The University Of Sheffield.

COMPRESSED SENSING FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA

submitted by

Wattanit Hotrakool

A thesis submitted in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Engineering

Department of Electronic and Electrical Engineering

November, 2016

ABSTRACT

This thesis addresses the possibility of applying the compressed sensing (CS) framework to Functional Magnetic Resonance Imaging (fMRI) acquisition. The fMRI is one of the non-invasive neuroimaging technique that allows the brain activity to be captured and analysed in a living body. One disadvantage of fMRI is the trade-off between the spatial and temporal resolution of the data. To keep the experiments within a reasonable length of time, the current acquisition technique sacrifices the spatial resolution in favour of the temporal resolution. It is possible to improve this trade-off using compressed sensing.

The main contribution of this thesis is to propose a novel reconstruction method, named Referenced Compressed Sensing, which exploits the redundancy between a signal and a correlated reference by using their distance as an objective function. The compressed video sequences reconstructed using Referenced CS have at least 50% higher in terms of Peak Signal-to-Noise Ratio (PSNR) compared to state-of-the-art conventional reconstruction methods. This thesis also addresses two issues related to Referenced CS. Firstly, the relationship between the reference and the reconstruction performance is studied. To maintain the high-quality references, the Running Gaussian Average (RGA) reference estimator is proposed. The reconstructed results have at least 3dB better PSNR performance with the use of RGA references. Secondly, the Referenced CS with Least Squares is proposed. This study shows that by incorporating the correlated reference, it is possible to perform a linear reconstruction as opposed to the iterative reconstruction commonly used in CS. This approach gives at least 19% improvement in PSNR compared to the state of the art, while reduces the computation time by at most 1200 times.

The proposed method is applied to the fMRI data. This study shows that, using the same amount of samples, the data reconstructed using Referenced CS has higher resolution than the conventional acquisition technique and has on average 50% higher PSNR than state-of-the-art reconstructions. Lastly, to enhance the feature of interest in the fMRI data, the baseline independent (BI) analysis is proposed. Using the BI analysis shows up to 25% improvement in the accuracy of the Referenced CS feature.

ACKNOWLEDGEMENTS

Writing a thesis is like a journey along a long, unending road. For this challenging task to come to completion, I would like to express my deepest appreciation to the following people:

First and foremost, I would like to thank my dear parents for their supports in everything I do and without them, I wound not be where I am now. I would like to thank Dr. Charith Abhayaratne, my supervisor, for his invaluable supports, guidance, and comments. It has been such an honour to pursue my degree under his supervision. I would like to thank Miss Thanrada Mungthanya, my partner, for all the great moral supports during these past several years from the other side of the globe. I would like to thank Dr. Wei Liu for his supervision of the Doctoral Development Program. I would like to thank Dr. Aneurin Kennerley for providing the fMRI dataset and useful discussions about fMRI analysis. I also like to thank my fellow colleagues in our research group, particularly Dr. Deepayan Bhowmik and Dr. Matthew Oakes for their guidance during my early year, Mr. Sandipan Pal and Miss Yanxiang Wang—the fellow "thesis writers"—for reassuring each other during the final stage of this journey. Lastly, I would like to thank the staffs at the University of Sheffield and all my friends in Sheffield, Thailand, and the rest of the world. Without these kind people, it would be impossible for this thesis to complete.

Contents

List of Symbols and Acronyms

Symbols

Symbols	Descriptions
α	learning rate of Running Gaussian Average
γ	Gyromagnetic ratio
δ	distance of the correlated reference
$\boldsymbol{\mu}$	vector of mean
$\boldsymbol{\varsigma}$	vector of variance
$\boldsymbol{\Phi}$	sensing matrix
$\boldsymbol{\Psi}$	sparsifying basis
ω	Lamour Frequency
$\mathcal{B}$	orthonormal basis
C	collection of signals
$\mathbf{C}$	circulant matrix
$\mathcal{C}$	Normalised Cross-correlation coefficients
$\mathbb{C}$	complex number space
$\mathbf{D}$	sparse dictionary
$\mathbf{e}$	error/residue vector

$E(\hat{\mathbf{x}}, \mathbf{x})$	reconstruction error between $\hat{\mathbf{x}}$ and $\mathbf{x}$
$\mathcal{E}_{\mathbf{s}}$	energy of the signal $\mathbf{s}$
$\mathbf{f}$	feature set
$\mathrm{Gx}, \mathrm{Gy}, \mathrm{Gz}$	Gradient coils in x-, y-, and z-axis
$\mathbf{i}$	vectorised representation of an image
$\mathbf{i_r}$	vectorised representation of a reference image
$\mathbf{I}$	image (2-dimensional signal in spatial domain)
$\mathbf{I}^{(t)}$	image of frame t in a sequence
$\mathbf{I_r}$	reference image in spatial domain
$\mathcal{M}$	net magnetisation vector
P	pool of sample signals
$\mathcal{P}$	projection operation
$\mathbf{r}$	correlated reference
$\mathfrak{R}(\mathbf{\Phi}, \mathbf{y})$	reconstruction operation using a sensing matrix $\mathbf{\Phi}$ and a compressive measurements $\mathbf{y}$
$\mathbb{R}$	real number space
$\mathbf{s}$	1-dimensional dense vector
$\mathbf{S}$	2-dimensional dense matrix
$\mathcal{S}$	set of sparse supports
$\mathfrak{S}(\mathbf{\Phi}, \mathbf{x})$	sensing operation using a sensing matrix $\mathbf{\Phi}$ and a signal $\mathbf{x}$
T	reference reset period
$\mathbf{v}$	voxel temporal magnitude signal
$\mathbf{v}_b$	baseline of voxel magnitude signal $\mathbf{v}$
$\mathbf{v}_p$	profile of voxel magnitude signal $\mathbf{v}$
$\mathbf{V}$	fMRI volume
$\mathbf{V_r}$	reference fMRI volume
$\mathbf{x}$	1-dimensional sparse vector

$\mathbf{X}$	2-dimensional sparse matrix
$\mathbf{y}$	compressive measurements vector

Acronyms

Acronyms	Descriptions
ANOVA	Analysis of Variance
BI	Baseline independent analysis
BOLD	Blood-oxygen-level dependent
CoSaMP	Compressed Sensing Matching Pursuit
CS	Compressed sensing
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EEG	Electroencephalography
EPI	Echo-planar Imaging
FDG	flurodeoxyglucose
FID	Free Induction Decay
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
FSE	Fast Spin Echo
GLM	General Linear Model
HRF	Haemodynamic Response Function
IFFT	Inverse Fast Fourier Transform
IHT	Iterative Hard Thresholding algorithm
ISA	Infeasible Subgradient Algorithm
LDPC	Low-density parity-check code
LS	The Least Squares approximation

MEG	Magnetoencephalography
MRI	Magnetic Resonance Imaging
NCC	Normalised Cross-correlation
NMR	Nuclear Magnetic Resonance
OMP	Orthogonal Matching Pursuit
PET	Positron Emission Tomography
PSNR	Peak Signal-to-Noise Ratio
RF	Radio-frequency pulse (refer to the MRI excitation pulse)
RGA	Running Gaussian Average
RIP	Restricted Isometry Property
ROMP	Regularised Orthogonal Matching Pursuit
SP	Subspace Pursuit
SPM	Statistical Parametric Mapping
SSIM	Structural Similarity Measure index
TCS	Temporal Compressed Sensing
TE	Echo time
TR	Repetition time

List of Figures

List of Tables

Chapter 1

Introduction

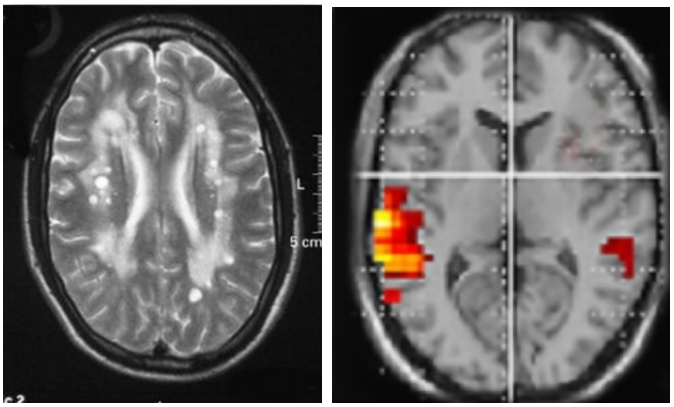
One of the most advanced neuroimaging techniques available in the present is Functional Magnetic Resonance Imaging (fMRI). fMRI allows the study of the inner working of the brain by mapping the information regarding the activation region of the brain as a signal. Despite all its wonders, however, the fMRI technique suffers from many challenging engineering issues including the trade-off between the spatial resolution and the length of its experiment. One approach to tackling this issue is to employ the compressed sensing [106]. The aim of this thesis is to study the possibility of applying the compressed sensing method to the fMRI acquisition scheme. It shows that, by using a novel compressed sensing reconstruction method called Referenced Compressed Sensing, it is possible to obtain accurate fMRI data using a fraction of time and samples while retaining important clinical features.

1.1 Functional Magnetic Resonance Imaging

fMRI is a functional neuroimaging technique primarily used to detect and monitor activities of the brain [123]. fMRI has many applications in both research and clinical fields. It is based on the concept of Magnetic Resonance Imaging (MRI) [123], a non-invasive medical imaging technique used for scanning the cross-sectional parts of the body. The principle of fMRI is similar to that of MRI in general. The subject of study is exposed to a strong external magnetic field, which resonates with the atoms of the human body. Once the atoms are excited by an external radio-frequency pulse, they will emit back signals also in the form of radio-frequency waves. Because of various properties of each tissue affect the absorption and emission rate differently, the signals received will create a contrast image of the cross-sectional body. This technique allows radiologists and practitioners to examine the internal parts of a subject without the need of an operation.

The difference between the standard MRI and fMRI techniques is that the final output of fMRI is the brain activity map. The brain activity is detected by using the relationship between the cerebral blood flow (blood supplied to brain tissue) and the neuronal activation responsible for the activity. This is due to the fact that when a part of the brain is in use, it will require more oxygen to be supplied to, thus increasing the blood flow. The increment of cerebral blood volume results in an increased MR signal acquired from the active region of the brain, thus increasing the voxel intensity of the region. This phenomenon is called the Blood-Oxygen-Level Dependent contrast (BOLD) and is at the heart of fMRI

(a) MRI scanner



(b) Anatomical MRI (c) Functional MRI

Figure 1.1: Example of (a) an MRI scanner, (b) a slice of high-resolution anatomical MRI data, and (c) a slice of function MRI data showing activated regions of the brain.

[121].

As fMRI is based on MR tomography, the scanned data is a series of 3-dimensional cross-sectional volumes with a relatively high spatial resolution. Unlike most MRI applications, where a single volume scan—a tomography—is required, fMRI requires a series of successive volumes to be scanned. The reason is that the BOLD contrast appears only in the temporal changes of voxel's intensity. The number of volumes required per experiment is a big challenge in the practical fMRI experiment design [85]. It is not uncommon that each session could consist of hundreds or even thousands of successive volumes. The sheer number of volumes

(a) EEG sensors (b) PET scanner

Figure 1.2: Example of (a) the Electroencephalography (EEG) sensors and (b) a Positron Emission Tomography (PET) scanner

required, combined with the fact that the MRI acquisition rate is slower than other medical imaging techniques, results in one of the biggest drawbacks of fMRI: poor temporal resolution.

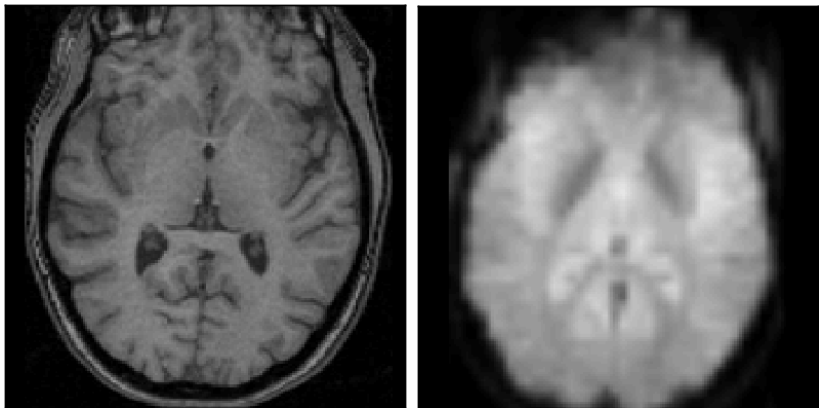
There are other alternatives to fMRI for the functional experiment. Electroencephalography (EEG) [120] and its counterpart, Magnetoencephalography (MEG) [76], are other popular non-invasive functional techniques. These techniques are polar opposite to fMRI, as they are rich in temporal resolution but poor in spatial resolution. For example, the spatial resolution of EEG—a technique based on the measurement of the brain's electrical activity—is governed by the number of electrodes placed on subject's scalp. The electrodes limit the spatial resolution of EEG to be only a few data points (less than 30), compared to a few thousands of fMRI [147]. Also, since all the data points are measured from the surface, EEG requires lots of interpretation regarding the spatial location of each activation. However, EEG has a very high temporal resolution, using the sampling rate varying from 250Hz to 20000Hz to record the brain activity. The temporal resolution of EEG is, therefore, in the order of milliseconds rather than seconds as of fMRI.

Another alternative to fMRI is Positron Emission Tomography (PET) [115]. PET relies on the same physiology as fMRI that the blood volume is increased in the activated region of the brain. Unlike fMRI, the activation signal of PET is detected using a tracer such as Fluorodeoxyglucose (FDG)[30]. The high-density concentration of the tracer, signifying the region with high blood volume, can show the activated region precisely. However, the use of tracer exposes the subject to ionising radiation in the form of gamma rays [30].

Because each of the imaging technique has its own strength and weaknesses, it is possible to combine several techniques in one study [148, 1]. The combination of EEG and fMRI in particular is a common multimodal technique used in psychology studies [140, 80], which can measures both high spatial resolution data (using fMRI) and high temporal resolution (using EEG) simultaneously. Nevertheless, fMRI remains one of the most useful non-invasive functional imaging techniques that can deliver the brain activity map with high spatial resolution without exposure to ionising radiation from tracers or contrast agents.

1.2 Challenges in fMRI acquisition

One of the biggest disadvantages of MRI in general is its low temporal resolution. Compared to other imaging techniques, such as CT, the speed of MRI is hundreds of time slower. This drawback has a significant impact on the functional application in particular because fMRI consists of not one, but a series of images acquired repeatedly. To capture the brain activity signals, it is essential to acquire these images as fast as possible. Not only the fast acquisition can improve the temporal resolution of fMRI, but it can also reduce the negative mental effect on the sub-



(a) Anatomical image (b) Echo Planar Image

Figure 1.3: Examples of (a) a high resolution anatomical scan and (b) Echo Planar Image (EPI) scan of the same slice

ject in the machine. Because of this, rapid acquisition techniques such as Fast Spin Echo (FSE) and Echo-planar Imaging (EPI) are developed and frequently employed in fMRI experiments [85].

EPI, in particular, is the most popular acquisition scheme for the fMRI data. Compared to normal Cartesian trajectory, EPI can provide a much faster acquisition time at the cost of its spatial resolution. Figure 1.3 shows the examples of EPI data as opposed to the typical high-resolution scan commonly used for diagnosis. In general, EPI acquires only a small portion of the entire scan space in a single excitation, focusing only on the low frequency coefficients. The working of EPI is discussed in details in Section 2.1.3.

In addition to the issue of acquisition time, another important challenge of the fMRI acquisition is the image quality—specifically in terms of the signal-to-noise ratio [85, 77]. The challenge is the fact that, in reality, the BOLD signal picked up by fMRI is very small. In most cases, the BOLD signal has a peak magnitude at most no more than 5% of the image's total dynamic range. The signal-to-noise ratio is what limits the spatio-temporal resolution of the acquisition. To increase the spatial res-

olution without a negative impact on the temporal resolution will result in a lower signal-to-noise ratio and *vice-versa*. To maintain a high signal-to-noise ratio, it suffices to say that the spatial and temporal resolutions have an inversely proportional relationship.

In conclusion, one of the biggest challenges for the fMRI acquisition is to achieve a high temporal resolution, a high spatial resolution, and a high signal-to-noise ratio simultaneously. Whilst EPI can satisfy the high temporal resolution requirement, it suffers from the low spatial resolution and low signal-to-noise ratio. This is in direct contrast to the conventional acquisition scheme which yields the results with high resolution and high signal-to-noise ratio with a very poor temporal resolution.

1.3 Compressed Sensing

One of the techniques which have the potential to solve many of the aforementioned shortcomings of fMRI is the compressed sensing (CS) [29]. It is a recently developed signal acquisition scheme in the field of signal processing. Compressed sensing combines many mathematical principles together, including the stochastic statistic and the mathematical optimisation, to become a framework that can efficiently acquire and reconstruct a full-length signal from its highly undersampled measurements. Compressed sensing enables us to obtain a sampling rate lower than that limited by Shannon/Nyquist sampling theorem [29, 52].

The first work presenting the compressed sensing is the reconstruction of partial Fourier data by Candès [28]. The formal framework, comprising of both the sensing and reconstruction parts, is proposed by Donoho [52]. In essence, compressed sensing consists of 2 main parts: the sensing and the reconstruction parts. While each author has individual notations

and terms for these mechanisms, in this thesis, they are referred to as *operations*. The sensing operation, an acquisition of the undersampled signal from the physical world, is represented as a sensing operator. The reconstruction operation, a process to recover the acquired undersampled signal into a full-length, fully-determined signal normally obtained using conventional sensing scheme, is represented as a reconstruction operator.

