac{3}{4}}\psi\left(a^{-1}x_1, a^{-\frac{1}{2}}x_2\right),$$

using scaling matrices A_a with $a \in \mathbb{R}_+$, given by

$$A_a = \begin{pmatrix}a & 0\\0 & a^{1/2}\end{pmatrix}.$$

The use of different dilation factors in the two directions enables shearlets to effectively capture anisotropic features, such as edges and singularities along curves. To properly orient the support of the generating functions, we need an orthogonal operator. Although rotations might appear to be the simplest choice, they disrupt the integer lattice structure of $\mathbb{Z}^2$ unless the rotation angles are restricted to $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$, or 2π. This issue complicates their application in discrete settings. As a more effective alternative, shearlet systems use shearing matrices, which preserve the integer lattice structure, defined as:

$$S_s = \begin{pmatrix}1 & s\\0 & 1\end{pmatrix}, \quad s \in \mathbb{R}.$$

The shearing operators are defined by

$$D_{S_s}\psi(x_1, x_2) := |\det S_s|^{-\frac{1}{2}}\psi\left(S_s^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) = \psi(x_1 - sx_2, x_2).$$

The translation operator is defined for any $t \in \mathbb{R}^2$ as

$$T_t\psi := \psi(\cdot - t).$$

As previously introduced, the shearlet systems are given by a combination of these three operators. Therefore, given a function $\psi \in L^2(\mathbb{R}^2)$, one can associate a corresponding shearlet system to it.

Definition 4.11. *For any $a \in \mathbb{R}_+$, $s \in \mathbb{R}$, and $t = (t_1, t_2) \in \mathbb{R}^2$, we define*

$$\psi_{a,s,t}(x_1, x_2) = T_t D_{A_a} D_{S_s}\psi(x_1, x_2) = a^{-\frac{3}{4}}\psi\left(A_a^{-1}S_s^{-1}\begin{pmatrix}x_1\\x_2\end{pmatrix} - \begin{pmatrix}t_1\\t_2\end{pmatrix}\right).$$

For $\psi \in L^2(\mathbb{R}^2)$, the continuous shearlet system is defined by

$$SH(\psi) := \{\psi_{a,s,t} : a \in \mathbb{R}_+, s \in \mathbb{R}, t \in \mathbb{R}^2\}.$$

In analogy to the wavelet case, the continuous shearlet transform maps a function $f \in L^2(\mathbb{R}^2)$ to its components along the shearlet system $\mathrm{SH}(\psi)$.

Definition 4.12. *For $\psi \in L^2(\mathbb{R}^2)$, the continuous shearlet transform of $f \in L^2(\mathbb{R}^2)$ is defined by:*

$$\mathcal{SH}_\psi f(a, s, t) := \langle f, \psi_{a,s,t}\rangle$$

where $a \in \mathbb{R}_+$, $s \in \mathbb{R}$, and $t \in \mathbb{R}^2$.

Of particular importance are the conditions on ψ under which the continuous shearlet transform is a multiple of an isometry, since this is automatically associated with a reconstruction formula. To formalize this, we introduce the notion of an admissible shearlet, analogous to the concept of admissibility defined for wavelets in (4.6).

Definition 4.13. *A function $\psi \in L^2(\mathbb{R}^2)$, with $\psi \neq 0$, is referred to as an admissible shearlet if*

$$\int_{\mathbb{R}^2} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2}\, d\xi_2\, d\xi_1 < +\infty.$$

The notion of admissible shearlets allows us to state sufficient conditions for a reconstruction formula in $L^2(\mathbb{R}^2)$.

Theorem 4.14. *Let $\psi \in L^2(\mathbb{R}^2)$ be an admissible shearlet. Define*

$$C_\psi^+ := \int_0^{+\infty} \int_{\mathbb{R}} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2} \, d\xi_2 \, d\xi_1,$$

$$C_\psi^- := \int_{-\infty}^{0} \int_{\mathbb{R}} \frac{|\hat{\psi}(\xi_1, \xi_2)|^2}{\xi_1^2} \, d\xi_2 \, d\xi_1.$$

If $C_\psi := C_\psi^+ = C_\psi^- < +\infty$, then for every $f \in L^2(\mathbb{R}^2)$, the shearlet transform $\mathcal{SH}_\psi$ satisfies

$$\int_{\mathbb{R}^2} \int_{\mathbb{R}} \int_{\mathbb{R}_+} |\mathcal{SH}_\psi f(a, s, t)|^2 \, \frac{da}{a^3} \, ds \, dt = C_\psi \int_{\mathbb{R}^2} |f(x_1, x_2)|^2 \, dx_1 \, dx_2. \tag{4.9}$$

In particular, if $C_\psi = 1$, then SH_ψ is an isometry.

The theorem provides sufficient conditions under which the shearlet transform becomes a multiple of an isometry. From this, a reconstruction formula analogous to the one for wavelets given in (4.7) can be derived.

It is quite straightforward to construct examples of admissible shearlets. Essentially, any function ψ such that its Fourier transform $\hat{\psi}$ is compactly supported away from the origin qualifies as an admissible shearlet.

A notable class of admissible shearlets satisfying the assumption of the theorem is the class of the so-called *classical shearlets*. These shearlets are band-limited, indicating that they have compact support in the frequency domain, as illustrated in Figure 4.3.

Definition 4.15. *A function $\psi \in L^2(\mathbb{R}^2)$ is said to be a classical shearlet if it is defined by*

$$\widehat{\psi}(\xi_1, \xi_2) = \widehat{\psi}_1(\xi_1) \widehat{\psi}_2 \left(\frac{\xi_2}{\xi_1} \right),$$

where $\psi_1 \in L^2(\mathbb{R})$ satisfies

$$\sum_{j \in \mathbb{Z}} \left| \widehat{\psi}_1(2^{-j}\xi) \right|^2 = 1 \quad \text{for a.e. } \xi \in \mathbb{R},$$

with $\widehat{\psi}_1 \in C^\infty(\mathbb{R})$ and $\operatorname{supp} \widehat{\psi}_1 \subseteq \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right]$, and $\psi_2 \in L^2(\mathbb{R})$ satisfies

$$\sum_{k=-1}^{1} \left| \widehat{\psi}_2(\xi + k) \right|^2 = 1 \quad \text{for a.e. } \xi \in [-1, 1],$$

with $\widehat{\psi}_2 \in C^\infty(\mathbb{R})$ and $\operatorname{supp} \widehat{\psi}_2 \subseteq [-1, 1]$.

If ψ is a classical shearlet, then the associated shearlet transform $\mathcal{SH}_\psi$ is an isometry.

Lemma 4.16. *Let $\psi \in L^2(\mathbb{R}^2)$ be a classical shearlets, then $C_\psi^+ = C_\psi^- = 1$.*

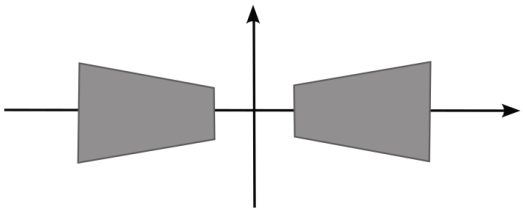


Figure 4.3: Support of the Fourier transform of a classical shearlet.

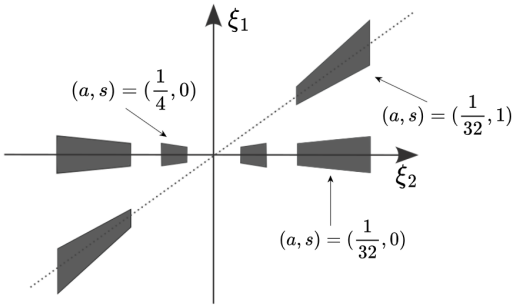


Figure 4.4: Support of classical shearlets in the frequency domain for different values of a and s.

4.4 Cone-Adapted Continuous Shearlet Systems

In practice, classical shearlet systems are not frequently used due to their inherent directional bias. To understand this, consider a classical shearlet ψ. As shown in Figure 4.3, classical shearlets are supported in the frequency domain within a pair of trapezoidal regions that are symmetric with respect to the origin. For a shearlet $\psi_{a,s,t}$, these trapezoids are oriented along a line with slope s, and as the scale parameter a approaches zero, the support of $\hat{\psi}_{a,s,t}$ becomes increasingly narrow. As s tends to $+\infty$, $\psi_{a,s,t}$ aligns with the vertical axis, but this alignment results in a highly elongated support, as suggests Figure 4.4. This elongation diminishes the ability of shearlets to accurately capture vertical information, representing a significant limitation for some applications.

For instance, if we consider a function f concentrated along the vertical axis in the frequency domain, the shearlet components increasingly aligns with the energy of f as the shearing parameter s increases towards $+\infty$. Hence, given the discussion above, the shearlets are not able to capture the information of the signal f effectively. Additionally, from a numerical perspective, this poses a challenge since the shearing parameter cannot take on values from an unbounded set in practical discretization.

