1 Introduction

Images are everywhere. Whether on a social media page, a doctor's desk to aid diagnosis or a scientist's computer screen to help study a chemical, physical or biological process, the modern world is awash with digital images.

All images are produced by acquiring measurements using a physical device. The list of different acquisition devices is long and varied. It includes simple devices found in most homes, such as digital cameras; specialist medical imaging equipment found in hospitals, such as a Magnetic Resonance Imaging (MRI) or X-ray Computed Tomography (CT) scanner; or scientific devices, such as electron microscopes, found in laboratories.

The concern of this book is the task of image *reconstruction*. This is the algorithmic process of converting the raw data (the measurements) into the final image seen by the end user. The overarching aim of image reconstruction is to achieve the four following, and competing, objectives:

Objective #1 (accuracy): to produce the highest-quality images.

Accuracy is of course paramount. High-quality images are desirable in virtually all applications. However, in direct competition with this is:

Objective #2 (sampling): to use as few measurements as possible.

Acquiring more measurements usually comes at a cost. It could mean an additional outlay of time, power, monetary expense or risk, depending on the application at hand. Reducing the number of measurements is often the primary goal of image reconstruction. For example, in MRI, taking more measurements involves a longer scan time, which can be unpleasant and challenging for the patient – especially in paediatric MRI. It also makes the measurements acquired more susceptible to corruptions due, for instance, to patient motion. In X-ray CT, the number of measurements loosely corresponds to the amount of radiation to which the patient is exposed. Acquiring fewer measurements per scan opens the door for more frequent scans, which in turn allows for more effective treatment monitoring.

Objective #3 (stability): to ensure that errors in the measurements or in the numerical computation do not significantly impact the quality of the recovered image.

All imaging systems introduce error in the measurements, due to noise, corruptions or modelling assumptions. There are also round-off errors in the numerical computations performed by image reconstruction algorithms. It is vital that reconstruction algorithms

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

be robust to such perturbations, so that small errors do not have a deleterious effect on the output image.

Objective #4 (efficiency): to recover the image in reasonable computing time without significant computing power and memory.

In many applications, images need to be reconstructed rapidly. After all, a doctor does not want to wait a long time for an image to be generated after a scan has concluded. In other imaging systems, notably portable systems, computing resources may be severely limited. Since modern images are often comprised of tens of millions of pixels, practical image reconstruction algorithms must have the ability to scale to large problem sizes without suffering a blow-up in computing time or memory requirements.

1.1 Imaging and Inverse Problems

Mathematically, the simplest way to model an image reconstruction problem is as the following discrete, linear inverse problem:

$$\text{Given the measurements } y = Ax + e, \text{ recover } x. \tag{1.1}$$

Here $y \in \mathbb{C}^m$ is the vector of *measurements* produced by the sensing device, $A \in \mathbb{C}^{m \times N}$ is the *measurement matrix* representing the acquisition process, $e \in \mathbb{C}^m$ is a vector of measurement noise and $x \in \mathbb{C}^N$ is the (vectorized version of the) unknown image to be recovered. The integer m is the number of measurements, and N is the number of pixels in the image. Designing an image reconstruction procedure means constructing a *reconstruction map*

$$\mathrm{R} \colon \mathbb{C}^m \to \mathbb{C}^N, \tag{1.2}$$

that takes input y and outputs an approximation $\mathrm{R}(y)$ to the true image x.

It is important to note that (1.1) is derived through mathematical modelling of the sensing device, a process which usually involves a series of assumptions. While a finite-dimensional, linear model such as (1.1) may be appealing in its simplicity, it may result in *mismatch* between the model and the true physics, which in turn may lead to additional errors in the reconstruction. One common error of this type arises from *discretization*: the conversion of a continuous problem into a discrete one such as (1.1). To avoid subsequent errors in the reconstruction, it may be beneficial to consider an infinite-dimensional model:

$$\text{Given the measurements } y = \mathcal{A}f + e, \text{ recover } f. \tag{1.3}$$

Here $f \colon [0,1]^d \to \mathbb{C}$ is a function representing the image, $\mathcal{A}$ is a linear operator representing the acquisition process, $y \in \mathbb{C}^m$ is the vector of measurements and $e \in \mathbb{C}^m$ is measurement noise, as before. Of course, there may well be other modelling issues beyond discretization. The sensing process could be nonlinear, for example, in which case (1.3) will still result in model mismatch.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

1.2 What is Compressive Imaging?

For the sake of simplicity, consider the discrete problem (1.1). If $m = N$ and A is invertible, then reconstructing x is, in principle, straightforward. In the absence of noise, we simply solve the linear system $Ax = y$. However, this situation is rare in practice. Because of Objective #2, it is often the case that the number of measurements m is much smaller than the problem size N. This renders the linear system highly *underdetermined*, making *exact* recovery of x from y generally impossible.

The classical way to *approximately* recover x is to apply a left inverse of A to y: for example, its pseudoinverse $A^{\dagger}$. This is a simple, linear recovery procedure (the map $\mathrm{R}\colon y \mapsto A^{\dagger}y$ is a linear map) that is often computationally efficient and robust. It also has a simple interpretation: the computed image $\hat{x} = A^{\dagger}y$ has the smallest ℓ^2-norm amongst all those that fit measurements. Unfortunately, this process generally leads to low-quality reconstructions. Figure 1.1(c) illustrates this phenomenon for a synthetic MRI experiment involving the classical Shepp–Logan phantom. The recovered image is a poor reconstruction of the original image. As we observe, it exhibits substantial *artefacts*.

What to do? The way forward is to realize that images are not just arbitrary arrays of pixels: they have characteristic features such as edges and textures. Mathematically, this means that natural images can be modelled as objects in low-dimensional, nonlinear spaces embedded in the high-dimensional vector space $\mathbb{C}^N$. This on its own is by no means a new concept. It lies at the heart of modern lossy image compression standards such as JPEG-2000 and MPEG. Yet it was not until the mid-2000s that researchers began to develop mathematical tools for exploiting such structure in the context of solving image reconstruction problems such as (1.1). This has led to a topic in its own right, termed *compressed sensing* (also known as *compressive sensing*, *compressed sampling* or *compressive sampling*), with applications not only in imaging but also many other problems in computational science and engineering.

This brings us to the topic of this book.

> This book is about *compressive imaging*: the development, analysis and application of sampling strategies and (nonlinear) reconstruction procedures that exploit the low-dimensional structure of images to achieve substantially better image recovery than with classical techniques.

The growth of compressive imaging has wrought a profound change on practical image reconstruction over the past decade. In many applications, not least MRI, classical linear recovery procedures have been replaced by a new generation of techniques.

It is not hard to see why. Figure 1.1(d) shows what happens when the procedure used in Fig. 1.1(c) is replaced by a compressive imaging procedure based on compressed sensing. Both procedures use exactly the same data, and in particular, the same number of measurements. Yet the compressed sensing recovery is significantly better. It has none of the artefacts that plagued the classical reconstruction and appears to reconstruct the image perfectly – at least to the human eye.

The purpose of this book is to explain how compressive imaging makes this possible. It aims to describe how Fig. 1.1(d) was computed, why it offers such a significant

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

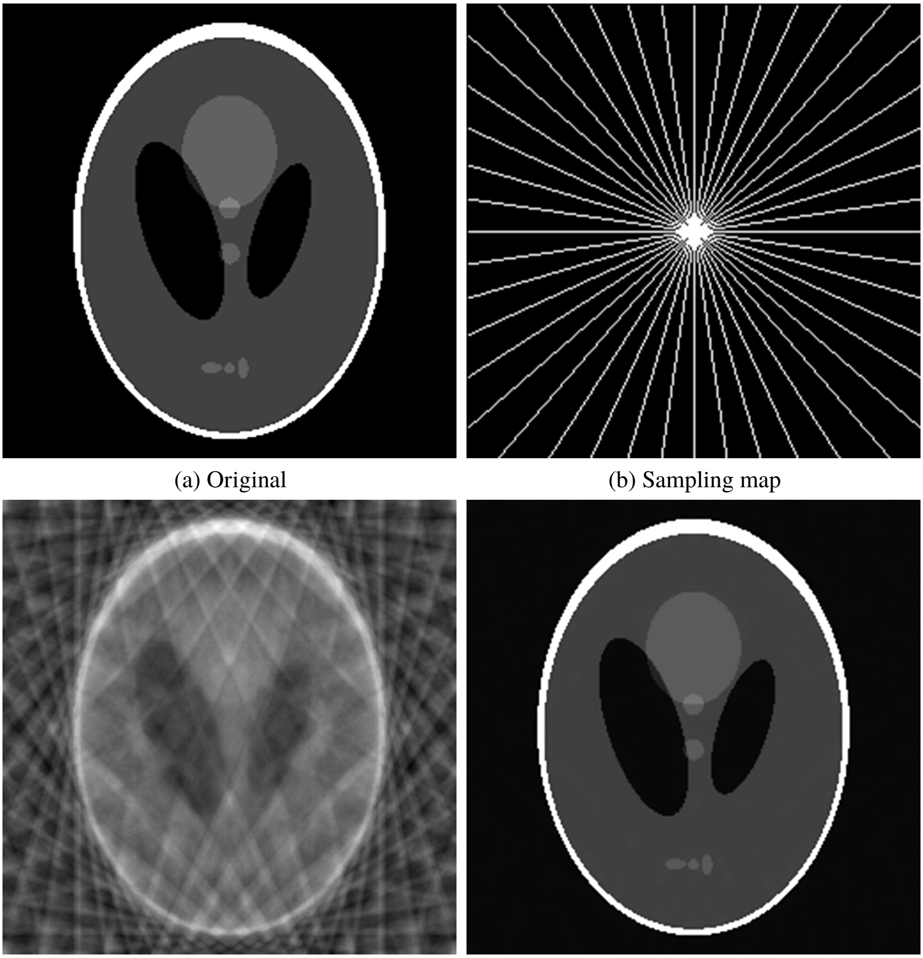


Figure 1.1 Reconstruction of the Shepp–Logan phantom image from discrete Fourier measurements. (a) Original image of size $256 \times 256 = 65,536$. (b) Sampling map in frequency space with $m = 5481$. Each white dot represents a frequency sampled. (c) Classical reconstruction using the zero-padded inverse DFT (this is equivalent to $A^{\dagger}$). (d) Compressed sensing reconstruction using TV minimization.

improvement over Fig. 1.1(c) and how performance, in the sense of Objectives #1–#4, can be even further improved.

