2 Images, Transforms and Sampling

We commence this chapter with a short introduction to images in §2.1. We define continuous and discrete images, and the relation between the two. Next, in §2.2, §2.3 and §2.4 we introduce the Fourier, Radon and Walsh transforms, respectively. We first formulate each transform in the continuous setting, before addressing its discretization. For the first and third, we also discuss the relation to the Discrete Fourier and Discrete Walsh–Hadamard Transforms (DFT and DHT), respectively.

Much of this chapter serves to fix concepts and notation that will be used throughout the rest of the book. In doing so, we will assume a number of standard mathematical concepts. Further details are given in the Appendices.

2.1 Images

We consider *continuous* and *discrete* images. A *continuous* image is a function $f\colon [0,1]^d \to \mathbb{C}$. We allow f to be complex-valued, since it is relevant for some applications and mathematically no more difficult to treat. Depending on the application, the values f can take may be restricted. For instance, in the real case, f may be nonnegative ($f(x) \geq 0$), nonnegative and bounded ($0 \leq f(x) \leq T$ for some T), or even binary ($f(x) \in \{0,1\}$). We shall not discuss such considerations in any detail.

Typically, d is equal to one, two or three. We somewhat euphemistically refer to a function $f\colon [0,1] \to \mathbb{C}$ as an image, even though it is one-dimensional. Where needed, we will also consider f as a function over the whole of $\mathbb{R}^d$ that takes the value zero outside $[0,1]^d$.

This setup applies to *greyscale* images, which are our primary focus. Colour images can be represented as vector-valued functions; for example, $f\colon [0,1]^d \to \mathbb{R}^3$ in the case of an RGB (Red, Green, Blue) image.

A *discrete* image is a d-dimensional array of numbers:

$$X = (X_{i_1,\dots,i_d})_{i_1,\dots,i_d=1}^{N} \in \mathbb{C}^{N\times\cdots\times N}. \tag{2.1}$$

For simplicity, we consider only square arrays. We refer to N as the *resolution* of the image. Often, it will be convenient to reshape X into a vector. We do this in the standard way via *lexicographical ordering*. Formally, this is the bijection $\varsigma\colon \{1,\dots,N^d\} \to \{1,\dots,N\}^d$ whose inverse is given by

$$\varsigma^{-1}(i) = N^{d-1}(i_1 - 1) + N^{d-2}(i_2 - 1) + \cdots + i_d, \qquad i = (i_1,\dots,i_d). \tag{2.2}$$

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

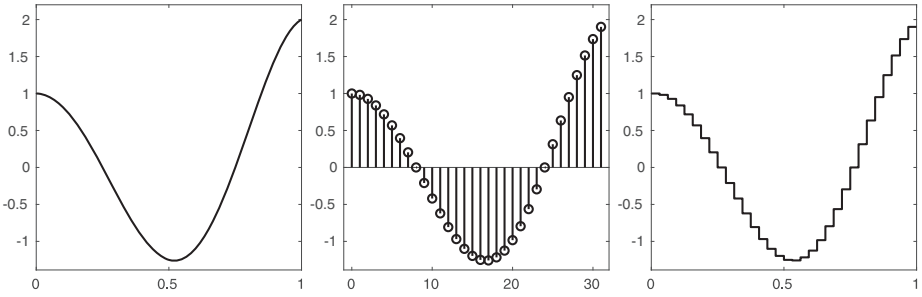

Figure 2.1 Left: A one-dimensional continuous image f. Middle: Its discretization $x \in \mathbb{C}^N$ at resolution $N = 32$. Right: The continuous image g given by (2.4).

Given X of the form (2.1), we define its *vectorization*

$$\mathrm{vec}(X) = x = (x_i)_{i=1}^{N^d} \in \mathbb{C}^{N^d}, \qquad x_i = X_{\varsigma(i)}. \tag{2.3}$$

Likewise, given $x \in \mathbb{C}^{N^d}$, we write $X = \mathrm{vec}^{-1}(x) \in \mathbb{C}^{N \times \cdots \times N}$ for its corresponding *anti-vectorization*.

A discrete image X can be converted to a continuous image g by

$$g = \sum_{j=1}^{N^d} x_j \chi_j, \qquad x = \mathrm{vec}(X). \tag{2.4}$$

We assume throughout that $\{\chi_j\}_{j=1}^{N^d}$ is the *pixel basis*. To be precise, if $\varsigma(j) = (i_1, \ldots, i_d)$, then χ_j is the indicator function of the corresponding cell

$$[(i_1 - 1)/N, i_1/N) \times \cdots \times [(i_d - 1)/N, i_d/N) \subset [0, 1]^d.$$

Conversely, a continuous image f can be *discretized*, giving a discrete image X, by sampling it on the corresponding equispaced grid of points:

$$X_{i_1,\ldots,i_d} = f((i_1 - 1)/N, \ldots, (i_d - 1)/N), \quad i_1, \ldots, i_d \in \{1, \ldots, N\}. \tag{2.5}$$

Throughout, we refer to X as the *discretization of f at resolution N*. Figure 2.1 gives an illustration of this process for $d = 1$.

There are other ways to convert between continuous and discrete images. For instance, one may replace the pixel basis by a smoother, but still local basis, or take local averages in (2.5) instead of pointwise samples. Different procedures lead to somewhat different *discretization errors*. However, for our purposes it will be sufficient to keep in mind that all discretizations commit errors, without worrying unduly about their specific nature.

2.1.1 Test Images

We use a number of different test images throughout this book to examine the various reconstruction procedures. These are shown in Fig. 2.2. The first two, 'SL phantom' and 'GLPU phantom', are synthetic brain images used to benchmark methods for MRI