To adress this problem, the *cone-adapted* construction has been developed, which leads to shearlet systems in which both the scale parameter a and the shearing parameter s are restricted over compact sets.

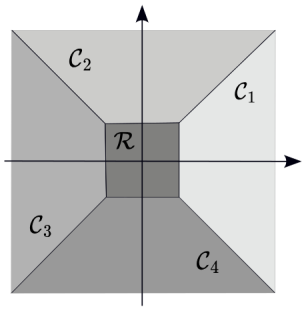


Figure 4.5: Partition of the frequency plane.

Let $f \in L^2(\mathbb{R}^2)$. The idea is to introduce a second shearlet system that covers the vertical cone in the frequency plane, rather than having only one system in the horizontal one. Additionally, a square region centered at the origin is considered to ensure full coverage in the frequency domain.

Therefore, the frequency plane is partitioned as illustrated in Figure 4.5. This approach restricts the shearing variable s to a limited range within each cone, ensuring that the resulting elements are more uniformly distributed in their orientations.

Thus, we define the following variant of the continuous shearlet systems.

Definition 4.17. *For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, the cone-adapted continuous shearlet system $SH(\phi, \psi, \tilde{\psi})$ is defined by*

$$SH(\phi, \psi, \tilde{\psi}) = \Phi(\phi) \cup \Psi(\psi) \cup \tilde{\Psi}(\tilde{\psi}),$$

where

$$\begin{aligned}
\Phi(\phi) &= \{\phi_t = \phi(\cdot - t) : t \in \mathbb{R}^2\}, \\
\Psi(\psi) &= \{\psi_{a,s,t} = a^{-\frac{3}{4}}\psi(A_a^{-1}S_s^{-1}(\cdot - t)) : a \in (0,1], |s| \leq 1 + a^{\frac{1}{2}}, t \in \mathbb{R}^2\}, \\
\tilde{\Psi}(\tilde{\psi}) &= \{\tilde{\psi}_{a,s,t} = a^{-\frac{3}{4}}\tilde{\psi}(\tilde{A}_a^{-1}S_s^{-T}(\cdot - t)) : a \in (0,1], |s| \leq 1 + a^{\frac{1}{2}}, t \in \mathbb{R}^2\},
\end{aligned}$$

and $\tilde{A}_a = \begin{pmatrix} a^{\frac{1}{2}} & 0 \\ 0 & a \end{pmatrix}$.

The function ϕ will be chosen to have compact frequency support near the origin, ensuring that the system $\Phi(\phi)$ is associated with the low-frequency region $\mathcal{R} = \{(\xi_1, \xi_2) : |\xi_1|, |\xi_2| \leq 1\}$. By selecting ψ according to the conditions specified in Definition 4.17, the system $\Psi(\psi)$ is associated with the horizontal cones $\mathcal{C}_1 \cup \mathcal{C}_3 = \{(\xi_1, \xi_2) : \left|\frac{\xi_2}{\xi_1}\right| \leq 1, |\xi_1| > 1\}$. Similarly, the shearlet $\tilde{\psi}$ can be chosen with the roles of ξ_1 and ξ_2 reversed, i.e., $\tilde{\psi}(\xi_1, \xi_2) = \psi(\xi_2, \xi_1)$. Then the system $\tilde{\Psi}(\tilde{\psi})$ is associated with the vertical cones $\mathcal{C}_2 \cup \mathcal{C}_4 = \{(\xi_1, \xi_2) : \left|\frac{\xi_2}{\xi_1}\right| > 1, |\xi_2| > 1\}$.

As for the continuous shearlet systems, an associated transform can be defined for cone-adapted continuous shearlet systems.

Definition 4.18. *Set*

$$\mathbb{S}_{cone} = \{(a, s, t) : a \in (0, 1],\ |s| \leq 1 + a^{\frac{1}{2}},\ t \in \mathbb{R}^2\}.$$

Then, for $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, the cone-adapted continuous shearlet transform of $f \in L^2(\mathbb{R}^2)$ is defined by

$$\mathcal{SH}_{\phi,\psi,\tilde{\psi}} f(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) = (\langle f, \phi_{t'} \rangle, \langle f, \psi_{a,s,t} \rangle, \langle f, \tilde{\psi}_{\tilde{a},\tilde{s},\tilde{t}} \rangle),$$

where $(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) \in \mathbb{R}^2 \times \mathbb{S}^2_{cone}$.

As previously observed, the continuous wavelet transform effectively characterizes the singular support of one-dimensional functions and distributions but falls short in capturing the geometry of singularities due to its isotropic nature. In contrast, the anisotropic shape of elements of cone-adapted continuous shearlet system enables the continuous shearlet transform to very precisely characterize the geometric properties of the singularity set. Indeed, it can be proven that the cone-adapted continuous shearlet transform precisely resolves the wavefront set for distributions ([21], [14]).

For illustration purposes, let us examine the linear delta distribution $\mu_p(x_1, x_2) = \delta(x_1 + px_2)$, $p \in \mathbb{R}$, defined by

$$\langle \mu_p, f \rangle = \int_{\mathbb{R}} f(-px_2, x_2)\, dx_2,$$

as a simple model for a distributed singularity. For simplicity, we assume that $|p| \leq 1$, so the singularity lies along a line in the frequency horizontal cone. Letting ϕ be a scaling function and $\psi, \tilde{\psi}$ be classical shearlets, the asymptotic analysis of its cone-adapted continuous shearlet transform $\mathcal{SH}_{\phi,\psi,\tilde{\psi}}\mu_p$ shows that this transform precisely identifies both the position and the orientation of the singularity by its decay behaviour at fine scales. In particular, the decay is slow if the shearlet's position and orientation align with a singularity. Specifically, we have the following result, whose proof can be found in [21].

Proposition 4.19. *Let $t' \in \mathbb{R}^2$ and $(\tilde{a}, \tilde{s}, \tilde{t}) \in \mathbb{S}_{\text{cone}}$ be fixed. For $t = (t_1, t_2)$ with $t_1 = -pt_2$ and $s = p$, we have*

$$\mathcal{SH}_{\phi,\psi,\tilde{\psi}}\mu_p(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) \sim a^{-\frac{1}{4}} \text{ as } a \to 0.$$

In all other cases, $\mathcal{SH}_{\phi,\psi,\tilde{\psi}}\mu_p(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t}))$ decays rapidly as $a \to 0$; that is, for all $N \in \mathbb{N}$, there is a constant C_N such that

$$\mathcal{SH}_{\phi,\psi,\tilde{\psi}}\mu_p(t', (a, s, t), (\tilde{a}, \tilde{s}, \tilde{t})) \leq C_N a^N \text{ as } a \to 0.$$

This property of resolving the wavefront set of a signal will be useful in the reconstruction algorithm discussed in Section 5.2.

4.5 Discrete Shearlet Systems

By sampling the continuous shearlet systems, various discrete shearlet systems can be constructed. This approach allows for the derivation of both standard discrete shearlet systems and cone-adapted variants. Given that the cone-adapted shearlet system is more commonly used in practical applications, our focus will be on this type.

Let $\phi \in L^2(\mathbb{R}^2)$ be a scaling function, and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be shearlet generators. Then the cone-adapted discrete shearlet system $\mathrm{SH}(\phi, \psi, \tilde{\psi})$ is defined by the union of the following sets:

$$\Phi(\phi) = \left\{\psi_{0,0,m,0} = \phi(\cdot - m) : m \in \mathbb{Z}^2\right\},$$
$$\Psi(\psi) = \left\{\psi_{j,k,m,1} = 2^{\frac{3}{4}j}\psi(S_k A_{2^j} \cdot - m) : j \geq 0, k \in \mathbb{Z}\ |k| \leq \lceil 2^{j/2} \rceil,\ m \in \mathbb{Z}^2\right\},$$
$$\tilde{\Psi}(\tilde{\psi}) = \left\{\psi_{j,k,m,-1} = 2^{\frac{3}{4}j}\tilde{\psi}(S_k^T \tilde{A}_{2^j} \cdot - m) : j \geq 0, k \in \mathbb{Z},\ |k| \leq \lceil 2^{j/2} \rceil,\ m \in \mathbb{Z}^2\right\},$$

with $A_{2^j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\frac{j}{2}} \end{pmatrix}$, and $\tilde{A}_{2^j} = \begin{pmatrix} 2^{\frac{j}{2}} & 0 \\ 0 & 2^j \end{pmatrix}$.

Notice that the system Φ is associated to the low frequency region, and Ψ and $\tilde{\Psi}$ are associated with the conic regions $\mathcal{C}_1 \cup \mathcal{C}_3$ and $\mathcal{C}_2 \cup \mathcal{C}_4$, respectively.

The corresponding discrete shearlet transform maps functions to the sequence of shearlet coefficients, hence it is simply the analysis operator.