Key Point #1. Image reconstruction has been revolutionized in the last decade by the emergence of compressive imaging in tandem with compressed sensing.

Remark 1.1 Figure 1.1 has an important historical context. It replicates an experiment performed by Candès, Romberg & Tao in their seminal 2006 paper that introduced

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

compressed sensing in tandem with the work of Donoho. The impact that this dramatic proof-of-concept experiment had on the imaging community is hard to overstate.

1.3 Terminology

In the remainder of this chapter, we discuss some of the main themes arising in this book. But, first, a word on terminology. Compressive imaging and compressed sensing are not synonymous. The latter is a mathematical theory for exploiting low-dimensional structures in abstract problems of the form (1.1), i.e. where x need not be an image and A need not arise from an imaging device. Unsurprisingly, the former deals exclusively with imaging. However, compressive imaging arguably owes its existence to the latter, imaging being both a primary motivation for much of compressed sensing research (see Remark 1.1) and one of the areas where it has been most successfully applied.

Yet compressive imaging is also not limited to compressed sensing techniques. Nowadays, it is beginning to see the increasing use of tools from machine learning, such as neural networks and deep learning. Part V of this book considers these approaches. Although closely related, they are not compressed sensing approaches per se.

Compressive imaging can also be seen as a subset of the larger, and rather older, field of *computational imaging*. Here the general goal is to enhance image quality through the design of better reconstruction algorithms and the availability of more powerful computing resources, rather than through hardware improvements in the sensor itself – the latter being ever increasingly harder to achieve due to physical limits.

1.4 Imaging Modalities

In order to exhibit some of the main aspects of the book, we now describe some of the different sampling processes that arise in typical imaging problems.

1.4.1 Integral Transforms

Measurements of an image are often acquired by sampling with an integral transform. The Fourier transform is an important example of this process. If f is a function representing a continuous image, then its Fourier transform is defined by

$$\mathcal{F}f(\omega) = \int_{\mathbb{R}^d} f(x) \mathrm{e}^{-\mathrm{i}\omega \cdot x} \,\mathrm{d}x, \qquad \omega \in \mathbb{R}^d.$$

In *Fourier imaging* the (noiseless) measurements $y \in \mathbb{C}^m$ correspond to samples

$$\{\mathcal{F}f(2\pi\omega) : \omega \in \Omega\}$$

of $\mathcal{F}f$ at a set of m frequencies $\Omega = \{\omega_1, \ldots, \omega_m\} \subset \mathbb{R}^d$. MRI is an important example of Fourier imaging.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

The Radon transform is another key integral transform found in imaging. It models the acquisition process in *tomographic* modalities such as X-ray CT. Here the measurements correspond to line integrals through the object being scanned.

It is important to note that the integral transform in a given problem is generally fixed. The acquisition process in MRI is modelled by the Fourier transform; in CT it is modelled by the Radon transform. Neither can be easily changed. However, there may be substantial freedom to choose the samples themselves, i.e. the set of frequencies Ω in Fourier imaging. Hence a key challenge in imaging with integral transforms is to understand how the choice of Ω affects the recovery of f, and then to use this insight to design sets Ω that lead to the highest-quality reconstructions.

Key Point #2. Many imaging modalities involve sampling via an integral transform. An important challenge is determining values at which the integral transform should be sampled so as to produce the best reconstructions.

Since this is also an inherently infinite-dimensional problem, a second challenge is to devise suitable discretizations to render the problem amenable to computations.

1.4.2 Binary Sampling

Binary sampling occurs when an image is measured by taking inner products with a function or vector that takes values in $\{+1,-1\}$. In the discrete setting (1.1), a single, noiseless measurement of the image $x = (x_j)_{j=1}^N \in \mathbb{C}^N$ takes the form

$$\langle x, a \rangle = \sum_{j=1}^{N} x_j a_j, \qquad a = (a_j)_{j=1}^N \in \{-1,+1\}^N. \tag{1.4}$$

After repeating this process m times, one obtains a collection of measurements

$$\langle x, a_1 \rangle, \ldots, \langle x, a_m \rangle,$$

and a binary measurement matrix $A \in \{+1,-1\}^{m \times N}$ in which the ith row is the ith binary measurement vector $a_i \in \{-1,1\}^N$.

Binary sampling arises in many *optical imaging* applications. Examples include lensless imaging and the so-called single-pixel camera, as well as fluorescence microscopy and numerous others. Usually in these applications a mask is placed in front of the object to be imaged. This selectively illuminates and obscures different pixels, thus effecting a binary measurement of the image. Because of this setup, there is often significant freedom to *design* the measurement vectors $a_1, \ldots, a_m$ to maximize the quality of the reconstructed images.

Key Point #3. Many optical imaging modalities involve binary sampling. This often affords significantly more flexibility than imaging with integral transforms in terms of the choice of measurements. In the discrete setting, choosing which measurements to acquire is equivalent to choosing the whole binary measurement matrix A.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

1.4.3 Sampling with Orthonormal Vectors

Many discrete imaging problems can be cast as sampling an image $x \in \mathbb{C}^N$ by taking inner products with respect to an orthonormal basis $\{u_i\}_{i=1}^N \subset \mathbb{C}^N$. If $\Omega = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, N\}$ is a set of size $|\Omega| = m$, then the noiseless measurements in this case take the form

$$y = (y_j)_{j=1}^m, \qquad y_j = \langle x, u_{i_j} \rangle. \tag{1.5}$$

Let $U = (u_1 | \cdots | u_N) \in \mathbb{C}^{N \times N}$ be the matrix whose ith column is the ith vector u_i. This matrix is unitary, $U^* U = I$, since $\{u_i\}_{i=1}^N$ is an orthonormal basis. The measurement matrix A corresponding to (1.5) is known as a *subsampled unitary matrix*. It is constructed by selecting the rows of U^* corresponding to the indices in Ω.

Sampling with orthonormal vectors arises in standard discrete formulations of Fourier imaging problems, in which case $U^* = F$ is the *Fourier matrix*. This type of discretization is highly convenient. Matrix–vector multiplications with F are equivalent to *Discrete Fourier Transforms (DFTs)*, which can be implemented efficiently with *Fast Fourier Transforms (FFTs)*. Sampling with orthonormal vectors is also useful in binary imaging. In this case, $U^* = H$ might be chosen as the *Hadamard* matrix – the binary analogue of the Fourier matrix F – which can be implemented efficiently via the *Discrete Walsh–Hadamard Transform (DHT)* and its associated fast transform.

As with integral transforms, a key issue when sampling with orthonormal vectors is how to choose the sampling vectors, or equivalently, the index set Ω. A secondary issue is how well this discrete setup models the true acquisition process, and what effect any possible model mismatch may have on the resulting reconstruction.

1.5 Conventional Compressed Sensing

Since it lies at the heart of compressive imaging, in this section we briefly depart from the world of imaging to introduce some of the main facets of standard compressed sensing theory. A more thorough treatment is given in Chapter 5.

1.5.1 Sparsity and Compressibility

As noted, compressed sensing aims to solve the reconstruction problem (1.1) in an abstract sense, where $x \in \mathbb{C}^N$ is a vector and $A \in \mathbb{C}^{m \times N}$ is a matrix, neither of which may be strictly related to an imaging problem.