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

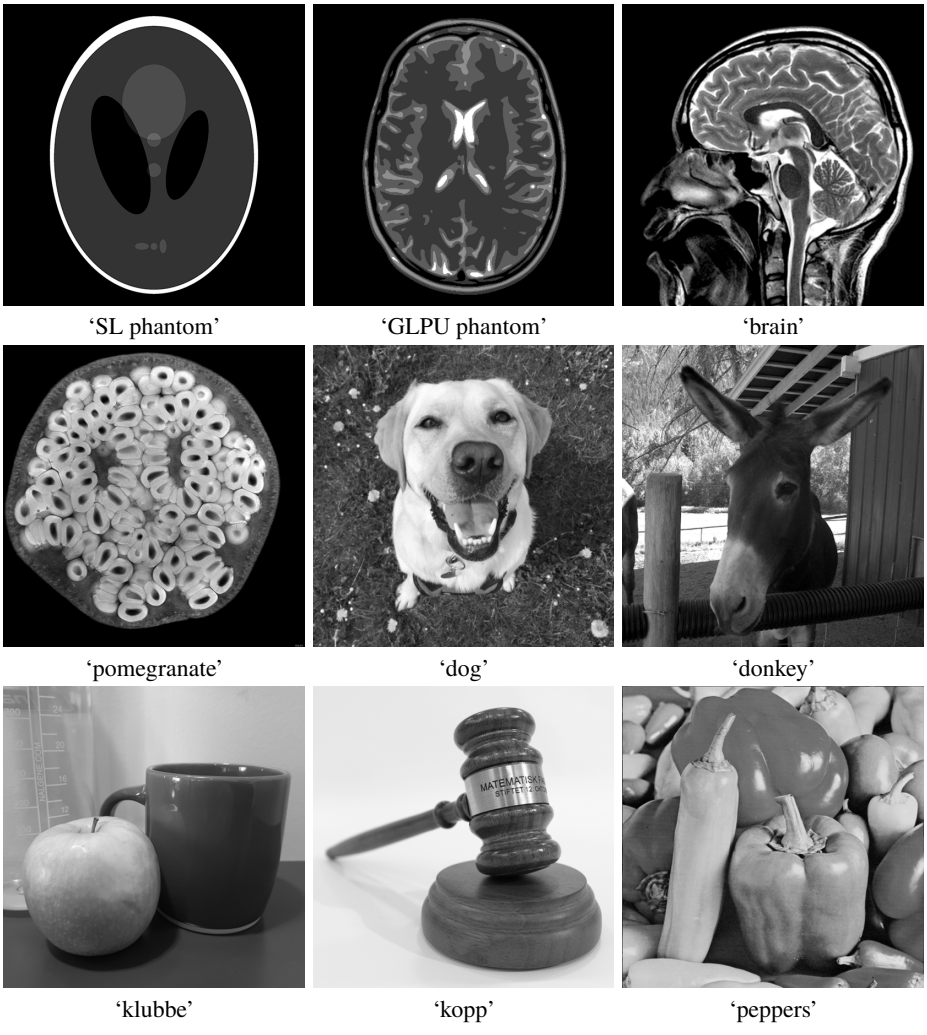

Figure 2.2 Test images. The original 'dog' image was provided by courtesy of Elizabeth Sawchuk, the 'kopp' and 'klubbe' images courtesy of Vegard Antun, the 'pomegranate' image courtesy of Andy Ellison, Boston University Medical School, and the 'brain' image courtesy of General Electric Healthcare. See the Notes section for more information about these and the other test images.

or X-ray CT reconstruction. The former is the famous *Shepp–Logan phantom* image, ubiquitous in imaging literature. They are both continuous images. The third and fourth, 'brain' and 'pomegranate', are discrete images obtained from actual MRI scans. The remaining images are natural images, also discrete, which will be used primarily to test reconstructions from binary measurements.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

2.1.2 Assessing Image Quality

The quality of a recovered image can be assessed in many different ways. In this book, we use either a simple visual comparison (informally known as the 'eyeball metric') or the *Peak Signal-to-Noise Ratio (PSNR)*. PSNR is defined as

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}_X}{\sqrt{\text{MSE}}}\right), \tag{2.6}$$

where MSE is the *Mean Squared Error*:

$$\text{MSE} = \frac{1}{N^d} \sum_{i_1,\ldots,i_d=1}^{N} |X_{i_1,\ldots,i_d} - \widehat{X}_{i_1,\ldots,i_d}|^2.$$

Here X is the true image (often called the *ground truth*), $\widehat{X}$ is its reconstruction and MAX_X is the maximum possible pixel value of the image. Often, $\text{MAX}_X = 255$. Pixel values are commonly represented as unsigned 8-bit integers, and hence they belong to the range $\{0, \ldots, 255\}$. PSNR is usually stated in decibel (dB) units, with a higher PSNR meaning better image quality. Typically, for good visual quality one wants the PSNR to be at least 30dB.

While PSNR is a widely used measure of image quality, it does have several limitations. It can be highly affected by, for instance, small rotations or shifts, even though such an operation barely changes the visual quality of the image. It also does not distinguish between different types of artefacts. Blurry artefacts, which are often more visually appealing, can easily carry the same PSNR as less visually appealing noisy artefacts. The design of more advanced image quality metrics is a subject in itself, and one we will not consider. The Notes section contains several references for the interested reader.

2.2 Sampling with the Fourier Transform

Having discussed images, we now move on to sampling and our first of three transforms, the Fourier transform. In this section, we make use of material in Appendix E.

2.2.1 The Fourier Reconstruction Problem

As noted in Chapter 1, a key problem in imaging is to recover a continuous image f from samples of its Fourier transform $\{\mathcal{F}f(2\pi\omega) : \omega \in \Omega\}$, where $\Omega \subset \mathbb{Z}^d$ is some finite index set. In the absence of noise, this is a problem of the form (1.3). Specifically:

$$\text{Given the data } \{\mathcal{F}f(2\pi\omega) : \omega \in \Omega\}, \text{ recover } f. \tag{2.7}$$

This is an infinite-dimensional inverse problem, which needs to be discretized in order to be solved numerically. We describe a standard discretization below.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

2.2.2 Fourier Orderings

Before doing so, we need a short discussion on orderings. Let $\{\upsilon_\omega\}_{\omega\in\mathbb{Z}^d}$ denote the d-dimensional Fourier basis (E.3), and observe that $\mathcal{F}f(2\pi\omega) = \langle f, \upsilon_\omega\rangle_{L^2}$. The Fourier basis is naturally indexed over $\mathbb{Z}^d$. However, to convert (2.7) into a standard discrete inverse problem of the form (1.1), it is useful to reindex it over $\mathbb{N}$. We write $\{\upsilon_i\}_{i=1}^\infty$ for the corresponding basis, and note that such reindexing is permitted since the Fourier series (E.4) of f converges unconditionally. To distinguish between the two indexings, we refer to $\omega \in \mathbb{Z}^d$ as the *frequency* and $i \in \mathbb{N}$ as the *index*.

It is useful to formalize, at least partially, how such reindexing is performed. When $d = 1$, we do this via the bijection

$$\varrho\colon \mathbb{N} \to \mathbb{Z},\ i \mapsto (-1)^i \lfloor i/2 \rfloor. \tag{2.8}$$

Note that this bijection satisfies

$$\varrho(\{2L+1,\dots,2M\}) = \{-M+1,\dots,-L\} \cup \{L+1,\dots,M\}, \tag{2.9}$$

for all $0 \le L < M$. In particular, for even N,

$$\varrho(\{1,\dots,N\}) = \{-N/2+1,\dots,N/2\}. \tag{2.10}$$

In other words, it gives a correspondence between the *first* N indices $i = 1,\dots,N$ and the *lowest* N frequencies $\omega = -N/2+1,\dots,N/2$.