Compressed sensing shows that if a signal $\mathbf{x}$ of length n is a sparse signal, it is possible to obtain only a small portion of its samples using the sensing operation then reconstruct the full-length signal $\hat{\mathbf{x}}$ of length n back accurately. The sensing operator $\mathfrak{S}(\mathbf{\Phi}, \mathbf{x})$ is

$$\mathbf{y} = \mathfrak{S}(\mathbf{\Phi}, \mathbf{x}) = \mathbf{\Phi}\mathbf{x}, \tag{1.1}$$

where the vector $\mathbf{y}$ of length m is the undersampled measurements of $\mathbf{x}$, and $m \ll n$. The matrix $\mathbf{\Phi}$ is referred to as the *sensing matrix*, which is required to be incoherent to the signal $\mathbf{x}$ [27].

The reconstruction operation—the counterpart of the sensing operation—is generally a non-linear operation with the goal to obtain a reconstructed signal $\hat{\mathbf{x}}$, *i.e.*,

$$\hat{\mathbf{x}} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}), \tag{1.2}$$

that makes $\hat{\mathbf{x}} \simeq \mathbf{x}$. The reconstruction operation of compressed sensing has a root as the problem of the under-determined system of linear equations, a much older problem compared to compressed sensing. Most signal processing operations—such as signal transforms—are in the category of the fully-determined problems and can be solved for a single solution. In an under-determined problem, however, its feasible solutions are infinitely many, thus, it cannot be solved directly.

It is possible, however, to find the most desirable solution among the space of infinitely many using a heuristic. By introducing an objective function to the solution space and optimising it, it is possible to find a solution with specific characteristics. To be specific, given *a priori* that the solution is sparse, we can introduce a sparsity-promoting objective function such as l_1-norm. Minimising l_1-norm leads to the sparsest solution among the solution space. This leads to the field known as the sparse signal reconstruction or the sparse solution to inverse linear problems [59]. The reconstruction operation of compressed sensing is developed based on this concept of the sparse signal reconstruction.

The more detailed discussion of both the sensing and reconstruction operators will follow in Section 2.2.2. Because compressed sensing allows a signal to be acquired using fewer samples, it has become of interest as a way to improve the resolution-acquisition time ratio of MRI. By using compressed sensing to reduce the number of K-space samples required per slice, it is possible to acquire high-resolution MRI data with the acquisition time comparable to the high-speed technique such as EPI [106].

1.4 Research Question

The main research question of this thesis is *to study the possibility of applying the compressed sensing framework to the Functional Magnetic Resonance Imaging (fMRI) acquisition.*

To achieve this goal, a set of sub-questions is to be studied and addressed, as followed:

1. The gap of the current state of the arts in applying the compressed sensing to fMRI data (addressed in Chapter 2).

2. The possibility to view the fMRI data as a form of spatio-temporal signal (addressed in Chapter 2).

3. The performance of the available compressed sensing reconstruction algorithms for spatio-temporal signals (addressed in Chapter 2).

4. To propose a generic and flexible compressed sensing reconstruction algorithm for spatio-temporal signals (addressed in Chapter 3).

5. To verify the applicability of the propose algorithm to the fMRI data (addressed in Chapter 5).

6. To address the practical requirements of the algorithm, including the complexity (in Chapter 4) and the impact to the clinical analysis (in Chapter 5).

1.5 Structure of the Thesis

The structure of this thesis is organised as follows:

Introduction: this current chapter. Here we have introduced fMRI and compressed sensing, discussed the motivation of this study, the research questions, and the contributions.

Background and the Literature Review: this chapter addresses the gap in the current state of the art, assesses the possibility of treating the fMRI data as a spatio-temporal signal, and evaluate the performance of the state-of-the-art reconstruction algorithms. This chapter is divided into three separate sections of different themes.

1. **Functional MRI**: this section reviews the physics, the signal acquisition scheme, and the state-of-the-art fast acquisition techniques currently available for fMRI.

2. **Compressed sensing**: the background and the state-of-the-art literature regarding compressed sensing in general are reviewed here.

3. **Compressed sensing MRI**: this section exclusively reviews the literature involving the application of compressed sensing to MRI.

Referenced Compressed Sensing: this chapter proposes a novel compressed sensing reconstruction method that exploits the redundancy in a signal, such as images, video sequences, and the MRI data.

Extensions of Referenced Compressed Sensing: this chapter presents the study involving the correlated references used by Referenced Compressed Sensing, as well as proposing the low complexity variation that allows a rapid signal reconstruction of compressed sensed measurements.

Referenced Compressed Sensing and fMRI: the performance of the Referenced Compressed Sensing for reconstructing the fMRI data is evaluated in this chapter. While it is aimed specifically to preserve the brain activity maps and clinical features in the reconstructed data, both visual quality and feature accuracy are evaluated. The baseline-independent analysis is also introduced in this chapter.

Conclusions: the conclusion of this thesis.

1.6 Contributions

The main contributions of this thesis are given in the following list.

1. The Referenced Compressed Sensing is introduced in Chapter 3. The Referenced Compressed Sensing is an l_1-norm minimisation-based reconstruction algorithm for spatio-temporal signals, such

as video sequences and the MRI data. By exploiting the redundancy between the signal and arbitrary references, the Referenced Compressed Sensing outperforms the state-of-the-art l_1-norm minimisation algorithms by more than 50%.

2. The performance of the Referenced Compressed Sensing depends on the conditions of the references. The Gaussian Running Average-estimator for the correlated references is proposed in Chapter 4,Section 4.1. By using this estimator, the performance of the Referenced Compressed Sensing is improved by at least 3dB compared to using the naive reference scheme, which much less noisy results.

3. The complexity issue of the Referenced Compressed Sensing is addressed in Chapter 4, Section 4.2.1. This section shows that by exploiting the correlated reference, it is possible to perform the Referenced Compressed Sensing using the l_2-norm minimisation instead of l_1-norm while maintaining the reconstruction accuracy in the acceptable level. Using l_2-norm instead of l_1-norm reduces the computation time at most 1200 times. It also has at least 19% improvement in reconstruction accuracy compared to l_1-norm minimisation algorithm.

4. The study of the reconstruction of compressively sampled fMRI data using the Referenced Compressed Sensing is presented in Chapter 5. The study involves several settings and parameters related to the application of the compressed sensing to the fMRI reconstruction. It suggests that, by using the lossless pre-scan as an initial reference, the fMRI data that has high visual quality can be obtained using Referenced Compressed Sensing with either l_1-norm

or l_2-norm objective functions. The final clinical features can be obtained from the sampling rate as low as 30%.

5. Lastly, Chapter 5 also shows that most of the errors in the reconstruction results from the Referenced Compressed Sensing are due to the loss of dynamic range in the baseline. Here, it is shown that by performing the baseline-independent analysis of the data, the accuracy of the reconstructed data can be increased up to 25%.

1.7 List of Publications

Parts of the work in this thesis have been published in the following publications:

1. W. Hotrakool and C. Abhayaratne, "Fast compressed sensing reconstruction using the least squares and signal correlation," *in Proceeeding of the Intelligent Signal Processing Conference (IET) 2013, IET*, 2013.

2. W. Hotrakool and C. Abhayaratne, "Running Gaussian reference-based reconstruction for video compressed sensing," *in Proceeding of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

3. W. Hotrakool and C. Abhayaratne, "Efficient Reconstruction of Functional Magnetic Resonance Imaging (fMRI) Data From Compressive Measurements Based On Correlated Reference", *in Proceeding of the 10th IMA International Conference on Mathematics in Signal Processing*, 2014.

4. W. Hotrakool and C. Abhayaratne, "An optimal learning parameter for running Gaussian-based referenced compressive sensing", *in Proceeding of the Intelligent Signal Processing Conference (ISP) 2015, IET*, 2015.

1.8 Symbols and Conventions

In this thesis, we will work with various types of closely related, but different signals. To avoid confusion and make the text easy to follow, we will distinguish between each type using a different set of symbols and conventions.

All signals mentioned in this thesis are treated as strictly discrete, which can be represented perfectly using vectors and matrices without the need to consider the factor of the analogue-to-digital conversion.

The most basic distinction is between scalars and vectors (as well as matrices). Whenever a single value of scalar is referred, a regular italic letter is used; for example, a pixel x_1. On the other hand, when a vector is referred, a bold letter is used instead; for example, a vector $\mathbf{x}$. A vector is always denoted using a lower case bold letter while a matrix of any dimension larger than 1 is donated using a capital letter; for example, a 1-dimensional sparse signal $\mathbf{x}$ versus a 2-dimensional sparse matrix $\mathbf{X}$.

We also make a distinction between a general (non-sparse or spatial) signal, denoted by $\mathbf{s}$ or $\mathbf{S}$, and a sparse signal, denoted by $\mathbf{x}$ or $\mathbf{X}$. A spatial image, which is a special class of 2-dimensional non-sparse signal, is denoted exclusively as $\mathbf{I}$. A reconstructed signal is denoted by a hat symbol, *e.g.*, a reconstructed sparse vector $\hat{\mathbf{x}}$. Moreover, for a sequence of signals, each signal in the sequence is denoted by its index. For example, a signal at time t is denoted as $\mathbf{x}^{(t)}$.

In the setting of fMRI, we denoted each volume of MR data as $\mathbf{V}$. The lower case v refers to each voxel in $\mathbf{V}$, while $\mathbf{v}$ denotes a temporal plot of a voxel v in the time domain.

The complete list of all symbols used in this thesis can be found in the *List of Symbols*.

1.9 Summary

In this chapter, we have introduced fMRI and compressed sensing. The main challenge of fMRI is the trade-off between its spatial resolution and temporal resolution—as well as the length of its experiment. Compressed sensing could tackle this issue by reconstructing the full fMRI data from its compressively sampled measurements. The aim of this thesis is to study the possibility of doing so, to reconstruct the fMRI data that retains accurate clinical features worthy of neurological studies.

To do this, the thorough understanding of the working and limitation of fMRI is needed. The next chapter addresses these, along with the state of the arts of compressed sensing in general, as well as the previous attempts to apply compressed sensing to MRI.

Chapter 2

Background and the Literature Review

This chapter is divided into three separate sections. Section 2.1 describes the background and working of fMRI in details. To understand the need of the compressed sensing fMRI, firstly the thorough understanding of fMRI and the MRI system has to be developed. Afterwards, the concept of compressed sensing is introduced. Section 2.2 presents the background and the state of the arts of compressed sensing in general. Finally, to combine these concepts together, the review of literature involving both compressed sensing and MRI is presented in Section 2.3.

2.1 Overview of Functional Magnetic Resonance Imaging

This section presents an overview of fMRI, from the physics of the MRI system to the principle of the fMRI. It also reviews the fast acquisition techniques and modalities currently used in clinical fMRI experiments.

2.1.1 The physics of Magnetic Resonance Imaging

We will first overview the physics and working of MRI in general, which is common between anatomical and functional MRI. The processes involved in MRI system can be distinguished into two parts: the signal generation and the signal encoding. The signal generation is the part where the MR signal is created from the subject body, while the signal encoding handles the sampling and spatial mapping of the created signal, creating output data.

MR signal generation

The principle behinds the MRI signal generation is the concept of Nuclear Magnetic Resonance (NMR), proposed in 1938 by Isidor Rabi. NMR is a quantum phenomenon where atomic nuclei, when induced by a magnetic field, absorb and emit energy [77]. MRI is one of the direct applications of NMR, along with other applications such as the NMR spectroscopy.

Every atomic nucleus has a quantum property called *spin*, which is the phenomenon where each nucleus spins around a fixed axis. Even though this atomic spin is a vector quantity, quantum mechanic dictates that the spin can only take up two distinct values: down (clockwise) and up (counter-clockwise). When a nucleus with odd-number protons spins, without any counterpart to neutralise it, a surrounding magnetic field will be induced. This induced magnetic field is represented by a *magnetisation vector* (Figure 2.1).

In normal circumstances, magnetisation vectors from atoms in any object arrange themselves randomly, cancelling each other out completely. In other words, the *net magnetisation vector*, denoted $\mathcal{M}$, is zero. However, when there is a presence of a strong external magnetic field B_0, the

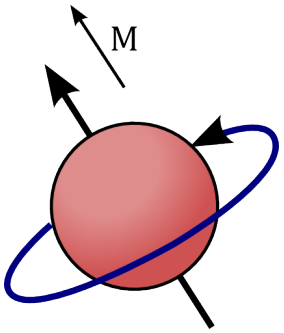
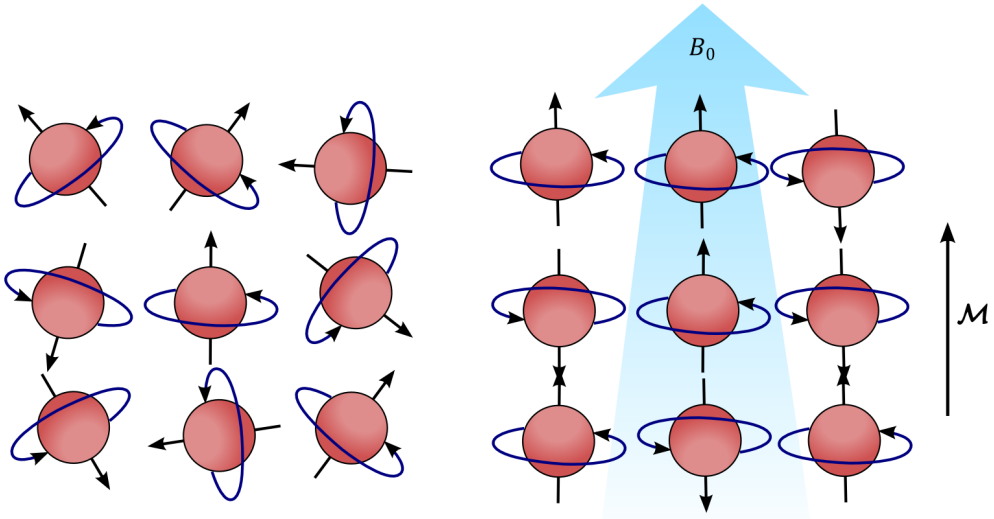


Figure 2.1: Atomic spin showing up spin (counter-clockwise) and its magnetisation vector M

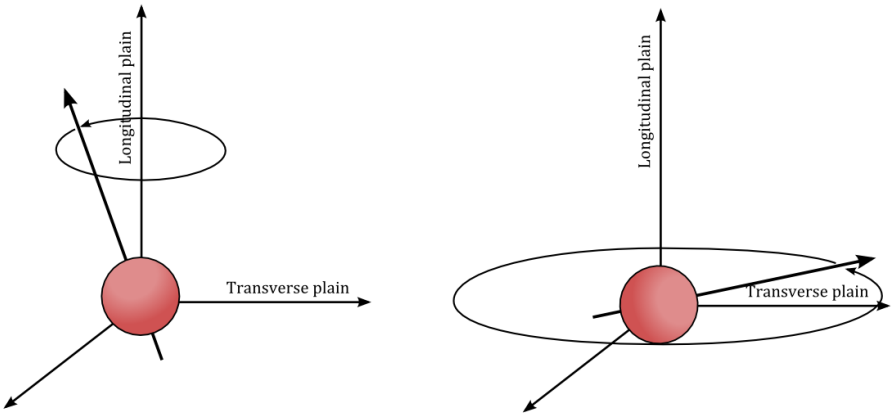


(a) Without B_0 (b) With B_0

Figure 2.2: Effect of an external magnetic field B_0 to the net magnetisation vector $\mathcal{M}$

nuclei will align their magnetisation vectors parallel to B_0, either along or against it depending on their energy states (Figure 2.2). This is the magnetic resonance phenomenon, hence the name of the technique itself, which creates a net magnetisation vector $\mathcal{M} > 0$ that represents the total magnetic field of the entire object.

An interesting property of the magnetisation vector is the *precession*[77]. When nuclei align themselves with B_0, their magnetisation vectors do not steadily point parallel to the direction of B_0. Instead, their magnetisation vectors precess—rotate their points like a spinning



(a) Low energy atom (b) High energy atom

Figure 2.3: The precession of the magnetisation vector M. The longitudinal component of M is larger in the low energy state (a). However, after the atom absorbs energy, M flips into the transverse plain (b).

top—around the axis of B_0 (Figure 2.3). The frequency of the precession is known as the Larmor Frequency, denoted by ω, and is governed by the Larmor Equation:

$$\omega = \gamma B_0. \tag{2.1}$$

The Larmor Frequency ω depends only on the strength of the external field B_0 and a property of atom called Gyromagnetic ratio γ, which is a unique constant for each element.

The Larmor Frequency has an important role in signal generation. In a normal state, the precession angles of nuclei are very narrow; thus, the magnetisation vectors mostly consist of their components in the *longitudinal plane* (a plane along the external field B_0, shown in Figure 2.3a). However, the nuclei can be excited by a radio frequency (RF) pulse—a short burst of electromagnetic wave—whose frequency matches with the nuclei's Larmor Frequency. The nuclei will absorb the energy and their precession angles will begin to widen, increasing the magnetisation vec-

tors' component in the *transverse plane* (a plane perpendicular to B_0, as shown in 2.3b). The longer the RF pulse, the more energy the nuclei absorb until, eventually, the magnetisation vectors flip completely into the longitudinal plane pointing against B_0. The advantage of this technique is that the nuclei will only absorb the energy and flip their magnetisation vectors if and only if the frequency of RF pulse matches exactly with the nuclei's Larmor Frequency, which allows the precise control over which type of atom to be excited.

