

# Chapter 11

## Magnetic Resonance Imaging

### 11.1 Introduction

Magnetic resonance imaging (MRI) is based on the science of nuclear magnetic resonance (NMR). Magnetic resonance states that certain atomic nuclei (such as protons in the water molecules) can absorb and emit radio frequency energy when placed in an external magnetic field. The emitted energy is proportional to important physical properties of a material such as proton density. Therefore in physics and chemistry, magnetic resonance is an important method for studying structures of chemical substances, and its discovery was awarded the Nobel Prize in 1952.

Later in 1970's, Paul Lauterbur and Peter Mansfield discovered that by introducing spatial gradients in the magnetic field, it is possible to create two-dimensional images of the structures, now known as magnetic resonance imaging (MRI). MRI soon proved to be extremely useful for medical diagnosis as it provided an accurate and non-invasive method for imaging internal organs of the human body. Unlike the X-rays or computed tomography (CT) scans, MRI does not exert ionizing radiation, hence is much less harmful. Today, MRI has become a routine medical examination in hospitals worldwide, especially for examining the brain and the spinal cord. For their contributions to MRI, Lauterbur and Mansfield were awarded a Nobel Prize in Physiology or Medicine in 2003.

Nevertheless, MRI machines can be rather expensive, and the acquisition process of MRI is considerably time-consuming as it needs to densely sample the magnetization responses with many different gradient fields. In

order to lower the cost of MRI and improve patient comfort or safety,<sup>1</sup> in recent years, techniques from compressive sensing have proven to be extremely effective in improving the efficiency of MRI [Lustig et al., 2008b], which was briefly highlighted in Chapter 2 as one of the heralding successes.

In this chapter, we explain in more technical detail why MRI is particularly suitable for techniques in compressive sensing. First, a high-level review of the physics of MRI in Section 11.2 reveals that the MRI imaging process is amenable to compressive sampling as it naturally take spatially encoded samples of the image in the frequency domain. Secondly, medical images of human organs are naturally structured and mostly piecewise-smooth. We verify empirically that such images are highly compressible/sparse in a properly chosen transform domain, and introduce several effective sampling schemes in Section 11.3. Finally, we introduce in Section 11.4 some customized fast algorithms that can efficiently reconstruct the image from such compressive samples with high fidelity, despite imaging noise and other artifacts.

## 11.2 Formation of MR Images

In medical applications, MRI is based on measurements of a radio frequency (RF) signal, known as the *transverse magnetization*, generated by protons which exist in abundance as the hydrogen nuclei in the molecules of water and fat in human tissues. The signal measured is largely proportional to the density of protons at each spatial location, which indicates the presence or absence of such molecules. This information can then be used by physicians for diagnostic purposes. Here, we give a simplified mathematical model that captures the essence of this process. For a more detailed description of the physical process, one may refer, e.g., to [Wright, 1997] or textbook introductions to MRI.

### 11.2.1 Basic Physics

It is known in quantum physics that each proton spins along an axis that creates an angular momentum. In the absence of any external magnetic field, the angular momenta of the protons are oriented randomly in their neutral state, hence collectively the protons (in the body tissue) do not produce any measurable magnetization. However, when a strong external magnetic field, denoted as  $\mathbf{B}_0$ , is applied to the tissue mass, it polarizes the protons and aligns their spins along the direction of  $\mathbf{B}_0$  and produces a

---

<sup>1</sup>For young pediatric cancer patients, frequent exposures to strong magnetic fields for long periods of time can still be unsafe.

net magnetization, denoted as  $\mathbf{M}$ .  $\mathbf{B}_0$  is also called the *primary magnetic field*, and its strength typically can be ranged from 1.5 to 3 Tesla.<sup>2</sup>

Following conventional notations in physics, we use  $(\mathbf{i}, \mathbf{j}, \mathbf{k})$  to denote the three unit vectors in the  $x, y, z$  axes of a (local) Cartesian frame. Without loss of generality, we may assume the direction of the external static magnetic field  $\mathbf{B}_0$  aligns with the  $z$ -axis, that is,  $\mathbf{B}_0 = B_0 \mathbf{k}$ . In general the magnetization  $\mathbf{M}$  takes the form  $\mathbf{M} = M_x \mathbf{i} + M_y \mathbf{j} + M_z \mathbf{k}$ . If the external magnetic field is static,  $\mathbf{M}$  will eventually reaches an equilibrium magnetization of the form  $M_0 \mathbf{k}$ .

Although protons can respond very quickly to the external magnetic field, the polarization itself does not yield any RF signal that can be measured by the machine. The key is that the magnetization  $\mathbf{M}_{xy} = M_x \mathbf{i} + M_y \mathbf{j}$  in the *transverse plane* orthogonal to the primary axis undergoes very different dynamics and can be exploited for measurement. This transverse magnetization precesses about  $\mathbf{B}_0$  according to the so called *Bloch equation*:

$$\frac{d\mathbf{M}_{xy}}{dt} = \gamma \mathbf{M}_{xy} \times \mathbf{B}_0, \quad (11.2.1)$$

where  $\gamma$  is a constant. From this equation, we see that the precession frequency is  $\omega_0 = \gamma B_0$ , known as the *Larmor frequency*. This rotating magnetic moment radiates an electromagnetic signal which is picked up by the MRI machine.

So in order to produce a precessing magnetization orthogonal to  $\mathbf{B}_0$ , in the second step of MRI imaging, one applies a second time-varying magnetic field  $\mathbf{B}_1$  in the  $xy$  plane transverse to  $\mathbf{B}_0 = B_0 \mathbf{k}$ . An MRI machine usually has three RF coils along the  $x, y, z$  axes respectively, as shown in Figure 11.1, and can produce a magnetic field in any direction by running electric currents through respective coils. Typically  $\mathbf{B}_1$  is chosen to

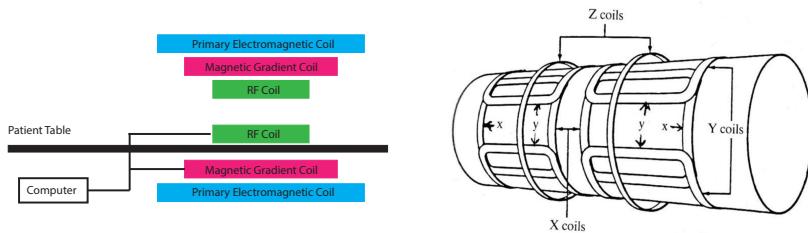


Figure 11.1. **Left:** Key components of a basic MRI machine. **Right:** The three-axis gradient coils.

be  $\mathbf{B}_1 = \cos(\omega_0 t) \mathbf{i} + \sin(\omega_0 t) \mathbf{j}$  which rotates around  $\mathbf{B}_0$  at the radian

---

<sup>2</sup>In comparison, the magnitude of the Earth's natural magnetic field only ranges from 25 to 65 micro Tesla.

frequency  $\omega_0$ . This magnetic field excites the protons to a higher energy state.