Definition 4.20. *Set*

$$\Lambda := \{(j, k, m, \iota) : j \in \mathbb{N}, k \in \mathbb{Z}, |\iota| j \geq j \geq 0, |k| \leq |\iota| 2^{\lceil \frac{j}{2} \rceil}, m \in \mathbb{Z}^2, \iota \in \{-1, 0, 1\}\}.$$

Further, let $SH(\phi, \psi, \tilde{\psi})$ be a shearlet system as defined above. Then the associated cone-adapted discrete shearlet transform of $f \in L^2(\mathbb{R}^2)$ is the mapping defined by

$$f \in L^2(\mathbb{R}^2) \mapsto SH_{\phi,\psi,\tilde{\psi}} f(j, k, m, \iota) = \left(\langle f, \psi_{j,k,m,\iota} \rangle\right)_{(j,k,m,\iota) \in \Lambda}.$$

It has been shown in [16] that one can find band-limited shearlet generators $\psi, \tilde{\psi}$ and a shearlet scaling function ϕ such that the associated discrete one-adapted shearlet system defines a Parseval frame of $L^2(\mathbb{R}^2)$.

4.6 Compactly Supported Shearlets

Shearlet theory has impacted various applications for which sparse encoding or analysis of anisotropic features is crucial. However, while classical and band-limited shearlets have advantageous functional analytic properties, many applications require high spatial localization, making compactly supported shearlets

more desirable. This thesis does not delve into the construction of these shearlets, but a more detailed treatment can be found in [19]. It has been shown that there exist compactly supported generators $\psi, \tilde{\psi}$ such that the discrete cone-adapted shearlet system forms a frame in $L^2(\mathbb{R}^2)$. Moreover, although presumably Parseval frames can not be derived, the frame bounds are within a numerically stable range.

It was shown in [23] that there exist compactly supported shearlet generators $\psi, \tilde{\psi}$ and a shearlet scaling function ϕ, such that the best $N-$term approximation of a function f in the class of cartoon like images obeys

$$\|f - f_N\|^2 = O(N^{-2} \log^3(N)),$$

which is the optimal decay rate achievable, up to logarithmic factors. This provides a mathematical justification for the superiority of shearlet systems over wavelets, especially given that, as mentioned in Section 4.2, the approximation rate using wavelets is $O(N^{-1})$.

Chapter 5

Reconstruction algorithms for limited-angle computed tomography

Before delving into methods developed for image reconstruction from limited-angle tomography data, it is useful to introduce the notation that will be used in this chapter to describe the underlying model representing the problem. Starting with the Radon transform, we can exploit its symmetry property to allow the angle θ to vary within the interval $[-\pi/2, \pi/2)$, as this range is sufficient to capture all projections. We further assume that the Radon transform of f is known only within a subinterval of $[-\pi/2, \pi/2)$, specifically within $[-\gamma, \gamma]$ where $0 < \gamma < \pi/2$. Denoting this restricted transform as $\mathcal{R}_\gamma f = \mathcal{R}f|_{[-\gamma,\gamma]\times\mathbb{R}}$, the problem can be expressed as the recovery of f from

$$y = \mathcal{R}_\gamma f + \eta,$$

with η representing the error.

5.1 Deep learning approaches

As discussed in Chapter 3, the problem of image reconstruction from limited-angle computed tomography is highly ill-posed, making traditional methods inadequate for accurate recovery. This challenge has driven the exploration of alternative approaches, particularly with the development of artificial neural networks. While neural network-based algorithms are not the focus of this thesis, their widespread use and strong performance make it valuable to briefly introduce them, in order to clarify why the development of traditional algorithms remains important despite the advances in neural networks.

Artificial neural networks were first introduced in 1943 by McCulloch and Pitts as a means to develop learning algorithms that mimic the human brain.

Their initial goal was to create a theoretical foundation for artificial intelligence. However, early efforts were hampered by limited data and insufficient computing power, which prevented the training of networks with many layers.

Today, these limitations have been overcome. The availability of vast amounts of training data and significant advances in computing power have made it possible to train deep neural networks effectively. This progress has led to a resurgence of neural networks, yielding impressive results in various applications ([11], [2]).

From a mathematical perspective, a deep neural network in an idealized form is a high-dimensional function $\mathcal{NN}\colon \mathbb{R}^n \to \mathbb{R}^d$ of the form

$$\mathcal{NN}(x) = W_L(\sigma(W_{L-1}(\sigma(\dots(\sigma(W_1(x)))\dots)))),$$

where W_j are affine-linear functions and $\sigma\colon \mathbb{R} \to \mathbb{R}$ is the (non-linear) activation function applied componentwise. In essence, the goal of deep learning is to approximate an unknown structural relation between the input space $\mathbb{R}^n$ and the output space $\mathbb{R}^d$ using $\mathcal{NN}$. This is accomplished by determining the affine-linear functions W_j from the knowledge of training examples $(x_i, y_i)_{i=1}^N \subset \mathbb{R}^n \times \mathbb{R}^d$ that follow the underlying relation.

Despite the lack of complete theoretical understanding, deep learning has begun to penetrate many areas of applied mathematics, particularly in inverse problems within imaging sciences, where convolutional neural networks (CNNs) are the most prominent architecture. In fact, sophisticated model-based approaches are increasingly being outperformed by deep learning methods (see [1]).

Some approaches directly train deep neural networks for inversion from noisy measurements based on a collection of training samples following the forward model. Other methods aim to explicitly incorporate knowledge of the forward model into the reconstruction process.

In the case of limited-angle tomography, there are several techniques based on a direct inversion, followed by a post-processing procedure using neural networks to remove artifacts (see, for instance, [18], [15] and [36]).

Despite the impressive results yielded by these methods, they have potential drawbacks. Indeed, the lack of a well-defined mathematical framework for deep learning methods raises concerns about the reliability of reconstructions, particularly in medical contexts.

5.2 Learning the invisible

Given the limitations of neural networks, particularly the lack of control we have over the data-driven algorithms, combining them with classical methods can be more effective. This concept underpins the algorithm developed in [4], which will be detailed in this section. For further information we refer to the cited article.

The algorithm presented in [4] is a hybrid deep learning-shearlet framework designed for limited-angle computed tomography reconstruction. In particular, it integrates model-based sparse regularization using shearlets with a data-driven deep neural network to address the challenges posed by the ill-posed nature of the problem.

Inspired by Quinto's visibility analysis through microlocal analysis, introduced in Chapter 3, this framework leverages shearlet decomposition to separate the wavefront set of the image into a visible part, which can be recovered using classical methods, and an invisible part, missing from the measured data. The shearlet coefficients corresponding to the invisible boundaries are estimated using a trained neural network, with accuracy depending on the precision of the reconstruction of the visible part.

The algorithm employs cone-adapted and band-limited shearlets and consists of three steps.

First step - Recover the visible

The first step of the algorithm relies on Tikhonov regularization, where the penalty term is the ℓ_1-norm of the shearlet coefficients, promoting sparsity in the shearlet domain. This approach follows a common principle in regularization theory, which assumes that each class of data has a sparsifying representation system. In our case, sparsity is imposed in the shearlet domain because, as discussed in Chapter 4, shearlets are specifically designed for multivariate data and excel at capturing anisotropic structures, providing an optimal sparse approximation of cartoon-like images. Additionally, they unify continuous and discrete settings, making them well-suited for practical implementation.

Based on these observations, the following optimization problem is naturally formulated to obtain an approximation f^* from the measurements y:

$$f^* \in \arg\min_{f \geq 0} \left\| \mathrm{SH}_{\phi,\psi,\tilde{\psi}}(f) \right\|_{1,w} + \frac{1}{2} \left\| \mathcal{R}_\gamma f - y \right\|_2^2, \tag{5.1}$$

where $\mathrm{SH}_{\phi,\psi,\tilde{\psi}}$ is the discretized shearlet transform defined in Definition 4.20. The term $\|x\|_{1,w} = \sum_j w_j |x_j|$ denotes the weighted ℓ^1-norm of $x \in \ell^1(\Lambda)$, where Λ is as defined in Definition 4.20, and the weight vector $w \in \ell^2(\Lambda, \mathbb{R}^+)$ balances the influence between the shearlet regularizer and the ℓ^2-data fidelity term. This effectively incorporates the regularization parameter.

Additionally, in many tomographic problems, it is often known a priori that the desired image f is non-negative. Imposing this non-negativity constraint in equation (5.1) has been shown to lead to superior reconstruction results.

There are various possible methods to solve the optimization problem (5.1). In [4], the Alternating Direction Method of Multipliers (ADMM) is employed, and the resulting numerical reconstruction is denoted by $\boldsymbol{f}^*$.