Standard compressed sensing is based on the *sparsity model*. Specifically, we assume that there is a fixed orthonormal basis $\{\phi_i\}_{i=1}^N \subset \mathbb{C}^N$ such that

$$x = \sum_{i=1}^N d_i \phi_i, \tag{1.6}$$

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

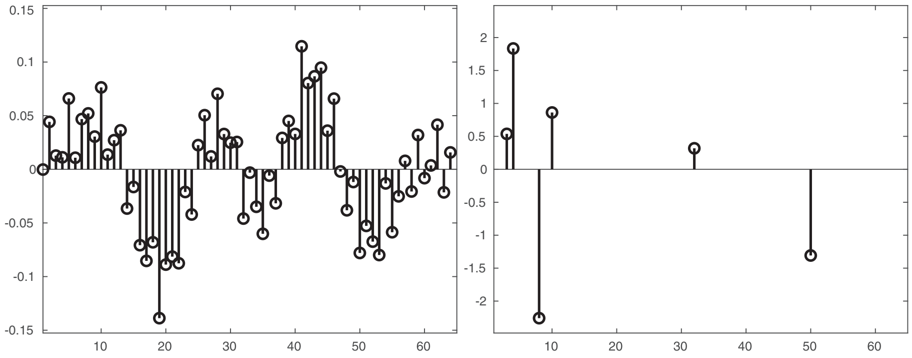


Figure 1.2 Left: Discrete signal $x \in \mathbb{R}^{64}$. Right: The 6-sparse vector of coefficients $d = \Phi^* x \in \mathbb{R}^{64}$, where Φ is a DCT.

where the vector of *coefficients* $d = (d_i)_{i=1}^N$ has only s nonzero entries for some number $s \ll N$. This number is known as the *sparsity* of d, and d is said to be *s-sparse*. The vector x is said to have an *s-sparse representation* in the basis $\{\phi_i\}_{i=1}^N$.

Observe that the set of s-sparse vectors does not constitute a subspace of $\mathbb{C}^N$. Adding two s-sparse vectors produces a vector that may be at most $2s$-sparse. In fact, this set is a union of s-dimensional subspaces of $\mathbb{C}^N$. Sparsity is therefore a type of nonlinear, low-dimensional structure.

Let $\Phi = (\phi_1 | \cdots | \phi_N) \in \mathbb{C}^{N \times N}$ be the unitary matrix corresponding to the basis $\{\phi_i\}_{i=1}^N$. Then (1.6) is equivalent to the expression $x = \Phi d$. Since Φ is unitary we also have $d = \Phi^* x$. We commonly refer to $\{\phi_i\}_{i=1}^N$ as the *sparsity basis* and Φ as the *sparsifying transform*. The latter highlights the fact that applying Φ^* to x yields the sparse vector d.

Note that Φ could be the identity matrix, in which case the vector x is itself s-sparse. This is relatively uncommon in imaging applications. Typical examples in imaging include the *Discrete Cosine Transform (DCT)* and *Discrete Wavelet Transform (DWT)*. Figure 1.2 illustrates a vector that is 6-sparse with respect to the DCT sparsifying transform.

Unfortunately, real-world objects such as images are never exactly sparse. However, they are typically *compressible*, or *approximately sparse*. This means that their coefficients $d = \Phi^* x$ can be accurately approximated by an s-sparse vector for some $s \ll N$. Figure 1.3 demonstrates this property when the DWT is applied to a natural image. In this case, 95% of its coefficients can be discarded, and the resulting sparse image is indistinguishable from the original – at least to the human eye.

Key Point #4. Conventional compressed sensing concerns the recovery of vectors that are approximately sparse in a fixed orthogonal sparsifying transform.

Remark 1.2 Sparsity and compressibility predate compressed sensing. Notably, they lie at the heart of modern lossy image compression algorithms such as JPEG-2000 and MPEG. In lossy compression, the coefficients $d = \Phi^* x$ of an image x are first computed by applying Φ^*, then all but the largest s (in absolute value) are set to zero. This yields an s-sparse vector $\tilde{d}$ that can be stored more efficiently. The number s can be viewed as the

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

(a) Original (b) Wavelet coefficients (c) Compressed

Figure 1.3 (a) Original image x. (b) Its wavelet coefficients $d = \Phi^* x$. Light values correspond to large coefficients and dark values to small coefficients. (c) Compressed image $\tilde{x} = \Phi \tilde{d}$, where 95% of the wavelet coefficients have been set to zero.

compression factor: if s is small, storing the s coefficients and their locations requires significantly less memory than storing the full array of N coefficients. When the image is needed, one simply applies Φ to $\tilde{d}$ to obtain an approximation $\tilde{x} = \Phi \tilde{d}$ to x. Figure 1.3 shows an example of this process.

1.5.2 Recovery

Suppose x is compressible in a fixed sparsifying transform Φ. We now aim to use this information to recover accurately (Objective #1) and stably (Objective #3) from its noisy measurements

$$y = Ax + e. \tag{1.7}$$

Clearly, we lose all hope of high-quality recovery if the noise e is too large. Hence we now assume that e is bounded, and that its energy

$$\|e\|_{\ell^2} \leq \eta, \tag{1.8}$$

for some known and small $\eta \geq 0$. Note that other noise models are also possible – for instance, assuming that e follows some specific distribution – but (1.8) is typically most common in compressed sensing.

Consider (1.7). The question is how to exploit the compressibility of x in order to effect a good reconstruction. While numerous approaches have been proposed, arguably the most popular involves solving a convex optimization problem that minimizes the ℓ^1-norm of the coefficients $\Phi^* x$. A common example is *Quadratically Constrained Basis Pursuit (QCBP)*. This takes the form

$$\min_{z \in \mathbb{C}^N} \|\Phi^* z\|_{\ell^1} \text{ subject to } \|Az - y\|_{\ell^2} \leq \eta. \tag{1.9}$$

In other words, it finds a vector $\hat{x}$ for which the coefficients $\Phi^* \hat{x}$ have the smallest ℓ^1-norm amongst all vectors that fit the measurements up to the noise bound η. The choice

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

of the ℓ^1-norm here is crucial. For reasons that are discussed in more detail in Chapter 5, the ℓ^1-norm *promotes* sparsity of x in the sparsifying transform Φ. It also renders the QCBP problem convex, thus making it amenable to efficient algorithms.

However, (1.9) is by no means the only possible approach, even within the realm of convex optimization-based approaches. Another common choice is the *Least Absolute Shrinkage and Selection Operator (LASSO)*. This is the unconstrained problem

$$\min_{z \in \mathbb{C}^N} \lambda \|\Phi^* z\|_{\ell^1} + \|Az - y\|_{\ell^2}^2. \tag{1.10}$$

Here, the term $\|Az - y\|_{\ell^2}^2$ promotes good fitting of the data and the term $\|\Phi^* z\|_{\ell^1}$ promotes sparsity. These two terms are balanced via a positive parameter $\lambda > 0$, the optimal choice of which, unsurprisingly, depends on the sparsity s and the noise level η. Note that (1.10) is the so-called *unconstrained* LASSO; we often write U-LASSO for clarity. We discuss the constrained version, the C-LASSO, in Chapter 6.

1.5.3 Stable and Accurate Recovery

Having chosen a sparsifying transform and recovery procedure, we now return to Objectives #1 and #2, as well as #3. These can be posed as the following question. How many measurements, and of what type, are sufficient for accurate and stable recovery of x through, for instance, the QCBP problem (1.9)? This is, in essence, the main question compressed sensing theory seeks to answer.

It does this by devising conditions on the measurement matrix A which are sufficient for recovery. There are various different conditions, several others of which we review in Chapter 5, but arguably the most well known is the *Restricted Isometry Property (RIP)*. A matrix $A \in \mathbb{C}^{m \times N}$ satisfies the RIP of order s if there is a constant $0 < \delta < 1$ such that

$$(1 - \delta)\|x\|_{\ell^2}^2 \leq \|Ax\|_{\ell^2}^2 \leq (1 + \delta)\|x\|_{\ell^2}^2,$$

for every s-sparse vector x. In other words, the *energy* of an s-sparse vector is approximately preserved via the measurement process $x \mapsto Ax$.

As we see in Chapter 5, the RIP is sufficient for stable and accurate recovery. An important result in compressed sensing states the following. Suppose the product $A\Phi \in \mathbb{C}^{m \times N}$ has the RIP of order $2s$ for sufficiently small constant δ. Then for all measurements of the form $y = Ax + e$ with $\|e\|_{\ell^2} \leq \eta$, every minimizer $\hat{x}$ of (1.9) satisfies the error bound

$$\|\hat{x} - x\|_{\ell^2} \lesssim C\mathcal{S}_s(x, \eta), \qquad C\mathcal{S}_s(x, \eta) = \frac{\sigma_s(\Phi^* x)_{\ell^1}}{\sqrt{s}} + \eta. \tag{1.11}$$

Here $\sigma_s(\cdot)_{\ell^1}$ is the ℓ^1*-norm best s-term approximation error*, defined by

$$\sigma_s(d)_{\ell^1} = \min\{\|z - d\|_{\ell^1} : z \text{ is } s\text{-sparse}\}.$$

In other words, $\sigma_s(\Phi^* x)_{\ell^1}$ measures how compressible x is in the sparsity basis Φ. Note that $\sigma_s(\Phi^* x)_{\ell^1} = 0$ if x happens to be exactly s-sparse in Φ.

We refer to (1.11) as a compressed sensing *error bound*, and the statement 'the RIP implies (1.11)' as a *recovery guarantee*. Crucially, it asserts accurate and stable recovery

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

of x via QCBP. Accuracy is measured in terms of the best s-term approximation error $\sigma_s(\Phi^* x)_{\ell^1}$ and stability in terms of the noise bound η. In the idealized case where x is exactly s-sparse in Φ and the measurements are noiseless (i.e. $\eta = 0$), the bound (1.11) asserts that x is recovered *exactly*.

