CAMBRIDGE

Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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.

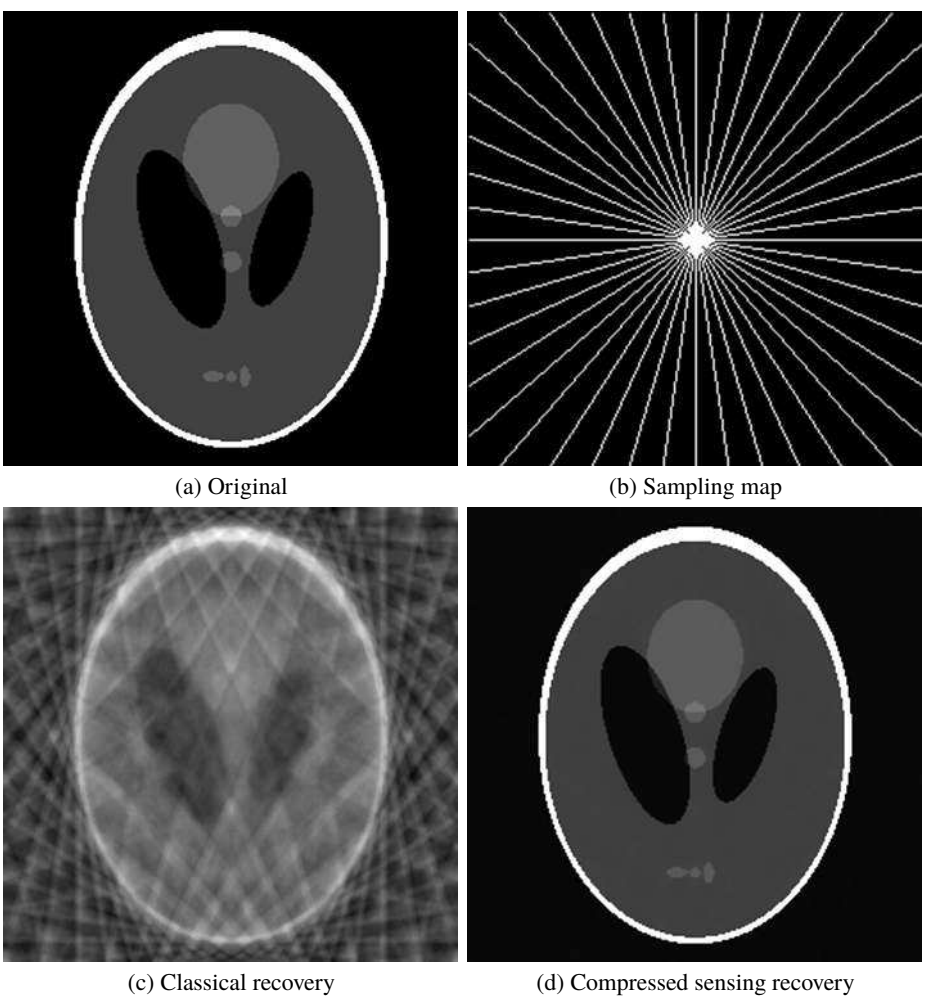
The growth of compressive imaging has wrought a profound change on practical image reconstruction over the past decade. In many applications, not least MRI, classical linear recovery procedures have been replaced by a new generation of techniques.

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information



(a) Original (b) Sampling map

(c) Classical recovery (d) Compressed sensing recovery

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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 $\Omega = \{i_1, \dots, i_m\} \subseteq \{1, \dots, N\}$ 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

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

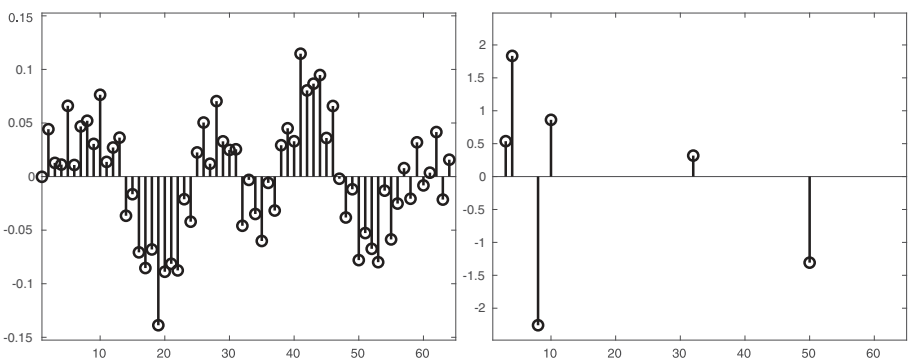


Figure 1.2 Left: Discrete signal $x \in \mathbb{R}^{64}$. Right: The 6-sparse vector of coefficients $d = \Phi^* x \in \mathbb{R}^{64}$, where Φ is a DCT.

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

(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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-42161-4 — Compressive Imaging: Structure, Sampling, Learning
Ben Adcock , Anders C. Hansen
Excerpt
More Information

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

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

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

© in this web service Cambridge University Press www.cambridge.org