According to Theorem 3.9 and the visibility principle introduced in Section 3.2, the boundaries of the image that are not tangent to any line in the limited-

angle data set are challenging to reconstruct, corresponding to what are known as invisible singularities. Consequently, the variational approach in (5.1) should only reconstruct boundaries that are visible within the angular range covered by the limited-angle data. In terms of shearlet coefficients, this implies that the coefficients corresponding to shearlets aligned with invisible boundaries of the solution f^* are negligible. It is important to note that the visible boundaries are entirely determined by the angular range $[-\gamma, \gamma]$ of the measurements, which can be conveniently expressed in the frequency domain using the Fourier slice theorem. This leads to the following definition, particularly relevant to band-limited shearlet constructions, where $\hat{\psi}$ has compact support.

Definition 5.1. *Let $SH_{\phi,\psi,\tilde{\psi}}$ denote a bandlimited, cone-adapted discrete shearlet system, and let $\gamma \in (0, \pi/2)$. The visible wedge is defined as:*

$$W_\gamma := \left\{\xi \in \mathbb{R}^2 : \xi = r \cdot (\cos\omega, \sin\omega)^T, r \in \mathbb{R}, |\omega| \leq \gamma\right\}.$$

Invisible shearlet indices are defined by

$$I_{inv} := \left\{(j,k,m,\iota) \in \Lambda : supp\, \hat{\psi}_{j,k,m,\iota} \cap W_\gamma = \emptyset\right\}, \tag{5.2}$$

and the visible indices by $I_{vis} := \Lambda \setminus I_{inv}$.

Since in the Fourier space the support of the invisible shearlets, i.e., $\psi_{j,k,m,\iota}$ with $(j,k,m,\iota) \in I_{\text{inv}}$, is disjoint from the visible cone, it follows from the Fourier slice theorem that such shearlets are contained in the kernel of the Radon transform in the Fourier space. From this, it can be deduced that

$$\mathcal{R}_\gamma \psi_{j,k,m,\iota} = 0 \quad \text{for} \quad (j,k,m,\iota) \in I_{\text{inv}}.$$

For shearlets where $\operatorname{supp} \hat{\psi}_{j,k,m,\iota} \not\subset W_\gamma$ and $\operatorname{supp} \hat{\psi}_{j,k,m,\iota} \cap W_\gamma \neq \emptyset$, the visibility classification may be less clear. Although $\mathcal{R}_\gamma \psi_{j,k,m,\iota} \neq 0$ in such cases, the contribution may still be negligible if most of the support lies outside the visible wedge W_γ. In numerical experiments, the condition in (5.2) is relaxed by classifying a shearlet as visible if its orientation, determined by shearing and anisotropic scaling, aligns with a visible direction of $\mathcal{R}_\gamma$.

This allows for a clear understanding of which shearlets align with the invisible directions, indicating the shearlet coefficients that do not provide reliable information.

The concept of invisible coefficients was introduced by Frikel in [9] for curvelet frames, which are similar to shearlets but based on rotation rather than shearing. It was demonstrated that invisible curvelet elements $\psi_j \in L^2(\mathbb{R}^2)$ satisfy $\psi_j \in \ker \mathcal{R}_\gamma$. Therefore, considering the synthesis-based ℓ_1-regularization

$$z^* \in \arg\min_z \|z\|_{1,w} + \frac{1}{2}\left\|\mathcal{R}_\gamma\left(\sum_{j\in\Lambda} \psi_j z_j\right) - y\right\|_2^2,$$

where Λ denotes the index set of the curvelet frame $(\psi_j)_{j\in\Lambda}$, it has been shown that coefficients associated with invisible curvelets satisfy $z_j^* = 0$, leading to an equivalent, lower-dimensional problem. Due to the similarities between shearlets and curvelets, these results extend to shearlet frames, as noted in [9]. Hence, it is expected that a similar statement holds true for the analysis-based minimization of (5.1), i.e. for an invisible shearlet index $(j,k,m,\iota) \in I_{\text{inv}}$ of a solution f^* of (5.1) it holds $\langle f^*, \psi_{j,k,m,\iota}\rangle \approx 0$.

Therefore, the shearlet coefficients $\text{SH}_{\phi,\psi,\tilde{\psi}}(f^*)$ from (5.1) can be divided into two sets:

(a) The visible coefficients, which are the coefficients corresponding to the indices I_{vis}, i.e., $\text{SH}_{\phi,\psi,\tilde{\psi}}(f^*)|_{I_{\text{vis}}}$, contain reliable information about edges that can be reconstructed based on the visibility principle outlined in Section 3.2.

(b) The invisible coefficients $\text{SH}_{\phi,\psi,\tilde{\psi}}(f^*)|_{I_{\text{inv}}}$ are penalized by the ℓ_1-norm and do not contain relevant information.

In summary, solving the minimization problem outlined in (5.1) yields a reconstruction f^* that lacks accuracy in the invisible regions, which correspond to the missing information from the limited-angle data. This leads to a distinction between well-reconstructed visible shearlet coefficients and invisible coefficients that are small and do not carry significant information.

At this point, we can observe that cone-adapted shearlets are capable of resolving the wavefront set of distributions, as described in Section 4.4, by analyzing the decay properties of the continuous shearlet transform. In the context of the discrete shearlet system, the rapid decay of the continuous transform manifests as sparsity in the associated shearlet coefficients, and when sorted properly, the coefficients belonging to one particular scale are reminiscent of a discretized version of the wavefront set of f^*, obeying similar structural properties in the phase space $\mathbb{R}^2 \times \mathbb{P}^1$. Thus, the structure of the shearlet coefficients of f^* reflects the wavefront set, but with gaps corresponding to the coefficients associated with the invisible shearlets.

Second Step - Learning the invisible

The gaps just discussed, corresponding to the invisible shearlet coefficients, can be estimated by an artificial deep neural network using the previously reconstructed visible coefficients. The network employed in this algorithm, called PhantomNet, is primarily based on U-Net [34], a specific type of convolutional neural network (CNN).

The core principle, as just mentioned, is that the shearlet coefficients of natural images adhere to specific structural rules, similar to those of a wavefront set in phase space. During training on a particular class of images, the parameter vector φ learns these general structural properties in the shearlet domain. When

applied to new data, the neural network estimates the invisible coefficients based on these learned patterns. This process can also be interpreted as a 3D inpainting task, where the neural network $\mathcal{NN}_\varphi$ inpaints the invisible shearlet coefficients. Therefore, denoting by $\boldsymbol{f}^*$ the reconstruction obtained by numerically solving the problem in (5.1), and by $\mathbf{SH}(\boldsymbol{f}^*)$ the corresponding digital shearlet coefficients, we have that

$$\boldsymbol{F} = \mathrm{NN}_\varphi(\mathbf{SH}(\boldsymbol{f}^*))$$

approximates the invisible coefficients of the image.

Third step - Combine both Parts

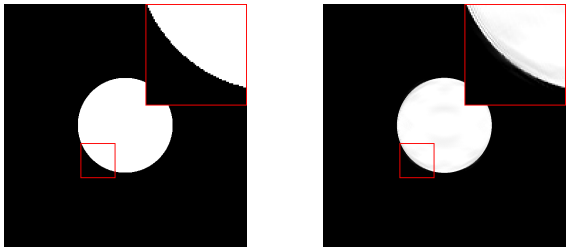
The final step involves combining the visible and estimated invisible components and mapping them back to the image domain. Hence, the final result is given by

$$\boldsymbol{f}_{\mathrm{LtI}} = \mathbf{SH}^T \left(\mathbf{SH}(\boldsymbol{f}^*)|_{I_{\mathrm{vis}}} + \boldsymbol{F}\right),$$

where $\mathbf{SH}^T$ denotes the synthesis operator and $\mathbf{SH}(\boldsymbol{f}^*)|_{I_{\mathrm{vis}}}$ denotes the visible digital shearlet coefficients. This process is supported by the fact that, under mild assumptions on the generator, a discrete cone-adapted shearlet system defines a Parseval frame of $L^2(\mathbb{R}^2)$. This means that a function $f \in L^2(\mathbb{R}^2)$ can be reconstructed as

$$f = \sum \langle f, \psi_{j,k,m,\iota} \rangle \psi_{j,k,m,\iota}.$$

The paper [4] presents several numerical tests demonstrating that this algorithm not only provides high-quality reconstructions but also adapts well to different image structures. In this thesis, however, only the reconstruction of a circular shape is included, using data acquired over a restricted angular range, as introduced in Example 3.1. This reconstruction is shown in Figure 5.1, alongside the original target image. It is evident that the boundary of the invisible part is reconstructed very well, and the image is almost completely cleared of artifacts.



(a) (b)

Figure 5.1: (a) Target image. (b) Reconstruction obtained using the algorithm.

To conclude, the deep neural network is employed solely to infer the invisible shearlet coefficients, focusing exclusively on estimating the truly invisible

boundary information. Indeed, the ℓ_1-minimization of (5.1), a well-established model-based method, directly contributes to the formation of the final image. Hence, by clearly defining the role of neural networks, this hybrid reconstruction framework gains clarity. Furthermore, since the network operates on a relatively accurate reconstruction of the visible coefficients, the estimation of the invisible information becomes easier.