1.5.4 Measurement Matrices

The previous result leads naturally to the question of finding matrices that satisfy the RIP. Identifying explicit classes of matrices that have this property is a focal point of compressed sensing theory.

A *Gaussian random matrix* is a matrix whose entries are independent, normal random variables with mean zero and variance one. Such matrices have a prominent place in the compressed sensing canon, due to their particularly elegant theoretical properties. A celebrated result is that if Φ is any unitary matrix and $A = m^{-1/2}\tilde{A}$, where $\tilde{A}$ is a Gaussian random matrix, then $A\Phi$ satisfies the RIP with high probability, provided

$$m \gtrsim s \cdot \log(\mathrm{e}N/s). \tag{1.12}$$

The same result also holds if $\tilde{A}$ is replaced by a *Bernoulli random matrix* – that is, a matrix whose entries are independent random variables taking the values $+1$ and -1 with equal probability.

This result exemplifies the substantial potential of compressed sensing. It states that there are choices of measurements for which, according to (1.12), *all* vectors that are approximately s-sparse in a fixed, but arbitrary orthonormal sparsifying transform can be recovered using a number of measurements that scales linearly in s, and logarithmically in the ambient dimension N. Since $s \ll N$ in practice (recall Fig. 1.3), this represents a substantial saving over what one might classically think was needed, namely $m = N$, which of course is necessary for invertibility of A. In fact, (1.12) is *optimal* up to the constant: it is impossible, regardless of the type of matrix A and the reconstruction procedure used, to recover s-sparse vectors stably from asymptotically fewer than $s \cdot \log(\mathrm{e}N/s)$ measurements.

Key Point #5. Gaussian or Bernoulli random matrices are theoretically optimal for stable recovery of s-sparse vectors, regardless of orthonormal sparsity basis.

However, while theoretically elegant, these matrices are generally impractical for imaging problems. They are dense and unstructured, thus computationally infeasible at large scale (recall Objective #4). Perhaps surprisingly, they also offer relatively poor performance in practical imaging problems, despite being theoretically optimal for the recovery of sparse vectors. We will explain this seeming contradiction in §1.6.

Remark 1.3 Randomness is a crucial component in constructing matrices that satisfy the RIP. All constructions we see in this book involve some form of randomness, although typically less than in the case of the Gaussian random matrix, wherein each entry is an independent random variable. Despite significant research, practical constructions of deterministic matrices satisfying the RIP have remained elusive.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

1.5.5 Algorithms for Compressed Sensing

We are now nearly ready to return to the problem of image reconstruction. However, before we do so, the reader will have noticed that there is one ingredient lacking. The QCBP or LASSO problem (1.9) or (1.10) defines a reconstruction map $\mathrm{R}: \mathbb{C}^m \to \mathbb{C}^N$, $y \mapsto \hat{x}$, where $\hat{x}$ is a minimizer of the corresponding problem.[1] This is not an algorithm. To actually compute $\mathrm{R}(y)$ for a given y, we need methods that solve these optimization problems up to some tolerance in finitely many arithmetic operations. Moreover, the resulting algorithms should be efficient and scalable to large problems (Objective #4). This is the topic of Chapter 7.

Interestingly, but also surprisingly, problems such as (1.9) are generally *uncomputable*. For certain classes of inputs (A, y), no algorithm can compute a minimizer of (1.9) to arbitrary accuracy in finitely many arithmetic operations. This is in stark contrast to many classical image reconstruction techniques.

Is compressed sensing therefore doomed to failure? Is it merely an elegant theoretical idea that fails to translate to the murky world of real computations? Fortunately, the answer is no. But the reason is rather subtle. While minimizers are not computable to arbitrary accuracy, they are computable to the accuracy needed – namely, the error term $C\mathcal{S}_s(x, \eta)$ in (1.11). Moreover, this accuracy can be achieved with efficient first-order optimization algorithms capable of handling large-scale imaging problems. Chapter 8 considers this issue.

Key Point #6. Compressed sensing succeeds in practice because minimizers of problems such as QCBP, while generally uncomputable to arbitrary accuracy, are computable up to the accuracy needed. Moreover, this can be done efficiently.

1.6 Imaging with Compressed Sensing

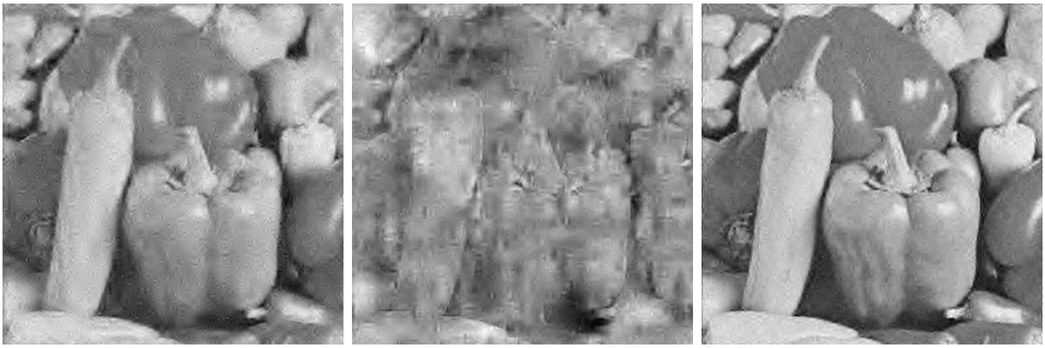
We are now ready to make an initial foray into the application of compressed sensing to imaging. We do this through a series of examples.

1.6.1 Binary Imaging

Consider a discrete binary imaging problem. As described in §1.4.2, this can be modelled by the equation $y = Ax$, where $A \in \{+1, -1\}^{m \times N}$ is a binary matrix that can often be chosen arbitrarily (Key Point #3). Computational concerns aside, conventional compressed sensing suggests that we should take A as a Bernoulli random matrix. After all, this is optimal for the recovery of s-sparse vectors (Key Point #5). Figure 1.4(a) shows the recovery of an image from these measurements. So-called DB4 wavelets are used as the sparsifying transform (these are introduced formally in Chapter 9; see Remark 9.17 therein). The reconstruction is based on the QCBP problem (1.9).

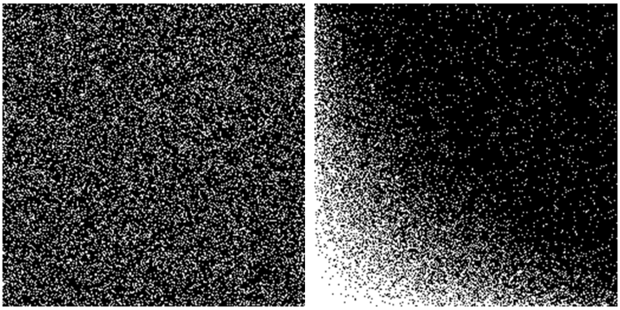
[1] Such a minimizer may not be unique. Hence R is generally a multivalued map. Fortunately, this technicality does not cause an issue, since the error bound (1.11) is guaranteed to hold for *all* minimizers.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press



(a) Bernoulli (b) Walsh, uniform random (c) Walsh, variable density

Figure 1.4 Reconstructions from 20% measurements for a 256×256 image (i.e. $m = 0.2 \times 256^2 = 13,107$). (a) Random Bernoulli measurements. (b) Walsh measurements with uniform random sampling. (c) Walsh measurements with variable-density sampling. Figure 1.5 shows the sampling patterns used in (b) and (c).



Uniform random Variable density

Figure 1.5 The Walsh sampling strategies used in Fig. 1.4(b,c). Each white dot represents a (Walsh) frequency sampled. The zero frequency is located in the bottom left corner.

On the other hand, suppose we consider the approach discussed in §1.4.3 and form A by selecting m rows of the $N \times N$ Hadamard matrix H. Because it has an associated fast transform, this is a better candidate for large-scale problems. This begs the question: which rows should we choose? Given our previous statement about randomness (Remark 1.3), a natural first choice is to select m rows uniformly at random – a strategy known as *uniform random sampling*. Figure 1.4(b) shows the reconstruction that results from these measurements. Unfortunately, the recovery quality is terrible!