When $d \ge 2$, there are many different ways to define a bijection $\varrho = \varrho^{(d)}\colon \mathbb{N} \to \mathbb{Z}^d$. Often, we do not worry about writing down the bijection explicitly as we did for the $d = 1$ case in (2.8). We do, however, assume that such a bijection takes the form

$$\varrho^{(d)}(i) = (\varrho(\varsigma(i)_1),\dots,\varrho(\varsigma(i)_d)), \quad i \in \mathbb{N}, \tag{2.11}$$

where ϱ denotes the one-dimensional bijection (2.8), ς is an arbitrary bijection $\mathbb{N} \to \mathbb{N}^d$ and $\varsigma(i)_k$ is the kth component of $\varsigma(i)$.

2.2.3 Discrete Fourier Measurements and the Fourier Matrix

Let $f\colon [0,1] \to \mathbb{C}$. Then

$$\mathcal{F}f(2\pi\omega) = \int_0^1 f(x)\mathrm{e}^{-2\pi\mathrm{i}\omega x}\,\mathrm{d}x \approx \frac{1}{N}\sum_{j=1}^N x_j \mathrm{e}^{-2\pi\mathrm{i}\omega(j-1)/N}, \tag{2.12}$$

where $x_j = f((j-1)/N)$ and $x = (x_j)_{j=1}^N$ is the discretization of f at resolution N. Assume that N is even. Restricting ω to the values $\{-N/2+1,\dots,N/2\}$, the expression (2.12) gives an approximation to $\mathcal{F}f$ at the lowest N frequencies. Using the bijection (2.8), and noting (2.10), we can write this as

$$(\langle f, \upsilon_{\varrho(i)}\rangle)_{i=1}^N \approx \frac{1}{N}Fx. \tag{2.13}$$

Here the matrix $F \in \mathbb{C}^{N\times N}$ is defined as follows:

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

Definition 2.1 (Fourier matrix) The one-dimensional *Fourier matrix* is the matrix $F = F^{(1)} \in \mathbb{C}^{N\times N}$ with entries

$$F_{ij} = \exp(-2\pi \mathrm{i}\varrho(i)(j-1)/N), \quad i,j = 1,\dots,N.$$

For $d \geq 2$, the d-dimensional Fourier matrix is given by

$$F^{(d)} = F^{(1)} \otimes \cdots \otimes F^{(1)} \in \mathbb{C}^{N^d \times N^d},$$

where $\otimes$ denotes the Kronecker product.

When clear, we simply write F for the d-dimensional Fourier matrix. Note that $N^{-d}F^*F = I$; that is, the Fourier matrix is unitary up to the scaling $1/\sqrt{N^d}$. As we discuss in Remark 2.2, F is closely related, although not identical, to the matrix of the *Discrete Fourier Transform (DFT)*.

Much as in the one-dimensional case (2.13), the d-dimensional Fourier matrix gives an approximation to the lowest N^d frequencies of a d-dimensional continuous image $f\colon [0,1]^d \to \mathbb{C}$. Let ς be the lexicographical ordering (2.2) and $\varrho = \varrho^{(d)}\colon \{1,\dots,N^d\} \to \{-N/2+1,\dots,N/2\}^d$ be the bijection defined by

$$\varrho^{(d)}(i) = (\varrho(\varsigma(i)_1),\dots,\varrho(\varsigma(i)_d)), \quad i = 1,\dots,N^d, \tag{2.14}$$

where ϱ is as in (2.8) and $\varsigma(i)_k$ denotes the kth component of $\varsigma(i)$. Let X be the discretization of f at resolution N and $x = \mathrm{vec}(X) \in \mathbb{C}^{N^d}$. Then

$$\begin{aligned}
\langle f, \upsilon_{\varrho(i)}\rangle &= \int_{[0,1]^d} f(x)\mathrm{e}^{-2\pi\mathrm{i}\varrho(i)\cdot x}\,\mathrm{d}x \\
&\approx N^{-d}\sum_{j=1}^{N^d} f\left(\frac{\varsigma(j)_1 - 1}{N},\dots,\frac{\varsigma(j)_d - 1}{N}\right)\exp\left(-2\pi\mathrm{i}\sum_{k=1}^{d}\frac{\varrho(\varsigma(i)_k)(\varsigma(j)_k - 1)}{N}\right) \\
&= N^{-d}\sum_{j=1}^{N^d} x_j \prod_{k=1}^{d} F^{(1)}_{\varsigma(i)_k,\varsigma(j)_k} = N^{-d}\sum_{j=1}^{N^d} F^{(d)}_{ij}x_j.
\end{aligned}$$

Here in the last step we used (A.21). This gives

$$\frac{1}{N^d}Fx \approx \left(\langle f, \upsilon_{\varrho(i)}\rangle\right)_{i=1}^{N^d}.$$

In other words, $N^{-d}Fx$ is an approximation to the Fourier transform $\mathcal{F}f(2\pi\omega)$ at the lowest N^d frequencies $\omega \in \{-N/2+1,\dots,N/2\}^d$.

Remark 2.2 The form of the Fourier matrix used in Definition 2.1 is convenient for several reasons. First, it arises directly in the discretization of the Fourier integral (2.12). Second, the ordering of its rows is useful for the mathematical analysis performed in later chapters. However, it is somewhat unconventional, and therefore worth explaining how it relates to the DFT $\mathcal{F}\colon \mathbb{C}^N \to \mathbb{C}^N$ (Definition E.1). The connection is as follows. For any $x \in \mathbb{C}^N$,

$$Fx = Q\mathcal{F}(Dx),$$

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

where $D \in \mathbb{C}^{N\times N}$ is the diagonal matrix with diagonal entries

$$D_{jj} = \exp(2\pi \mathrm{i}(N/2 - 1)(j-1)/N), \quad j = 1, \dots, N,$$

and $Q \in \mathbb{R}^{N\times N}$ is the permutation matrix with

$$Q_{ij} = \delta_{\varrho(i)+N/2-j}, \quad i, j = 1, \dots, N.$$

Hence Fx is the DFT of a phase-shifted version of x, followed by a permutation to account for the different indexing of frequency space. As a result, the matrix–vector multiplications $x \mapsto Fx$ and $y \mapsto F^* y$ can be implemented efficiently, in $O(N \log(N))$ arithmetic operations, using the *Fast Fourier Transform (FFT)*. A similar relation, which we will not discuss explicitly, also holds in two or more dimensions.