This approach is particularly important in medical applications, where relying on a "black-box" CNN to remove artifacts without control over the modifications can be unsatisfactory.

For even greater reliability, an optimal approach would involve replacing the second step with a classical method for reconstructing the invisible part. This would completely eliminate dependence on neural networks and enhance the overall reliability of the reconstruction by relying entirely on well-understood and interpretable techniques. However, it is challenging to develop an algorithm that does not depend on neural networks while still delivering high-quality results. Work in this direction has been done in the paper [12], where a variational regularization framework is introduced, combining the advantages of different regularizers, particularly the ℓ^1-norm applied to curvelet coefficients and the TV norm. The algorithm developed in this thesis adopts a similar approach, though it is strongly based on the algorithm proposed in [4].

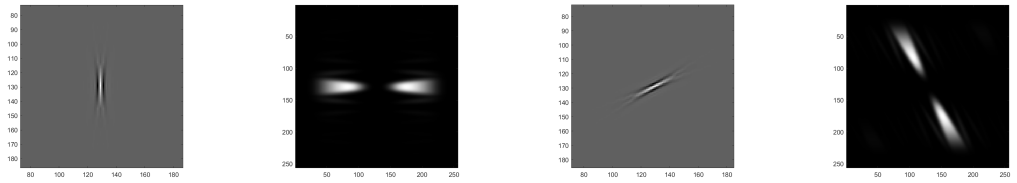
5.3 Developed algorithm based on a variational approach

In this section, we introduce an algorithm that replace neural networks with classical techniques for reconstructing the invisible part of the image. This study provides clearer insight into how intuitive, theoretically grounded techniques generally fail to achieve the same high-quality results obtained with the aid of neural networks. Consequently, it highlights the challenges in identifying efficient classical methods for this reconstruction problem.

The developed method builds upon the algorithm discussed in the previous section and, consequently, on the idea of differentiating between visible and invisible shearlet coefficients.

In this algorithm, cone-adapted and compactly supported shearlets are employed, which are ideal for analyzing directional features in images. Indeed, the compact support in the spatial domain allows them to precisely capture localized details such as edges and textures. However, due to the uncertainty principle, a function with compact support in space cannot also have compact support in the frequency domain. Despite this limitation, it is possible to construct cone-adapted and compactly supported shearlets that achieve good frequency localization. This

is feasible thanks to the design of the shearlet generator function and the application of a 2D fan filter in the Fourier domain, which creates a wedge-shaped support that enables them to efficiently focus on specific directions and frequencies. The digital shearlets used for numerical experiments, which are implemented in the `MATLAB` package ShearLab, are precisely based on this concept. In Figure 5.2, some shearlets of this type are shown both in the time and frequency domains.



(a) (b) (c) (d)

Figure 5.2: (a)-(c) Shearlets in the spatial domain. (b)-(d) Shearlets in the frequency domain.

Thanks to this construction, the concept of invisible shearlets can still be meaningfully applied, although it requires a more relaxed definition compared to Definition 5.1 for band-limited shearlets. Specifically, as is done when classifying semi-visible shearlets in Section 5.2, we classify a shearlet as invisible if its orientation aligns with an invisible direction in the frequency domain. In fact, in this case, its contribution to the visible wedge would be negligible because the support of the shearlet in Fourier space is concentrated along the invisible direction, with minimal overlap in the visible region. This is because, although these shearlets do not have compact support in the Fourier domain, their construction ensures that their energy is primarily concentrated in specific directions, making their influence outside these regions quite small (see Figure 5.2, (b) and (d)).

It is essential to note that the distinction between visible and invisible coefficients does not apply to the coefficients associated with low-frequency shearlets. Indeed, the shearlet scaling function, which is shown in Figure 5.3, does not clearly belong to either the visible or invisible categories, regardless of the specific missing wedge in the Radon transform. As a result, the coefficients tied to these shearlets contain information about both categories and should be treated with caution. This is particularly important because low-frequency shearlet coefficients reflect the overall structure of the signal and are generally associated with larger values.

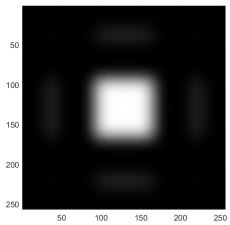


Figure 5.3: Low frequency shearlet in the frequency domain.

First step

The first step maintains the same underlying idea that forms the basis of the algorithm described in Section 3.2. However, in this case, the ℓ^1 norm used is not weighted but is the classical ℓ^1 norm. Hence, the minimization problem is given by:

$$f^* \in \arg\min_{f \geq 0} \frac{1}{2} \left\| \mathcal{R}_\gamma f - y \right\|_2^2 + \alpha \left\| \mathrm{SH}_{\phi,\psi,\tilde{\psi}}(f) \right\|_1,$$

where here $\mathrm{SH}_{\phi,\psi,\tilde{\psi}}$ denotes the discrete cone-adapted shearlet transform associated with the compactly supported shearlets, and $\alpha \in \mathbb{R}_+$ is the regularization parameter.

Second step

The second step builds upon the idea of differentiating between visible and invisible shearlet coefficients, as outlined in [4]. The idea is to maintain a certain fidelity to the visible shearlet coefficients from the reconstruction in the first step, given their reliability, while focusing on the reconstruction of the invisible coefficients.

This step also builds on the idea that the total variation (TV) norm has been proved to be a good prior for completing missing data.

The definition of total variation in the continuous case is provided.

Definition 5.2. *For a bounded domain $\Omega \subset \mathbb{R}^n$ and $u \in L^1(\Omega)$, the total variation of u is given by:*

$$\|u\|_{TV} = \sup \left\{ \int_\Omega u(x) \ \mathrm{div}\, v(x)\, dx : v \in C_0^1(\Omega), \|v\|_\infty \leq 1 \right\}.$$

For functions in the Sobolev space $W^{1,1}(\Omega)$, the existence of their weak derivative in $L^1(\Omega)$, allows the total variation to be expressed as

$$\|u\|_{TV} = \int_\Omega \|\nabla u(x)\|_1 \, dx.$$

This can be seen as the L^1 norm of the magnitude of the signal gradient, which captures all directional intensity changes. Penalizing this norm allows for discontinuities, making it effective for preserving sharp edges in reconstructions while

reducing noise. The total variation defined above is referred to as the *anisotropic* total variation, and it will be used in this algorithm. However, it is worth noting that other definitions of total variation exist. Specifically, given $u \in W^{1,1}(\Omega)$ the magnitude of the signal gradient can also be measured using the L^2 norm, which leads to what is known as the *isotropic* total variation:

$$\|u\|_{TV}^{\text{iso}} = \int_{\Omega} \|\nabla u(x)\|_2 \, dx.$$

The second step can be approached through two distinct formulations: analysis and synthesis. In the analysis formulation, the focus is on recovering the signal by analyzing its structural properties, working directly in the image domain. This leads to a lower-dimensional problem and naturally allows the inclusion of important constraints, such as non-negativity, which is crucial in tomography applications.

In contrast, the synthesis formulation reconstructs the unknown by synthesizing a solution that explains the observed data, operating in the shearlet domain. However, this approach results in a higher-dimensional problem compared to the analysis formulation, which reduces computational efficiency. Moreover, incorporating non-negativity constraints becomes particularly challenging. For these reasons, the analysis formulation is chosen, closely aligning with the algorithm developed in [4], on which this method is based.

The idea behind the second step consists of imposing a fidelity term on the visible coefficients and minimizing the total variation of the function to be reconstructed. Additionally, a certain level of sparsity is required for the invisible coefficients that need to be reconstructed. The problem can be formulated in the following way:

$$\arg\min_{f \geq 0} \frac{1}{2} \left\| \mathcal{M}_{\text{vis}} \text{SH}_{\phi,\psi,\tilde{\psi}} f - \mathcal{M}_{\text{vis}} \text{SH}_{\phi,\psi,\tilde{\psi}} f^* \right\|_2^2 + \beta_1 \|f\|_{TV} + \beta_2 \left\| \mathcal{M}_{\text{inv}} \text{SH}_{\phi,\psi,\tilde{\psi}} f \right\|_1.$$

Here, $\mathcal{M}_{\text{vis}}$ is an operator that maps $\ell^1(\Lambda)$ to $\ell^1(I_{\text{vis}})$, with Λ as defined in Definition 4.20 and I_{vis} given by the identification of visible shearlets outlined at the beginning of the section. Specifically, this operator extracts the visible coefficients from the shearlet representation of a function. On the other hand, $\mathcal{M}_{\text{inv}}$ performs the same operation for the invisible coefficients. Additionally, β_1 and β_2 are the regularization parameters that need to be selected appropriately.