Uniform random sampling is clearly a poor *sampling strategy* for Walsh measurements. Is there a better choice? It turns out that there is. The Walsh–Hadamard transform is a binary analogue of the Fourier transform. Therefore, much like with the Fourier matrix, the rows of the Hadamard matrix correspond to certain (Walsh) frequencies. Images tend to have more energy at low frequencies. Hence it makes sense to sample the low frequencies more densely and the high frequencies less densely. This is known as a *variable-density* sampling strategy. In Fig. 1.4(c) we do exactly that, with the strategy used being shown in Fig. 1.5. The improvement over uniform random sampling is dramatic. While the reconstruction shown in Fig. 1.4(b) is heavily polluted by artefacts, the reconstruction in Fig. 1.4(c) accurately recovers all the key image features.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

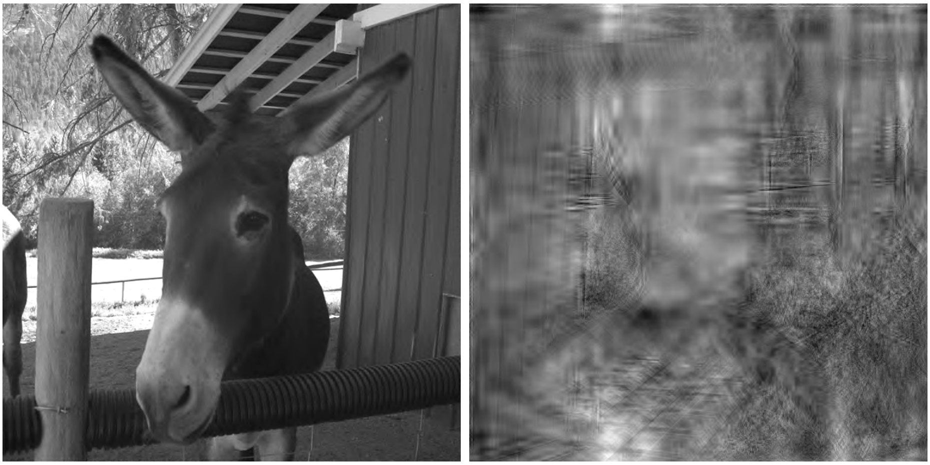


Figure 1.6 Left: A 768×768 image x which is s-sparse in the DB4 wavelet basis with $s = 58,982$. Right: The image $\tilde{x} = \Phi P \Phi^* x$ formed after a random permutation P of its wavelet coefficients $d = \Phi^* x$. Note that $\tilde{x}$ is exactly s-sparse in the DB4 wavelet basis but it is a highly nonphysical image.

1.6.2 Images and Structured Sparsity

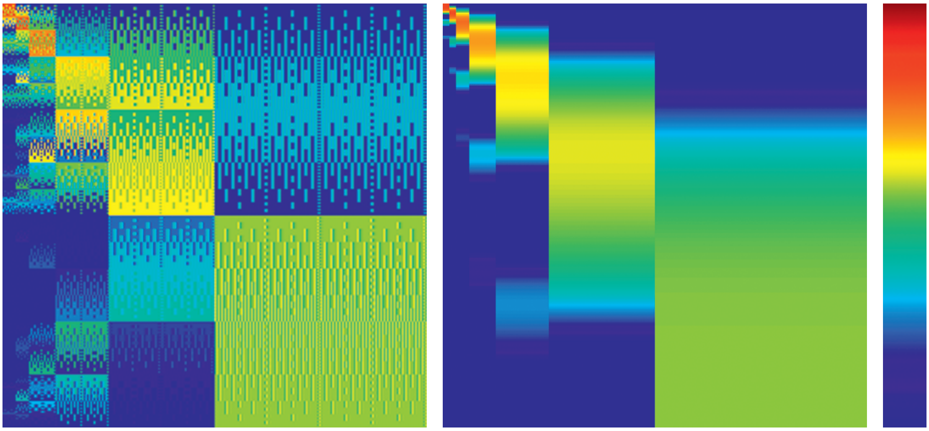
But this is not all. The variable-density Walsh sampling scheme used in Fig. 1.4(c) also gives a significantly better reconstruction than the Bernoulli measurement matrix that was used in Fig. 1.4(a).

What has happened? Sampling with a Bernoulli random matrix is optimal for sparse recovery, yet it has been handsomely beaten in this example. This is no accident, and the reason can be traced to the underlying assumption: sparsity. Sparsity in a wavelet basis is just one model for natural images. And like all models, it is imperfect. While most natural images have approximately sparse wavelet coefficients, the converse is not true. As can be seen quite dramatically in Fig. 1.6, many (in fact, most) sparse vectors of wavelet coefficients do not correspond to natural images.

In fact, wavelet coefficients of natural images possess additional *local* sparsity structure. This can already be seen in Fig. 1.3(b), where the majority of the large coefficients tend to cluster in the upper-left corner of the matrix, with progressively fewer large coefficients as one transitions to the bottom right. As we explain later, this corresponds to moving from the *coarsest* to the *finest* wavelet *scales*.

Key Point #7. Natural images are approximately sparse in wavelet bases. However, they also possess additional local sparsity structure across their wavelet scales.

So how does this relate to Fig. 1.4? Mathematically, this means that natural images are a subset of the set of vectors with s-sparse wavelet coefficients. While a Bernoulli random matrix is optimal for recovering s-sparse vectors, it ceases to be optimal over this subclass. It is essentially too conservative: its ability to recover *every* sparse vector means it is incapable of recovering those with additional structure more efficiently.



(a) Hadamard

(b) Fourier

Figure 1.7 The absolute values of the matrix (a) $H\Phi$, where H is the Hadamard matrix, and (b) $F\Phi$, where F is the Fourier matrix. In both cases, Φ corresponds to the DB4 wavelet basis. The approximate block diagonality means that wavelets at a fixed scale (a single block of columns) are approximately concentrated in a band of Walsh or Fourier frequencies (a single block of rows).

Conversely, Walsh sampling is capable of exploiting this local structure. Wavelets at different scales are essentially compactly supported in *dyadic* regions of (Walsh) frequency space. Hence the sampling strategy can be designed to efficiently target the underlying sparsity structure of the wavelet coefficients.

This can be understood a little better by examining the entries of the matrix $H\Phi$. Here H is the Hadamard matrix and Φ is the DWT matrix, and therefore the (i, j)th entry of $H\Phi$ corresponds to the ith Walsh frequency of the jth wavelet. Figure 1.7(a) plots $H\Phi$. As is clear, this matrix has a distinct block structure. The column blocks correspond to the wavelet scales, going from coarsest (left) to finest (right). The row blocks correspond to the dyadic regions of frequency space, going from lowest (top) to highest (bottom).

Evidencing the (essential) compact support of wavelets in frequency space, the diagonal blocks of this matrix contain the largest values, with the entries in the off-diagonal blocks being much smaller in magnitude. Understanding this behaviour in a precise mathematical way leads to Walsh sampling strategies such as that used in Fig. 1.4(c) that are capable of exploiting local sparsity structure. How, exactly, this is done is the focal point of Chapters 15 and 16.

Key Point #8. The local sparsity structure of natural images can be efficiently captured by variable-density Walsh sampling schemes. For a suitably chosen scheme, this significantly outperforms random Bernoulli sampling.

1.6.3 Fourier Imaging

Now consider Fourier imaging. For simplicity, we assume the discrete model based on §1.4.3, where the measurements of the image $x \in \mathbb{C}^N$ correspond to rows of the Fourier

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

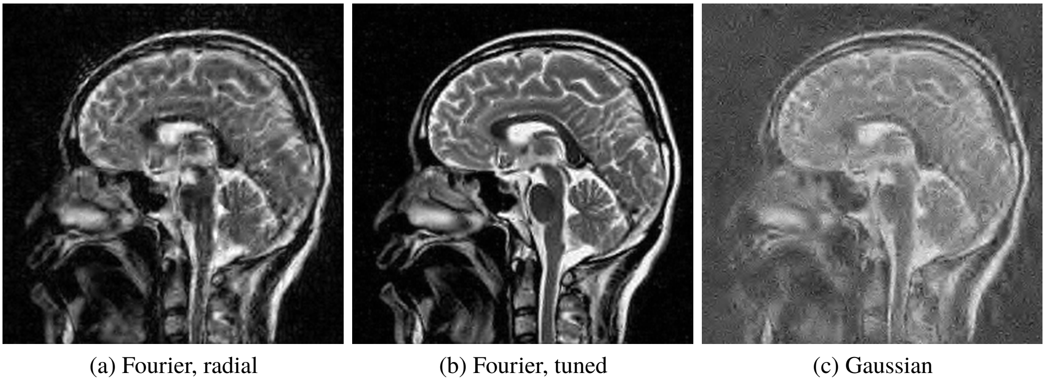


Figure 1.8 Reconstructions from 20% measurements for a 256×256 image. (a) Fourier measurements with radial sampling. (b) Fourier measurements with tuned variable-density sampling. (c) Gaussian random measurements.

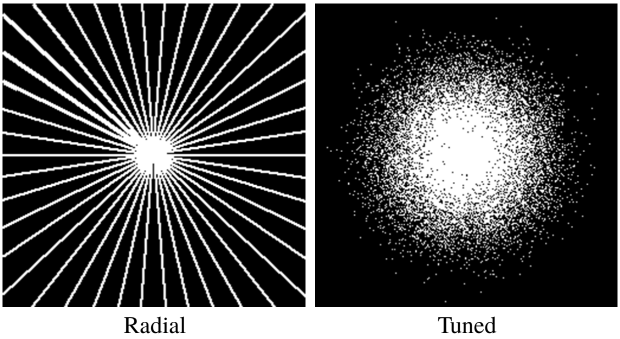


Figure 1.9 The Fourier sampling strategies used in Fig. 1.8(a,b). Each white dot represents a frequency sampled. The zero frequency is located in the centre.

matrix F. A similar story occurs in this case. Wavelets are essentially concentrated in Fourier frequency space, a fact which can again be visualized by plotting the matrix $F\Phi$. This is shown in Fig. 1.7(b). There is a similar dyadic structure to the Walsh case, which once more suggests the use of a variable-density sampling scheme taking more measurements at low frequencies and fewer at high frequencies.