2.2.4 The Discrete Fourier Reconstruction Problem

Consider the problem (2.7) and let N be such that $\Omega \subseteq \{-N/2+1, \dots, N/2\}^d$. In other words, the maximal frequency sampled is at most $N/2$. Let $P \in \mathbb{R}^{m\times N^d}$ be the *row selector* matrix that picks the rows of the d-dimensional Fourier matrix F corresponding to the values

$$\{\varrho^{-1}(\omega) \colon \omega \in \Omega\} \subseteq \{1, \dots, N^d\},$$

where ϱ is as in (2.14), and write $A = PF \in \mathbb{C}^{m\times N^d}$. We now arrive at the following discrete counterpart of the problem (2.7):

$$\text{Given the data } y = Ax, \text{ recover } x. \tag{2.15}$$

Since Ω usually corresponds to a subset of the N^d lowest frequencies, we refer to the matrix A as a *subsampled Fourier matrix*. Because of Remark 2.2, matrix–vector multiplications with A and A^* can be performed efficiently.

It is worth stressing that (2.7) and (2.15) are related, but distinct problems. The former involves measurements of the continuous Fourier transform of a continuous image f, while the latter involves measurements of the discrete Fourier transform of its discretization $X = \mathrm{vec}^{-1}(x)$. The corresponding *discretization error* and its effect will be discussed further in the next chapters.

2.3 Sampling with the Radon Transform

While the Fourier transform arises in many imaging modalities, arguably even more modalities are based on the so-called *Radon transform*. This transform and its variants underlie *tomographic* imaging, which is found widely in applications. We now give a short introduction to this transform.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

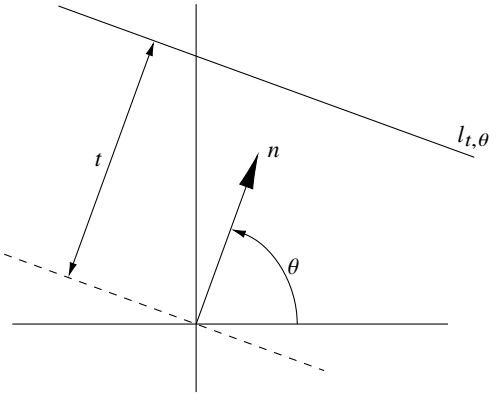

Figure 2.3 The parametrization (t, θ) of lines in the plane.

2.3.1 Definition

The Radon transform is an operator taking a function $f\colon \mathbb{R}^2 \to \mathbb{C}$ to the set of values corresponding to its integrals along arbitrary lines in the plane. In order to define it, we first need a parametrization of such lines. Every line in the plane can be expressed in terms of an angle $0 \le \theta < 2\pi$ measured counterclockwise from the x-axis that defines a normal vector $n = (\cos(\theta), \sin(\theta))$, and a perpendicular displacement $t \in \mathbb{R}$ from the origin. This is shown in Fig. 2.3. Given $(t, \theta) \in \mathbb{R} \times [0, 2\pi)$ we write $l_{t,\theta}$ for the corresponding line. Note that $l_{-t,\theta} = l_{t,\theta+\pi}$, meaning this parametrization is not one-to-one. To enforce uniqueness we may, for instance, consider angles $\theta \in [0, \pi)$ only.

Definition 2.3 (Radon transform) The *Radon transform* $\mathcal{R}f$ of $f\colon \mathbb{R}^2 \to \mathbb{R}$ is

$$\mathcal{R}f(t,\theta) = \int_{l_{t,\theta}} f \, \mathrm{d}s = \int_{-\infty}^{\infty} f(t\cos(\theta) - s\sin(\theta), t\sin(\theta) + s\cos(\theta)) \, \mathrm{d}s,$$

for $t \in \mathbb{R}$ and $0 \le \theta < \pi$, where, in the first integral, $\mathrm{d}s$ is the arclength measure along the line $l_{t,\theta}$.

Note that $\mathcal{R}f$ need not be defined for arbitrary f and arbitrary (t, θ). We shall not discuss *existence* conditions in any depth, except to say that $\mathcal{R}f$ is defined everywhere when, for example, f is piecewise continuous and compactly supported. This is a suitable model for the images in this book. Observe that the Radon transform is linear, i.e. $\mathcal{R}(af + bg) = a\mathcal{R}(f) + b\mathcal{R}(g)$ whenever the corresponding transforms are defined.

2.3.2 The Radon Reconstruction Problem

Similar to the Fourier transform, the image reconstruction problem from Radon measurements takes the following form:

$$\text{Given the data } \{\mathcal{R}f(t,\theta) : (t,\theta) \in \Theta\}, \text{ recover } f. \tag{2.16}$$

Here $\Theta \subset \mathbb{R} \times [0, \pi)$ corresponds to a finite subset of lines in the plane.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

This problem also requires discretization – an issue we pursue in a moment. First, however, we develop the intimate relationship between the Radon and Fourier transforms, the so-called *Fourier-slice* theorem.

2.3.3 The Fourier-Slice Theorem

The Fourier-slice (also *projection-slice* or *central-slice*) theorem states (under mild conditions on $f : \mathbb{R}^2 \to \mathbb{C}$) that the one-dimensional Fourier transform of $\mathcal{R}f(t, \theta)$ with respect to the displacement t is equal to the two-dimensional Fourier transform of f expressed in polar coordinates. Specifically,

$$\mathcal{F}_1 \mathcal{R} f(\omega, \theta) = \mathcal{F} f(\omega \cos(\theta), \omega \sin(\theta)), \quad \omega \in \mathbb{R},\ \theta \in [0, 2\pi), \tag{2.17}$$

where $\mathcal{F}$ denotes the two-dimensional Fourier transform and $\mathcal{F}_1$ is the one-dimensional Fourier transform with respect to the first component.

From the perspective of measurements, the Fourier-slice theorem means that sampling the Radon transform of f at a fixed angle and a full set of displacements $t \in \mathbb{R}$, i.e. acquiring the data

$$\{\mathcal{R} f(t, \theta) : t \in \mathbb{R}\}, \tag{2.18}$$

is equivalent to sampling the Fourier transform of f along the *radial line* with angle θ, i.e. the data

$$\{\mathcal{F} f(\omega \cos(\theta), \omega \sin(\theta)) : \omega \in \mathbb{R}\}.$$

Note that (2.18), or more precisely, the function $g_\theta(t) = \mathcal{R}f(t, \theta)$, is often referred to as the *Radon projection* of f at angle θ. Hence, acquiring the Radon projection is equivalent to sampling the Fourier transform of f along the corresponding radial line. More generally, acquiring k Radon projections of f at k distinct angles is equivalent to sampling $\mathcal{F} f$ along the corresponding k radial lines.