After the RF pulse is turned off, the nuclei will begin to release their absorbed energy in various forms. This process is called the *relaxation*, as nuclei relax to their normal, low energy state. There are several different types of relaxation, characterised by the source of the energy lost: T1, T2, and T2* relaxation. T1 relaxation, also known as the spin-lattice relaxation, is due to the loss of nuclei's energy to the surrounding lattice. T1 relaxation governs the rate of the return of the net magnetisation vector $\mathcal{M}$ back into the longitudinal plane. In other words, T1 relaxation rate is the growth rate of the longitudinal component of $\mathcal{M}$ after the RF pulse is off. On the contrary, T2 relaxation, or spin-spin relaxation, governs the rate of the decay of the transverse component of $\mathcal{M}$ after the RF pulse is off. The T2 relaxation is caused by the dephasing of each atom in the precession, resulting in their magnetisation vectors cancelling out each others. Both T1 and T2 relaxation rates are the unique property of each tissue. The T2 relaxation rate is much higher than T1, and, as a result, the decay of the transverse component is much faster than the restoration of the longitudinal component. In addition, T2* relaxation increases the decay rate of the transverse component even higher. T2* relaxation, unlike T1 and T2, is caused by the inhomogeneity of external

magnetic field B_0 rather than being a constant property of the tissue. Many factors such as the imperfection of the magnet, the air pockets in a subject, or even the interference from the surrounding environment can lead to a much higher decay rate of transverse magnetisation vectors.

The transition between the longitudinal plane and transverse plane of the net magnetisation vector $\mathcal{M}$ is the source of the MRI signal. In most MRI scanners, there is a set of coils that functions as both an RF pulse emitter and a signal receiver. The spinning magnetisation vector will induce an electric current in the receiver coils, which is the observed MRI signal. This induced signal is known as the *Free Induction Decay*(FID) signal, as it decays due to the relaxation. As mentioned, the T1 and T2 relaxation rates of each tissue are different, thus, when the RF pulse is turned off, each tissue will relax at a different rate. Normally a small amount of time, called *Echo Time*(TE), is waited after the RF pulse is off before measuring the MR signal. During this time, each tissue relaxes at a different rate, which makes the received signals generated from different tissues to have different magnitudes. For example, in T2-weighted images, the higher the T2 relaxation rate of tissue, the lower its magnitude of the magnetisation vector, thus, the lower magnitude of the received signal. By utilising these relaxation rates, it allows several types of image which focus on different types of tissue to be taken.

There are also other factors that affect the received MRI signal other than the relaxation rates and the echo time. A property called the *magnetic susceptibility* affects the homogeneity of the magnetic field. Specifically, most tissues in the human body are *diamagnetic* that repel the external field through them, thus, behaving as they are exposed to a slightly weaker magnetic field. On the contrary, parts of the body are

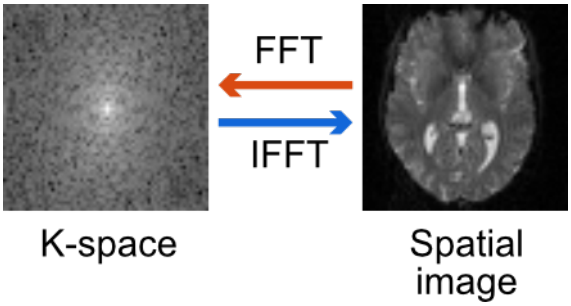


Figure 2.4: In practice, the spatial image is obtained from K-space by applying Inverse Fast Fourier Transform.

paramagnetic which increase the local magnetic field's strength. The changing of B_0 due to these susceptibilities affect the relaxation time directly. Another important parameter affecting MRI signals is the proton density. Clearly, the amount of emitted MRI signal from any given volume is directly proportional to the amount of nuclei in the volume. This is the main reason why hydrogen is used as the main source of the signal in MRI, because hydrogen can be found in abundance in every part of the human body. Using hydrogen gives us the strongest MRI signal possible.

MR signal encoding

The other important aspect of MRI is the signal encoding. Because the emitted signals gathered by the receiver coils are the induced electric currents, the recorded MR signals are in the form of the magnitude and phase coefficients of such currents. This information is recorded into the discrete data space, known as *K-space*, which resembles the frequency domain of the spatial image representing the part of the body the signal originated from. The spatial image can be obtained easily by performing the inverse Fourier transform upon the K-space, shown in Figure 2.4.

In order to construct K-spaces, another physical mechanism of the MRI scanner is required: the *gradient coils*(Figure 2.5). The gradient

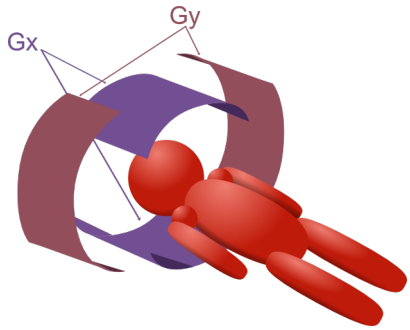
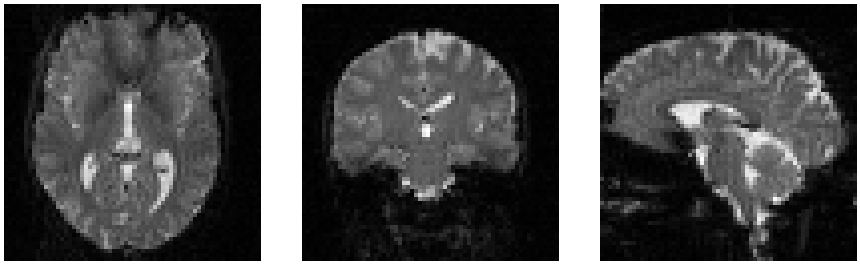


Figure 2.5: The gradient coils. This figure shows the gradient coils in x-axis (Gx) and y-axis (Gy).



(a) Axial (b) Coronal (c) Sagittal

Figure 2.6: Examples of spatial viewing plains showing a) axial plain, b) coronal plain, and c) sagittal plain

coils consist of multiple electromagnetic coils that generate weak magnetic fields—far weaker than the external field B_0—to create artificial magnetic inhomogeneity to the subject. The coils are arranged in x-, y-, and z-axis (denoted Gx, Gy, and Gz respectively). Their functions are interchangeable depending on the design and requirement of the scanner's operator. Unless the scanned data is a complete 3-dimensional tomography, the coils' functions are also dictated by the viewing plain of the spatial image Figure 2.6.

Each K-space can represent only a limited part of the body. Each image created from a K-space, therefore, represents a single slice of the subject's body in any given plane. A process called *slice selection* selects a part of the body to include in a K-space. To select a slice, a magnetic gradient is applied to the subject along the direction perpendicular to the

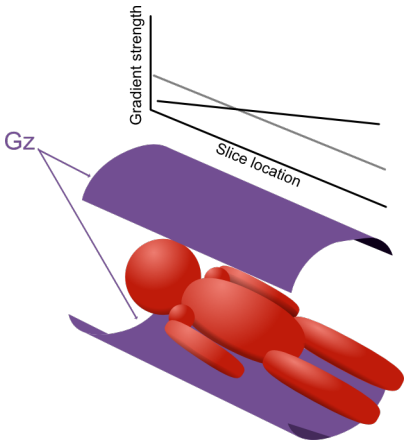


Figure 2.7: Slice selection using z-axis gradient (Gz) making magnetic field B_0 varies along the length of the body

slice's plane. For example, in the transverse (x-y) plane, the slice selection is done using z-axis gradient coil (Figure 2.7). This slice selection gradient is applied during the excitation, which makes each slice of the body to be subjected to a magnetic field of the different magnitude; thus, each slice has a different Larmor Frequency. By exciting the subject using RF pulses of a specific frequency, it is possible to locally excite only the narrow slice of interest without affecting the other parts of the body. The received MRI signal, therefore, contains only the signals originated from the chosen slice.

Still, the received MRI signal is a mixture of all signals originated from every *voxel* in the chosen slice. The process called *spatial encoding* or *spatial mapping* handles the separation of these signals to match their respective spatial positions. The two remaining gradient coils are used in spatial encoding. One of the coils functions as the frequency encoding coil while the other works as the phase encoding coil. In the transverse plane, for example, the frequency encoding coil is the x-axis coil while the y-axis coil takes the role of the phase encoding coil in general. The y-direction (row) of the transverse K-space is phase encoded while the

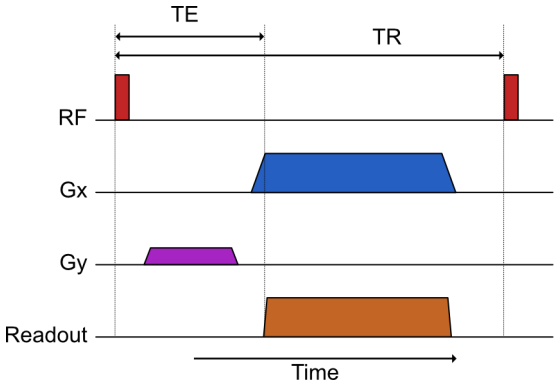


Figure 2.8: Examples of a pulse sequence of one TR cycle

x-direction (column) is frequency encoded.

The frequency encoding works by applying a gradient to the subject during the readout (the sampling of the receiver coils). This makes each voxel along the frequency encoding axis to experience a different magnetic field's strength, thus, the signal emitted from each voxel has different frequency and their spatial information is recorded. On the contrary, the phase encoding gradient is applied for a short period before the readout. It makes each row along the phase encoding axis to precess in a different phase, even though they still have the same frequency. By applying phase encoding and frequency encoding gradients accordingly, the spatial position of voxels is encoded as coefficients in the K-space. The described signal encoding method, however, allows only a single row of K-space (one phase distribution) to be sampled per excitation. Because the FID signal decays rapidly, to maintain a sufficient signal-to-noise ratio, a new excitation is required to readout each row.

The *pulse sequence* (Figure 2.8) describes the timing and magnitude of the RF pulse and gradients require to encode the K-space, with one cycle of pulse sequence per row. Therefore, to obtain a complete K-space, several cycles of pulse sequence are required. Moreover, the pulse sequence also shows the timing of readout. The readout time, or echo

time, as mentioned before, controls the contrast of each tissue on the final images. After the readout, a relatively long period of time is waited before the next RF excitation can take place. This time period between each excitation is called the *repetition time* (TR). This is to allow the tissue to be fully relaxed, thus, TR is typically longer than T1 relaxation rate. This TR is a major constraint of the acquisition speed, scanning more rapidly than TR may cause the undesired contrast distortion. Generally, given a total of n rows in K-space, if TR is t seconds, the time required to obtain a single slice is $t \times n$.

2.1.2 Functional Magnetic Resonance Imaging

fMRI utilises the MR imaging technique described earlier to record the activities of the brain. The fundamental mechanism that allows the brain activity to be captured by the MRI scanner is the theory of *Blood Oxygenation Level Dependent*(BOLD) contrast, proposed by S. Ogawa in 1990 [122]. The effect of BOLD increases the intensity of voxels where the brain activities occur.

The origin of the BOLD contrast is the relationship between the brain activity and the cerebral blood volume—the amount of blood flow in the brain. Human brains consist of billions of small neuron cells which handle all the thinking, memory, and all of the psychological activities. Despite their special functions, these neurons are, nevertheless, cells; they require energy to work like any other cell in the body. As the neurons work, they require more energy to be consumed. The metabolism process that generates energy for these neurons requires the consumption of glucose and oxygen. As more energy is needed, more oxygen supply has to be transported to the neuron via blood. As a consequence, more blood will

be supplied to the activated region to increase the amount of oxygen. This results in a rapid increase of blood flow in that region.

The oxygen is carried by blood using a protein called haemoglobin, which binds itself with oxygen molecules to carry them from lungs to any part of the body. Haemoglobin has an interesting property that it changes its magnetic susceptibility when it carries oxygen. The oxyhaemoglobin—the red haemoglobin that is bound with oxygen—is diamagnetic. However, when the oxygen is released, the deoxyhaemoglobin is strongly paramagnetic. This magnetic susceptibility of haemoglobin is the origin of BOLD contrast. Despite the paramagnetic deoxyhaemoglobin creating more magnetic homogeneity, thus reducing the magnitude of MR signal, the greatly increased amount of blood flow into the activated region overcomes this effect and increases the observed MR signal in such region.

The BOLD response has a unique characteristic which can be observed as the changing of the contrast of a region of activity over time. First, when the neural activity starts, the intensity of the signal at the activated region will drop slightly due to the increase of deoxyhaemoglobin as more oxygen are being consumed. After few seconds, more blood is sent into the region, which significantly increases the intensity of the signal. Lastly, when the activity stops, the blood level in the region will drop down. The intensity of signal during this period is observed to be lower than the normal baseline intensity for a few seconds before returning to the baseline, even though the exact mechanism behind this phenomenon is yet to be explained. This characteristic change in the intensity of MR signal is known as the *Haemodynamic Response* which is the feature of interest for the localisation of brain activity. Figure 2.9 shows the unique characteristics of the Haemodynamic Response Function.

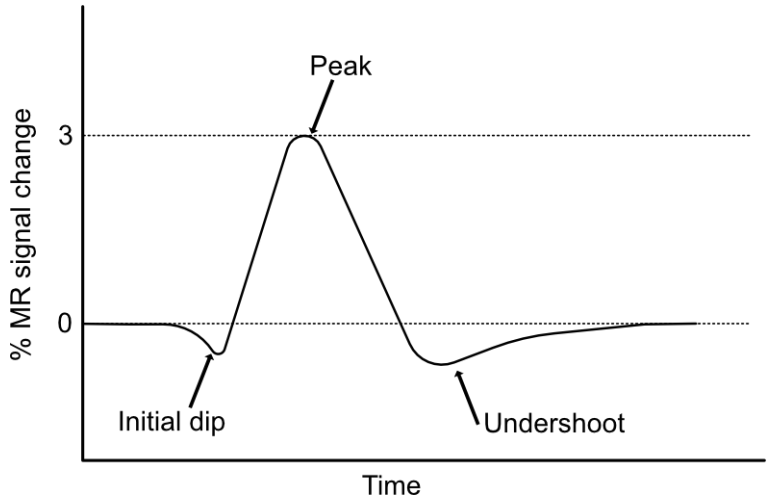


Figure 2.9: Haemodynamic Response function for a stimulus impulse at time 0.

The haemodynamic response is observed in the time domain—along the temporal axis. Therefore, in order to observe the haemodynamic response, a time-series of scanned data is required. This is one of the biggest distinctions, and disadvantages, of fMRI in that, like dynamic MRI technique, it requires a series of consecutively scanned data in order to obtain the desired information.

The BOLD contrast is not the only contributor to the changes in the intensity of a voxel in the temporal axis. Many sources of noise also contribute to the temporal variation of the signal, for example, machine noise, coil imperfection, and external magnetic interference *etc.* One of the biggest challenges for computational neuroscientists is to extract the BOLD contrast signal from the noise and identify it as a haemodynamic response. If this is the case, the region of the brain that contains such response is labelled as an activated region or a region of activity.

There are many techniques currently employed to analyse the data in order to separate the BOLD contrast from the noise and identify it as a haemodynamic response. The most notable techniques are as follow:

- **Subtraction technique**: This is the easiest technique for a two-state experiment, where a task—the activity currently under study—is simply turn on and off. This technique simply subtracts the average of the images acquired during the on state of the task with the average of those acquired during the off state [9, 178]. However, it cannot cope with the overlapping responses and is sensitive to many factors.

- **Correlation technique**: This technique computes the correlation between the data and the predicted response which is estimated from the task's stimulus. This technique is less sensitive to movement and uncontrolled physiological changes than the subtraction technique, though it requires the assumption of the response function to work correctly [71, 37, 170, 169].

- **General Linear Model**: The General Linear Model (GLM) is the generalised version of the previous correlation technique. The GLM is also based on fitting the predicted response to the data; however, the GLM also works with an experiment that contains multiple tasks [68, 67]. Not only the GLM can fit the data to the multiple-task responses, but it can also yield the statistical analysis over joint-task (multiple stimuli applied simultaneously) and group experiment (joint analysis of data from several subjects over the same stimuli). The GLM is the analysis method used in a software package Statistical Parametric Mapping (SPM), which is one of the most widely used fMRI analysis software[66].

- **Analysis of Variance (ANOVA)**: This technique is different from the previously mentioned techniques that it does not require

the assumption of the shape of the response function. ANOVA works by computing a variance based on sliding window technique across the entire signal. This variance is used to identify the active response. Any changes in the data that have local variance bigger than the global variance are labelled as active responses [109, 72]. ANOVA (along with other techniques) is implemented in FSL package, another widely used fMRI analysis software [84].

2.1.3 Challenges of fMRI

There are two main challenges regarding fMRI from the engineering point of view. The first challenge is the low signal-to-noise ratio of the fMRI data. The reason for this is that the variation in the magnitude of the fMRI data from the BOLD contrast is not much different from that of noise. The peak variation of the haemodynamic response is usually no bigger than 3 to 5 percent of the baseline intensity. This is a physical limitation of the scanner and the only way to increase the signal-to-noise ratio further is to increase the strength of the magnetic field B_0. However, due to the negative effect of the strong magnetic field to organic bodies, most clinical scanners only use weak field with the field strength between 1.5T to 3T. The use of very strong field MRI scanners (those having the field strength of more than 6 Tesla) is usually limited to research purposes.