The excitation stops after a short period, and the protons gradually fall back to its equilibrium state. This relaxation process lasts from milliseconds to several seconds. As the transverse magnetization  $\mathbf{M}_{xy}$  precesses, it induces an electromagnetic force in the RF coils. The magnitude, phase, and relaxation time of the signal represent different properties of the matter that can be recorded in different types of MR images.

### 11.2.2 Selective Excitation and Spatial Encoding

One question remains with regard to the imaging process, namely, how does the MRI machine isolate and measure the RF signal from different parts of the body, because if the body is affected by a single static magnetic field, then all the protons will be aligned homogeneously. Several clever (Nobel prize worthy) techniques are required to address this issue, including the so-called *selective excitation* and *spatial encoding*. The goal is to be able to sample and measure magnetization signals from any spatial location  $(x, y, z)$  (up to certain resolution).

As the transverse magnetization  $\mathbf{M}_{xy}$  is of interest, we may choose to excite the magnetic field in a thin slice along the  $z$ -axis, say around  $z_0$ . That is, we are interested in the plane  $(x, y, z_0)$ . The selective excitation can be achieved by first making the Larmor frequency varying linearly in the  $z$ -direction with the magnetic field

$$\mathbf{B}(z) = (B_0 + G_z(z - z_0))\mathbf{k}.$$

We then apply an additional RF excitation pulses with energy over a limited range of frequency bandwidth  $\Omega$  corresponding to the Larmor frequency at the slice  $z_0$ . Typically, we could choose the pulse to be

$$\mathbf{B}_1(t) = \text{sinc}(\Omega t)(\cos(\omega_0 t)\mathbf{i} + \sin(\omega_0 t)\mathbf{j}),$$

which has a rectangular (energy) distribution around the temporal frequency  $\omega_0$ . As result, only protons around the slice  $(x, y, z_0)$  may resonate with the excitation and reach a high magnetization level, say  $M_{xy}(x, y) \doteq M_{xy}(x, y, z_0)$ . This is why the whole process is called *magnetic resonance imaging*.

The remaining question is how to image magnetization of different spatial locations inside the plane  $(x, y, z_0)$ . As we see from the above, spatial selectivity (along the  $z$ -axis) can be achieved through introducing a spatially varying excitation with a gradient ( $G_z$ ) in the  $z$  direction. Hence, we could generalize this idea by introducing an additional magnetic field  $\mathbf{B}$  with a gradient  $G_x$  and  $G_y$  in the  $x$  and  $y$  directions, respectively. Moreover, we could vary the gradients as a function of time  $t$ :

$$\mathbf{B} = (B_0 + G_x(t)x + G_y(t)y)\mathbf{k}.$$

After the selective excitation, the magnetic field in the transverse plane  $(x, y, z_0)$  is  $M_{xy}(x, y)$ . Once the slice is subject to the above magnetic field,  $M_{xy}$  precesses according to the Bloch equation and we can measure the electromagnetic signal generated by it. Assume that the magnitude  $|M_{xy}|$  remains relatively constant during the acquisition period, then from the Bloch equation, we have:

$$M_{xy}(x, y, t) = |M_{xy}(x, y)| e^{-j\omega_0 t} e^{-j\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y) d\tau}.$$

From this equation, we can ascertain the true reason for introducing a gradient magnetic field: It allows us to manipulate the phase of the transverse magnetic field  $M_{xy}$  so as to encode the spatial information about  $M_{xy}$  that we needed in the first place.

To see this, note that the actual signal we measure is the collective effect of all  $M_{xy}$  in the  $xy$ -plane. To simplify the notation, let us define

$$k_x(t) \doteq \gamma \int_0^t G_x(\tau) d\tau, \quad k_y(t) \doteq \gamma \int_0^t G_y(\tau) d\tau.$$

In the MRI literature, the so-defined quantities  $(k_x, k_y)$  index a two-dimensional space called  $k$ -space. We then have the measured signal, say  $s(t)$ , as

$$s(t) = \exp^{-j\omega_0 t} \int_x \int_y |M_{xy}(x, y)| e^{-j(k_x(t)x + k_y(t)y)} dx dy.$$

Notice that this measured signal  $s(t)$ , once with the  $e^{-j\omega_0 t}$  component demodulated, is essentially a *2D spatial Fourier transform* of  $|M_{xy}(x, y)|$  at the spatial frequency  $(k_x(t), k_y(t))$ :

$$S(k_x, k_y) = \int_x \int_y |M_{xy}(x, y)| e^{-j(k_x x + k_y y)} dx dy. \quad (11.2.2)$$

In the MRI literature, this technique is called *spatial frequency encoding*. So, in principle, once we have collected measurements of  $S(\cdot)$  at sufficiently many spatial frequencies  $(k_x, k_y)$ 's, we could recover  $|M_{xy}(x, y)|$  simply from its *inverse Fourier transform*:

$$|M_{xy}(x, y)| \propto \int_{k_x} \int_{k_y} S(k_x, k_y) e^{j(k_x x + k_y y)} dk_x dk_y, \quad (11.2.3)$$

which can be visualized as a 2D image, say  $I(x, y)$ , on the  $xy$ -plane (at  $z_0$ ).

### 11.2.3 Sampling and Reconstruction

We have described above in a nutshell the physical and mathematical models of MR imaging. In short, we see that the value of the measured signal  $S$  at any given time  $t$  is essentially the 2D Fourier transform of the image of interest  $I(x, y)$  (or  $|M_{xy}(x, y)|$ ) at a particular spatial frequency  $(k_x, k_y)$ . For any given gradient field generated by  $(G_x(t), G_y(t))$ ,

if we measure the signal  $S$  at a sequence of time  $\{t_1, t_2, \dots\}$ , we obtain the samples of the Fourier transform of  $I(x, y)$  at different frequencies  $\{(k_x(t_1), k_y(t_1)), (k_x(t_2), k_y(t_2)), \dots\}$  in the transform domain (the  $k$ -space).

In practice, we are interested in recovering the image up to certain spatial resolution. That is, instead of a function on a continuous domain (the entire  $xy$ -plane), we consider the image  $I(x, y)$  is a function on a finite Cartesian grid (say of size  $N \times N$ ). We denote the coordinates of the pixels as a vector  $\mathbf{v} = (x, y)$ . In this case, the measurements can be viewed as discrete Fourier transform of the image, which lie on a Cartesian grid (of size  $N \times N$ ) in the  $k$ -space. We denote the coordinates of the frequencies as a vector  $\mathbf{u} = (k_x, k_y)$ . We collect all measurements as a vector  $\mathbf{y} \in \mathbb{R}^m$  with  $m = N^2$ . That is, each entry of  $\mathbf{y}$  is of the form:

$$y_i = \sum_{\mathbf{v}} I(\mathbf{v}) e^{-j\mathbf{u}_i^* \mathbf{v}} \Delta\mathbf{v}, \quad i = 1, \dots, m$$

where the sum is over the grid and  $\Delta\mathbf{v}$  is the grid step size. If we also view the image  $I(\mathbf{v})$  as a vector of dimension  $m$ , then we have:

$$\mathbf{y}(\mathbf{u}) = \mathcal{F}[I(\mathbf{v})](\mathbf{u}), \quad (11.2.4)$$

where  $\mathcal{F}$  is an  $m \times m$  matrix representing the discrete (2D) Fourier transform. As the matrix  $\mathcal{F}$  is invertible, the MR image  $I$  can be simply recovered from such Cartesian samples as:

$$I(\mathbf{v}) = \mathcal{F}^{-1}[\mathbf{y}(\mathbf{u})](\mathbf{v}). \quad (11.2.5)$$

Figure 11.2 shows an example of recovered MR image from such Cartesian sampling scheme.