As previously mentioned, special attention must be paid to the low-frequency coefficients, as these cannot be clearly distinguished as visible or invisible. A reasonable initial approach is to consider these coefficients as visible. However, this choice is not ideal, as evidenced by the results of the numerical tests presented

in Table 5.2 and Figure 5.5, which show that, despite an improvement in reconstruction compared to first step, the results remain suboptimal. This is partly due to the fact that considering the low-frequency coefficients as visible results in imposing a certain level of fidelity even on the poorly reconstructed areas, which are affected by artifacts. This issue is significant because the shearlet coefficients associated with low frequencies have a significant weight due to their much higher values compared to other coefficients.

One initial idea could be to exclude these coefficients from being considered visible. However, this approach has relevant drawbacks, as it would result in a lack of control over the rest of the image, leading to a loss of the reconstruction quality in the visible areas. A less drastic solution might involve including these coefficients in the fidelity term of the minimization problem while assigning them a weight that is less than one. This approach would allow for some control over the reconstruction while mitigating the impact on the "invisible area".

However, reducing the weight assigned to these coefficients is expected to result in less control over the visible part of the reconstruction. Coherently, the numerical experiments revealed that the results are highly sensitive to the choice of regularization parameters, which must be kept small to achieve satisfactory outcomes. Although this adjustment to the weight has resulted in some improvement in the metrics, the enhancements are not significant.

Numerical experiments and results

After appropriate discretization, we obtain the finite-dimensional measurement vector

$$\boldsymbol{y} = \boldsymbol{\mathcal{R}}_\gamma \boldsymbol{f} + \boldsymbol{\eta},$$

where $\boldsymbol{f} \in \mathbb{R}^{N^2}$ denotes the unknown discrete and vectorized image, the operator $\boldsymbol{\mathcal{R}}_\gamma \in \mathbb{R}^{M \times N^2}$ represents a discretized version of $\mathcal{R}_\gamma$, and $\boldsymbol{\eta} \in \mathbb{R}^M$ accounts for the discretized measurement noise. The algorithms described above have been implemented in `MATLAB`; however, before delving into the details, it is useful to provide information about the numerical setup.

In these tests, we specifically set $N = 256$ and $M = 44040$. The value of M is determined by the equation $M = J \cdot n$, where J represents the number of angle samples and n corresponds to the number of samples of the linear parameter s of the Radon transform. In the numerical tests we assume that the range of angles for the acquired lines is limited to $-\frac{\pi}{3}$ to $\frac{\pi}{3}$, sampling 120 angles uniformly within this interval. Additionally, we consider 367 samples for the linear parameter s, leading to the calculated value of M.

To prevent encountering the issue of an *inverse crime* (see [28]), special care is taken in how the data is generated and reconstructed. An inverse crime occurs when the data used for reconstruction is too closely tied to the assumptions or algorithms of the reconstruction method itself, leading to artificially perfect results.

For example, if the same computational grid is used for both the data simulation and the reconstruction, or if idealized, noise-free data is used, the results may show excellent reconstructions that are not representative of real-world problems, where noise and modeling errors are always present. To avoid this pitfall, the data used for reconstruction is not simply the sinogram obtained from the Radon transform of the original image. Instead, the tomographic data is generated by considering the image on a grid that is twice as fine as the original. This ensures that the simulated data does not resonate too closely with the reconstruction algorithm, providing a more realistic challenge and preventing overly optimistic results. Additionally, a Gaussian noise is added to the sinogram.

To continue the analysis from Chapter 3, the same two-dimensional unit circle image is used for the tests, consisting of values of 1 inside the circle and 0 outside, with a resolution of 256×256 pixels. This simplified case is chosen to evaluate the performance of different reconstruction methods, using various criteria to assess the quality of the results. The relative difference between the original and reconstructed images is evaluated using both ℓ^1 and ℓ^2 norms, where lower values indicate better reconstructions. Additionally, the Structural Similarity Index (SSIM) and Mean Structural Similarity Index (MSSIM) are considered. These metrics measure similarity by comparing textures and structures, with scores ranging from 0 to 1, where 1 indicates perfect similarity. HaarPSI (Haar Perceptual Similarity Index) is also used, which assess the perceptual similarity betweeen two images with respect to a human viewer (see [33]). Similar to SSIM, HaarPSI values range from 0 to 1, with higher values indicating better structural preservation and overall reconstruction quality.

As previously mentioned, the shearlets used in these tests are provided by the `MATLAB` package ShearLab. These shearlets are compactly supported and cone-adapted, making them ideal for analyzing directional features in images.

Since ShearLab 3D provides an implementation of the digital shearlet transform, it can be effectively employed to compute both the forward and inverse transforms of signals. For further details on the specific implementation, we refer to [24].

A 2D shearlet system is constructed by setting the values of the vector

$$\texttt{shearLevels} = [d_1, \ldots, d_s],$$

where the length of the vector s defines the number of scales. This vector specifies, for each scale, the fineness of the partitioning of the corresponding ring-like passband in the frequency domain.

More precisely, if d_j represents the j-th component of `shearLevels`, the range of the shearing parameter k at scale j is given by $|k| \leq 2^{d_j}$ for each cone. This generates $2(2 \cdot 2^{d_j} + 1)$ shearlets at each scale.

As each shearlet coefficient corresponds to one shearlet at a specific scale, shearing, and translation, the total number of coefficients computed by the shear-

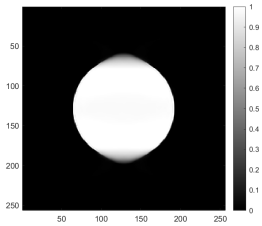


Figure 5.4: Recontruction obtained from Step 1.

let decomposition is $X \cdot Y \cdot R$, where X and Y denote the size of the signal (i.e., the number of translations), and R represents the redundancy of the shearlet system. This redundancy, which includes the low-frequency component, is given by:

$$R = 1 + \sum_{j=j_0}^{j_0+s-1} 2(2 \cdot 2^{d_j} + 1),$$

where j_0 represents the coarsest scale for the shearlet transform. Thus, the entire set of shearlet coefficients can be organized in a tensor with R layers along the third dimension. Fixing a "slice" of this tensor along the j-th scale yields a matrix. This matrix represents the shearlet coefficients at that particular scale and shearing, where each entry corresponds to a shearlet translated to a specific location. In this way, the position of the coefficient within the matrix reflects the translation of the shearlet in the spatial domain, while the j index determines the scale and shearing associated with that set of coefficients. In our case, the dimensions of the image $\boldsymbol{f}$ satisfy $X = Y = N$. Denoting by $\mathbf{SH}$ the digital version of $\mathrm{SH}_{\phi,\psi,\tilde{\psi}}$ we have $\mathbf{SH}(\boldsymbol{f}) \in \mathbb{R}^{N\times N\times R}$. In these tests, $N = 256$, and two scales are considered, with `shearLevels` $= [2, 1]$, meaning 2^2 shearings occur at the first scale and 2 at the second scale. Given this choice, the redundancy R is 25. Therefore, $\mathbf{SH}(\boldsymbol{f})$ corresponds to a tensor of dimensions $256 \times 256 \times 25$. The algorithms under study in this section are now described from a computational perspective.

First step

For clarity, the finite-dimensional version of the problem defining Step 1 is reported:

$$\boldsymbol{f}^* \in \arg\min_{\boldsymbol{f}\geq 0} \frac{1}{2} \left\|\boldsymbol{\mathcal{R}}_\gamma \boldsymbol{f} - \boldsymbol{y}\right\|_2^2 + \alpha \left\|\mathbf{SH}(\boldsymbol{f})\right\|_1, \tag{5.3}$$

where $\mathbf{SH}(\boldsymbol{f})$ is vectorized before applying the norm. The VMILA algorithm is used to solve the minimization problem (5.3). This algorithm will also be applied in the second step, as will be explained later. A brief overview of the VMILA algorithm is provided in the Appendix.

The reconstruction obtained from the sinogram of the circle image in the first step was found by conducting a high number of tests and identifying the parameter α that yielded the best reconstruction. Specifically, it was determined that this occurs with the parameter $\alpha = 5.3367$. The reconstruction obtained is shown in Figure 5.4. As expected from the theory, the circle is reconstructed quite accurately, except for the top and bottom areas where the boundary lines are tangent to the missing lines. In Table 5.1, the values of the metrics related to the first step reconstruction are reported.