A key question, which we now examine a little further, is which variable-density strategy gives the best reconstruction. In Fig. 1.8(a,b), we compare radial sampling – a standard sampling strategy in modalities such as MRI – with a variable-density scheme that has been tuned to give good performance. The schemes themselves are shown in Fig. 1.9. While both strategies qualitatively do the right thing – namely, they take more measurements at low frequencies and fewer at high frequencies – it is clear that the latter offers a significant improvement. A key topic in this book is explaining why this scheme performs better, and indeed, how it was designed in the first place. This is the aim of Chapters 15 and 16.

Figure 1.8 also compares Fourier sampling with a Gaussian random measurement matrix. This approach is not feasible in most Fourier imaging modalities – as discussed, the integral transform governing the sampling is usually fixed. Even if it were, though, doing

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

so would still not be desirable. Like with Bernoulli measurements, Gaussian measurements do not target the local sparsity structure, and correspondingly yield significantly inferior reconstructions.

Key Point #9. The benefit of compressed sensing in Fourier imaging lies not only with the fact that images are approximately sparse, but that they possess additional sparsity structure that can be exploited by carefully designed Fourier sampling strategies.

Fourier imaging modalities such as MRI have been some of the key beneficiaries of compressed sensing. This success is often attributed to sparsity. But this is not the full story. Had it been the case, we would have expected similar reconstructions in Fig. 1.8(b,c). In fact, the reason for this success is the following serendipity: the integral transform that arises in such modalities just happens to be the right one for efficiently exploiting the local sparsity structure inherent to natural images.

1.6.4 Beyond Wavelets: Total Variation and X-lets

Up to this point we have focused on wavelet sparsifying transforms. Another widely used strategy in compressive imaging involves taking $\Phi = \nabla$ to be the discrete gradient operator. This leads to the *Total Variation (TV) minimization* problem

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \text{ subject to } \|Az - y\|_{\ell^2} \le \eta,$$

where $\|\cdot\|_{TV} = \|\nabla \cdot\|_{\ell^1}$ is the TV semi-norm. Images have approximately sparse gradients, with large gradients corresponding to edges, and minimizing the ℓ^1-norm of the gradient promotes this structure. Yet gradient sparsity is also highly structured, since edges form piecewise smooth curves. Unsurprisingly, variable-density Walsh or Fourier sampling are also well suited to efficiently exploiting this local sparsity structure. Chapter 17 considers TV minimization in depth.

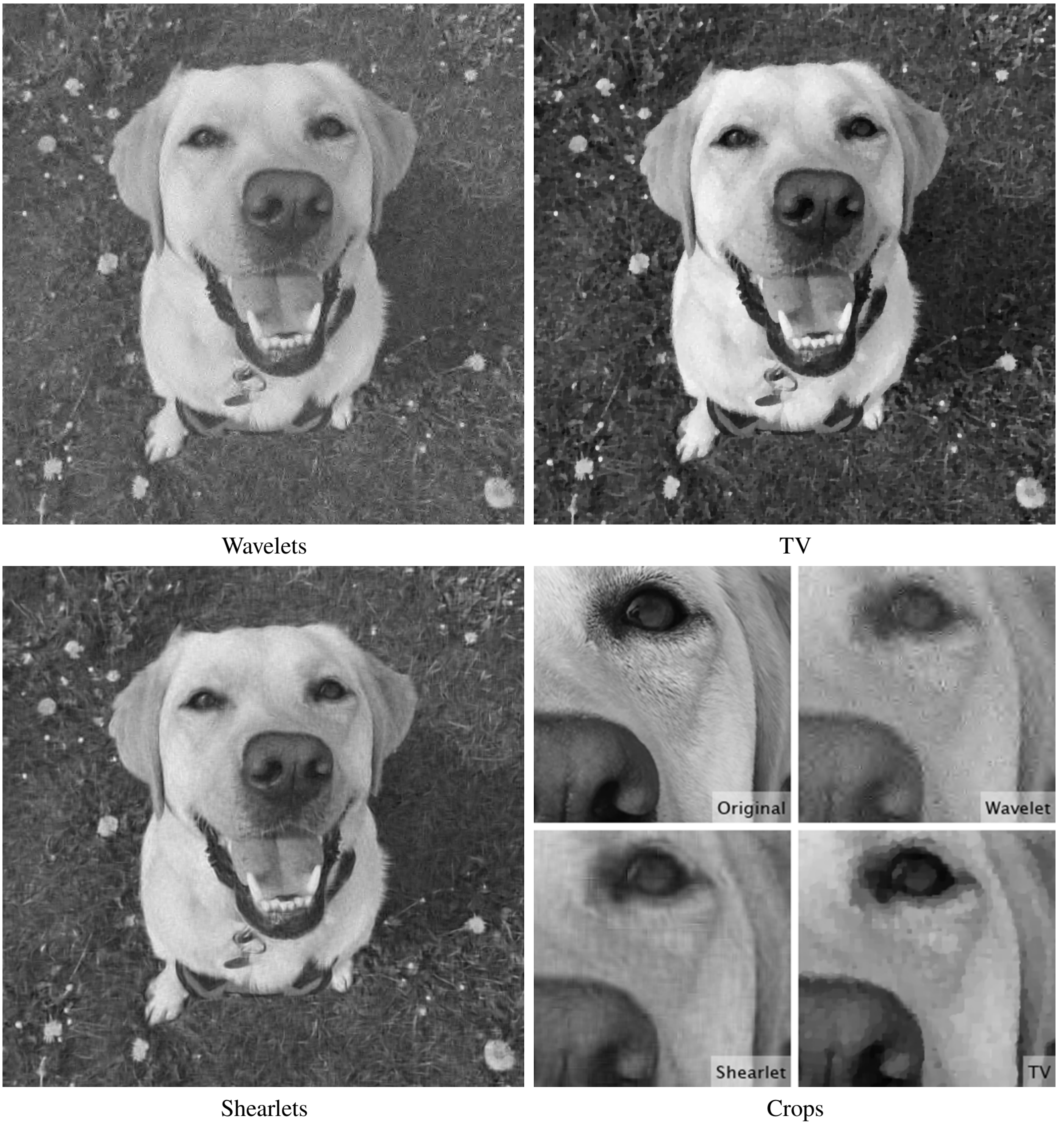
Other alternatives to wavelets are the various ‘X-let’ families – curvelets, shearlets and suchlike. These are generalizations of wavelets that aim to more efficiently capture geometric properties of images to produce higher-quality reconstructions. Images also possess similar local sparsity in such transforms, which once more is exploitable by the use of variable-density Walsh or Fourier sampling.

Figure 1.10 compares several of these transforms for Walsh sampling on a high-resolution image. Notice they all recover the image well. Yet the shearlet reconstruction exhibits fewer artefacts than both the wavelet and TV reconstructions.

1.7 Neural Networks and Deep Learning for Compressive Imaging

As noted in Key Point #1, compressed sensing ushered in a revolution in image reconstruction that led to the development of compressive imaging. Parts II–IV of this book are devoted to this topic, and in particular, the key issues raised in the previous sections. Yet, arguably, a second revolution is now taking place. Looking to push performance even further, compressive imaging strategies based on neural networks and deep learning

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press



Wavelets

TV

Shearlets

Crops

Figure 1.10 Reconstruction from 6% Walsh measurements for a 1024×1024 image, using a similar variable-density strategy to that of Fig. 1.5.

have begun to emerge in the last several years. As the reader is no doubt aware, deep learning has been used with tremendous effect in computer vision tasks such as image classification and recognition. It is therefore unsurprising that it has begun to percolate into the domain of image reconstruction. The aim of Part V of this book is to give an overview of this nascent topic and its relation to compressed sensing techniques.

1.7.1 From Model-Based to Data-Driven

Compressed sensing is primarily a *model-based* approach. The image model (i.e. structured sparsity in a fixed sparsifying transform) and reconstruction procedure (e.g. solving QCBP or LASSO) are both carefully designed in advance. However, in many imaging applications one has access to large databases of training images. This raises the following question: why not *learn* the reconstruction procedure instead of crafting it from a fixed model? Such approaches are often referred to as *data-driven*.

This is by no means a new idea. In the well-established field of *dictionary learning*, rather than using a fixed transform such as wavelets, one seeks to compute a better sparsifying transform Φ from the training data. In essence, an image model specific to the problem is learned *explicitly* from the data. This can then be incorporated into the QCBP or LASSO optimization problem to give a new recovery map based on the learned Φ. We discuss dictionary learning briefly in Chapter 4.

In some senses, the goal of deep learning is to go even further. Rather than just learning a better sparsifying transform, one instead seeks to learn the whole reconstruction map $\mathrm{R} \colon \mathbb{C}^m \to \mathbb{C}^N$. In other words, an image model is now *implicitly* learned in the process of learning R. The key idea is to choose R as a *neural network*.