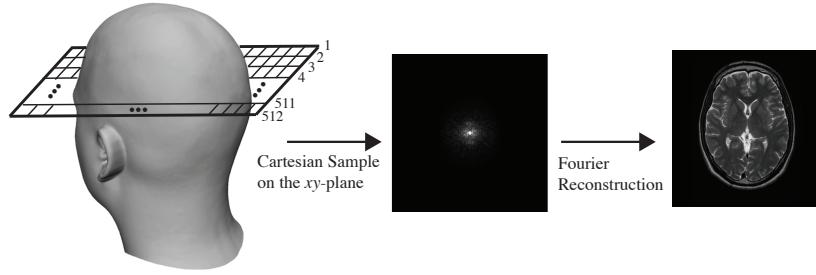


Figure 11.2. A Cartesian sample of the human brain (left) and its reconstructed MRI image (right). The sampling resolution in this example is  $m = 512 \times 512$ .

At first sight, sampling the entire Cartesian grid in the transform domain seems natural, and the reconstruction via inverse Fourier transform is straightforward. However, for practical images, it is too redundant to get all

the  $m = N^2$  samples as illustrated in the example in Figure 11.2. Conventional signal processing techniques have been applied to reduce the number of samples. For instance, if the image largely consists of low-frequency components and has a cutoff bandwidth  $f_{\max}$ , then we only need to sample the transform domain on a sub grid according to the Nyquist rate

$$f_{\text{Nyquist}} \geq 2f_{\max}.$$

Nevertheless, the number of samples required by the Nyquist rate is still very large,<sup>3</sup> which makes the conventional MRI imaging process very time-consuming. For the rest of this chapter, we will see that by harnessing additional structures (e.g., sparsity) of the MR image, one can significantly reduce the number of samples needed. We will first discuss the sparsity of MR images, and then introduce a few effective compressive sampling schemes. Finally, we will discuss numerical methods for reconstructing MR images from small sets of samples, since, in this undersampled regime we can no longer simply rely on the inverse Fourier transform.

## 11.3 Sparsity and Compressive Sampling of MR Images

### 11.3.1 Sparsity of MR Images

In order to improve the sampling efficiency of MR images, we need to leverage additional structure of the target image  $I$ . We know from Chapters 2–3 that *sparsity* is a very powerful structural assumption, which, when present can substantially reduce the number of measurements that are required to reconstruct a signal of interest. However, MR images are not sparse – most of the pixels are nonzero! On the other hand, MR images *are* structured: they tend to be approximated as piecewise smooth functions with relatively few sharp edges. We will see that this type of structure actually leads to a form of sparsity, in an appropriately chosen transform domain.

From signal processing and harmonic analysis, we know that piecewise smooth functions are compressible (nearly sparse) when represented in terms of appropriate basis functions – say, wavelets. There is a deep theory associated with wavelets and related 2D signal representations. Here, we only sketch these constructions at a loose, operational level.

A wavelet transform  $\Phi$  maps an  $N \times N$  image  $I$  to a collection of  $N^2$  coefficients  $\mathbf{x} = \Phi[I]$ . The inverse transform  $\Psi = \Phi^{-1}$  maps the coefficients  $\mathbf{x}$  to an image  $I = \Psi[\mathbf{x}]$ . The inverse mapping can be interpreted as

---

<sup>3</sup>For images with sharp edges and contours, its cutoff bandwidth may be very high.

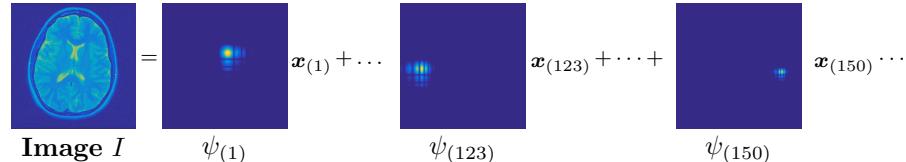


Figure 11.3. **Wavelet representation of an image.** The image  $I$  is expressed as a superposition of basis functions  $\psi_i$ , with coefficients  $x_i$ . In this figure, we order the coefficients by magnitude, in descending order:  $x_{(1)}$  is the largest magnitude coefficient, and  $\psi_{(1)}$  its corresponding basis function,  $x_{(2)}$  the second largest, and so on. The largest coefficients capture low-frequency structure, as well as high-frequency structure around edges. Notice, e.g., that  $\psi_{(123)}$  and  $\psi_{(155)}$  are located near sharp edges at the left and right side of the brain.

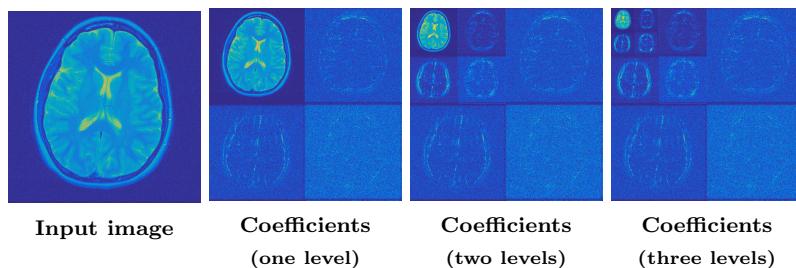


Figure 11.4. **Wavelet coefficients of an image.** From left to right, an original image, and the coefficients of one level, two level and three level wavelet decompositions using the Daubechies db4 wavelet. The level one coefficients are organized as LL (upper left), LH (upper right), HL (lower left) and HH (lower right). The detail coefficients (high frequency) are concentrated near sharp edges.

expressing the image as a superposition of basis function  $\psi_1, \dots, \psi_n$ :

$$I = \Psi[\mathbf{x}] = \sum_{i=1}^{N^2} \psi_i x_i. \quad (11.3.1)$$

Figure 11.3 visualizes several of the basis functions associated with a particular two-dimensional wavelet transform.<sup>4</sup>

The coefficients  $\mathbf{x}_i$  have a very nice interpretation. To transform the image  $I$ , we split the image into four bands, which capture vertical and horizontal frequency content at different spatial locations in the image. The low frequency band, typically labeled LL, contains low frequency in both directions, while the high frequency band HH contains high frequency con-

---

<sup>4</sup>There are a variety of wavelets, leading to a variety of different transforms. In the experiments in this chapter, we adopt the Daubechies db4 wavelet. Other choices of separable wavelets lead to different transforms, but their behavior is qualitatively similar.