The second challenge is the long experimental time required for fMRI. As the haemodynamic response is observed along the temporal axis, a large amount of scanned data has to be collected over a long period of time. In a typical fMRI experiment, several volumes of brain data are scanned while the tasks are given to the subject, a session of this experi-

ment could last up to 30 minutes or more. As discussed in Section 2.1.1, each volume of data contains several slices of the image, each constructed from a K-Space. Using the typical MRI encoding strategy, only one row of the K-Space can be recorded per one cycle of the pulse sequence. To obtain a volume of the size $m \times n \times z$, at least $n \times z$ cycles are required, given that a slice has the resolution of $m \times n$ and z is the number of slices in a volume. Because the length of pulse sequence's cycle is limited by the value of TR (as shown in Section 2.1.1), there is a limit to how many volumes can be obtained during a single session.

In fMRI, unlike most structural MRI techniques that prefer high spatial resolution, the high temporal resolution is more desirable than high spatial resolution. In the data with high temporal resolution, it is easier to distinguish between the haemodynamic responses from the noise. It is also necessary for capturing short haemodynamic responses, those that only last for a few seconds after the stimuli are applied. Therefore, it is preferable to have the temporal resolution to be as high as possible in fMRI experiments.

As outlined earlier in Section 2.1.1 and Section 2.1.2, there are many factors that affect the acquisition speed and signal-to-noise ratio of the fMRI experiments. Among these, TR is the most prominent restriction. Because TR itself is the parameter that controls the contrast of the acquired image—changing this parameter will results in a change of tissue's contrast—it is a controlled parameter that cannot be changed for the sake of the acquisition speed. Thus, the length of a pulse sequence is fixed by TR and cannot be shortened. Instead, to improve the acquisition speed, radiologists came up with various methods which utilise the scanner more efficiently. The most popular acquisition techniques employed nowadays

in real clinical applications are Fast Spin Echo and Echo Planar Imaging.

Fast Spin Echo

In Fast Spin Echo, it is possible to acquire several readouts from the receiver coils in a single pulse sequence. As mentioned in Section 2.1.1, one of the mechanisms that reduce the FID signal over time is the T2 relaxation, where the magnetisation vector in transverse plane loses its magnitude due to the de-phasing of nuclei. A standard technique to cope with the T2 relaxation is known as Spin Echo. In a Spin Echo sequence, the second excitation by RF pulse is applied before the readout. This second excitation is known as an 180-degree excitation, as its purpose is to flip the direction of the magnetisation vectors. This, in effect, makes the de-phasing vectors start to be in-phase again. This pulse is also commonly known as the refocusing pulse.

In Fast Spin Echo, instead of using just one refocusing pulse per a sequence, a series of refocusing pulses is used to refocus the magnetisation vectors time and time again. Each time the vectors are focused, the received signal is at the maximum magnitude consisting solely of T1 relaxation, without any effect from T2 relaxation. This allows a readout every time the vectors are in phase, thus, several readouts can be acquired in a single pulse sequence.

Fast Spin Echo is currently the most commonly used acquisition sequence in the MRI experiments. It is versatile and can be used with many types of MRI data. It can improve up to 32 times the acquisition speed compared to the normal Spin Echo technique. The disadvantages of Fast Spin Echo, however, are the edge blurring artefact of the scanned data and—most importantly for fMRI—its long TR period [101]. By ex-

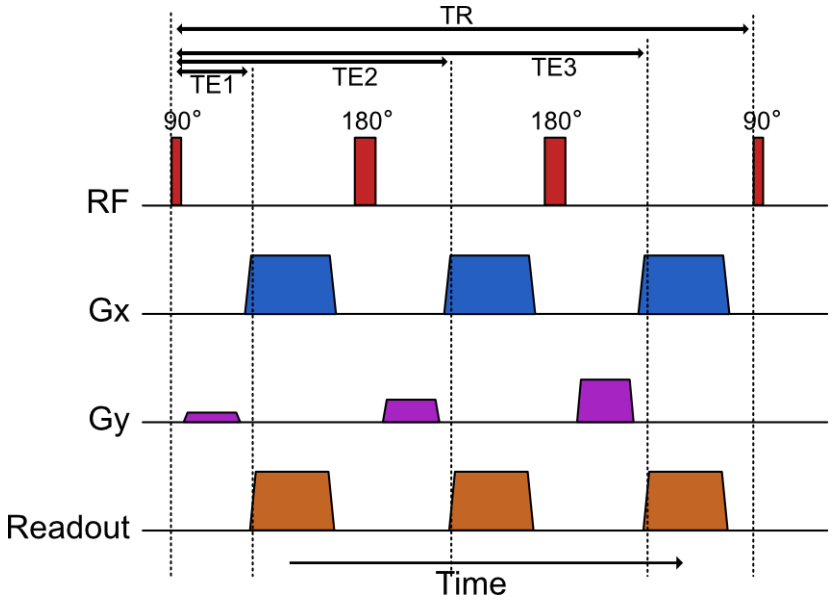
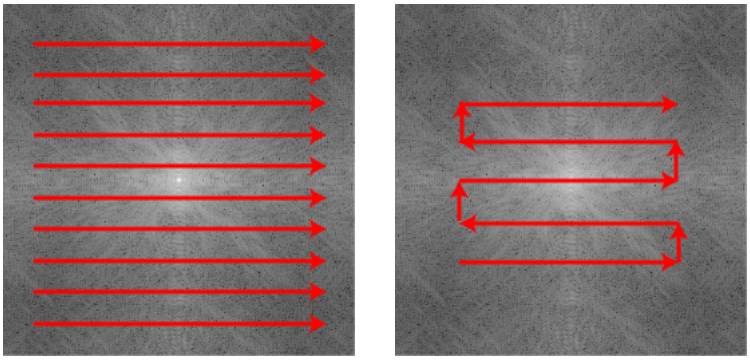


Figure 2.10: Pulse sequence for Fast Spin Echo shows 3 readouts using 180-degree refocusing radio frequency (RF) pulses after a 90-degree excitation pulse, along with the x- and y-gradient coil (Gx and Gy).

citing the subject repeatedly with the refocusing pulses, the TR period is much longer, thus, the final speed improvement is not significant.

Echo Planar Imaging

Echo Planar Imaging (EPI) is the fastest technique available for the fast acquisition of MRI data. Recall that in the conventional MR encoding (discussed in Section 2.1.1), to get a complete K-Space, several readouts—each with a different phase encoding—are acquired, which in turns requires several pulse sequences. On the contrary, using EPI, a complete K-Space can be acquired within a single pulse sequence. This leads to a dramatic improvement in acquisition speed over the conventional MR encoding scheme. The time required to acquire one volume depends on the number of samples, the TR period, and the desired contrast; in typical Spin Echo MRI, it generally takes about 380 seconds per volume. In EPI, the same volume could be acquired within 150 millisec-



(a) Conventional trajectory (b) EPI trajectory

Figure 2.11: Scan trajectory of Echo Planar Imaging versus the conventional trajectory

onds; thousands of times faster than the conventional technique [130]. This dramatic speed of EPI is the reason EPI is the most commonly used acquisition technique for the fMRI experiment.

EPI does this by using the rapidly switching readout (frequency encoding) gradient. Within the period of a single pulse sequence, EPI rapidly applies the phase encoding gradient while reading MRI signal to and fro simultaneously (shown in Figure 2.11 and Figure 2.12). This rapid switching gradient requires special hardware to operate, thus, only some scanners that are fitted with this specially built gradient coils can perform EPI.

The main drawback of EPI is the fact that using the high strength gradient field gives a narrower bandwidth compared to the conventional trajectory, especially along the phase encoding axis. The narrower bandwidth results in the loss of fine details in the data acquired using EPI, coupling with the loss of spatial resolution. The loss of the spatial resolution and details are the biggest disadvantages of EPI. Though EPI can deliver data with a good temporal resolution within a reasonable length of session, the loss of the spatial resolution limits the accuracy of the

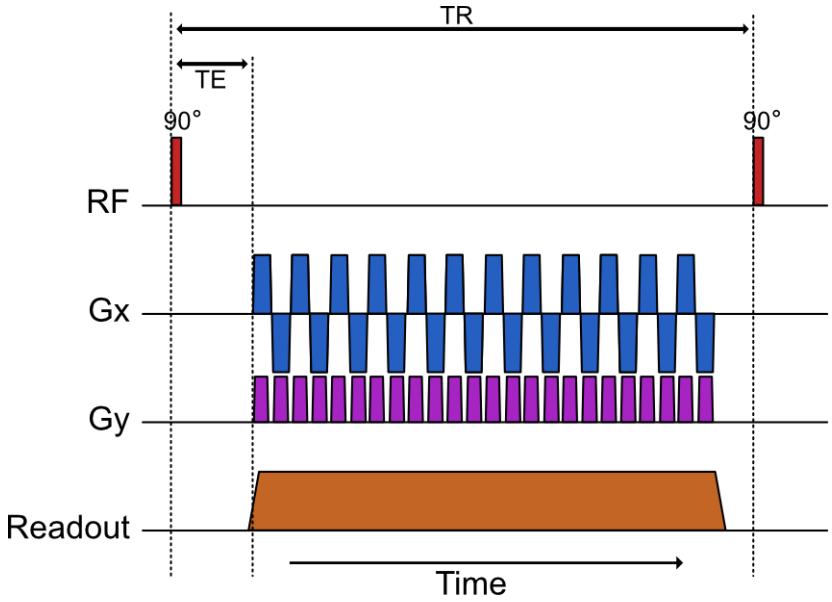


Figure 2.12: Pulse sequence of Echo Planar Imaging. The notations are similar to Figure 2.10.

studies of the data.

From these two commonly used acquisition techniques, we can see the inversely proportional relationship between the temporal resolution and the spatial resolution. One approach to improve the ratio of this trade-off, *i.e.*, to increase the spatial resolution without reducing the temporal resolution, or *vice versa*, is to incorporate in the compressed sensing.

2.2 Compressed Sensing

As introduced in Section 1.3, compressed sensing is a signal acquisition scheme emerged in the field of signal processing during the last decade that allows the acquisition of a full-length signal from underdetermined samples. In this section, we will explore the basic background of the compressed sensing, along with the state of the arts in the field. The works of literature we survey here include those regarding compressed sensing in general, on both the sensing operation and the reconstruction

operation. The works regarding the use of compressed sensing to MRI will be covered specifically in Section 2.3.

2.2.1 Sparse Representation of Signal

Even though compressed sensing has been introduced in 2006, the root of it can be traced back more than just a decade ago. It can be said that the starting point of compressed sensing is the concept of the sparse representation of signals and its reconstruction.

Sparse Representation of Signal

Consider a natural signal $\mathbf{s} \in \mathbb{R}^n$ of length n. The natural signal $\mathbf{s}$ is some unknown function in either time or spatial domain. There is no *a priori* characteristic that can be said more specifically about this signal. Generally, such signal is *dense* in the sense that most of its element have some (non-zero) value, *i.e.*, there are $\mathbf{s}_i \neq 0 | i = 1, ..., k$ and $k \approx n$.

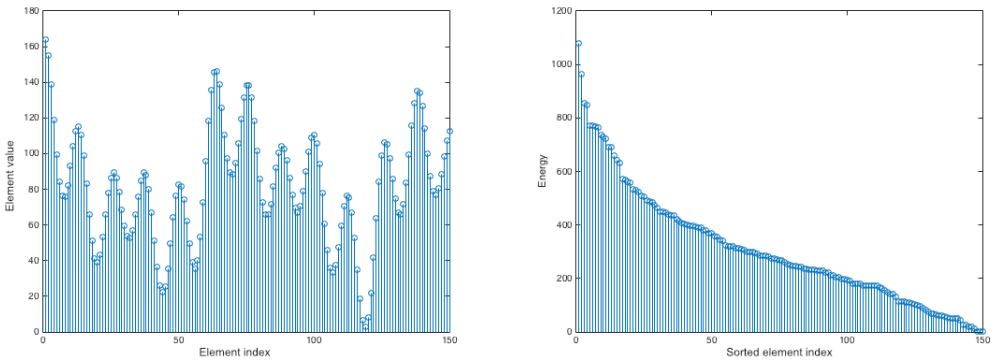
To define the dense signal precisely, we have to consider the distribution of its energy. The energy $\mathcal{E}_\mathbf{s}$ of a discrete signal $\mathbf{s}$ is defined as

$$\mathcal{E}_\mathbf{s} = \sum_{i=0}^{n} |\mathbf{s}_i|^2. \tag{2.2}$$

The energy of each element $|\mathbf{s}_i|^2$ contributes to the sum $\mathcal{E}_\mathbf{s}$ differently and can be plotted into a distribution plot. We can now define the dense signal precisely.

Definition 1. *A dense signal* $\mathbf{s}$ *is a signal which sees the distribution of its energy spreads over the entire signals.*

Figure 2.13 shows an example of a dense signal and its energy distribution.



(a) Dense signal (b) Signal energy

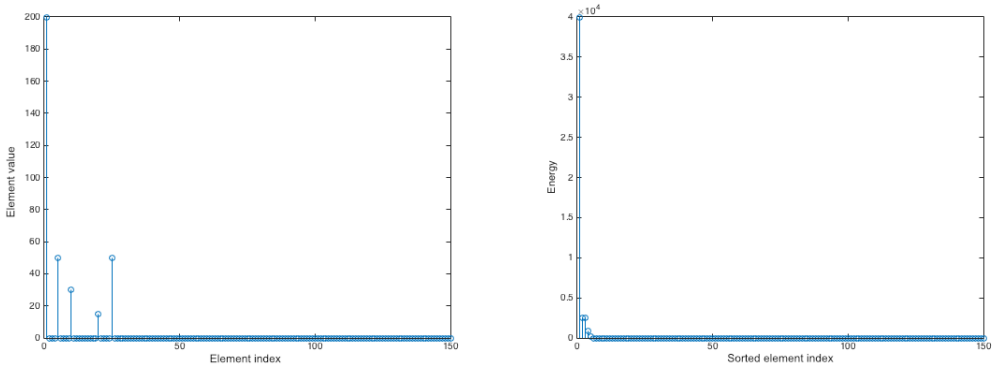
Figure 2.13: Example of a dense signal and its ordered energy distribution

It is well known that signals in the time domain can be efficiently processed in some transform domains. Indeed, the use of transform domains is the heart of signal processing. Many transform domains result in transformed coefficients that are *sparse*. Strictly speaking, a sparse signal $\mathbf{x} \in \mathbb{R}^n$ has most of its elements as zero elements, *i.e.*, there are $\mathbf{x}_i \neq 0 | i = 1, ..., k$ and $k \ll n$. Such signal is referred to as a *strictly sparse* signal. Some signals are not strictly sparse but, however, most of their elements are relatively negligible compared to a few of their much larger elements. Such signals are called *approximately sparse* signals. To define our sparse signals to include both the strictly and approximately sparse signals, we can employ the same concept of energy distribution:

Definition 2. *A sparse signal* $\mathbf{x}$ *is a signal which has the distribution of its energy concentrates only on few of its elements.*

Figure 2.14 shows an example of the sparse signal (the discrete cosine coefficients of the signal in Figure 2.13) and its energy distribution.

Following the *Definition 2*, a strictly sparse signal is a sparse signal where only a few high energy, non-zero elements exist, whereas all other elements are zeros and have no energy. On the other hand, an approxi-



(a) Sparse signal (b) Signal energy

Figure 2.14: Example of a sparse signal and its ordered energy distribution

mately sparse signal is a sparse signal where there are many low energy, non-zero elements exist alongside a few of high energy elements. It is worth noting that while the sparse signals generally considered in early compressed sensing literature are strictly sparse, most sparse representations from the transformation of real-world signals are approximately sparse.

Sparsifying Basis of Signal

To transform a dense signal $\mathbf{s}$ to its sparse representation $\mathbf{x}$, a sparsifying basis $\mathbf{\Psi}$ is required. The sparsifying basis is the matrix representation of a signal transformation. The sparse representation of a dense signal $\mathbf{s}$ is, therefore,

$$\mathbf{x} = \mathbf{\Psi s}. \tag{2.3}$$

Any transformations that aim to achieve a more compact energy distribution of signal can be used as a sparsifying basis [59]. The most commonly used bases are the Fourier Transform basis, Discrete Cosine Transform (DCT) basis, and Wavelet Transform basis. For examples, by

using the Discrete Fourier Transform as a sparsifying basis $\mathbf{\Psi} \in \mathbb{C}^{N \times N}$, it is defined as

$$\mathbf{\Psi}_{\text{DFT}} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix}, \tag{2.4}$$

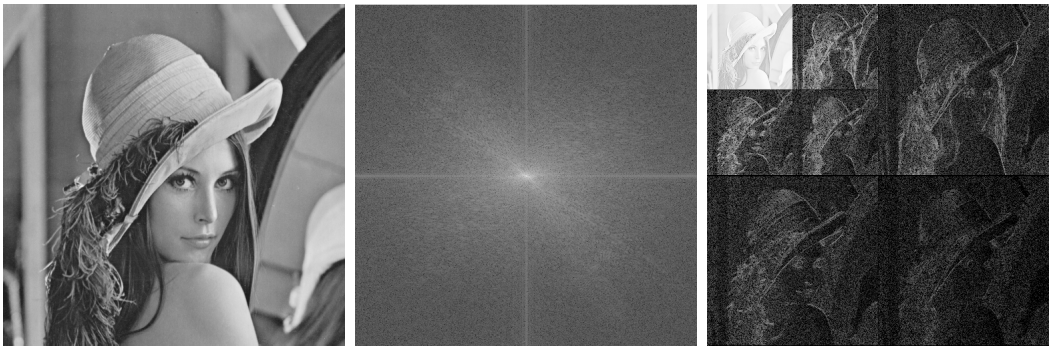
where $\omega = e^{-\frac{2\pi i}{N}}$. Fourier basis is one of the most commonly used sparsifying bases.