Table 5.1: Values of the metrics for the Step 1 reconstruction

Metric	**SSIM**	**MSSIM**	**HaarPSI**	ℓ^1 **rel err**	ℓ^2 **rel err**
Value	0.9581	0.9777	0.3829	0.0010	0.1136

Second step

The finite-dimensional formulation is expressed as follows:

$$\arg\min_{\boldsymbol{f}\geq 0} \frac{1}{2} \left\| \boldsymbol{\mathcal{M}}_{\text{vis}} \mathbf{SH} \boldsymbol{f} - \boldsymbol{\mathcal{M}}_{\text{vis}} \mathbf{SH} \boldsymbol{f}^* \right\|_2^2 + \beta_1 \left\| \boldsymbol{f} \right\|_{TV} + \beta_2 \left\| \boldsymbol{\mathcal{M}}_{\text{inv}} \mathbf{SH} \boldsymbol{f} \right\|_1.$$

Here, $\boldsymbol{\mathcal{M}}_{\text{vis}}$ is an operator mapping from $\mathbb{R}^{N\times N\times R}$ to $\mathbb{R}^{N\times N\times I_{\text{vis}}}$, while $\boldsymbol{\mathcal{M}}_{\text{inv}}$ maps from $\mathbb{R}^{N\times N\times R}$ to $\mathbb{R}^{N\times N\times I_{\text{inv}}}$. Here, I_{vis} and I_{inv} are determined using the same principle proposed at the beginning of the section, but they are subsets of $\{1,\dots,R\}$. The tensors involved are vectorized to compute the norms.

In the discrete case, the TV norm is approximated using finite differences along the horizontal and vertical directions. In this thesis, we adopt the anisotropic TV norm, which is given by

$$\left\| \boldsymbol{f} \right\|_{TV} = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left| D_x \boldsymbol{f}_{i,j} \right| + \left| D_y \boldsymbol{f}_{i,j} \right| \right),$$

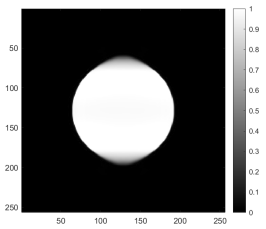
where $D_x \boldsymbol{f}_{i,j} = \boldsymbol{f}_{i+1,j} - \boldsymbol{f}_{i,j}$ and $D_y \boldsymbol{f}_{i,j} = \boldsymbol{f}_{i,j+1} - \boldsymbol{f}_{i,j}$ represent finite differences in the horizontal and vertical directions, respectively. This formulation, compared to the isotropic version, simplifies computation while still preserving sharp edges in reconstructed images, and it provides better control over regularization in each direction.

To solve this minimization problem, it is advantageous to use the VMILA algorithm. However, determining the regularization parameters presents a significant challenge, since the behavior of the solution varies irregularly with different parameter choices. Consequently, the values chosen were identified by conducting multiple tests and evaluating reconstructions based on the selected metrics. Specifically, 400 tests were conducted using various combinations of the parameters β_1 and β_2. As previously mentioned, the reconstruction quality is measured

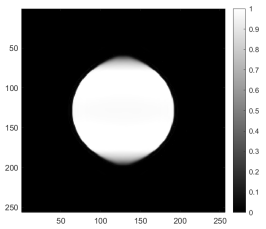
using SSIM, HaarPSI, MSSSIM, and ℓ^1 and ℓ^2 error. The parameters yielding the best results are presented in Table 5.2, with the best metric values highlighted in bold. The corresponding reconstructions are shown in Figure 5.5.

β_1	β_2	SSIM	MSSSIM	HaarPSI	ℓ^1 rel err	ℓ^2 rel err
0.0311	0.0084	0.9587	0.9790	**0.3676**	0.0558	0.1141
0.0437	0.0042	0.9560	**0.9792**	0.3569	0.0560	0.1142
0.0163	0.0147	**0.9604**	0.9785	0.3607	0.0551	**0.1131**
0.0311	0.0032	0.9571	0.9790	0.3659	**0.0549**	0.1133

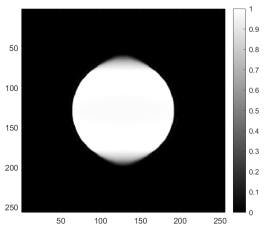
Table 5.2: Reconstruction metrics for different combinations of β_1 and β_2.



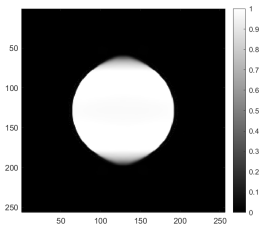
$\beta_1 = 0.0311, \beta_2 = 0.0084$



$\beta_1 = 0.0437, \beta_2 = 0.0042$



$\beta_1 = 0.0163, \beta_2 = 0.0147$



$\beta_1 = 0.0311, \beta_2 = 0.0032$

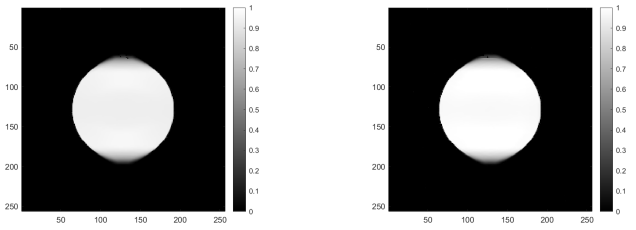
Figure 5.5: Reconstruction results for different combinations of β_1 and β_2.

The reconstructions obtained do not show a significant improvement compared to those from the first step, as the region corresponding to the "invisible" part remains inadequately reconstructed. However, it can be observed that there is an enhancement in the range of values in the image. As previously explained, it is reasonable to try to modify the minimization problem, particularly the fidelity term. The idea is to impose a weight lower than 1 to the shearlet coefficients associated to the low frequency region. Several tests were conducted, and as expected, lowering the weight given to these coefficients reduces control over the visible part of the result, making the outcomes very sensitive to the choice of regularization parameters, which need to be kept small to achieve good results.

Based on these tests, a weight of $\frac{1}{\sqrt{10}}$ was selected for the "low-frequency level" of the operator $\boldsymbol{\mathcal{M}}_{\text{vis}}$. This corresponds to assigning a weight of $\frac{1}{10}$ to the squared difference between each of the low-frequency shearlet coefficients of the function to be recovered and those from the reconstruction at the first step. As before, 400 tests were conducted to find appropriate parameters. The reconstructions are illustrated in Figure 5.6 and the associated metrics can be found in Table 5.3.

β_1	β_2	SSIM	MSSSIM	HaarPSI	ℓ^1 rel err	ℓ^2 rel err
0.0153	0.0058	**0.9681**	**0.9828**	**0.3925**	0.0914	0.1326
0.0153	0.0001	0.9666	0.9822	0.3866	**0.0518**	**0.1162**

Table 5.3: Reconstruction metrics for different combinations of β_1 and β_2.



$\beta_1 = 0.0153, \beta_2 = 0.0058$ $\beta_1 = 0.0153, \beta_2 = 0.0001$

Figure 5.6: Reconstruction results for different combinations of β_1 and β_2.

This modification of the weight has led to improved results in terms of metrics, although the improvements are not significant. However, this result confirms the theoretical insight that requiring fidelity to the low-frequency coefficients of the reconstruction from the first step can lead to issues with reconstruction quality. Despite this, it is not immediately clear how to resolve these issues, as these coefficients also contain important information related to the well-reconstructed part.

In general, in both approaches, the image quality does not improve significantly between the first and second steps, suggesting a limited capacity for improvement in the second step.

Appendix

A VMILA algorithm

In this section, we provide an overview of the VMILA algorithm, presented in [3], which is used in the algorithm described in Section 5.3.

We consider the problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{A.1}$$

with $f = f_1 + f_0$. It is assumed that f_1 is a proper, convex, lower semicontinuous function and f_0 is smooth on an open subset Ω_0 of $\mathbb{R}^n$ containing

$$\mathrm{dom}(f_1) = \{x \in \mathbb{R}^n : f_1(x) < +\infty\}.$$

It is also assumed that f_1 is bounded from below and that $\mathrm{dom}(f_1)$ is nonempty and closed.

If f_1 is the indicator function of a convex set Ω, denoted by ι_Ω, with

$$\iota_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega \\ +\infty & \text{if } x \notin \Omega, \end{cases}$$

then a simple and well-studied algorithm for the solution of (A.1) is the gradient projection method. Several variants of this algorithm has been proposed to accelerate the convergence, which for the basic implementation can be very slow. Particularly, reliable acceleration techniques have been proposed for the so-called gradient projection method with line-search along the feasible direction, whose iteration consists of

$$x^{(k+1)} = x^{(k)} + \lambda^{(k)} \left(y^{(k)} - x^{(k)} \right),$$

where $y^{(k)}$ is the Euclidean projection of the point $x^{(k)} - \nabla f_0(x^{(k)})$ onto the feasible set Ω, and $\lambda^{(k)} \in [0, 1]$ is a step length parameter ensuring the sufficient decrease of the objective function.