2.3.4 Filtered Back-Projection

While not strictly connected to the question of discretization, since it follows naturally from the Fourier-slice theorem, we now briefly discuss the matter of inversion of the Radon transform.

The Fourier transform is invertible and its inverse has a simple expression. By contrast, inverting the Radon transform is a more delicate affair. The so-called *filtered back-projection* formula gives an explicit expression for a certain left inverse of $\mathcal{R}$. In order to state this, we first introduce the back-projection operator:

Definition 2.4 (Back-projection) Let $g = g(t, \theta)$ be a function on $\mathbb{R} \times [0, \pi)$. The *back-projection* of g, $\mathcal{B}g$, is

$$\mathcal{B} g(x) = \frac{1}{\pi} \int_0^\pi g(x_1 \cos(\theta) + x_2 \sin(\theta), \theta)\, \mathrm{d}\theta, \qquad \forall x = (x_1, x_2) \in \mathbb{R}^2.$$

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

Suppose that $g = \mathcal{R}f$ is the Radon transform of some function f. Then the back-projection $\mathcal{B}g(x)$ at a point x simply averages the value of the line integrals of f over all lines $l_{t,\theta}$ passing through the point x.

It is tempting to think that back-projection might invert the Radon transform. Unfortunately, applying $\mathcal{B}$ to the Radon transform $\mathcal{R}f$ of a function f does not return f itself, but rather a blurred version of it. We will illustrate this later in Fig. 3.7. To obtain an inversion formula, one first needs to filter $\mathcal{R}f$ in a suitable way before applying $\mathcal{B}$. This leads to the well-known *filtered back-projection* formula:

$$f = \frac{1}{2}\mathcal{B}\left(\mathcal{F}_1^{-1}\left(|\omega|\mathcal{F}_1\left(\mathcal{R}f\right)(\omega,\theta)\right)\right). \tag{2.19}$$

Observe that if the factor $|\omega|$ were omitted, the right-hand side would be precisely $\mathcal{B}\mathcal{R}f$, and therefore not equal to f. This factor provides the filtering needed – specifically, suppressing low frequencies and amplifying high frequencies – to counteract the blurring effect of the back-projection operator $\mathcal{B}$.

2.3.5 The Discrete Radon Reconstruction Problem 1: Radial Fourier Sampling

There are several ways to discretize the Radon transform. The first way is to use the Fourier-slice theorem followed by the Fourier discretization of §2.2.3 to obtain discrete Fourier measurements along radial lines. This approach assumes that the set of samples

$$\Theta = \{(t,\theta_i) : t \in \mathbb{R}, i = 1,\dots,k\}$$

consists of a finite set of angles $0 \le \theta_1 < \cdots < \theta_k < \pi$ and all possible displacements $t \in \mathbb{R}$. That is, one acquires the full Radon projections of f at each angle θ_i. The Fourier-slice theorem then gives that (2.16) is equivalent to the Fourier reconstruction problem (2.7) with

$$\Omega' = \{(\omega\cos(\theta_i), \omega\sin(\theta_i)) : \omega \in \mathbb{R},\ i = 1,\dots,k\}.$$

To convert this to a discrete problem, we fix a resolution N and then approximate Ω' by a finite index set of integer frequencies $\Omega \subseteq \{-N/2+1,\dots,N/2\}^2$. This can be done, for instance, by replacing each point $(\omega\cos(\theta_i), \omega\sin(\theta_i)) \in \Omega' \cap (-N/2, N/2]^2$ with its nearest integer neighbour, a process known as *nearest-neighbour gridding* (also known as *regridding*). Having done this, the discrete analogue of (2.16) is then simply the discrete Fourier reconstruction problem (2.15) with the index set Ω.

This discretization procedure commits several errors. There is the error due to discretizing the continuous Fourier transform as in §2.2.3. There is also the error due to gridding. Moreover, in order to set up the discrete problem, one needs to take the one-dimensional Fourier transform $\mathcal{F}_1$ of the projection $\mathcal{R}f(\cdot,\theta)$. In fact, one usually cannot acquire $\mathcal{R}f(t,\theta)$ at all displacements $t \in \mathbb{R}$ in practice. Instead, the displacements are sampled on a finite, but sufficiently fine grid, and then the continuous Fourier transform $\mathcal{F}_1$ is replaced by a DFT. This, however, gives rise to another source of error.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

2.3.6 The Discrete Radon Reconstruction Problem 2: Algebraic Formulation

Discretization errors aside, another issue with this approach is that it requires one to acquire the full Radon projection at each angle. As we explain in Chapter 3 in the context of X-Ray CT, this may not be desirable or even possible in practice.

We now describe a simpler and more flexible discretization procedure. Suppose that the measurements correspond to an arbitrary set of lines

$$\Theta = \{(t_i, \theta_i) : i = 1, \ldots, m\} \subset \mathbb{R} \times [0, \pi).$$

We cannot appeal to the Fourier-slice theorem in this case. Instead, we discretize simply by replacing f with its discretization $x \in \mathbb{C}^{N^2}$ at resolution N. This gives

$$\mathcal{R}f(t_i, \theta_i) \approx \sum_{j=1}^{N^2} x_i \mathcal{R}(\chi_j)(t_i, \theta_i),$$

where $\{\chi_j\}$ is the pixel basis (see §2.1). Hence we may replace (2.16) with the discrete problem

$$\text{Given the data } y = Ax, \text{ recover } x, \tag{2.20}$$

where $A \in \mathbb{R}^{m \times N^2}$ is the matrix with (i, j)th entry $\mathcal{R}(\chi_j)(t_i, \theta_i)$. Note that these entries are straightforward to compute since $\{\chi_j\}$ is the pixel basis.

This discretization is not only simpler and more flexible than the previous approach, it also commits only one discretization error, instead of three. A downside is that FFTs cannot be used for the matrix–vector multiplications. Yet, the matrix A is usually sparse, since the vast majority of pixels do not intersect a given line. Hence sparse matrix tools can be employed to accelerate computations.