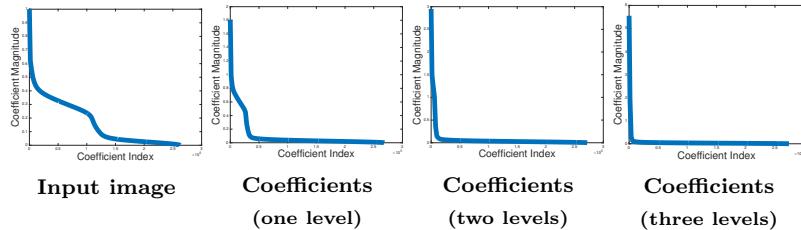


Figure 11.5. **Decay of the wavelet coefficients.** Left: the magnitudes of the image pixel values, plotted in descending order. Right: the magnitudes of the wavelet coefficients  $\mathbf{x}$  in descending order, for one, two and three level wavelet transforms. The wavelet coefficients decay much more rapidly than the original pixel values.

tent in both directions. Two other bands **HL** and **LH** contain high frequency content in one direction and low-frequency content in the other direction. Figure 11.4 illustrates this operation. Notice that most of the significant entries occur in the **LL** band. By repeating this operation to the **LL** band, we obtain a two-level transform which captures localized frequency content at multiple scales in the image. We can continue in this manner. Figure 11.4 illustrates the three level to five level transforms of this image.

MR images tend to be piecewise smooth, with only a few sharp edges. The **HL**, **LH** and **HH** coefficients concentrate around edges, and so they tend to be quite sparse. Indeed, classical results in harmonic analysis can be paraphrased as arguing that the one dimensional version of this representation is nearly optimal for representing one-dimensional functions which are piecewise smooth with only a few discontinuities.<sup>5</sup> Figure 11.5 plots the sorted magnitudes of the coefficients  $\mathbf{x}$ , for each level  $l$ , from  $l = 0$  (the original image) to  $l = 3$ . Notice that as we increase the number of levels in the transform, the coefficients become increasingly compressible.

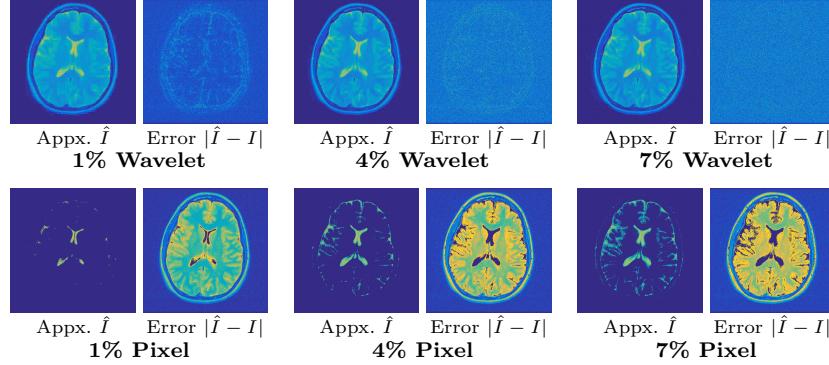
Because the wavelet coefficients are nearly sparse, we can accurately approximate the input image using just a few wavelet coefficients. Let  $J = \{i_1, \dots, i_k\}$  denote the indices of the  $k$  largest coefficients  $x_i$  (across all scales) in absolute value. We can form the *best k-term approximation*

$$\hat{I} = \sum_{i \in J} \psi_i x_i, \quad (11.3.2)$$

by retaining only these largest coefficients. Figure 11.6 visualizes approximations with the best 1%, 4% and 7% of the coefficients, respectively. It also visualizes the approximation error  $|I - \hat{I}|$ . Notice that the approximation errors are almost entirely populated with noise. For comparison, Figure 11.6 (bottom) shows approximations using the best 1%, 4% and 7%

---

<sup>5</sup>For piecewise smooth functions on a two-dimensional domain, the situation is more complicated, and there is a large literature of image representations.



**Figure 11.6. Wavelet Reconstructions.** Top: Approximations of the brain image using the most significant wavelet coefficients. We plot reconstructions  $\hat{I}$  using the largest 1%, 4% and 7% of the wavelet coefficients, as well as the approximation error  $|\hat{I} - I|$ . Retaining roughly 7% of the wavelet coefficients captures most of the important structure in the image; what remains is mostly noise. Bottom: For comparison purposes, we plot reconstructions and errors using the 1%, 4% and 7% largest image pixels. These approximations are very inaccurate: the image is nearly sparse in the wavelet domain, but not in the original pixel domain.

of the original image pixels. The wavelet approximations are dramatically more accurate than pixel approximations.

Of course, there is no reason to believe that wavelet sparsity captures *all* of the structure in an MR image. Other structural assumptions may lead to sparser representations, which can be leveraged to sample even more efficiently. The literature is rich with alternatives, including representations that capture oriented edges, nonlocal representations that capture repeated structure, and learned representations that adapt to the specific classes of images. We will return to this point in Section 11.4, where we sketch one means of further reducing the sampling burden for MRI, by leveraging an additional form of sparsity. For now, we turn to the question of how we can use the knowledge that the wavelet coefficients are sparse to sample more efficiently.

### 11.3.2 Compressive Sampling of MR Images

Although a wavelet transform is able to sparsify the MR image  $I$ , notice that we cannot have access to the wavelet coefficients  $\mathbf{x}$  unless we have acquired the entire image  $I$  (and then apply the transform  $\Phi$ ). Hence in conventional image processing, wavelet transforms have mostly been used in the post-processing of an image after it has been acquired, such as for compression. Now the question is how can we exploit the fact the MR

image is sufficiently sparse in certain (wavelet) domains so that we can significantly reduce the number of measurements sampled in the acquisition time and still recover the image with good quality?

First we notice that the relationship between the measurements (Fourier coefficients)  $\mathbf{y} \in \mathbb{C}^{N^2}$  and the (sparse) wavelet coefficients  $\mathbf{x} \in \mathbb{R}^{N^2}$  is given by:

$$\mathbf{y} = \mathcal{F}[\Psi\mathbf{x}].$$

From the physical model we have described above, we can directly measure any subset of the Fourier coefficients  $\mathbf{y}$  or any linear superpositions of them. For convenience we denote the image  $I$  as a vector  $\mathbf{z} \doteq I \in \mathbb{R}^{N^2}$ :

$$\mathbf{z} = \Psi\mathbf{x}.$$

Suppose, instead of taking all the  $N^2$  Fourier coefficients, we measure only  $m \ll N^2$  samples of (linear superpositions) of the Fourier coefficients. Then the transform from  $\mathbf{z}$  to the  $m$  partial measurements  $\mathbf{y}$  can be represented as an  $m \times N^2$  matrix, denoted as  $\mathcal{F}_U \in \mathbb{C}^{m \times N^2}$ . Hence we have:

$$\mathbf{y} = \mathcal{F}_U[\Psi\mathbf{x}] \doteq \mathbf{A}\mathbf{x}, \quad (11.3.3)$$

where we denote  $\mathbf{A} \doteq \mathcal{F}_U\Psi \in \mathbb{C}^{m \times N^2}$ .