Another example of a commonly used basis is the Haar Wavelet basis, which is defined as

$$\mathbf{\Psi}_{\text{W}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}. \tag{2.5}$$

Figure 2.15 shows the example of sparse representation using these sparsifying bases.

The sparsifying basis is a square matrix. Thus, the sparse representation of a signal always has the same length as its original dense counterpart.



(a) Spatial image (b) DFT Coefficients (c) DWT Coefficients

Figure 2.15: Examples of sparse representation of an image using Discrete Fourier Transform and Discrete Wavelet Transform. All coefficients are presented in logarithmic scale.

Application of Sparse Representation of Signal

It is possible to represent the sparse signal further using a matrix called a *dictionary* $\mathbf{D} \in \mathbb{R}^{m \times n}$. This dictionary is different depending on each application, however, every dictionary shares the same characteristic. Unlike the sparse bases, which are square, dictionaries are frequently underdetermined, *i.e.*, $m < n$. Each element of $\mathbf{D}$ is referred to as an *atom*. Using the dictionary, a feature set $\mathbf{f}$ of a sparse signal $\mathbf{x}$ can be obtained from

$$\mathbf{f} = \mathbf{D}\mathbf{x}. \tag{2.6}$$

The use of the sparse dictionary and sparse representation has been employed in a wide range of applications. For example, image deblurring [40, 64], denoising[61, 60], and imprinting[62]. Figure 2.16 shows examples of these application.

Sparse Signal Reconstruction

The use of the dictionary is paired with the inverse process to recover the processed sparse signal $\hat{\mathbf{x}}$ from the feature set $\mathbf{f}$. We refer to this inverse

(a) Deblurring application showing a blurry image (left) is deblurred (right) using a dictionary based on DWT

(b) Denoising application showing a noisy image (left) is denoised (right) using a dictionary based on DCT

Figure 2.16: Examples of the applications of the sparse representation of signal [59]

process as a *reconstruction operator*,

$$\hat{\mathbf{x}} = \mathfrak{R}(\mathbf{D}, \mathbf{f}). \tag{2.7}$$

Unfortunately, because the feature set $\mathbf{f}$ is underdetermined, the linear solutions to the inverse problem are infinitely many. To solve this recovery problem, the *a priori* fact that the sparse signal has only a few non-zero elements is exploited. The problem to recover the sparse signal has become an optimisation problem to optimise the sparsity of the recovered signal.

Many objective functions can be used to promote the signal sparsity. The most straightforward objective function that promotes sparsity is the minimisation of l_0-norm. Strictly, l_0-norm is not a proper norm because it does not follow the homogeneity property; however, the l_0-norm of a signal $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_0 = \#\{i | x_i \neq 0, i \in \mathbb{I}^+\}. \tag{2.8}$$

In words, it is the number of non-zero elements in $\mathbf{x}$ [59].

Using the l_0-norm as the objective function, we can define the reconstruction operation as an optimisation problem

$$(P_0) \qquad \min \|\hat{\mathbf{x}}\|_0 \text{ subject to } \mathbf{D}\hat{\mathbf{x}} = \mathbf{f}. \tag{2.9}$$

The sparsest solution of the (P_0) problem is the desired solution with high probability.

However, this (P_0) problem is NP-hard for arbitrary dictionaries, which has no tractable algorithm to solve it [59]. In [59], it is shown

that the solution to the (P_1) problem,

$$(P_1) \qquad \min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{D}\hat{\mathbf{x}} = \mathbf{f}, \tag{2.10}$$

where l_1-norm of $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \tag{2.11}$$

is the same as that of the (P_0) problem of sparse signals. In other words, both (P_0) and (P_1) problems are equivalent when $\mathbf{x}$ is a sparse signal. The (P_1) problem, however, can be solved using various available optimisation algorithm. Reconstruction of sparse signal by solving (P_1) problem is referred to as the l_1-norm minimisation—or l_1-minimisation in some literatures. The details regarding the reconstruction algorithms will be discussed later in Section 2.2.4.

2.2.2 Compressed Sensing background

In 2006, Candès [28] and Donoho [52] independently came up with the concept of compressed sensing (the term itself coined by Donoho in his paper). Built on the available concept of the recovery of the sparse representation of signals, both Candès and Donoho showed that if the dictionary satisfies a certain property, it is possible to acquire samples in the underdetermined fashion and reconstruct the full-length signal perfectly.

The broad idea of compressed sensing looks very similar to the recovery of sparse representation. As we once introduced in Section 1.3, compressed sensing can be viewed as consisting of two operations: the sensing operation and the reconstruction operation. The reconstruction

is done in the same way as the sparse recovery by using the optimisation algorithms. The main difference between compressed sensing and the sparse recovery is, however, the sensing operation.

Given a signal $\mathbf{s} \in \mathbb{C}^n$ of length n, the undersampled measurements—or samples—$\mathbf{y} \in \mathbb{R}^m$, where $m \ll n$, can be obtained from the sensing operation

$$\mathbf{y} = \mathbf{A}\mathbf{s}. \tag{2.12}$$

We can compare the matrix $\mathbf{A}$ in the sensing operation of compressed sensing to the sparse dictionary $\mathbf{D}$ of the sparse representation in Section 2.2.1. While the dictionary $\mathbf{D}$ is a matrix consisting of sparse atoms, the matrix $\mathbf{A}$ consists of two parts, *i.e.*,

$$\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}. \tag{2.13}$$

In other words, the undersampled measurements $\mathbf{y}$ is obtained from

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}. \tag{2.14}$$

The first part of $\mathbf{A}$ is the sparsifying basis $\mathbf{\Psi} \in \mathbb{C}^{n \times n}$. The task of the sparsifying basis is to get a sparse representation of $\mathbf{s}$. As discussed in Section 2.2.1, there are many sparsifying bases available, with the Fourier basis being one of the popular choices. Indeed, the paper of Candès [28] specifically proposes the use of the Fourier basis as a necessary condition. However, in his later works [27] and Donoho's work, it is shown that other sparsity bases can also serve as $\mathbf{\Psi}$. In imaging applications, Discrete Cosine Transform and Wavelet Transform are also popular choices as sparsity bases.

The other part of $\mathbf{A}$ is the incoherent sampling matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$.

This part is the heart of compressed sensing. Candès showed that in order to reconstruct the full-length signal successfully from the undersampled measurements, the sampling matrix $\mathbf{\Phi}$ has to satisfy the Restricted Isometry Property (RIP) [22]. It is found that, in general, most random matrices—especially the Gaussian matrix—satisfy the RIP property and thus can be used as the sampling matrix successfully [52]. The use of the undersampled Gaussian matrix as a sampling matrix has become the most common setting of compressed sensing. The more detailed discussion regarding the sensing matrix $\mathbf{A}$ will follow in Section 2.2.3.

In order to obtain the full-length signal $\hat{\mathbf{s}}$ from the undersampled measurement $\mathbf{y}$, the reconstruction operation is employed. The reconstructed signal $\hat{\mathbf{s}}$ can be obtained from its reconstructed sparse representation $\hat{\mathbf{x}}$ easily as

$$\hat{\mathbf{s}} = \mathbf{\Psi}^{-1}\hat{\mathbf{x}}. \tag{2.15}$$

The reconstructed sparse signal $\hat{\mathbf{x}}$, in turn, is obtained from

$$\hat{\mathbf{x}} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}). \tag{2.16}$$

We can see that the reconstruction operation works very similarly in compressed sensing to what we previously discussed in Section 2.2.1. It is the same sparse reconstruction problem, with the aim to maximise the sparsity of $\hat{\mathbf{x}}$. In this setting, the reconstruction operation is done by solving the optimisation problem:

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}. \tag{2.17}$$

Candès proved that the solution to Eq. (2.17) is the exact reconstruction of $\hat{\mathbf{x}}$ if $\mathbf{x}$ is exactly sparse and the sampling matrix $\mathbf{\Phi}$ satisfies the RIP

property. Eq. (2.17) can be solved by various optimisation algorithms, which will be discussed in Section 2.2.4.

2.2.3 Sensing Matrices

One of the two major parts of compressed sensing is the compressive acquisition by the sensing operation. The goal of the compressive acquisition is to obtain measurements of a signal such that the length of measurements is much smaller than the number of elements of the signal. That is, given a signal $\mathbf{s} \in \mathbb{R}^n$, we want to obtain the measurement $\mathbf{y} \in \mathbb{C}^m$ such that $m \ll n$. As introduced in Section 2.2.2, the measurement $\mathbf{y}$,

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}, \tag{2.18}$$

where $\mathbf{\Phi}$ is the downsampling sensing matrix and $\mathbf{\Psi}$ is a sparsifying basis. We have already discussed the sparsifying basis $\mathbf{\Psi}$ in Section 2.2.1. The focus of this section is on the sensing matrix $\mathbf{\Phi}$.

Since the early development of compressed sensing, the traditional non-deterministic, non-adaptive, random sensing matrices are most commonly used. This is evident in the original works by Donoho and Candès [52, 28], and other works thereafter [138, 129, 142]—including the famous single pixel camera, the first hardware implemented using compressed sensing [56]. The random matrix with the Gaussian distribution, in particular, is the most popular choice as a sensing matrix. The random Gaussian matrix is a matrix $\mathbf{\Phi}$ where each element ϕ_i is drawn from the Gaussian distribution $\mathcal{N}(\phi_i, \mu, \sigma^2)$,

$$\mathcal{N}(\phi_i, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\phi_i - \mu)^2}{2\sigma^2}}, \tag{2.19}$$

with some value of average μ—commonly $\mu = 0$—and variance σ^2. The main reason for this popularity is because the random Gaussian matrix has low coherence with nearly all sparsifying bases [93, 65]. Even though, most of the time, random matrices are not the most incoherence choice of the sensing matrix for any particular basis, using random matrices guarantees to satisfy the RIP (details follow in the next subsection). However, using a traditional random matrix as a sensing matrix has many drawbacks and limitations, namely:

- the low incoherence of random matrices reduces the accuracy of the reconstruction operation,

- random matrices are not efficient in term of storage memory and the transmission bandwidth. For example, to compressively sense an image with the resolution of 256×256 at 25% sampling rate, as the image has to be vectorised into a 65536-element vector, the sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{16384 \times 65536}$ is needed, containing the total of 1.075×10^9 elements. With the usual 64-bit double precision per element used in most computational packages, this sensing matrix would require approximately 8 gigabytes of memory to store. Moreover, unless the construction of $\mathbf{\Phi}$ is based on a seeded pseudorandom generator, the whole matrix has to be stored and transmitted across the network in many applications. It is clear that this restricts the practical applications of compressed sensing.

- Random matrices are application-independent, signal-independent, cost function-independent, and algorithm-independent. They are non-adaptive, work equally as good or as bad with any signals, costs, and algorithms.

These mentioned issues motivated many works on sensing matrices that are more efficient, sparsifying basis-specific, application-specific, and adaptive.

Incoherence of Random Sensing Matrices

To measure the undersampled measurements $\mathbf{y}$ that allows the full-length signal $\hat{\mathbf{x}}$ to be reconstructed with the reconstruction operation (Eq. (2.17)), the sensing matrix $\mathbf{\Phi}$ has to be incoherent to the sparse signal $\mathbf{x} = \mathbf{\Psi s}$. Candès proposed the formal notion to measure this incoherence, which is now known as the Restricted Isometry Property (RIP). It is demonstrated in [24, 22] that if the sensing matrix $\mathbf{\Phi}$ satisfies the RIP, the full-length signal $\hat{\mathbf{x}}$ can be reconstructed with overwhelming probability.

The RIP is a concept that guarantees the minimum incoherence between the sensing matrix $\mathbf{\Phi}$ and the sparse basis $\mathbf{\Psi}$ in any given sensing operation. The RIP is defined in terms of the restricted isometry constants.

Definition 3. *[22] For each integer $s = 1, 2, ...,$ define the isometry constant δ_k of a matrix $\mathbf{\Phi}$ as the smallest number such that*

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{2.20}$$

holds for all k-sparse vectors. A vector is said to be k-sparse if it has at most k nonzero entries.

It is shown in [22] that if the restricted isometry constant δ_k of $\mathbf{\Phi}$ is small, then the accurate reconstruction can be obtained. Let $\mathbf{x}_k$ be the $\mathbf{x}$ with all but the largest k nonzero elements set to zero, then:

Theorem 1. *[22] Assume that $\delta_{2k} < \sqrt{2}-1$, the solution $\hat{\mathbf{x}}$ to Eq. (2.17) obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq C_0 \|\mathbf{x} - \mathbf{x}_k\|_1 \tag{2.21}$$

and

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 k^{-1/2} \|\mathbf{x} - \mathbf{x}_k\|_1 \tag{2.22}$$

for some small constant C_0. In particular, if x is k-sparse, the reconstruction is exact.

Theorem 1 shows the relation between the sensing matrix and the reconstruction operation. In general, when $\delta_{2k} < 1$, then the l_0-norm optimisation has a unique k-sparse solution. For compressed sensing, the solution to the l_1-norm optimisation is the same as the l_0-norm solution when $\delta_{2k} < \sqrt{2} - 1$ [22].

In [10], it is shown that all random matrices satisfy the RIP property. Intensive studies also reveal that random matrices also satisfy other lessor-known incoherence guarantees such as the Nullspace Property, and are incoherent with almost every sparse basis [29, 46, 111, 152, 65]. This makes random matrices—particularly the random Gaussian matrix described in 2.2.3—become the *de facto* sensing matrices that work with a large variety of sparsifying bases.

Deterministic Sensing Matrices

To overcome the issues of random sensing matrices described earlier, the *deterministic sensing matrices* are developed. The deterministic approach to create sensing matrices allows the identical matrices to be recreated during both acquisition and reconstruction parts, thus removing the need to store and transmit large matrices. A deterministic sensing matrix

is created by using a generator function, and only a small set of parameters are required to replicate a specific matrix. While there are several approaches currently present to create deterministic sensing matrices, each application has its physical limitation that limits the availability of the approaches, as highlighted by Duarte in [57].

One of the simple deterministic matrices is Subsampled Incoherent Basis. Subsampled Incoherent Basis matrices are obtained by selecting a subset of the coefficients of an orthonormal basis that is incoherent with a given sparsity basis. To be precise, assume that an orthonormal basis $\mathcal{B} \in \mathbb{R}^{n \times n}$ is incoherent to a sparsity basis $\mathbf{\Psi}$. The sensing matrix $\Phi = \bar{\mathcal{B}}^{\mathrm{T}}$ is created from a column submatrix $\bar{\mathcal{B}} \in \mathbb{R}^{m \times n}$ of $\mathcal{B}$, $m < n$. The selection of the submatrix's columns is typically made randomly. In fact, one would argue that Subsampled Bases are not completely deterministic since there presents a certain number of randomness. However, they do have a clear structure and, thus, can be replicated if only a small number of indices are known.

The performance of the Subsampled Incoherent Bases are defined in term of the mutual coherence [24].

Definition 4. *The mutual coherence of the N-dimensional orthonormal bases $\mathcal{B}$ and $\mathbf{\Psi}$ is the maximum absolute value of the inner product between elements of the two bases:*

$$\mu(\mathcal{B}, \mathbf{\Psi}) = \max_{1 \leq i,j \leq N} |\langle \mathbf{b}_i, \psi_j \rangle|, \tag{2.23}$$

where $\mathbf{b}_i$ denotes the ith row of $\mathcal{B}$ and ψ_j denotes the jth column of $\mathbf{\Psi}$.

It is shown that the Subsampled Incoherent Basis provides a recovery guarantee.

Theorem 2. *[24] Let* $\mathbf{x} = \mathbf{\Psi s}$ *be a K-sparse signal in* **Psi** *with support* $\Omega \subset \{1, ..., N\}$. $|\Omega| = K$, *and with entries having signs chosen uniformly at random for the set of observed measurements* M. *Suppose that* $M \geq CKN\mu^2(\Phi, \Psi)\log(\frac{N}{\delta})$ *and* $M \geq C' \log^2(\frac{N}{\delta})$ *for fixed values of* $\delta < 1, C, C'$. *Then with probability at least* $1 - \delta$, $\hat{\mathbf{s}}$ *is the solution to the Eq. (2.17).*

Moreover, Rudelson and Vershynin also demonstrates the link between the mutual coherence and the restricted isometry constants [143].

Theorem 3. *Choose a subset* $\Gamma \subseteq \{1, ..., N\}$ *for the set of observed measurements, with* $M = |\Gamma|$. *Suppose that*

$$M \geq CK\sqrt{N}t\mu(\mathbf{\Phi}, \mathbf{\Psi})\log(tK\log N)\log^2 K \tag{2.24}$$

for a fixed value of C *and* $t > 1$. *Then with probability at least* $1 - 5e^{-t}$ *the matrix* $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Psi}$ *has the RIP with constant* $\delta_{2K} \leq \frac{1}{2}$.