2.4 Binary Sampling with the Walsh Transform

We now introduce our final transform of this chapter, the *Walsh* (or *Walsh–Hadamard*) transform. This is a binary analogue of the Fourier transform, and shares many of its properties. Additional details on this transform and proofs of the results stated in this section can be found in Appendix F.

2.4.1 Walsh Functions

The derivation of the Walsh transform begins with the definition of the so-called *Walsh functions*. These are based on dyadic representations of numbers:

Definition 2.5 The *dyadic expansion* of $x \in [0, 1)$ is the series

$$x = \sum_{i=1}^{\infty} x_i 2^{-i}, \tag{2.21}$$

where $(x_i)_{i \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$. The sequence $(x_i)_{i \in \mathbb{N}}$ is denoted $\dot{x}$.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

The dyadic expansion defines a mapping between the set of binary sequences $\{0,1\}^{\mathbb{N}}$ and the interval $[0,1)$. This map is not one-to-one. Indeed, a dyadic rational $x = k/2^j$ can be represented either by a finite expansion, or by an infinite expansion where $x_i = 1$ for all $i > j$. This problem is easily remedied, however, by simply choosing the finite expansion instead of the infinite expansion whenever the possibility arises. This yields a one-to-one correspondence between $[0,1)$ and the subset G of $\{0,1\}^{\mathbb{N}}$, defined by

$$G = \left\{ \dot{x} = (x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} : x_i = 0 \text{ infinitely often} \right\}.$$

We write $g\colon [0,1) \to G$ for the corresponding bijection. Observe that every $\dot{x} \in \{0,1\}^{\mathbb{N}} \backslash G$ has $x_i = 0$ for only finitely many i, i.e. $x_i = 1$ for all $i > j$ for some j.

Similarly, but more straightforwardly, we define the dyadic expansion of a nonnegative integer $n \in \mathbb{N}_0$ as

$$n = \sum_{i=1}^{\infty} n_i 2^{i-1}, \tag{2.22}$$

where $(n_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$. This gives a one-to-one mapping $\mathbb{N}_0 \to H \subset \{0,1\}^{\mathbb{N}}$, where H is the set of binary sequences with only finitely many nonzero terms.

Definition 2.6 (Walsh functions, Paley ordering) The *Paley-ordered Walsh functions* are defined by

$$\upsilon_n(x) = \upsilon_n^P(x) = (-1)^{\sum_{i=1}^{\infty} n_i x_i}, \quad x \in [0,1),\ n \in \mathbb{N}_0, \tag{2.23}$$

where $(x_i)_{i\in\mathbb{N}} \in G$ and $(n_i)_{i\in\mathbb{N}} \in H$ are the dyadic expansions of x and n, respectively.

Figure 2.4 plots the first 16 Walsh functions. Observe that each function takes values in $\{+1,-1\}$ and has finitely many sign changes in $[0,1)$. The pattern that these sign changes take is described further in Lemma F.1.

2.4.2 The Walsh Basis and Transform

As shown in Theorem F.7, the Walsh functions $\{\upsilon_n\}_{n\in\mathbb{N}_0}$ form an orthonormal basis of $L^2([0,1))$. Hence, we may write every $f \in L^2([0,1))$ as

$$f = \sum_{n=0}^{\infty} \mathcal{H}f(n)\upsilon_n,$$

where $\mathcal{H}f(n) = \langle f, \upsilon_n \rangle_{L^2}$ are the *(Paley-ordered) Walsh coefficients* of f. We refer to the operator

$$\mathcal{H} : L^2([0,1)) \to \ell^2(\mathbb{N}_0),\ f \mapsto (\langle f, \upsilon_n \rangle_{L^2})_{n=0}^{\infty} \tag{2.24}$$

as the (Paley-ordered) *Walsh* transform (also known as the *Walsh–Hadamard* transform).

We can extend the Walsh basis to higher dimensions via tensor products. Specifically, we define the system $\{\upsilon_n\}_{n\in\mathbb{N}_0^d}$, where υ_n is the d-fold tensor product

$$\upsilon_n = \upsilon_{n_1} \otimes \cdots \otimes \upsilon_{n_d}, \qquad n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d, \tag{2.25}$$

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

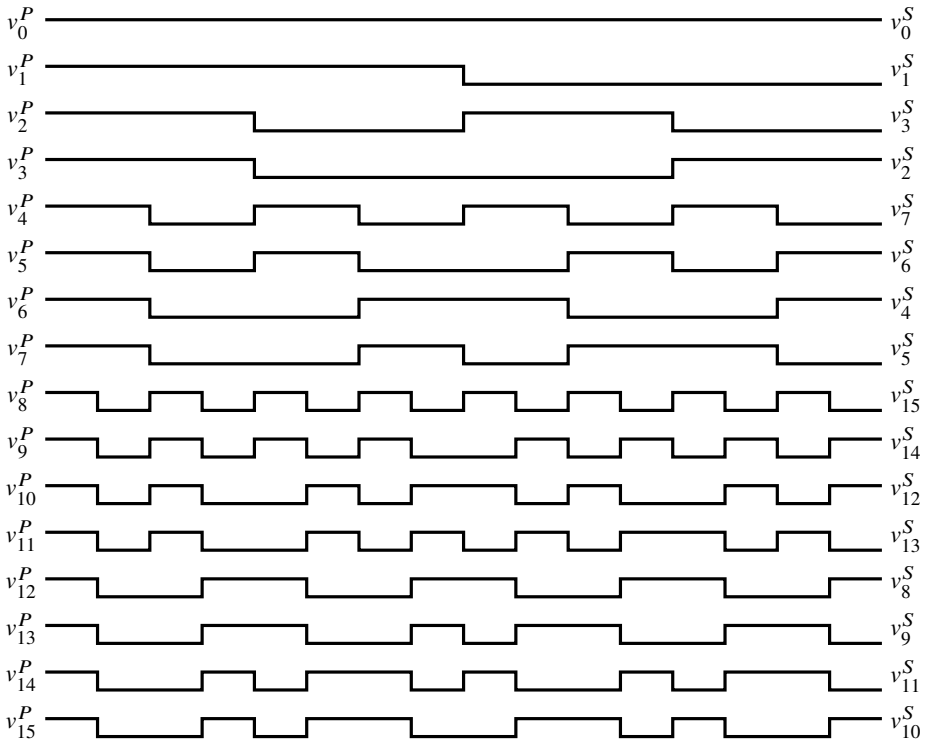

Figure 2.4 The first 16 Walsh functions with the Paley (v_n^P) or sequency (v_n^S) orderings.

of the univariate Walsh functions. This is an orthonormal basis of $L^2([0,1)^d)$ by construction. The d-dimensional Walsh transform $\mathcal{H}\colon L^2([0,1)^d) \to \ell^2(\mathbb{N}_0^d)$ is defined in the obvious manner.