As we have learned from early chapters of the book, if the overall sampling matrix  $\mathbf{A}$  is sufficiently *incoherent*, then we in principle can correctly recover all the sparse (wavelet) coefficients  $\mathbf{x}$  from significantly fewer  $m$  samples. To ensure the matrix  $\mathbf{A}$  is incoherent, we know from Chapter 3 (section 3.4.3) that randomly chosen partial submatrix of the Fourier (or wavelet) transform  $\mathcal{F}$  is incoherent. Hence, a conceptually simple compressive sampling scheme is to take some random measurements of the Fourier coefficients  $\mathbf{y}$ .

However, as we can notice in Figure 11.2, the most significant nonzero Fourier coefficients of a typical MR image are mainly in the low frequency region, and the coefficients in the high frequency region are already quite sparse and small. Hence a uniformly random sampling of the Fourier domain is not necessarily the most efficient. A more suitable sampling scheme for so-distributed coefficients is the *variable density random sampling*. It is designed specifically for 2D image objects where most of their energy is concentrated close the origin of the frequency domain. More specifically, although the locations of the samples are still randomly selected, it progressively gives higher chances for samples in lower frequencies to be selected than in the higher frequencies. Figure 11.7 shows one example of the variable density random sample pattern.

In practice, however, from the physical model of MRI that we have described above, we know that the MRI machine cannot take measurements at totally random locations from time to time. Instead, it produces a sequence of samples of the Fourier coefficients along a continuous trajectory

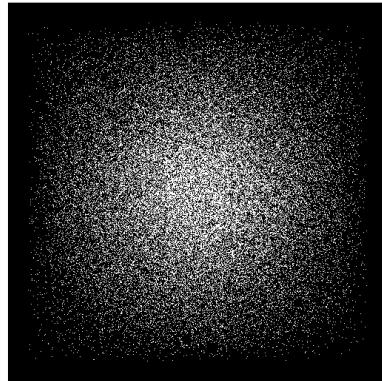


Figure 11.7. A variable density random sampling pattern in the Fourier domain.

$(k_x(t), k_y(t))$  in the  $k$ -space. Hence, the main challenge of compressive MRI is to design both practical and efficient sampling schemes in the Fourier domain for real MR images that are subject to the constraints of the physical process. To this end, some popular subsampling patterns have been proven (empirically) effective for MRI. Examples include a *radial* sampling pattern and a *spiral* pattern as shown in Figure 11.8. Clearly, both patterns are designed to have more coefficients densely sampled close to the origin and sparser coefficients far away from the origin.

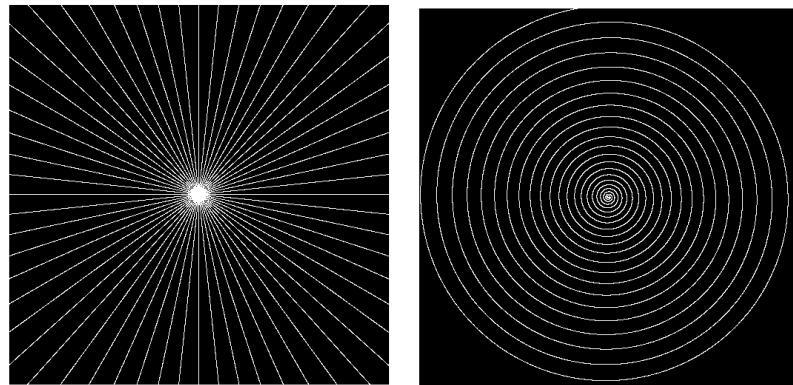


Figure 11.8. Examples of a radial sampling pattern and a spiral pattern.

To see the effectiveness of different sampling patterns, in Figure 11.9, we plot the PSNRs of the reconstructed images against different sample per-

centages.<sup>6</sup> To establish a baseline, we first calculate the PSNR values when the most significant nonzero wavelet coefficients in  $\mathbf{x}$  are given. The results are shown in the red curve. It clearly outperforms the other subsampling methods that do not have the knowledge of the ground truth sparse signal  $\mathbf{x}$ . Furthermore, compared to the deterministic radial and spiral patterns, the variable density random sampling initially achieves the worst reconstruction quality when the sampling percentage is low, namely, less than 20%. Then its performance increases significantly when the sampling percentage becomes higher, gradually surpassing the performance of the other subsampling patterns. The reader may refer to [Lustig et al., 2008b] for more discussions about compressive sampling of MRI.

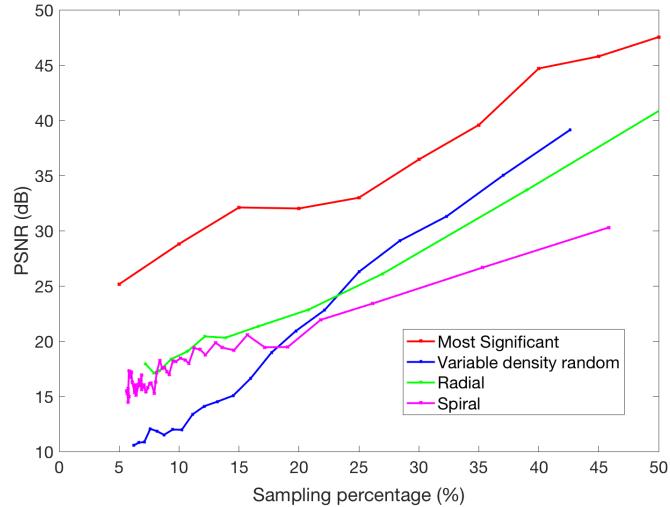


Figure 11.9. The reconstruction quality of the brain image using different subsampling patterns.

## 11.4 Algorithms for MR Image Recovery

In this section, we discuss algorithms for reconstructing MR images from a few measurements  $\mathbf{y}$ . This is conceptually straightforward: many of the methods described in Chapter 8 can be applied to reconstruct the sparse coefficients  $\mathbf{x}$  from the measurements  $\mathbf{y} = \mathbf{A}\mathbf{x}$ . However, there are several practical considerations that demand additional attention. First, MR measurements are subject to various nonidealities, including noise. Second,

---

<sup>6</sup>We will describe details of the reconstruction algorithm in the next section.

because it is so important to make the sampling scheme as efficient as possible, it is often helpful to leverage other structural information about the target image, beyond sparsity of its wavelet coefficients  $\mathbf{x}$ .