1.7.2 Deep Learning for Image Reconstruction

We now describe a standard setup. Consider a *training set*

$$\Theta = \{(y_i, x_i)\}_{i=1}^{K} \subset \mathbb{C}^m \times \mathbb{C}^N, \qquad y_i = Ax_i,$$

where the x_i are the images and the y_i their respective measurements. Let $\mathcal{N}$ denote a family of neural networks $\mathrm{N} \colon \mathbb{C}^m \to \mathbb{C}^N$. We forgo a formal definition for now (see Chapter 18) and simply remark that a neural network is a mapping formed by alternately composing affine maps with an elementwise nonlinear function, known as an *activation* function. Why do this? Again, we shall not discuss the details now. But the effectiveness of deep learning on computer vision tasks strongly suggests that neural networks are capable of encoding complex image structures. Hence this is a seemingly good choice for the related problem of image reconstruction.

Having fixed $\mathcal{N}$, our task is to find a suitable reconstruction map $\mathrm{N} \in \mathcal{N}$. This is usually done by solving an optimization problem that minimizes a so-called *cost* (or *loss*) function over the training data. A simple example is the ℓ^2-loss, in which case this problem takes the form

$$\min_{\mathrm{N} \in \mathcal{N}} \frac{1}{K} \sum_{i=1}^{K} \|\mathrm{N}(y_i) - x_i\|_{\ell^2}^2. \tag{1.13}$$

The reconstruction map $\widehat{\mathrm{N}}$ is taken as a minimizer of (1.13). Solving (1.13) is known as *training* the network.

So far, all seems straightforward. However, behind this simple description lie many complexities. For one, there is the question of how to choose the class of neural networks $\mathcal{N}$. This is often termed *choosing the architecture*. Neural networks come in all sorts of

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

flavours – feedforward, recurrent, fully connected, convolutional and so forth – and so there are countless ways to select $\mathcal{N}$. Quite unlike the case of model-based approaches, there is little theoretical guidance on which choice will work best in practice. Second, the optimization problem (1.13) – as well as the many variations that it, too, can have – is large-scale and nonconvex. It requires specialized techniques such as *backpropagation* and *stochastic gradients* to compute approximate minimizers.

Yet neither issue is a barrier to progress. There are heuristic principles for designing neural network architectures which work well in practice. Computations are also being made increasingly more manageable because of the development of specialized software and hardware (e.g., GPUs). Training neural networks for image reconstruction is becoming ever more feasible for practical problems.

1.7.3 The Accuracy–Stability Barrier

In Chapters 18 and 19 we present an overview of neural networks and deep learning, with a focus on the above questions. After this, we turn our attention to the following question. Supposing one has fixed an architecture and successfully trained a neural network, how good is this as a reconstruction map? In particular, is it stable, how accurate is it, and does it outperform state-of-the-art model-based techniques based on compressed sensing?

We shall see that some trained neural networks can indeed offer very high accuracy for certain image reconstruction tasks. Yet there is reason for concern. It is well known that trained neural networks for image classification problems are highly unstable, which makes them vulnerable to so-called *adversarial attacks*. A small perturbation of an image of a 'cat', often imperceptible to the human eye, can cause it to be mislabelled as a 'dog' by the network. Adversarial attacks and how to combat them is now an active area of research within the computer vision field.

A similar phenomenon also occurs in image reconstruction. Deep learning for compressive imaging has a tendency to produce networks that are unstable and highly susceptible to small perturbations in their input (the measurements y). Specifically, if R is a trained neural network reconstruction map, $x \in \mathbb{C}^N$ is a fixed image and $y = Ax \in \mathbb{C}^m$ its measurements, it is possible to find perturbations $r \in \mathbb{C}^N$ for which

$$\|\mathrm{R}(y + Ar) - \mathrm{R}(y)\|_{\ell^2} \gg 1, \qquad \|r\|_{\ell^2} \ll 1. \tag{1.14}$$

Figure 1.11 shows a typical instance of this phenomenon. The network recovers the image extremely well from the unperturbed measurements y. But the reconstruction from the perturbed measurements $y + Ar$ is badly affected by artefacts, even though the perturbed image $x + r$ is nearly indistinguishable from the original image x. As we explain in Chapter 20, this is not a rare event. Many trained networks are susceptible to perturbations, including, in some cases, completely random perturbations.

Much of Chapter 20 explores the theoretical reasons why this instability phenomenon occurs. The general idea is that the training procedure (1.13) can encourage the reconstruction map to *over-perform*, which forces it to become unstable. Specifically, the training procedure can lead to a neural network $\widehat{\mathrm{N}}$ that is able to reconstruct two distinct images x and x' (e.g. two elements of the training set) whose difference $x - x'$ lies

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

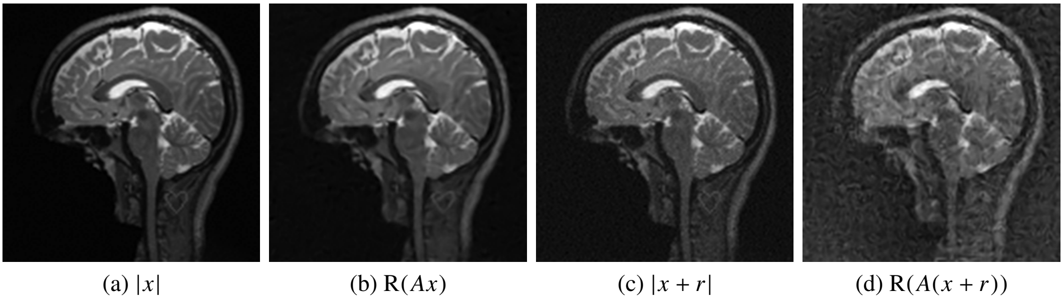


Figure 1.11 Reconstruction of a (complex-valued) image x via a deep learning strategy from Fourier measurements. For full details of this experiment, see Fig. 20.2. (a) Original image (absolute value is shown). (b) Reconstruction via the trained neural network reconstruction map R from unperturbed measurements $y = Ax$. (c) Perturbed image $x + r$ (absolute value is shown). (d) Reconstruction from perturbed measurements $y + Ar$. The perturbation is designed to ensure (1.14), i.e. to cause a significant effect on the output while being small in magnitude.

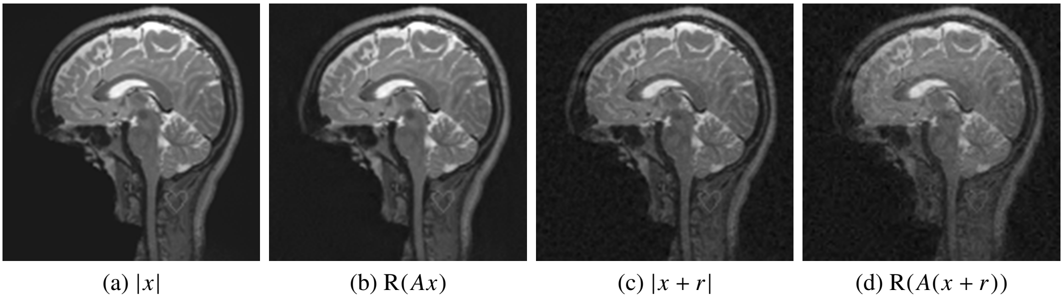


Figure 1.12 Reconstruction of the same image as in Fig. 1.11 from the same measurements using a stable neural network R. For full details of this experiment, see Fig. 21.1. (a) Original image (absolute value is shown). (b) Reconstruction from unperturbed measurements $y = Ax$. (c) Perturbed image $x + r$ (absolute value is shown). (d) Reconstruction from perturbed measurements $y + Ar$. The perturbation is designed to destabilize the network to the maximum extent possible, as in (1.14).

close to the null space of the measurement matrix A. In other words, $\widehat{\mathrm{N}}$ can reconstruct features that it has no right to recover from the measurements it sees. This lack of *kernel awareness* necessarily results in instabilities.

Key Point #10. Current deep learning strategies for compressive imaging are prone to instabilities, with certain small perturbations in the measurements leading to severe image artefacts. This arises because of the training procedure, which can cause the resulting network to over-perform.

1.7.4 Stable and Accurate Neural Networks for Compressive Imaging

This situation is rather unsatisfactory. While neural networks have produced stunning results in various imaging tasks, their use in compressive imaging appears hampered

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

by issues of instability. There is also an uncertainty surrounding their accuracy. Trained neural networks can sometimes produce very good reconstructions, yet there is little theoretical understanding as to when they do so and why.

All this raises the question: can stable and accurate neural networks be computed for compressive imaging problems? This book concludes by answering this question in the affirmative. Neural networks with the same accuracy and stability guarantees as compressed sensing are indeed computable. This is the objective of Chapter 21. To achieve this, one needs to develop a fundamental link between compressed sensing, optimization algorithms for problems such as QCBP and LASSO and neural networks. This is done through *unravelling*, the principle that a finite number of iterations of a standard optimization method for solving, for example, the QCBP problem (1.9) can be viewed as a neural network. Hence, by carefully choosing the number of iterations, one may assert the existence of a neural network with accuracy and stability matching that of compressed sensing. Through some further analysis, using similar ideas to those behind Key Point #6, one deduces that such a network can also be computed.