2.4.3 The Sequency Ordering

The Paley ordering is convenient for establishing many of the theoretical properties of the Walsh functions. However, as shown in Fig. 2.4, it has the disadvantage that the number of sign changes of the nth Walsh function is generally not equal to n. The number of sign changes is known as the *sequency* of the Walsh function, a property that can be thought of as its frequency. Fortunately, this issue can be remedied by switching to the so-called *sequency* ordering:

Definition 2.7 (Walsh functions, sequency ordering) The *sequency-ordered Walsh functions* are defined by

$$\upsilon_n(x) = \upsilon_n^S(x) = (-1)^{\sum_{i=1}^{\infty}(n_i+n_{i+1})x_i}, \quad x \in [0,1),\ n \in \mathbb{N}_0,$$

where $(x_i)_{i\in\mathbb{N}} \in G$ and $(n_i)_{i\in\mathbb{N}} \in H$ are the dyadic expansions of x and n, respectively.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

It will come as little surprise to the reader to learn that the sequency-ordered Walsh functions are ordered with increasing sequency. Specifically, for each n, $\upsilon_n = \upsilon_n^S$ has precisely n sign changes in $[0, 1)$. See Fig. 2.4. Since sequency is analogous to frequency, we commonly refer to the index $n \in \mathbb{N}_0$ of the sequency-ordered Walsh function as the *Walsh frequency*. In particular, figures in this book showing Walsh sampling schemes (see Fig. 1.5, as well as other examples in Chapters 3 and 4) are always done with respect to this ordering.

Figure 2.4 describes the relation between the first 16 Paley- and sequency-ordered Walsh functions. As shown in Lemma F.8, in the general case one has

$$\upsilon_n^S = \upsilon_{h(n)}^P, \quad n \in \mathbb{N}_0,$$

where $h(n) \in \mathbb{N}_0$ is known as the *Gray code* of $n \in \mathbb{N}_0$. Lemma F.9 also shows that

$$\{\upsilon_n^P : 2^j \le n < 2^{j+1}\} = \{\upsilon_n^S : 2^j \le n < 2^{j+1}\}. \tag{2.26}$$

In particular, the first 2^r Paley-ordered Walsh functions coincide with the first 2^r sequency-ordered Walsh functions, and in general, switching between the two orderings has no effect on the *dyadic* structure of the Walsh basis.

2.4.4 The Hadamard Matrix and the Discrete Walsh–Hadamard Transform

Let $N = 2^r$ and $x = (x_j)_{j=1}^N \in \mathbb{C}^N$ be the discretization of a continuous, one-dimensional image $f : [0, 1] \to \mathbb{C}$ at resolution N. As in §2.2.3, we can approximate the continuous Walsh transform of f by

$$\mathcal{H}f(n) \approx \frac{1}{N} \sum_{j=1}^{N} x_j \upsilon_n((j-1)/N).$$

Hence the lowest N Walsh frequencies can be approximated by

$$(\mathcal{H}f(n))_{n=0}^{N-1} \approx \frac{1}{N} Hx,$$

where $H \in \mathbb{R}^{N \times N}$ is the following matrix:

Definition 2.8 (Hadamard matrix) For $N = 2^r$, the one-dimensional *Paley- or sequency-ordered Hadamard matrix* is the matrix $H = H^{(1)} \in \mathbb{R}^{N \times N}$ with entries

$$H_{m+1,n+1} = \upsilon_m(n/N), \quad m, n = 0, \dots, N-1,$$

where υ_m are the Paley- or sequency-ordered Walsh functions. For $d \ge 2$, the d-dimensional Hadamard matrix is given by

$$H^{(d)} = H^{(1)} \otimes \cdots \otimes H^{(1)} \in \mathbb{R}^{N^d \times N^d}.$$

As before, we write H for $H^{(d)}$ whenever the meaning in clear. The matrix H has entries in $\{-1, +1\}$. It is also symmetric, $H = H^\top$, and unitary up to a constant with $N^{-d} H^\top H = I$ (see Proposition F.10). As in one dimension, if $f : [0, 1]^d \to \mathbb{C}$ and

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

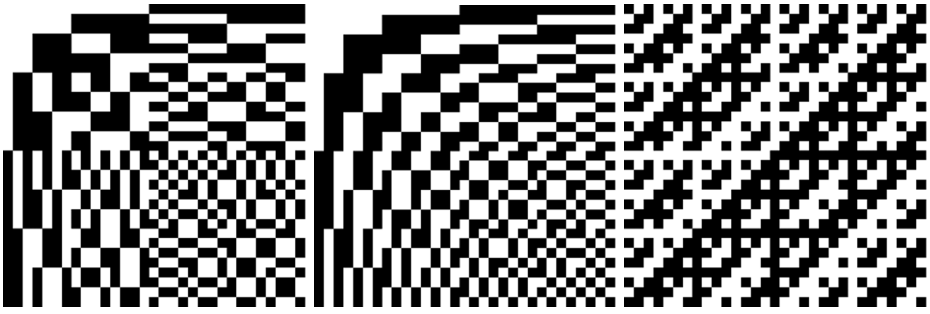

Figure 2.5 The 32×32 Hadamard matrix H with the Paley, sequency and ordinary orderings. White corresponds to a value of $+1$ and black to a value of -1.

$x \in \mathbb{C}^{N^d}$ is its discretization at resolution N, then $N^{-d}Hx$ is an approximation to the lowest N^d Walsh frequencies of f. Specifically,

$$\frac{1}{N^d} Hx \approx \left(\mathcal{H} f(\varrho^{-1}(i))\right)_{i=1}^{N^d},$$

where $\varrho = \varrho^{(d)} \colon \{1, \dots, N^d\} \to \{0, \dots, N-1\}^d$ is the bijection defined by

$$\varrho^{(d)}(i) = (\varsigma(i)_1 - 1, \dots, \varsigma(i)_d - 1), \quad i = 1, \dots, N^d, \tag{2.27}$$

and ς is as in (2.2).

Much like the discrete Fourier transform, the Hadamard matrix gives rise to a discrete transform $\mathcal{H} \colon \mathbb{C}^{N^d} \to \mathbb{C}^{N^d}$, the *Discrete Walsh–Hadamard Transform (DHT)*. It and its inverse are given by

$$\mathcal{H}(x) = Hx, \qquad \mathcal{H}^{-1}(y) = N^{-d}Hy, \qquad x, y \in \mathbb{C}^{N^d}. \tag{2.28}$$

Note that we use the letter 'H' here (rather than 'W') to avoid confusion with the Discrete Wavelet Transform (DWT) introduced in Chapter 9.