Variants of the basic scheme are obtained by introducing an additional variable step size parameter $\alpha^{(k)}$, which controls the step along the gradient, in combination with a variable choice of the underlying metric. In practice, the point $y^{(k)}$ can be defined as:

$$y^{(k)} = \arg\min_{y\in\Omega} \nabla f_0(x^{(k)})^T(y - x^{(k)}) + \frac{1}{2}\alpha^{(k)}(y - x^{(k)})^T D^{(k)}(y - x^{(k)}), \qquad \text{(A.2)}$$

where $\alpha^{(k)}$ is a positive parameter and $D^{(k)} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. The step sizes $\alpha^{(k)}$ and the matrices $D^{(k)}$ are considered as "free" parameters of the method, and a clever choice of them can lead to significant improvements in the practical convergence behavior.

The underlying idea of VMILA algorithm is to generalize the gradient projection scheme by introducing the concept of descent direction for the case where f_1 is a general convex function. In particular, the focus is on the case when the descent direction has the form $y^{(k)} - x^{(k)}$, with

$$y^{(k)} = \arg\min_{y\in\mathbb{R}^n} \nabla f_0(x^{(k)})^T(y - x^{(k)}) + d_{\sigma^{(k)}}(y, x^{(k)}) + f_1(y) - f_1(x^{(k)}),$$

where $d_{\sigma^{(k)}}(\cdot,\cdot)$ serves as a distance function, depending on the parameter $\sigma^{(k)}$ and is required to satisfy specific properties as outlined in [3]. It is evident that this formulation generalizes (A.2), which is recovered when $f_1 = \iota_\Omega$, by setting

$$d_\sigma(y, x) = \frac{1}{\alpha}(y - x)^T D(y - x), \qquad \text{(A.3)}$$

with $\sigma = (\alpha, D)$. The Euclidean norm associated with d_σ is defined as:

$$\|x\|_D = \sqrt{x^T Dx}.$$

We denote the extended real numbers set as $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and set $\Omega = \mathrm{dom}(f)$. For a given array of parameters $\sigma \in S \subseteq \mathbb{R}^q$, we define $h_\sigma \colon \mathbb{R}^n \times \mathbb{R}^n \to \bar{\mathbb{R}}$ as

$$h_\sigma(z, x) = \nabla f_0(x)^T(z - x) + d_\sigma(z, x) + f_1(z) - f_1(x) \quad \forall z, x \in \mathbb{R}^n.$$

It can be shown that, thanks to the conditions satisfied by d_σ, the function $h_\sigma(\cdot, x)$ is strongly convex and admits a unique minimum point for any $x \in \Omega$. Now, we introduce the following operator $p \colon \Omega_0 \to \Omega$ associated with any function h_σ of the form above:

$$p(x; h_\sigma) = \arg\min_{z\in\mathbb{R}^n} h_\sigma(z, x).$$

When d_σ is chosen as in (A.3), the operator becomes

$$p(x; h_\sigma) = \mathrm{prox}_{D,\alpha f_1}\left(x - \alpha D^{-1}\nabla f_0(x)\right),$$

where $\mathrm{prox}_{D,f}$ is the proximity or resolvent operator associated with a convex function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, in the metric induced by a symmetric positive definite matrix D:

$$\mathrm{prox}_{D,f}(x) = \arg\min_{z\in\mathbb{R}^n}\left(f(z) + \frac{1}{2}\|z - x\|_D^2\right) \quad \forall x \in \mathbb{R}^n.$$

At each iteration $x^{(k)}$, the value $y^{(k)}$ is given by $p(x^{(k)}; h_{\sigma^{(k)}})$. This minimizer may be difficult to compute in closed form. However, it is possible to compute an approximation of $y^{(k)}$. Specifically, the approach outlined in [3] has the advantage of offering an implementable condition for the approximate computation of the proximal point. Thus, the VMILA algorithm is based on a double-loop method, with the inner loop provided by an implementable stopping criterion. For further details, see [3].

Bibliography

[1] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

[2] Julius Berner, Philipp Grohs, Gitta Kutyniok, and Philipp Petersen. *The Modern Mathematics of Deep Learning*, pages 1–111. Cambridge University Press, 2022.

[3] Silvia Bonettini, Ignace Loris, Federica Porta, and Marco Prato. Variable metric inexact line-search-based methods for nonsmooth optimization. *SIAM J. Optim.*, 26(2):891–921, 2016.

[4] Tatiana A. Bubba, Gitta Kutyniok, Matti Lassas, Maximilian März, Wojciech Samek, Samuli Siltanen, and Vignesh Srinivasan. Learning the invisible: a hybrid deep learning–shearlet framework for limited angle computed tomography. *Inverse Problems*, 35(6):064002, 38, 2019.

[5] Ole Christensen. *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Cham, second edition, 2016.

[6] David Donoho. Sparse components of images and optimal atomic decompositions. *Constructive Approximation*, 17:353–382, 12 2001.

[7] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, Providence, RI, second edition, 2010.

[8] Gerald B. Folland. *Real analysis: Modern techniques and their applications*. John Wiley & Sons, Inc., New York, second edition, 1999.

[9] Jürgen Frikel. Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, 34(1):117–141, 2013.

[10] Jürgen Frikel and Eric Todd Quinto. Limited data problems for the generalized Radon transform in $\mathbb{R}^n$. *SIAM J. Math. Anal.*, 48(4):2301–2318, 2016.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[12] Simon Göppel, Jürgen Frikel, and Markus Haltmeier. Data-proximal complementary ℓ^1-TV reconstruction for limited data computed tomography. *Mathematics*, 12:1606, 2024.

[13] Karlheinz Gröchenig. *Foundations of time-frequency analysis*. Birkhäuser Boston, Inc., Boston, MA, 2001.

[14] Philipp Grohs. Continuous shearlet frames and resolution of the wavefront set. *Monatsh. Math.*, 164(4):393–426, 2011.

[15] Jawook Gu and J. C. Ye. Multi-scale wavelet domain residual learning for limited-angle ct reconstruction. *ArXiv*, abs/1703.01382, 2017.

[16] Kanghui Guo, Gitta Kutyniok, and Demetrio Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In *Wavelets and splines: Athens 2005*, Mod. Methods Math., pages 189–201. Nashboro Press, Brentwood, TN, 2006.

[17] Per Christian Hansen, Jakob Sauer Jørgensen, and William R. B. Lionheart, editors. *Computed tomography—algorithms, insight, and just enough theory*, volume 18 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2021.

[18] Eunhee Kang, Junhong Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical Physics*, 44:e360–e375, 2016.

[19] Pisamai Kittipoom, Gitta Kutyniok, and Wang-Q Lim. Construction of compactly supported shearlet frames. *Constr. Approx.*, 35(1):21–72, 2012.

[20] Venkateswaran P. Krishnan and Eric Todd Quinto. Microlocal analysis in tomography. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 847–902. Springer, New York, 2015.

[21] Gitta Kutyniok and Demetrio Labate. Resolution of the wavefront set using continuous shearlets. *Trans. Amer. Math. Soc.*, 361(5):2719–2754, 2009.

[22] Gitta Kutyniok and Demetrio Labate. Introduction to shearlets. In *Shearlets*, Appl. Numer. Harmon. Anal., pages 1–38. Birkhäuser/Springer, New York, 2012.

[23] Gitta Kutyniok and Wang-Q Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

[24] Gitta Kutyniok, Wang-Q Lim, and Rafael Reisenhofer. ShearLab 3D: faithful digital shearlet transforms based on compactly supported shearlets. *ACM Trans. Math. Software*, 42(1):Art. 5, 42, 2016.

[25] Demetrio Labate, Wang-Q Lim, Gitta Kutyniok, and Guido Weiss. Sparse multidimensional representation using shearlets. In *Wavelets XI*, volume 5914, pages 254–262, 2005.

[26] Giovanni Leoni. *A first course in Sobolev spaces*. American Mathematical Society, Providence, RI, 2009.

[27] Stéphane Mallat. *A wavelet tour of signal processing*. Academic Press, Inc., San Diego, CA, 1999.

[28] Jennifer L. Mueller and Samuli Siltanen. *Linear and nonlinear inverse problems with practical applications*. Computational Science & Engineering. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012.

[29] F. Natterer. *The mathematics of computerized tomography*, volume 32 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[30] Eric Todd Quinto. Singularities of the X-ray transform and limited data tomography in $\mathbf{R}^2$ and $\mathbf{R}^3$. *SIAM J. Math. Anal.*, 24(5):1215–1225, 1993.

[31] Eric Todd Quinto. An introduction to X-ray tomography and Radon transforms. In *The Radon transform, inverse problems, and tomography*, volume 63 of *Proc. Sympos. Appl. Math.*, pages 1–23. Amer. Math. Soc., Providence, RI, 2006.

[32] Eric Todd Quinto. Artifacts and visible singularities in limited data x-ray tomography. *Sensing and imaging*, 18(1):1–14, 2017.

[33] Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal processing. Image communication*, 61:33–43, 2018.

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[35] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[36] Hanming Zhang, Liang Li, Kai Qiao, Linyuan Wang, Bin Yan, Lei Li, and Guoen Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. *ArXiv*, abs/1607.08707, 2016.