The Subsampled Incoherent Basis matrices are applicable in a wide range of applications because of the freedom of choosing any pair of bases of choice. It is suitable for both applications where the choice of sparsity basis is freely available and the ones with limited choices of basis. The examples of the application without the basis restriction is the imaging application in [32, 176]. This kind of application requires a sensing matrix design that can be implemented easily in hardware. Interestingly, both examples employ the noiselet basis as their sensing matrices. Noiselet was proposed in [38] by Coifman *et al.* as a noise-like function in the sense that it is completely incompressible by orthogonal wavelet packet.

The noiselet function $f(t)$ is defined as a recursive function

$$f_1(t) = \chi_{[0,1)}(t), \tag{2.25}$$

$$f_{2n}(t) = (1-i)f_n(2t) + (1+i)f_n(2t-1), \tag{2.26}$$

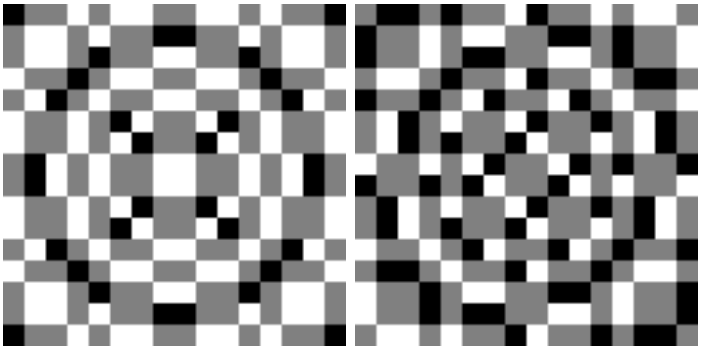
$$f_{2n+1}(t) = (1+i)f_n(2t) + (1-i)f_n(2t-1), \tag{2.27}$$

where

$$\chi_{[0,1)}(x) = \begin{cases} 1, & x \in [0,1) \\ 0, & otherwise. \end{cases} \tag{2.28}$$

Each Noiselet vector $f_i(t)$ corresponds to each row or column in a Noiselet matrix. Figure 2.17 shows an example of 16×16 Noiselet matrix where its rows are composed of the iterations f_{16} to f_{31} Noiselet vectors. Noiselet has become of interest because it is proven to have maximum incoherence with Haar wavelet basis, a popular sparsity basis employed in many imaging applications [159]. It can also be computed very efficiently, making it very desirable for most applications. Although it is best to pair the Subsampled Noiselet sensing matrix $\mathbf{\Phi}$ with Haar wavelet basis $\mathbf{\Psi}$, Noiselet basis is also highly incoherent with other sparse bases. Wen showed the combination of the quantised Discrete Cosine Transform with Noiselet basis in [168].

The other applications commonly seen the Subsampled Incoherent Basis matrices employed are those with restricted choices of the sparsity basis. The clearest examples of such applications are the MRI and the tomographic imaging. In these examples, their sparsity basis—limited by their hardware— is the 2-dimensional Fourier transform. Because the basic functions of the Fourier transform are sinusoidal, the most incoherent bases are those with localised supports, *e.g.*the wavelet basis



(a) Real part (b) Imaginary part

Figure 2.17: An example of a 16×16 Noiselet matrix

and Total Variation basis [24, 28]. More details on the choices of bases for MRI are discussed in Section 2.3.

Noiselet matrix is an example of structurally random matrices. These are pseudo-random matrices that are created in a deterministic fashion. Such matrices can then be subsampled either randomly or linearly while still maintain the properties of random matrices. Another example of the structurally random matrices is the matrix proposed by Thong Do in [49]. This matrix works by systematically scrambling the supports of a sparse signal's elements, as well as flipping their signs. The signal is subsampled after it is transformed into a sparse domain. This matrix satisfies the requirements of the structurally random matrices he previously laid out in [50], *i.e.*, a structurally random matrix should have universality (work with a variety of sparse bases), optimality (achieve optimal number of measurements required for exact reconstruction), and low complexity (can be constructed easily).

The next family of deterministic sensing matrices is the subsampled circulant matrices. The circulant matrix is a square matrix in the form

of

$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_n \\ c_n & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_n & c_1 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & c_4 & \cdots & c_1 \end{bmatrix}. \tag{2.29}$$

Circulant matrix is a special case of Toeplitz matrix—or diagonal-constant matrix—a matrix in which all elements in each diagonal are equal. The circulant matrix is a Toeplitz matrix that each row is a right-shifted of its previous row. The subsampled circulant matrix can be created by randomly subsampling the circulant matrix similarly to the Subsampled Incoherent Basis matrices. If the first row of the circulant matrix—known as a seed—is drawn from a random distribution, then the subsampled circulant matrix has the properties of the random matrix and, thus, satisfies the conditions as an incoherent sensing matrix.

The use of both Circulant matrix and Toeplitz matrix in compressed sensing first appears in communication applications, *e.g.*, channel estimation [8, 78, 144]. The advantage of using circulant sensing matrix is that it is easy to implement the circulant matrix into hardware as a convolution. This reduces the complexity of the sampling operation to just a linear subsampling. There are also other advantages of using a circulant and Toeplitz sensing matrix. For example, it draws only n random variables instead of $m \times n$, making the generation, transmission, and storage much more efficient. Also, the multiplication of Toeplitz matrix can be implemented efficiently using fast Fourier transform [8, 96]. Surprisingly, Valsesia and Magli demonstrated in [162] that it is possible to perform basic signal processing operations, *e.g.*, filtering and transforms, directly in the measurement domain via the circulant sensing matrices . This

leads to a potential of the measurement domain signal processing.

Another family of deterministic sensing matrices is the Separable Sensing Matrices. The sensing matrix in this family is generated from a set of smaller matrices, making it very computationally efficient for applications which have large signals [141]. Separable Sensing Matrices are normally created using Kronecker products. The Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} \mathbf{A}_{1,1}\mathbf{B} & \mathbf{A}_{1,2}\mathbf{B} & \cdots & \mathbf{A}_{1,q}\mathbf{B} \\ \mathbf{A}_{2,1}\mathbf{B} & \mathbf{A}_{2,2}\mathbf{B} & \cdots & \mathbf{A}_{2,q}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{p,1}\mathbf{B} & \mathbf{A}_{p,2}\mathbf{B} & \cdots & \mathbf{A}_{p,q}\mathbf{B} \end{bmatrix}. \tag{2.30}$$

If $\mathbf{A}$ has the size $p \times q$ and $\mathbf{B}$ has the size $r \times s$, then the product $\mathbf{A} \otimes \mathbf{B}$ is of size $pr \times qs$. A sensing matrix $\mathbf{\Phi}$ can be obtained from

$$\bar{\mathbf{\Phi}} = \mathbf{\Phi}_1 \otimes \mathbf{\Phi}_2 \otimes ... \otimes \mathbf{\Phi}_D, \tag{2.31}$$

where each $\mathbf{\Phi}_i, i \in \{1, 2, ..., D\}$ is a basis for each dimension for the D-dimensional signal. It is possible to exploit the sparsity of multiple dimensions simultaneously using Kronecker sensing matrices [55]. Since the sensing matrices created from Kronecker products are best suited for multi-dimension signals, particularly those with information of interest spans across many dimensions, they found many uses in various large-scale applications. Kronecker sensing matrices are used in applications involving video sequences and temporal sparsity [57].

Apart from the previously discussed families of deterministic matrices, over the years there were many attempts trying to create deterministic sensing matrices using other approaches. [151] proposed a determin-

istic matrix construction based on adjacency matrices of bipartite graphs which have a large girth (the length of the shortest cycle in the graph). [171] proposed a construction of sensing matrices by deterministically selecting rows of partial Fourier matrix with orthonormal columns, which is clearly an extension of the Subsampled Incoherent Bases. It replaces the last randomise aspect of random columns selection by a deterministic approach. In [97], Shuxing Li *et al.* proposed a deterministic construction of compressed sensing matrices by using algebraic curves over finite fields. His work shows that the sensing matrices constructed from Elliptic curves and Hermitian curves provide comparable recovery performance to the Gaussian random matrices, with a better perfect recovery percentage for small signal reconstruction. Li later extended his work to create deterministic binary sensing matrices derived from near orthogonal systems and Finite Geometry [99, 98]. The binary sensing matrix is a special class of deterministic matrix which will be discussed further on in Section 2.2.3.

Adaptive Sensing Matrices

All sensing matrices discussed so far, both random and deterministic, are non-adaptive. They are generic and work with any applications that satisfy their conditions. However, such non-adaptive matrices do not offer the optimality in term of application-specific performance. To tackle this problem, there are works of literature that proposed the use of adaptive sensing matrices. Instead of linearly sampling every element equally, adaptive sensing matrices aim to dynamically sample the signal in a way that maximises the signal's feature of interest. The adaptive sensing matrices are most excel in applications that deal with dynamic, undeter-

mined data.

The most prominent attempt to create adaptive sensing matrices is developed from the concept of Bayesian Compressed Sensing. Originally, the aim of Bayesian Compressed Sensing is not the adaptive sensing matrix, but rather the compressed sensing framework that can estimate the confidence of its reconstruction [86]. That is, not only Bayesian Compressed Sensing framework estimates the reconstructed signal $\hat{\mathbf{x}}$, it also estimates the error variance σ^2, *i.e.*, the range of uncertainty for the reconstruction. In [86], the error variance is obtained from the Bayesian inversion using Relevance Vector Machine. These error predictions provide the sense of confidence of the reconstruction—the larger the error, the less accurate the reconstruction. However, the concept of error prediction provided by Bayesian Compressed Sensing Framework created the possibility of the Adaptive Compressed Sensing matrices. Particularly, the error prediction helps to 1) adaptively select the sensing matrices such that the uncertainty is minimal, and 2) determine the number of measurements required that is enough for the accurate reconstruction [11].

From this beginning, the probability approach to compressed sensing developed further. The adaptive sensing matrices based on Bayesian approach is now a part of adaptive compressed sensing framework, along with the adaptive reconstruction as its counterpart.

A blind adaptive sensing for images was proposed in [45]. It is based on the proposed Statistical Pursuit, where the sampling is restricted by a set of masks. The masks are generated in a way such that they preserve the underlying statistical model of a target image. This method provides the sensing matrices that are adapted to image structures. The authors

of this work also proposed Blind Wavelet Sampling, a special case of Statistical Pursuit that exploits the statistical model of wavelet structure. An adaptive sensing matrix for temporal data was also proposed. In [125], a structurally random matrix based on a random Gabor system was proposed. This matrix was designed for time-frequency signal sensing, originally for the channel estimation application [126] and remote sensing [79], but later was generalised and shown to work with other types of temporal signals [127].

However, there are counterarguments that the use of Adaptive Sensing Matrices does not reach its potential of better reconstruction performance. Arias-Castro *et al.* shows that the improvement offered by the adaptive strategy, compared to the non-adaptive one, is minimal at best [6] and negligible at worst. Given the complexity of sophisticated adaptive strategies, for most applications, the traditional, non-adaptive matrices are the most accuracy-versus-complexity efficient. Nevertheless, adaptive sensing matrices are shown to give a clear improvement on specific problems where the structure of the signal is very distinct.

More recent works on Adaptive Sensing Matrices focus more on adaptively choosing the number of rows of the sensing matrix, effectively resulting in the compressed sensing with adaptive downsampling rate [177, 42]. This approach couples with the adaptive reconstruction allows the compressed sensing framework that sensed the signal at the optimal rate.

Binary Compressed Sensing Matrices

A special class of compressed sensing matrices is the binary compressed sensing matrices [17, 18]. Traditional sensing matrices are linear mapping

functions $\mathbf{\Phi} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$. Unlike the traditional sensing matrices, the binary sensing matrices map the real signal into a binary space, *i.e.*, mapping functions $\mathbf{\Phi}_b : \mathbb{R}^n \rightarrow \mathbb{F}_2^m$, where $\mathbf{\Phi}_b \in \mathbb{F}_2^{m \times n}$. The term binary here does not refer to the value of 0 and 1, but rather to the fact that there are only two possible values for the measurements. The most common type of binary sensing matrices is the sign matrix, where the elements of $\mathbf{\Phi}_b$ and $\mathbf{y} = \mathbf{\Phi}_b \mathbf{x}$ are $\{-1, 1\}$. It is also common that $\mathbf{\Phi}_b$ is preferred to be a orthonormal matrix.

The use of Binary Sensing Matrices simplifies the sensing operation, since both the sensing matrices and measurements are only 1 bit deep. It was first proposed in [17] as the 1-bit Compressed Sensing. Unlike later Binary Compressed Sensing, this work employs a traditional random sensing matrix. The random measurements, however, preserve only the sign information without any magnitude. It also shows that whilst it is possible to reconstruct 1-bit measurements using traditional methods, it is more beneficial to reformulate the reconstruction problem to use sign constraints rather than norm constraints. Binary Compressive Imaging was later proposed in [18], which is a framework designed to use the binary sensing matrix together with a binary reconstruction method on high dimension images [18].

Not only the Binary Sensing Matrices are more efficient in term of complexity, but it is also easier to be constructed in a deterministic fashion. Consider the example of Noiselet matrix in Figure 2.17. Both the real part and the imaginary part of the Noiselet matrix can be regarded as a binary sensing matrix. This leads to the state-of-the-art sensing matrices which are the Deterministic Binary Sensing Matrices.

It is found that Binary Sensing Matrices can be generated efficiently

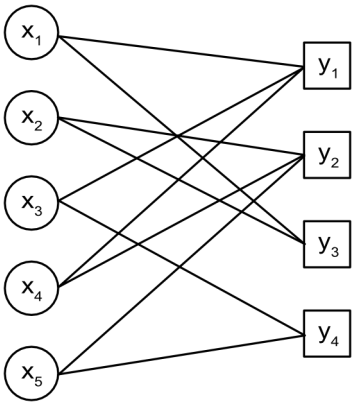


Figure 2.18: An example of a Low Density Frame with its Factor Graph

and deterministically by deriving from coding matrices. Amini and Marvasti first established the connection between the Orthogonal Optimal Codes and binary sensing matrices in [5]. This leads to a successful generation of binary (0,1), bipolar (-1,1), and ternary (-1,0,1) deterministic sensing matrices. It is now clear that many codewords in coding theory can be used to generate sensing matrices. One of the Binary Sensing Matrices derived from coding theory is the Low Density Frame, proposed in [3, 4]. The Low Density Frame is a matrix that has the zero elements in the majority of each row and column. It is directly derived from the parity-check matrix of the Low-Density Parity-Check code (LDPC). As with LDPC, the Low Density Frame can be represented by factor graphs. For example, a factor graph in Figure 2.18 represents the Low Density Frame

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{2.32}$$

So far we have surveyed the varieties of compressed sensing matrices.

There are many challenges around the use of random sensing matrices in early work of literature. The challenges of the complexity and storage of random matrices led to the development of Deterministic Sensing Matrices while the challenge of optimal performance in specific applications led to the development of Adaptive Sensing Matrices. The Binary Sensing Matrices are designed to reduce the complexity of both the sensing operation and reconstruction. Finally, the deterministic binary matrices such as Delsarte-Goethals Frame attempts to make the simplest, most efficient sensing matrices that suit the need of practical applications. Despite many disadvantages, however, the random sensing matrices—particularly the Gaussian random matrices—remain the most popular sensing matrices of choice. As such, there is still research in progress on the improvement of random sensing matrices. For example, in [95], the authors stepped away from the concept of global incoherence to a notion of local coherence, *i.e.*, to measure the correlation between each sensing vector and the sparse basis independently. The local coherence was shown to have more correlation to the reconstruction quality of images. [91] also demonstrated that an efficient random sensing matrices can be obtained by randomly selecting rows of scrambled Fourier basis. The sensing matrix obtained from subsampled scrambled Fourier basis is superior to the naively subsampled Fourier basis. It also shows that even when deterministically subsampled, the RIP bound is still retained.

2.2.4 Reconstruction Algorithms

Another main component of compressed sensing alongside the sensing operation is the reconstruction operation done using reconstruction algorithms. The primary goal of compressed sensing reconstruction is to

find the solution to Eq. (2.17). Historically, it is known that a signal with a sparse dictionary can be recovered from an undersampled vector (see Section 2.2.1 for more details). Such recovery is done by maximising the sparsity of the target signal. Ideally, the sparsity of a signal can be computed as a l_0-norm of the signal. By minimising the l_0-norm, one can maximise the sparsity of the signal (Eq. (2.9)). The optimisation of l_0-norm, however, is NP-hard and intractable.

To avoid the use of l_0-norm, there are three main approaches. The first approach is to solve the basis pursuit via convex optimisation. This is the most true-to-the-theory approach that was developed by using mathematical optimisation techniques to minimise the l_1-norm, the relaxed cost function of the l_0-norm, straightforwardly. Unlike l_0-norm, the minimisation of l_1-norm, known as the basis pursuit, is tractable [36]. It is shown that the basis pursuit (Eq. (2.17)) yields the approximate sparse solution that is very close to the solution of the l_0-norm minimisation Eq. (2.9) [53]. Indeed, the solution of the basis pursuit is exact if the sparse signal $\mathbf{x}$ is K-sparse with small K (refer to 2.2.3).

The second approach is a family of algorithms referred to as greedy algorithms. Most commonly, the greedy algorithms are variants of the Matching Pursuit, in which the sparse solution is searched for in a greedy fashion. That is, the location of the non-zero elements in the solution are to be found first, usually by maximum correlation to the measurements, then their values are estimated later [108]. Matching pursuit has much lower complexity than the basis pursuit, thus, resulting in a faster recovery. Basis pursuit, on the other hand, optimises the whole signal simultaneously, resulting in the more accurate reconstructions than the matching pursuit.