#### *Measurement Noise*

In practice, the measured MR image  $I$  can be degraded by thermal noise, so that the measurements  $\mathbf{y}$  are

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (11.4.1)$$

where  $\mathbf{n}$  is a noise term with bounded norm  $\|\mathbf{n}\|_2 < \varepsilon$  or assumed to be Gaussian for simplicity. One can accurately estimate the sparse coefficients  $\mathbf{x}$  by looking for the minimum  $\ell^1$  norm coefficients that agree with the observations  $\mathbf{y}$  up to the noise level (see Chapter 3 section 3.5):

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subj. to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon, \quad (11.4.2)$$

Once  $\mathbf{x}$  is recovered from solving this program, we can recover the image as  $\hat{\mathbf{z}} = \Psi[\hat{\mathbf{x}}]$ .<sup>7</sup>

#### *Gradient Sparsity*

The wavelet representation developed above is well-suited for representing piecewise smooth functions with smooth discontinuities. As we can see from the brain image, MR images may exhibit stronger properties than just piecewise smoothness: an image may be approximated as piecewise *constant* [Ma et al., 2008]. This means that the image value is constant away from a few sharp edges. The gradient of such an image is nonzero only at the edges, and hence is sparse.

To be more precise, let  $\nabla_1$  and  $\nabla_2$  represent finite-difference (differentiation) operators on the first ( $x$ ) and second ( $y$ ) coordinates of the image  $I$ , respectively. At any pixel  $i$ ,  $\nabla \mathbf{z}_i = (\nabla_1 \mathbf{z}, \nabla_2 \mathbf{z}) \in \mathbb{R}^2$  is its gradient vector. The  $\ell^2$  norm  $\|\nabla \mathbf{z}_i\|_2 = ((\nabla_1 z_i)^2 + (\nabla_2 z_i)^2)^{1/2}$  measures the length of this vector.

A piecewise constant image has relatively few pixels at which the gradient is nonzero:

$$\sum_i \mathbf{1}_{\|\nabla \mathbf{z}_i\|_2 \neq 0} \quad (11.4.3)$$

is small. This can be interpreted as a group sparsity assumption on the gradient vector field – see Chapter 6.

The number of points of nonzero gradient, (11.4.3), is conceptually simple, but is not well-suited to efficient computation. Following the intuition

---

<sup>7</sup>Here by abuse of notation, we here use  $\hat{\mathbf{z}}$  to denote both the 2D MR image  $I$  and its vectorized version as a vector in  $\mathbb{R}^{N^2}$ , as its meaning is clear from the context.

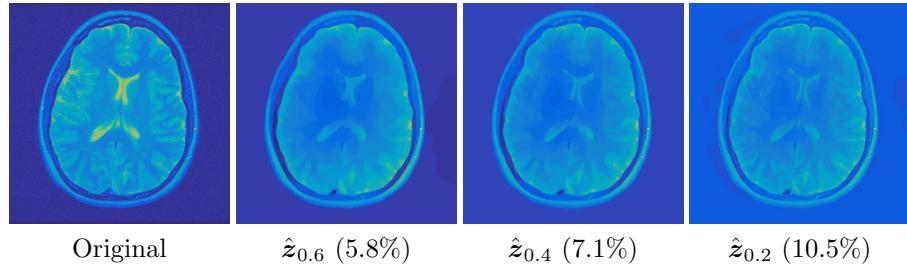


Figure 11.10. **Gradient-sparse approximations.** Left: target MR image  $\mathbf{z}$ . Right: gradient-sparse approximations computed using the proximal operator for the total variation,  $\text{prox}_{\lambda \text{TV}}(\mathbf{z})$  for  $\lambda = 0.6, 0.4, 0.2$ . For each approximation, we also display the fraction of pixels at which the gradient is nonzero.

for group sparsity in Chapter 6, we can define a convex relaxation of this function, known as the *total variation* of the image  $\mathbf{z}$ :

$$\|\mathbf{z}\|_{\text{TV}} \doteq \sum_i \|\nabla \mathbf{z}_i\|_2. \quad (11.4.4)$$

This is a convex function of  $\mathbf{z}$ .<sup>8</sup>

Using this convex function, we can verify experimentally that MR images are well-approximated as gradient-sparse. To do this, we take an image  $\mathbf{z}$ , and compute approximations to it using the proximal operator<sup>9</sup> for the total variation:

$$\hat{\mathbf{z}}_\lambda = \text{prox}_{\lambda \text{TV}}(\mathbf{z}) \doteq \arg \min_{\mathbf{x}} \lambda \|\mathbf{x}\|_{\text{TV}} + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2.$$

For each  $\lambda \geq 0$ , we have an approximation  $\hat{\mathbf{z}}_\lambda$ , whose gradient is sparse. The parameter  $\lambda$  trades off between gradient sparsity of  $\hat{\mathbf{z}}$  and fidelity to the original image  $\mathbf{z}$ .

Figure 11.10 shows approximations  $\hat{\mathbf{z}}_\lambda$  to  $\mathbf{z}$ , with  $\lambda = .6, .4, .2, .1$ . From the figure, we see that the image admits visually plausible approximations with roughly 10% of the nonzero gradient vectors. Here, the proximal operator  $\text{prox}_{\lambda \text{TV}}(\cdot)$  is computed using the alternating directions method of multipliers (ADMM); we will describe this in more generality below.

#### Combining Gradient Sparsity and Wavelet Sparsity

To encourage the recovered image to have sparse gradients, we can incorporate the total variation into the above stable sparse recovery program (11.4.2) as an additional regularization term. Notice that  $\mathbf{x} = \Phi \mathbf{z}$  and

<sup>8</sup>Strictly speaking, it is not a norm, because it is not positive definite.

<sup>9</sup>For more details on proximal operators, see Sections 1-3 of Chapter 8.

$A\mathbf{x} = \mathcal{F}_U \mathbf{z}$ . So the resulting program becomes:

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \alpha \|\Phi \mathbf{z}\|_1 + \beta \|\mathbf{z}\|_{\text{TV}} \quad \text{subj. to} \quad \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|_2 < \varepsilon, \quad (11.4.5)$$

where  $\alpha \in \mathbb{R}$  and  $\beta \in \mathbb{R}$  are two positive weight parameters. The use of total variation and  $\ell_1$  together for MRI recovery was originally introduced by the work of [Lustig et al., 2007a] and [Ma et al., 2008].

We can rewrite the above program in an unconstrained form as

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \alpha \|\Phi \mathbf{z}\|_1 + \beta \sum_i \|(\nabla \mathbf{z})_i\|_2 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2. \quad (11.4.6)$$

Since every term is a convex function of  $\mathbf{z}$ , the overall objective function is also convex. Such a program can be efficiently solved by the so-called fixed-point iteration (see [Ma et al., 2008] for more details).

#### Optimization Algorithm

We here introduce a simpler (and arguably faster) algorithm by exploiting the special structure of the program. We observe that the main challenge in solving the above program seems to be that the objective function contains two separate terms, one minimizing the  $\ell_1$ -norm of  $\Phi \mathbf{z}$  and the other minimizing the sum of  $\ell_2$  norms of the gradient of  $\mathbf{z}$ . If we optimize one of the two terms  $\|\Phi \mathbf{z}\|_1$  and  $\|\mathbf{z}\|_{\text{TV}}$  while treating the other constant, each of the two sub-problems will be a relatively easy optimization problem. Hence one may utilize the alternating direction minimization method (ADMM) introduced in Chapter 8 Section 8.5 to solve this program. This was first suggested by the work of [Yang et al., 2010]. We here give a brief description of the algorithm.