Key Point #11. While current training procedures can readily lead to instabilities, neural networks for compressive imaging that match the stability and accuracy of compressed sensing are computable.

To illustrate this, in Fig. 1.12 we show the same experiment as in Fig. 1.11, except using a stable neural network that was constructed in this way. The perturbation is once again computed with a view to causing a worst-case effect on the output. Yet because the network is stable, it does not lead to significant artefacts in the reconstruction. Note that this network also recovers the image from the unperturbed measurements with similar accuracy to the unstable network of Fig. 1.11.

1.7.5 Outlook

This book therefore ends on a positive note. Image reconstruction was first upended by the advent of compressed sensing. It now seems set to undergo a similar process through the introduction of neural networks and deep learning. These promising developments, however, come with pitfalls insofar as accuracy (Objective #1) and stability (Objective #3) are concerned. However, they certainly have substantial potential. As Key Point #11 makes clear, neural networks can at the very least perform as well as the current state-of-the-art. The questions of whether or not deep learning can do better, how it can do so, and by how much, are tantalizing ones on which to end this book.

1.8 Overview and Highlights

The aim of this book is to develop compressive imaging and the mathematics that underlies it. It is divided into five parts.

In Part I we cover the essentials of compressive imaging. We explain how popular imaging modalities can be formulated as linear inverse problems of the form (1.1) and

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

(1.3), and then how to apply compressed sensing techniques to them (Chapters 2 and 3). We also discuss sampling strategies and various methods for enhancing performance (Chapter 4). This part focuses on Key Points #1, #2, #3, #8 and #9.

Part II introduces the mathematics of compressive imaging via compressed sensing. We provide an overview of standard compressed sensing theory (Chapters 5 and 6), optimization for compressed sensing (Chapters 7 and 8), and wavelets and sparse approximation (Chapters 9 and 10). This part focuses on Key Points #4, #5, #6 and #7.

Next, in Part III, we develop a general compressed sensing framework that goes beyond that of Chapters 5 and 6 and, crucially, incorporates local sparsity structure (Chapters 11–13). This part of the book develops the tools needed in Part IV. We also consider the issue of discretization and develop so-called *infinite-dimensional compressed sensing* (Chapter 14) for overcoming the standard discretization errors in Fourier imaging.

In Part IV we use the techniques developed in Part III to examine the application of compressed sensing to imaging. In Chapters 15 and 16, we address the key question raised in §1.6: given the freedom, which (Fourier or Walsh) frequencies should one acquire? These chapters focus on wavelet sparsity. In Chapter 17 we consider gradient sparsity and TV minimization for imaging. This part focuses on Key Points #7, #8 and #9.

Finally, in Part V we address neural networks and deep learning. We first give a short introduction to the subject (Chapter 18), including an overview of its use in image classification. Next, we develop deep learning for compressive imaging (Chapter 19). We then conclude with two chapters examining accuracy and stability. Chapter 20 investigates how and why existing approaches tend to be unstable. On the other hand, Chapter 21 shows that neural networks that are as stable and accurate as compressed sensing are indeed computable. This part focuses on Key Points #10 and #11.

1.9 Disclaimers

1.9.1 What this Book Is and What it Is Not

Imaging is a vast topic. There are many interesting aspects of it that cannot be included in a single text. The guiding principle behind this book is to be both relevant to practice *and* mathematically rigorous, wherever possible. It aims to strike a delicate balance between methods that admit theory and those that are used in practice. Fortunately, these two sets have a large intersection! Moreover, many, if not most of the techniques used in practice are close cousins of those that can be theoretically analysed. This is arguably one of the most exciting aspects of compressed sensing research. This book therefore assumes the reader is interested not only in methods for compressive imaging, but also *how* and *why* they work. For readers only interested in finding an off-the-shelf, state-of-the-art reconstruction algorithm, this is not the best place to look. Having said that, in Chapter 4 we do give some recommendations for a good all-round compressive imaging strategy.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

1.9.2 Areas Not Covered

This book is about image reconstruction from *indirect* measurements. A related problem is image *restoration*, in which one directly samples a corrupted version of the image itself. *Denoising*, *deblurring* and *inpainting* are examples of image restoration problems. While these are important image processing tasks, they are outside the scope of this book.

The underlying inverse problem encountered in this book is either well-posed (e.g. the Fourier transform) or only mildly ill-posed (e.g. the Radon transform). Sparsity appears frequently in ill-posed inverse problems as a regularization tool, wherein it is often termed *sparse regularization*. This topic is similar, but not the same as compressed sensing. Generally speaking, compressed sensing is concerned with issues around sampling and the number of measurements needed to achieve good recovery. In sparse regularization, by contrast, the main concern is often with the nature and quality of the solutions obtained from a sparsity-promoting regularization procedure.

This book takes a deterministic approach to image reconstruction in inverse problems (with the caveat that the sampling strategies are themselves random). Statistical methodologies such as Bayesian inverse problems have their place in compressive imaging – especially with the latest innovations in deep learning – but they are outside our scope.

1.9.3 Topics Not Covered

Within compressive imaging, we have tended to exclude, or only lightly touch upon methods that lack theory. For example, much of the book focuses on wavelets or TV minimization. We spend more time on the former, not because we advocate for it strongly over the latter, but because compressed sensing theory for wavelets is more comprehensive. We recommend the practitioner try both.

We have also omitted in-depth descriptions of curvelets and shearlets, as well as generalizations of TV such as *Total Generalized Variation (TGV)*. The compressed sensing theory for these approaches is generally less mature than that for wavelets and TV, although one might expect an eventual theory to be qualitatively similar, at least when it comes to the issue of sampling and sparsity structure. We also largely avoid more sophisticated reconstruction procedures, including learned strategies and nonconvex approaches. In Chapter 4 we briefly overview these and other techniques for boosting practical performance. We also do not treat structured sparsity models such as *wavelet trees*. While interesting, it is arguably the case that such models have not found widespread use in compressive imaging.

In terms of imaging modalities, after Part I we largely ignore tomographic imaging via the Radon transform. This is primarily because this problem has so far resisted substantial mathematical analysis from the compressed sensing perspective. We also focus on static, single-image recovery problems. Compressed sensing techniques have been applied to many *dynamic* imaging problems. Arguably, though, the particularities of dynamic compressive imaging are sufficiently manifold for it to be worthy of a book in its own right. For this reason, we also do not discuss *low-rank matrix recovery*. While closely related to compressed sensing, its use in imaging has predominantly been in

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

dynamic imaging scenarios. We mainly treat standard, but ubiquitous, imaging models which can be recast as linear inverse problems of the form (1.1) or (1.3). We do not address nonlinear sampling operators. In particular, the *phase retrieval* problem, while it arises in some common imaging modalities, is outside the scope of this book. We also do not address calibration in any depth. While this is an important challenge in many practical imaging systems, it is also one that is often very specific to the modality.

Learning some part of the imaging process is an important topic in compressive imaging. As noted, a common approach is dictionary learning. Optimization parameters, sampling strategies and various other components of the imaging pipeline can also be learned. While much can and has been written about these 'classical' approaches to learning, we have chosen to forgo this. We have instead focused Part V of the book on the most contemporary approaches to learning in compressive imaging, namely, deep learning. Some discussion on dictionary learning is included in Chapter 4.

Note that many of the topics mentioned above are discussed in a little more detail in the Notes section of subsequent chapters, along with references for further reading.

Finally, we remark that we generally do not concern ourselves with constants in this book. Theoretically obtaining explicit constants in compressed sensing is a challenge, and one that arguably sheds little light on the underlying principles. We use the notation $\lesssim$ and $\gtrsim$ quite liberally throughout.

1.10 Reading this Book

We encourage everyone to read Part I to get a taste of compressive imaging. Readers familiar with conventional compressed sensing, optimization and wavelets may wish to skip elements of Part II. After Chapter 11, the remainder of Part III contains the main technical mathematics in the book. Part IV – specifically, Chapters 15 and 17 – answers one of the book's main questions: how should one sample? These chapters are intended to be read without necessarily having read Part III in detail. Finally, we recommend Part V to anyone who is interested in the latest innovations in compressive imaging.

1.10.1 Examples and Code

Aiming to make this book a practical guide for compressive imaging, we have included many numerical examples. A companion software library, *CIlib*, is available online:

`https://github.com/vegarant/cilib`

It is also accessible through the book's website

`www.compressiveimagingbook.com`

This library contains a broad set of functions and tools for compressive imaging, as well as code that reproduces many of the figures in the book. We encourage the reader to investigate this library. The website also contains further information about the experimental setup and software packages used.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press

1.10.2 Notes

Each subsequent chapter of this book concludes with a Notes section (or sections). These contain additional information, discussion and suggestions for further reading.

1.10.3 Prerequisites

In closing this chapter, we remark that this book assumes a level of mathematical background of a typical early-stage graduate student, but little specialist knowledge. The Appendices, while not comprehensive, contain some additional background material. The Notes sections at the ends of the chapters also give additional information and references.

https://doi.org/10.1017/9781108377447.002 Published online by Cambridge University Press