The third approach is the variants of hard thresholding algorithms, including the Iterative Hard Thresholding. This approach does not find the sparse solution by minimising the l_1-norm. Instead, the sparse solution is found by applying the inverse of residue and hard thresholding repeatedly [15]. This approach has the lowest complexity of all three discussed here. However, its performance is limited to the knowledge of the target sparse signal.

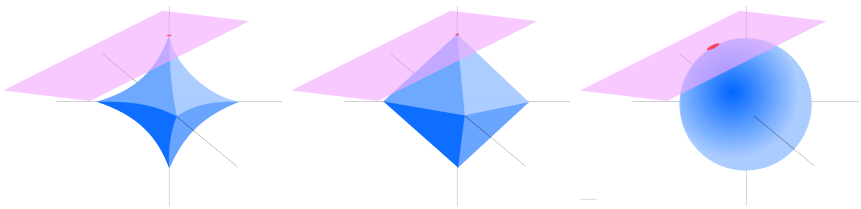
This section will discuss these three reconstruction approaches in more details. Here, reconstruction operations aim to reconstruct the full-length sparse signal $\hat{\mathbf{x}}$ from the undersampled measurements $\mathbf{y}$ of $\mathbf{x}$, obtained using the sensing operation, without *a priori* information regarding the structure of $\mathbf{x}$.

Basis Pursuit via Convex Optimisation

Recall the basis pursuit in Eq. (2.17), the basis pursuit is the l_1-norm minimisation problem in the form of

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}. \tag{2.33}$$

The basis pursuit is the most straightforward and most mathematically accurate reconstruction of compressed sensing. As Candès showed in *Theorem 1*, discussed in Section 2.2.3, the solution to Eq. (2.33) exactly matches the solution of the l_0-norm minimisation (Eq. (2.9)) if the sensing matrix $\mathbf{\Phi}$ satisfies the RIP, *i.e.*, $\mathbf{\Phi}$ is incoherent to the sparsity basis $\mathbf{\Psi}$ of $\mathbf{x}$. Basis pursuit, in essence, tries to find the optimal point in the convex set of feasible solutions. Figure 2.19 shows the geometric examples of several l_p-norm in 3-dimensional space. It is clear that when $0 < p \leq 1$, the 3-tuple solution is sparse. In contrast, when $p > 1$, the optimal



(a) $l_{0.5}$-norm (b) l_1-norm (c) l_2-norm

Figure 2.19: Examples of $l_{0.5}$-norm, l_1-norm, and l_2-norm objective functions

solution on the feasible set is no longer guaranteed to be the sparsest solution.

In early literature, the basis pursuit for CS reconstruction was solved by casting the basis pursuit problem into a linear programming problem. That is, we cast the basis pursuit problem in (Eq. (2.33)) into a linear programming problem in the form of

$$\min \mathbf{c}^{\mathrm{T}}\mathbf{x} \text{ subject to } \mathbf{\Phi x} = \mathbf{y}, \mathbf{x} \geq 0. \tag{2.34}$$

It is shown in [25, 34] that such conversion is possible, and resulting in a parametric linear programming problem:

$$\min \mathbf{1}^{\mathrm{T}}\mathbf{t} \text{ subject to } \mathbf{x} - \mathbf{t} \leq 0,\ \mathbf{x} + \mathbf{t} \geq 0,\ \mathbf{\Phi x} = \mathbf{y}. \tag{2.35}$$

By casting the basis pursuit into a linear programming problem, the solution to Eq. (2.35) can be found by generic linear programming algorithms, *e.g.*, the interior-point method and the simplex algorithm. All early works on compressed sensing employed the linear programming-based reconstruction algorithms. The l_1-Magic software [23] by Candès employs the simplex algorithm whilst the SparseLab [51] by Donoho and CVX package [74] employ the interior-point method.

It should be noted that the size of the linear programming problem in Eq. (2.35) can be reduced by the concept of dual problem. The solution to the dual problem, in the form of

$$\min \mathbf{b}^{\mathrm{T}}\mathbf{w} \text{ subject to } \mathbf{\Phi}^{\mathrm{T}}\mathbf{w} + \mathbf{x} = \mathbf{c}, \tag{2.36}$$

is equivalent to the primal problem in Eq. (2.34). The dual problem of Eq. (2.35) can be expressed in terms of dual variables $\mathbf{w}, \mathbf{v}, \mathbf{u}$ as

$$\min \mathbf{b}^{\mathrm{T}}\mathbf{w} \text{ subject to } \mathbf{\Phi}^{\mathrm{T}}\mathbf{w} - 2\mathbf{v} = -\mathbf{e}, 0 \leq \mathbf{v} \leq \mathbf{e}, \tag{2.37}$$

where $\mathbf{e} = \mathbf{v} + \mathbf{u}$ [34].

The reconstruction from basis pursuit discussed so far is theoretically exact only in the noiseless case, *i.e.*, when $\mathbf{y} = \mathbf{\Phi x}$ exactly. However, because any real-world measurement systems always subject to some degree of noise, the result of Eq. (2.33) is not reliable. When there is noise presents in the measurement system, *i.e.*, when $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$, where $\mathbf{n}$ is assumed to be independent and identically distributed noise, it is possible to solve the basis pursuit problem in the "approximate" fashion known as the basis pursuit denoising. Basis pursuit denoising relaxes the constraints of the feasible set to obtain the approximate solution $\hat{\mathbf{x}}'$ such that $\|\hat{\mathbf{x}}' - \hat{\mathbf{x}}\|_2 \leq \epsilon$, where ϵ is a small scalar threshold. The basis pursuit denoising solves the optimisation problem:

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \|\Phi\hat{\mathbf{x}} - \mathbf{y}\|_2 \leq \epsilon. \tag{2.38}$$

Further regularisation results in a regularised l_1-norm minimisation prob-

lem:

$$\min \|\mathbf{y} - \mathbf{\Phi x}\|_1, \tag{2.39}$$

which is equivalent to the linear programming problem [25]

$$\min \mathbf{1}^{\mathrm{T}}\mathbf{t} \text{ subject to } -\mathbf{t} \leq \mathbf{y} - \mathbf{\Phi x} \leq \mathbf{t}. \tag{2.40}$$

The more generalised form of basis pursuit designed to work with noisy measurement is the Dantzig Selector, proposed in [26]. Dantzig Selector is the convex optimisation problem in the form of:

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{\Phi}^*\mathbf{e}\|_\infty \leq \lambda_n \sigma \tag{2.41}$$

for some $\lambda_n > 0$, where

$$\|\mathbf{\Phi}^*\mathbf{e}\|_\infty = \sup_{1 \leq i \leq n} |(\mathbf{\Phi}^*\mathbf{e})_i| \tag{2.42}$$

and

$$\mathbf{e} = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}} \tag{2.43}$$

is the residuals vector.

The Dantzig Selector searches for the sparsest approximate solution $\hat{\mathbf{x}}$ that are consistent with the constraints. Unlike the basis pursuit denoising which tries to keep the solution close to the feasible set *globally* using the Euclidean norm, Dantzig Selector in Eq. (2.41) employs the l_∞-norm which essentially keep each element of the solution consistent *individually*. It is shown in *Theorem 4* that the Dantzig selector is very accurate in the presence of noise [26].

Theorem 4. *[26] Suppose* $\mathbf{x} \in \mathbb{R}^n$ *is any* K*-sparse vector of parameters*

obeying $\delta_{2K} + \theta_{K,2K} < 1$. *Choose* $\lambda_n = \sqrt{2 \log n}$ *in Eq. (2.41). Then with large probability,* $\hat{\mathbf{x}}$ *obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq C_1^2 (2 \log n) K \sigma^2, \tag{2.44}$$

with $C_1 = 4/(1 - \delta_K - \theta_{K,2K})$.

Apart from the Dantzig Selector, there are other generalisations of the basis pursuit. One attempt is to generalise the objective function as the Nuclear-norm Minimisation [139]. Nuclear-norm is the name given to the vector's element-wise norms, which also include the family of l_p-norm. This work aims to generalise the compressed sensing to work with 2-dimensional matrices instead of vectors. Here the vector sparsity is replaced by the matrix rank, with the goal to minimise the rank of the reconstructed matrix.

Apart from solving the basis pursuit by casting into the linear programming problem, there are many reconstruction algorithms proposed which aim to solve the basis pursuit directly. Unlike the generic solvers such as the interior-point method, these algorithms are tailored for the purpose of compressed sensing reconstruction. The following are the most notable variants of these algorithms.

- SPGL1 solver [163, 164]: SPGL1 solves the dual basis pursuit problem using the concept of Pareto curve to find the root of the dual solution.

- YALL1 solver [173]: YALL1 solves the primal basis pursuit using the concept of Alternating Direction Method of Multiplier (ADMM). In essence, ADMM solves the convex optimisation problem by splitting the problem into several smaller problems to be optimised

jointly.

- ISAL1 solver [102, 103]: ISAL1 is an example of a family of optimisation solver known as a gradient projection solver. This family of solvers traces the gradient of the cost function until the optimal solution is reached. Methods in this family are derived from the well-known method steepest gradient descent and the method of conjugate gradient. These conventional methods, however, are applicable to the fully determined problem.

- l_1-homotopy solver [7]: Homotopy method solves the basis pursuit by iteratively reweighting the optimisation problem, effectively trying to find the simplest sparse solution that fit the constraints.

As shown in [103], these tailored algorithms yield more accurate reconstructed signals than the generic optimisation solvers. Moreover, tailored algorithms are less complex and can handle problems on a larger scale. Among these algorithms, the SparseLab (Interior-point method), ISAL1 and l_1-homotopy are shown to be the most efficient in terms of the reconstruction complexity and accuracy and are chosen for further evaluation. The details of both these algorithms are also discussed in Section 2.2.5 and Section 3.4.

Matching Pursuit

The biggest disadvantage of the convex optimisation approach is its high complexity. This is the main motivation for another reconstruction approach that has lower computation complexity. Algorithms using this approach are known as greedy algorithms, or more commonly known—inspired by the most popular algorithm—as matching pursuits

algorithms. These algorithms are generally faster and easier to implement compared to the convex optimisation approach, thus, they are more practical in many applications that do not require highly accurate reconstructions [156].

Matching pursuit is a classical sparse approximation technique available for more than a few decades. As discussed in Section 2.2.1, matching pursuit algorithms aim to find the sparse signal $\hat{\mathbf{x}}$ that is the best match to the dictionary $\mathbf{D}$. Naturally, early works in compressed sensing adopted the use of matching pursuit in order to reconstruct the sparse signal $\hat{\mathbf{x}}$ from the sensing matrix Φ.

Algorithm 1 Orthogonal Matching Pursuit [59]

Input: the sensing matrix $\mathbf{\Phi}$, the measurements vector $\mathbf{y}$, and the error threshold ϵ_0.

Output: approximate solution of $\min \|\hat{\mathbf{x}}\|_0$ subject to $\mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}$.

1: Initialise the solution $\mathbf{x}^{(0)} := \mathbf{0}$, the residual $\mathbf{e}^{(0)} := \mathbf{y} - \Phi\mathbf{x}^{(0)} = \mathbf{y}$, support set $\mathcal{S}^{(0)} := \emptyset$, and $t = 0$.
2: **repeat** $t := t + 1$
3: Compute the error $\epsilon_j = \min_{z_j} \|\phi_j z_j - \mathbf{e}^{(t-1)}\|_2^2$ for every column j using $z_j = \phi_j^{\mathrm{T}}\mathbf{e}^{(t-1)}/\|\phi_j\|_2^2$.
4: Find $j_0 \notin \mathcal{S}^{(t-1)}$ and $\forall j, \epsilon_{j_0} \leq \epsilon_j$, then update $\mathcal{S}^{(t)} := \mathcal{S}^{(t-1)} \cup \{j_0\}$.
5: Compute $\mathbf{x}^{(t)}$ as the solution of $\min_{\mathbf{x}} \|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|_2^2$ where the supports of $\mathbf{x}^{(t)}$ is $\mathcal{S}^{(t)}$.
6: Compute $\mathbf{e}^{(t)} := \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{(t)}$.
7: **until** $\|\mathbf{e}^{(t)}\|_2 < \epsilon_0$

The earliest used matching pursuit algorithm for compressed sensing reconstruction is the Orthogonal Matching Pursuit (OMP) [59]. *Algorithm 1* describes the algorithm of the OMP, where ϕ_j is a jth column of $\mathbf{\Phi}$ [59]. In essence, the OMP finds the locations of the non-zero elements (known as sparse supports) in $\hat{\mathbf{x}}$ by selecting the indices where $\Phi\hat{\mathbf{x}}$ have to biggest magnitudes, one element per iteration while estimating their values using the least square approximation. This selection rule of the OMP is the simplest greedy rule. It is shown in [156, 157, 54] that the

OMP produces near-optimal sparse approximations, *i.e.*,

$$\|\mathbf{r}^t\|_2 \leq \sqrt{1+6t}\|\mathbf{x}-\mathbf{x}^{t*}\|_2, \tag{2.45}$$

where $\mathbf{x}^{t*}$ denotes the best l_2 approximation of $\mathbf{x}$ as a linear combination of t columns of $\mathbf{\Phi}$. The OMP, however, assumes that each column of the sensing matrix $\mathbf{\Phi}$ is orthogonal to one another, thus, only orthogonalised sensing matrix may be used. Another drawback of the OMP is that, by adding one element at a time, the accuracy of the reconstruction decreases as the sparsity level K increases along with its complexity.

Regularised Orthogonal Matching Pursuit (ROMP), a more refined version of the OMP, is proposed in [119]. The ROMP aims to combine the benefits of the OMP with a more accurate performance guarantee of convex optimisation approach. Currently, a large variety of matching pursuit algorithms have been proposed. There were several attempts to incorporate probabilistic approach into the support selection step of matching pursuit (using Bayesian approach [145, 11] and Expectation-maximisation [48, 47]). Some works addressed the bias in support selection, another drawback of matching pursuit. For example, in the OMP, once the support is added to the support set $\mathcal{S}$, it remains there without any chance of being removed. [149, 150] addressed this issue by reassessing the columns of $\mathbf{\Phi}$ in a cyclic manner. On the other hand, [114] tackled this by using splitting the sensing matrix into several random submatrices instead of one. This allows each column of the sensing matrix to be assessed several times. It is also possible to modify the matching pursuit algorithm to work with the Binary Compressed Sensing (discussed in Section 2.2.3) [172].

In [20], the authors proposed the use of an iterative tree search to

find the sparse approximation instead of searching through the entire space of possible support sets. The concept of reducing the space of support sets is employed in two nearly identical state-of-the-art algorithms, known as the Subspace Pursuit (SP) [41] and the Compressed Sensing Matching Pursuit (CoSaMP) [118]. These techniques add the following improvements to the conventional matching pursuit approach: 1) multiple columns are selected per iteration, 2) pruning the support sets to remove supports with too small magnitude, 3) performance guaranteed based on the restricted isometry property. *Algorithm 2* describes the CoSaMP algorithm. It is shown in [118, 157, 15] that CoSaMP algorithm has the error bound in such that, given $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$, the output $\hat{\mathbf{x}}$ of CoSaMP converges to satisfy

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq Ck^{-1/2}\|\mathbf{x} - \mathbf{x}^K\|_1 + C\|\mathbf{n}\|_2, \tag{2.46}$$

where $\mathbf{x}^K$ is $\mathbf{x}$ with all but K largest elements set to zero, and C is a constant.

Iterative Thresholding

The last approach covered in this section is the family of reconstruction algorithms known as the iterative thresholding. The distinct characteristic of algorithms in this family is that their aim is not to solve the basis pursuit by using optimisation techniques. Rather, these algorithms try to recover the target sparse signal $\hat{\mathbf{x}}$ directly by applying the residue inversion and using a threshold to reinforce the constraint.

The most well-known algorithm in this family is the Iterative Hard Thresholding [15]. This algorithm is shown to give near-optimal error guarantees using minimal observations. Given the measurements

Algorithm 2 Compressed Sensing Matching Pursuit (CoSaMP) [118]

Input: a signal $\mathbf{y}$, the sensing matrix $\mathbf{\Phi}$, target sparsity K, tuning parameter α, and the error threshold ϵ_0

Output: a K-sparse signal $\hat{\mathbf{x}}$

1: Initialise $\mathbf{x}^{(0)} := \mathbf{0}$, residual $\mathbf{e}^{(0)} := \mathbf{y}$, and $t := 1$.
2: **repeat** $t := t + 1$
3: Find α columns of $\mathbf{\Phi}$ that are most strongly correlated with $\mathbf{e}^{(t-1)}$:

$$\Omega \in \arg\min_{\mathcal{S}^t} \sum_{j \in \mathcal{S}^t} |\langle \mathbf{e}^{(t-1)}, \phi_j \rangle|.$$

4: $\mathcal{S}^{(t)} := \mathcal{S}^{(t-1)} \cup \Omega$.
5: Find the best coefficients that fit the residual with the given $\mathcal{S}^{(t)}$

$$\mathbf{w}^{(t)} = \arg\min_{\mathbf{w}} \|\mathbf{e}^{(t-1)} - \mathbf{\Phi}_{\mathcal{S}^{(t)}}\mathbf{w}\|_2.$$

6: Retain only the K largest elements: $\mathbf{x}^{(t)} := [\mathbf{w}^{(t)}]_K$.
7: update the residual: $\mathbf{e}^{(t)} := \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{(t)}$.
8: **until** $\|\mathbf{e}^{(t)}\|_2 < \epsilon_0$.