The first two terms of (11.4.6) both depend on  $\mathbf{z}$ . So to utilize ADMM, we need to separate the variables first. To this end, we introduce some auxiliary variables:  $\mathbf{x} \doteq \Phi \mathbf{z} \in \mathbb{R}^{N^2}$  for the (sparse) wavelet coefficients and  $\mathbf{v}_i \doteq (\nabla \mathbf{z})_i \in \mathbb{R}^2$  with  $i = 1, \dots, N^2$  for the (sparse) image gradients. With these auxiliary variables, the program (11.4.6) becomes

$$\begin{aligned} \min_{\mathbf{z}, \mathbf{x}, \mathbf{v}} \quad & \alpha \|\mathbf{x}\|_1 + \beta \sum_i \|\mathbf{v}_i\|_2 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|_2^2 \\ \text{subj. to} \quad & \mathbf{x} = \Phi \mathbf{z}; \quad \mathbf{v}_i = (\nabla \mathbf{z})_i \in \mathbb{R}^2 \quad \forall i. \end{aligned} \quad (11.4.7)$$

Consider the augmented Lagrangian formulation of (11.4.7). We define the two functions associated with these auxiliary variables:

$$g_1(\mathbf{z}, \mathbf{x}, \boldsymbol{\lambda}_1) \doteq \alpha \|\mathbf{x}\|_1 + \boldsymbol{\lambda}_1^*(\mathbf{x} - \Phi \mathbf{z}) + \frac{\mu_1}{2} \|\mathbf{x} - \Phi \mathbf{z}\|^2 \quad (11.4.8)$$

and

$$g_2(\mathbf{z}, \mathbf{v}_i, (\boldsymbol{\lambda}_2)_i) \doteq \beta \|\mathbf{v}_i\| + (\boldsymbol{\lambda}_2)_i^*(\mathbf{v}_i - (\nabla \mathbf{z})_i) + \frac{\mu_2}{2} \|\mathbf{v}_i - (\nabla \mathbf{z})_i\|^2. \quad (11.4.9)$$

The augmented Lagrangian function of (11.4.7) is given by

$$L(\mathbf{z}, \mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \doteq g_1(\mathbf{z}, \mathbf{x}, \boldsymbol{\lambda}_1) + \sum_i g_2(\mathbf{z}, \mathbf{v}_i, (\boldsymbol{\lambda}_2)_i) + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2. \quad (11.4.10)$$

Then the above constrained optimization program (11.4.7) is equivalent to the unconstrained one:

$$\min_{\mathbf{z}, \mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2} L(\mathbf{z}, \mathbf{x}, \mathbf{v}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2), \quad (11.4.11)$$

which can be optimized iteratively following the alternating direction method<sup>10</sup>:

$$\begin{cases} \mathbf{x}^{(k+1)} &= \arg \min_{\mathbf{x}} g_1(\mathbf{z}^{(k)}, \mathbf{x}, \boldsymbol{\lambda}_1^{(k)}), \\ \mathbf{v}_i^{(k+1)} &= \arg \min_{\mathbf{v}_i} g_2(\mathbf{z}^{(k)}, \mathbf{v}_i, \boldsymbol{\lambda}_2^{(k)}), \\ \mathbf{z}^{(k+1)} &= \arg \min_{\mathbf{z}} L(\mathbf{z}, \mathbf{x}^{(k+1)}, \mathbf{v}^{(k+1)}, \boldsymbol{\lambda}_1^{(k)}, \boldsymbol{\lambda}_2^{(k)}), \\ \boldsymbol{\lambda}_1^{(k+1)} &= \boldsymbol{\lambda}_1^{(k)} + \mu_1(\mathbf{x}^{(k+1)} - \Phi \mathbf{z}^{(k+1)}), \\ \boldsymbol{\lambda}_2^{(k+1)} &= \boldsymbol{\lambda}_2^{(k)} + \mu_2(\mathbf{v}^{(k+1)} - \nabla \mathbf{z}^{(k+1)}). \end{cases} \quad (11.4.12)$$

We notice that all terms of the Lagrangian function (11.4.10) are convex functions. Hence all the above sub-programs are convex optimization.

To solve the first subprogram in (11.4.12):

$$\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x}} g_1(\mathbf{z}^{(k)}, \mathbf{x}, \boldsymbol{\lambda}_1^{(k)}),$$

although  $g_1$  is non-differentiable with respect to  $\mathbf{x}$ , it has a closed-form solution in terms of the proximal operator for  $\ell_1$ -norm minimization:

$$\mathbf{x}^{(k+1)} = \mathcal{S}(\Phi \mathbf{z}^{(k)} - \boldsymbol{\lambda}_1^{(k)} / \mu_1, \alpha / \mu_1) \quad (11.4.13)$$

where  $\mathcal{S}(\cdot, \cdot)$  is the shrinkage operator

$$\mathcal{S}(x, \tau) \doteq \max\{|x| - \tau, 0\} \cdot \text{sgn}(x), \quad x \in \mathbb{R} \quad (11.4.14)$$

applied to the vector  $(\Phi \mathbf{z}^{(k)} - \boldsymbol{\lambda}_1^{(k)} / \mu_1)$  entry-wise. We leave the derivation of this as an exercise to the reader.

To solve the second subprogram in (11.4.12):

$$\mathbf{v}_i^{(k+1)} = \arg \min_{\mathbf{v}_i} g_2(\mathbf{z}^{(k)}, \mathbf{v}_i, \boldsymbol{\lambda}_2^{(k)}),$$

notice that  $g_2$  is essentially a 2D version of the 1D  $\ell_1$ -minimization

$$\min_v \beta |v| + \frac{\mu}{2} (v - x)^2.$$

It also has a closed-form solution in terms of a 2D shrinkage operator:

$$\mathbf{v}_i^{(k+1)} = \mathcal{S}_2((\nabla \mathbf{z}^{(k)})_i - (\boldsymbol{\lambda}_2^{(k)})_i / \mu_2, \beta / \mu_2) \quad (11.4.15)$$

---

<sup>10</sup>Here, different from the notation we have used in the optimization chapters, we will use superscript  $k$  to indicate iteration of the algorithm since the subscript  $i$  is already used for the pixels.

where  $\mathcal{S}_2(\cdot, \cdot)$  is the 2D shrinkage operator

$$\mathcal{S}(\mathbf{x}, \tau) \doteq \max\{\|\mathbf{x}\| - \tau, 0\} \cdot \mathbf{x}/\|\mathbf{x}\|, \quad \mathbf{x} \in \mathbb{R}^2. \quad (11.4.16)$$

Again, we leave the derivation of this as an exercise to the reader.