Finally, we remark that the Paley- and sequency-ordered Hadamard matrices are related to a third type of Hadamard matrix that is often found in the literature:

Definition 2.9 (Ordinary Hadamard matrix) Let $N = 2^r$. The *ordinary Hadamard matrix* $H = H_r \in \mathbb{R}^{N \times N}$ is defined by

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H_r = H_{r-1} \otimes H_1 = \begin{pmatrix} H_{r-1} & H_{r-1} \\ H_{r-1} & -H_{r-1} \end{pmatrix}.$$

The ordinary Hadamard matrix is simply a reordering of the rows of of the Paley- or sequency-ordered Hadamard matrix (Proposition F.11). These three forms of the Hadamard matrix are shown in Fig. 2.5.

A motivation for introducing the ordinary Hadamard matrix is because it easily admits a fast transform. Its dyadic structure means that matrix–vector multiplications can be computed efficiently using a divide-and-conquer approach. This approach is described in §F.5. Much like the FFT, it involves only $O(N \log(N))$ arithmetic operations. This

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

means the DHT (with respect to either the Paley or sequency ordering) can also be implemented in $O(N \log(N))$ time, since the Paley- or sequency-ordered Hadamard matrices are simply row permutations of the ordinary Hadamard matrix. We refer to the resulting procedure as the *Fast Walsh–Hadamard Transform (FHT)*.

2.4.5 The Continuous and Discrete Walsh Reconstruction Problems

Let $\Omega \subset \mathbb{N}_0^d$ be a finite set of Walsh frequencies (recall that this means we are considering the sequency ordering) and $f\colon [0,1]^d \to \mathbb{C}$ be a continuous d-dimensional image. The continuous Walsh reconstruction problem is

$$\text{Given the data } \{\mathcal{H}f(n) : n \in \Omega\}, \text{ recover } f.$$

To derive the discrete problem, let N be such that $\Omega \subseteq \{0, \dots, N-1\}^d$ and $P \in \mathbb{R}^{m \times N^d}$ be the row selector matrix that picks out the rows of the sequency-ordered Hadamard matrix $H \in \mathbb{R}^{N^d \times N^d}$ corresponding to the values

$$\{\varrho^{-1}(n) : n \in \Omega\} \subseteq \{1, \dots, N^d\},$$

where ϱ is as in (2.27). If $x \in \mathbb{C}^{N^d}$ is the discretization of f at resolution N, then the corresponding discrete reconstruction problem is simply

$$\text{Given the data } y = Ax, \text{ recover } x,$$

where $A = PH \in \mathbb{R}^{m \times N^d}$ is the corresponding *subsampled Hadamard matrix*.

Notes

The introduction to images given in §2.1 is brief, but sufficient for the remainder of this book. For a significantly more in-depth treatment, see for example [50]. See [50,273,491] for further discussion on the limitations of MSE and PSNR for assessing image quality.

The Shepp–Logan phantom ('SL phantom' in Fig. 2.2) is a famous test image in medical imaging, introduced in 1974 by Shepp and Logan [437]. It is a continuous, piecewise constant image defined as a sum of ten ellipses. In particular, its continuous Fourier and Radon transforms can be calculated exactly. Unfortunately, the SL phantom is rather too simple a test image in many cases. This motivated the development of the second image in Fig. 2.2, the 'GLPU phantom', introduced by Guerquin-Kern, Lejeune, Pruessmann & Unser [245]. Like the SL phantom it is a continuous, piecewise constant image with an analytic expression for its Fourier transform. Discretized versions of both images can, of course, be generated at any resolution. The 'brain' image is a 512×512 image obtained from an actual MRI scan. The resolution is limited by the scanner, and the object being scanned (a human brain). The 'pomegranate' image is a 2048×2048 image from an MRI scan of a pomegranate fruit, the higher resolution being possible in this case because the object is not a living subject.

The 'peppers' image is also a standard test image in image processing. It can be found in a database maintained by the University of Southern California's Signal and

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press

Image Processing Institute [1]. The original image is colour, and of size 512×512. The 'donkey' image is from the authors' collection and is of size 768×768.

Our cursory treatment of the Radon transform hardly does it justice. Whole books have been devoted to its mathematical properties and practical implementation. We have based §2.3 on [198, 204, 305, 370]. Derivations of the Fourier-slice theorem (2.17) and filtered back-projection formula (2.19) can be found in these sources. What makes the Radon transform challenging and interesting, both theoretically and computationally, is the question of its inversion. Filtered back-projection (2.19) is just one possible *left inverse*. However, there are many others. This stems from the fact that the range of the Radon transform is not the whole space. In contrast, the Fourier transform has a unique inverse on $L^2(\mathbb{R})$.

Discretization of the filtered back-projection formula, as described in §2.3.5, is the standard means to numerically invert the Radon transform in the classical setting – that is, where there is 'sufficient' data. For an in-depth treatment, see [198, Chpt. 10] and [370, Chpt. 5]. By and large, this remains the standard algorithm for commercial X-ray CT scanners [386] (see also the Notes section of Chapter 3). The other approach described in §2.3.6 is related to the so-called *Algebraic Reconstruction Technique (ART)*. It ignores properties of the Radon transform such as the Fourier-slice theorem and simply discretizes the continuous image. Classically, when the number of measurements m exceeds the number of degrees of freedom N^d this has been solved as an overdetermined (and possibly also regularized) least-squares problem. Fast solvers use iterative algorithms such as Kaczmarz's method. See [198, 370]. As we note in §3.3, this type of discretization is often preferred when applying compressed sensing techniques.

Walsh series were introduced by Walsh in 1923 [485]. They are far less well known than their famous Fourier cousin, and used far less widely in practice. Historically, Walsh series have found applications in coding theory and image and signal processing. Our treatment here and in Appendix F covers only the basic properties, which will be sufficient for the remainder of this book. We have based these sections on [54, 224, 232], all of which contain far greater detail. The various orderings of the Walsh basis are sometimes given different names in the literature. The sequency ordering is often referred to as the *Walsh–Kaczmarz* ordering [54, p. 17], and the Paley ordering as the *natural*, *normal* or *dyadic* ordering. Typically, the Paley ordering is preferred for theory, whereas the sequency ordering is favoured in practice, due to its analogy with frequency.

https://doi.org/10.1017/9781108377447.004 Published online by Cambridge University Press