Finally, to solve the third subprogram in (11.4.12):

$$\mathbf{z}^{(k+1)} = \arg \min_{\mathbf{z}} L(\mathbf{z}, \mathbf{x}^{(k+1)}, \mathbf{v}^{(k+1)}, \boldsymbol{\lambda}_1^{(k)}, \boldsymbol{\lambda}_2^{(k)}),$$

we notice that with  $\mathbf{x}^{(k+1)}, \mathbf{v}^{(k+1)}, \boldsymbol{\lambda}_1^{(k)}, \boldsymbol{\lambda}_2^{(k)}$  all being fixed, each term of the Lagrangian function  $L(\cdot)$  is a quadratic function in  $\mathbf{z}$ . As the optimal solution  $\mathbf{z}^{(k+1)}$  satisfies the condition  $\frac{\partial L}{\partial \mathbf{z}} \Big|_{\mathbf{z}^{(k+1)}} = 0$ , this gives

$$\mathbf{M}\mathbf{z}^{(k+1)} = \mathbf{b} \quad \text{or} \quad \mathbf{z}^{(k+1)} = \mathbf{M}^{-1}\mathbf{b}, \quad (11.4.17)$$

where

$$\begin{aligned} \mathbf{M} &= \mathcal{F}_U^* \mathcal{F}_U + \mu_1 I + \mu_2 \nabla^* \nabla, \\ \mathbf{b} &= \mathcal{F}_U^* \mathbf{y} + \Phi^*(\mu_1 \mathbf{x}^{(k+1)} + \boldsymbol{\lambda}_1^{(k)}) + \nabla^*(\mu_2 \mathbf{v}^{(k+1)} + \boldsymbol{\lambda}_2^{(k)}). \end{aligned}$$

Here,  $\nabla^*$  denotes the adjoint of the discrete derivative operator  $\nabla$ .<sup>11</sup>

One can show that as long as the step sizes  $\mu_1, \mu_2$  are chosen to be reasonably small, the above alternating minimizing scheme (11.4.12) will always converge to the optimal solution, starting from any initial conditions [Yang et al., 2010].

## 11.5 Notes and References

### Exercises

**11.1** (Compressive Sensing of Shepp-Logan Phantom\*). *Design and implement a pair of efficient encoder and decoder to encode the Shepp-Logan Phantom based on the principles of compressive sensing. To measure the performance of the encoder/decoder pair, plot the PSNR curve with respect to the dimension of the compressed signal.*

**11.2** (Sparse Gradient Approximation with Debiasing). *For each  $\lambda \geq 0$ , compute  $\hat{\mathbf{z}}_\lambda$ . Then let*

$$\hat{\mathbf{z}}_{\lambda, \text{debiased}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \quad \text{subject to} \quad \text{supp}(\|\mathbf{D}\mathbf{x}\|) \subseteq \text{supp}(\|\mathbf{D}\hat{\mathbf{z}}_\lambda\|)$$

*Plot the same curve as for option (ii). Debiasing improves fidelity to the observation  $\mathbf{z}$ , by removing shrinkage effects on the nonzeros. Note that  $\hat{\mathbf{z}}_{\lambda, \text{debiased}}$  can be computed from  $\hat{\mathbf{z}}_\lambda$  simply by solving a linear system of equations.*

---

<sup>11</sup>This is the linear operator that satisfies  $\langle \mathbf{g}, \nabla \mathbf{z} \rangle = \langle \nabla^* \mathbf{g}, \mathbf{z} \rangle$  for all  $\mathbf{g}, \mathbf{z}$ .

**11.3** (Proximal Operators). *What is the optimal solution to the following program:*

$$\min_v \beta|v| + \frac{\mu}{2}(v - x)^2 ? \quad (11.5.1)$$

*Based on this, prove that*

1. *The optimal solution for  $\mathbf{x}^{(k+1)}$  in (11.4.12) is given by (11.4.13).*
2. *The optimal solution for  $\mathbf{v}_i^{(k+1)}$  in (11.4.12) is given by (11.4.15).*

**11.4** (MRI Recovery with Anisotropic Total Variation [Wang et al., Birkholz, 2011, Block et al., 2007, Cruz et al., 2016]). *Sometimes, for simplicity, people also consider the anisotropic total variation (ATV) of the image  $I$ :*

$$\|\mathbf{z}\|_{\text{ATV}} \doteq \sum_i |\nabla_1 z_i| + |\nabla_2 z_i|.$$

*Notice that this is exactly the  $\ell_1$  norm of partial derivatives of the image at all pixels. Hence, minimizing  $\|\mathbf{z}\|_{\text{ATV}}$  would encourage the image to have a sparse partial derivatives. Let  $D$  be the (finite-difference) gradient operator  $(\nabla_1, \nabla_2)$  on the image  $\mathbf{z}$ . Then we have  $\|\mathbf{z}\|_{\text{ATV}} = \|D\mathbf{z}\|_1$ .*

*We may consider replacing the TV term in (11.4.5) with the ATV:  $\|\mathbf{z}\|_{\text{TV}} \rightarrow \|\mathbf{z}\|_{\text{ATV}}$ :*

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \alpha \|\Phi \mathbf{z}\|_1 + \beta \|\mathbf{z}\|_{\text{ATV}} + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2. \quad (11.5.2)$$

*The goal of this exercise is see how to derive a simpler algorithm for the ATV regulated problem using the ALM and ADMM method discussed in Section 8.5.*

*Using the operator  $D$ , the above program can be rewritten as:*

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \alpha \|\Phi \mathbf{z}\|_1 + \beta \|D\mathbf{z}\|_1 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2, \quad (11.5.3)$$

$$= \arg \min_{\mathbf{z}} \left\| \begin{pmatrix} \alpha \Phi \\ \beta D \end{pmatrix} \mathbf{z} \right\|_1 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2. \quad (11.5.4)$$

*If we denote  $\mathbf{W} \doteq \begin{pmatrix} \alpha \Phi \\ \beta D \end{pmatrix}$  and  $\mathbf{w} \doteq \mathbf{W} \mathbf{z} \in \mathbb{C}^{3m}$ , then the above program becomes*

$$\mathbf{z}^* = \arg \min_{\mathbf{z}, \mathbf{w}} \|\mathbf{w}\|_1 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|^2 \quad \text{subj. to} \quad \mathbf{w} = \mathbf{W} \mathbf{z}. \quad (11.5.5)$$

*Then using the Augmented Lagrange Multiplier method discussed in Chapter 8 for  $\ell_1$ -minimization,  $\mathbf{z}^*$  can be solved by alternatively minimizing  $\mathbf{z}$ ,  $\mathbf{w}$  and a Lagrange multiplier vector  $\boldsymbol{\lambda} \in \mathbb{R}^{3m}$  in*

$$\mathbf{z}^* = \arg \min_{\mathbf{z}, \mathbf{w}, \boldsymbol{\lambda}} \|\mathbf{w}\|_1 + \boldsymbol{\lambda}^* (\mathbf{w} - \mathbf{W} \mathbf{z}) + \frac{\mu}{2} \|\mathbf{w} - \mathbf{W} \mathbf{z}\|^2 + \frac{1}{2} \|\mathcal{F}_U \mathbf{z} - \mathbf{y}\|_2^2. \quad (11.5.6)$$

*We leave as an exercise to the reader to derive a detailed algorithm for (11.5.6).*