$\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, the reconstructed sparse signal $\hat{\mathbf{x}}$ can be obtained by using the iteration

$$\mathbf{x}^{(t+1)} = H_K(\mathbf{x}^{(t)} + \mathbf{\Phi}^{\mathrm{T}}(\mathbf{y} - \Phi\mathbf{x}^{(t)}), \tag{2.47}$$

where $t = 0, 1, 2, ...$ and $H_K(\mathbf{a})$ is the non-linear operator that sets all but the largest K elements of $\mathbf{a}$ to zero. It is shown in [16] that the iteration Eq. (2.47) converges to an optimal solution of the optimisation problenm

$$\min_{\hat{\mathbf{x}}} \|\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}\|_2^2 \text{ subject to } \|\hat{\mathbf{x}}\|_0 \leq K. \tag{2.48}$$

In other words, Eq. (2.47) converges to a K-sparse solution $\hat{\mathbf{x}}$ that approximately satisfies the constraint $\mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}$. *Algorithm 3* summarises the Iterative Hard Thresholding algorithm.

The error bound of the reconstruction using the Iterative Hard Thresholding is given in *Theorem 5.*

Theorem 5. *[15] Given a noisy observation* $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$*, where* $\mathbf{x}$ *is an*

Algorithm 3 Iterative Hard Thresholding (IHT) [15]

Input: a signal $\mathbf{y}$, the sensing matrix $\mathbf{\Phi}$, target sparsity K, and the error threshold ϵ_0

Output: a K-sparse signal $\hat{\mathbf{x}}$

1: Initialise $\mathbf{x}^{(0)} := \mathbf{0}$, and $t := 1$.
2: **repeat** $t := t + 1$
3: Compute the residual $\mathbf{e}^{(t)} := \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{(t-1)}$.
4: Update the signal $\mathbf{w}^{(t)} := \mathbf{x}^{(t)} + \Phi^{\mathrm{T}}\mathbf{r}^{(t)}$.
5: Update the signal $\mathbf{x}^{(t)} := H_K(\mathbf{w}^{(t)})$, with a hard threshold H_K.
6: **until** $\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^{(t)}\|_2 \leq \epsilon_0$.

arbitrary vector. Let $\mathbf{x}^K$ *be an approximation to* $\mathbf{x}$ *with no more than* K *non-zero elements for which* $\|\mathbf{x} - \mathbf{x}^K\|_2$ *is minimal. If* $\mathbf{\Phi}$ *has restricted isometry property with* $\delta_{3K} < 1/\sqrt{32}$*, then, at iteration* t*, Iterative Hard Thresholding will recover an approximation* $\mathbf{x}^{(t)}$ *satisfying*

$$\|\mathbf{x} - \mathbf{x}^{(t)}\|_2 \leq 2^{-t}\|\mathbf{x}^K\|_2 + 6\tilde{\epsilon}_K, \tag{2.49}$$

where

$$\tilde{\epsilon}_K = \|\mathbf{x} - \mathbf{x}^K\|_2 + \frac{1}{\sqrt{K}}\|\mathbf{x} - \mathbf{x}^K\|_1 + \|\mathbf{n}\|_2. \tag{2.50}$$

Furthermore, after at most

$$t^* = \lceil \log_2(\frac{\|\mathbf{x}^K\|_2}{\tilde{\epsilon}_K}) \rceil \tag{2.51}$$

iterations, Iterative Hard Thresholding estimates $\mathbf{x}$ *with accuracy*

$$\|\mathbf{x} - \mathbf{x}^{(t)}\|_2 \leq 7[\|\mathbf{x} - \mathbf{x}^K\|_2 + \frac{1}{\sqrt{K}}\|\mathbf{x} - \mathbf{x}^K\|_1 + \|\mathbf{n}\|_2]. \tag{2.52}$$

Despite the simplicity of this approach, *Theorem 5* shows that the performance guarantee of the Iterative Hard Thresholding is comparable to that of the Compressed Sensing Matching Pursuit (shown in Section 2.2.4).

Apart from the Iterative Hard Thresholding, many works using this approach are also present. In [82], the Iterative Hard Thresholding is extended to solve the Binary Compressed Sensing problem (more details in Section 2.2.3). Instead of using a blind hard thresholding, in [132], the authors developed the Generalised Expectation-Maximisation (GEM) hard thresholding that aims to reconstruct the signal in the presence of Gaussian noise. The same authors later proposed a more generalised version of their algorithm in [133] that works with a wider range of sensing matrices and sparse bases. A more extreme example of iterative thresholding approach can be seen in [58, 112] where the authors attempted to reconstruct an image from undersampled measurements by applying spatial filters to the inverse back-projection results repeatedly. The demonstrated results show the successful reconstruction from partial Fourier coefficients, however, it did not gain enough traction compared to a more generalised Iterative Hard Thresholding algorithm.

The biggest limitation of algorithms using this approach is that, much like most of the greedy algorithms, the level of sparsity has to be known in advance. The error bound in *Theorem 5*, for example, shows that the reconstructed signal $\hat{\mathbf{x}}$ is guaranteed to be close to $\mathbf{x}^K$. Thus, unless the sparsity level K of $\mathbf{x}$ is known, it is unlikely that $\hat{\mathbf{x}}^K$ will be close to $\mathbf{x}$.

2.2.5 Comparison of Reconstruction Algorithms

The fact that there are many reconstruction algorithms available makes it a challenge to compare for the best algorithm, especially when each has its own strength. In general, the convex optimisation based algorithms, such as l_1-homotopy and Interior-point method, yield more accurate reconstruction than greedy and iterative thresholding methods. However,

they also suffer from higher complexity, resulting in a larger computational time.

Because of the time and space constraint, it is impractical to evaluate the results of this study using every available reconstruction methods and approaches. Instead, in this section, the most optimal method in terms of reconstruction accuracy and complexity is identified and is used as the state-of-the-art benchmark for the rest of the study. Here the accuracy is given more priority than the complexity because it is vital for our focus, the fMRI data, to have an accurate reconstruction. Thus, the optimal reconstruction algorithm for other applications may vary.

To evaluate, the reconstruction algorithms are used to reconstruct a set of test signals and images from compressive measurements obtained using various sampling rate. The following reconstruction algorithms are evaluated:

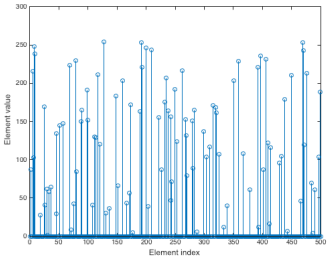
- Sparselab [51]
- ISAL1 [103]
- l_1-homotopy [7]
- Orthogonal Matching Pursuit (OMP) (*Algorithm 1*)
- Compressed Sensing Matching Pursuit (CoSaMP) [118]
- Iterative Hard Thresholding (IHT) [15]

It should be noted that the reconstruction accuracy and the computational time depend on the implementation of these algorithms. The original authors' supplied code and our own implementation have been tested and found that there are little effects between different implemen-

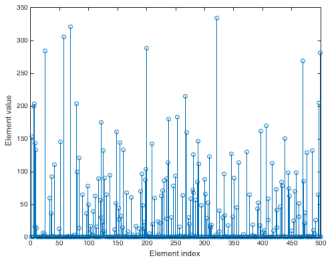
tations. To maximise their performance, the originally supplied code is used in the following evaluation.

The first test is the set of exact sparse signal randomly generated. Each signal is 1-dimensional, 500 elements long with $K = 100$ non-zero elements. Figure 2.20 shows the example of such a signal. These signals are the kind most compressed sensing literature work with, *i.e.*, they are short—only a few hundreds elements long—and very sparse. Figure 2.21 shows the average PSNR of the reconstructed signals versus the sampling rate using each algorithm. The results following the theoretical prediction discussed in Section 2.2.4 can be seen clearly. The convex optimisation based algorithms yield the best quality in terms of PSNR. The greedy algorithms give lower PSNR, with CoSaMP—the state-of-the-art greedy method—gives the results comparable to l_1-homotopy, down to the very low PSNR results given by OMP. On the other hand, Figure 2.22 shows a very interesting result. This is where the difference between the theoretical prediction and practice can be observed. It shows that, in practice, the convex optimisation based algorithms are several degree faster than greedy algorithms. The reason for this is that greedy algorithms, such as OMP and CoSaMP, optimise only one element per iteration. Despite having a much lower complexity—that is, simpler mathematical operations—in an iteration, these algorithms take a much higher number of iteration to converge to the optimal solution compared to optimisation based algorithms.

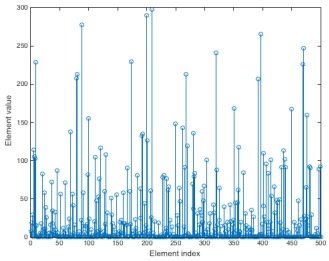
The contrast can be seen in the second test of real images. The set of test images consists of 64×64-pixel images, sparsified using Discrete Fourier basis. Unlike the previous test set, these images are only approximately sparse and are much larger in size. Figure 2.23 shows an example



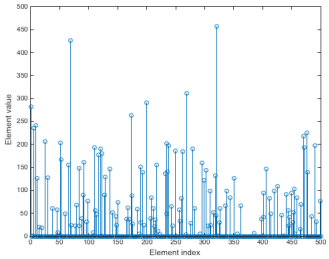
(a) original



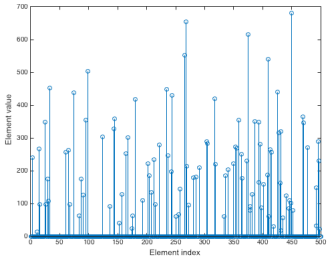
(b) SparseLab



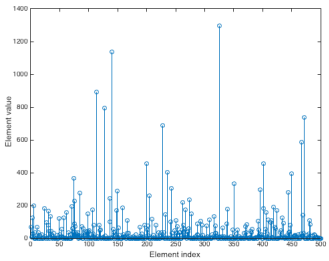
(c) ISAL1



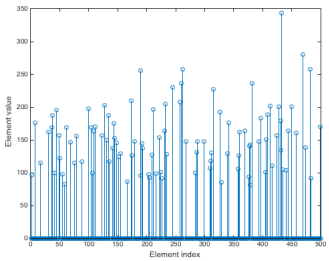
(d) l_1-homotopy



(e) IHT



(f) OMP



(g) CoSaMP

Figure 2.20: Example of a test sparse signal and its reconstruction from 30% samples

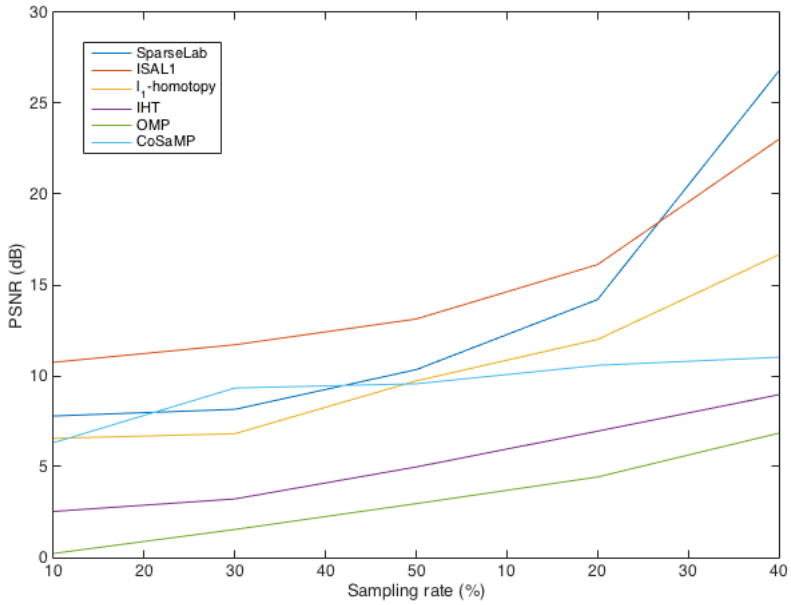


Figure 2.21: Average PSNR versus sampling rate comparison for exact sparse signals

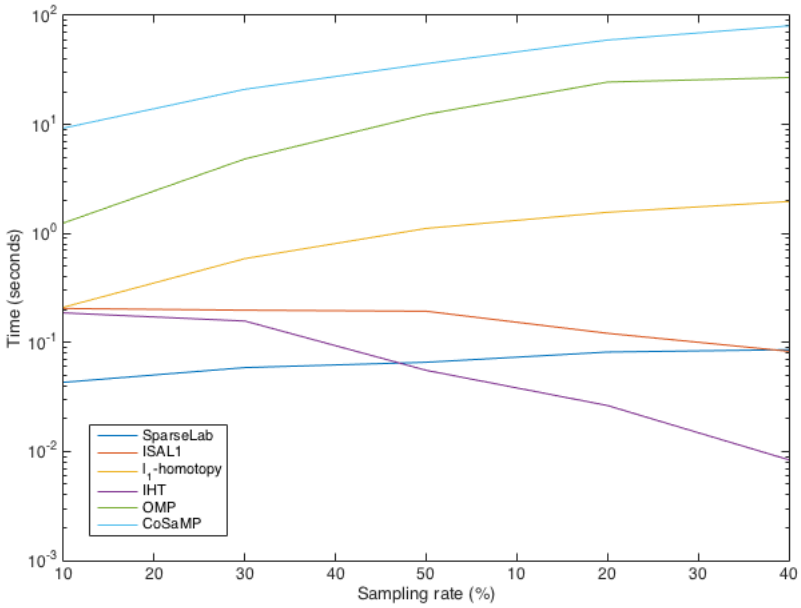
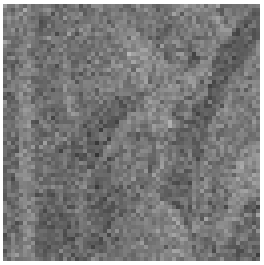


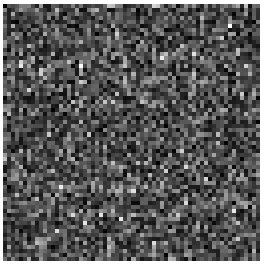
Figure 2.22: Average computation time versus sampling rate comparison for exact sparse signals

(a) original

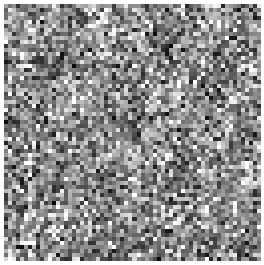


(b) SparseLab

(c) ISAL1

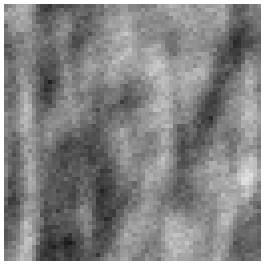


(d) l_1-homotopy



(e) IHT

(f) OMP



(g) CoSaMP

Figure 2.23: Example of a test image and its reconstruction from 30% samples

of the images and the reconstructed results. In this test, the PSNR plot, shown in Figure 2.24, shows that many of the algorithms do not work with large, approximately sparse signals. Interestingly, the OMP now yields results with high PSNR, outperforms CoSAMP, IHT, and even SparseLab. Because these algorithms work fine with short signals but not the larger ones, it is difficult to employ them in practical signal reconstruction such as images, video sequences, and—for our purpose in particular—the MRI data. To pinpoint exactly whether this phenomenon is a fault in their implementation or their theoretical limitation is beyond the scope of this study. Figure 2.25 shows the average computation time. The same observation about computation time as in Figure 2.22 can also be observed here. Now that the signals are much larger, it takes much longer time for the greedy algorithms to converge to the solution.

In conclusion, because the MRI data reconstruction gives a higher priority to accuracy, the ISAL1 algorithm is chosen—after taking both reconstruction accuracy and computational time for consideration— as the state-of-the-art reconstruction algorithm for l_1-norm minimisation.

2.3 Compressed sensing and MRI

So far, we have explored the general backgrounds and the state of the arts of compressed sensing framework, including both the sensing and reconstruction operations. In this section, we will explore the literature focuses specifically on applying the compressed sensing to MRI data.

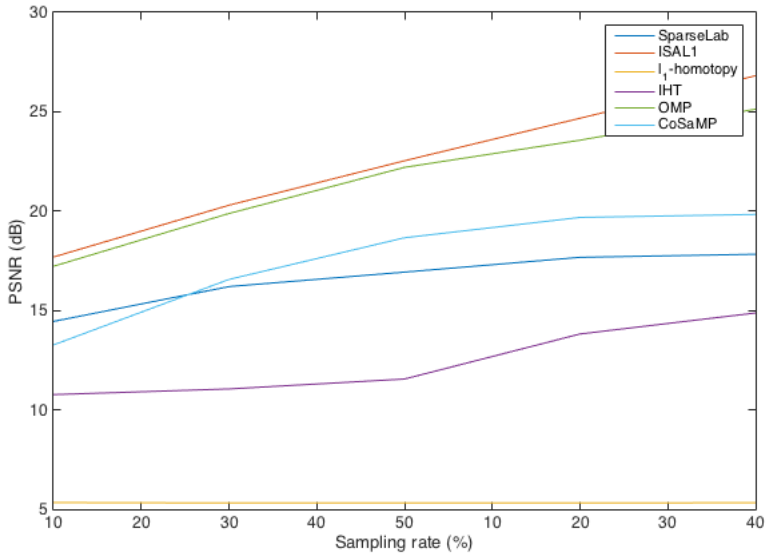


Figure 2.24: Average PSNR versus sampling rate comparison for test images

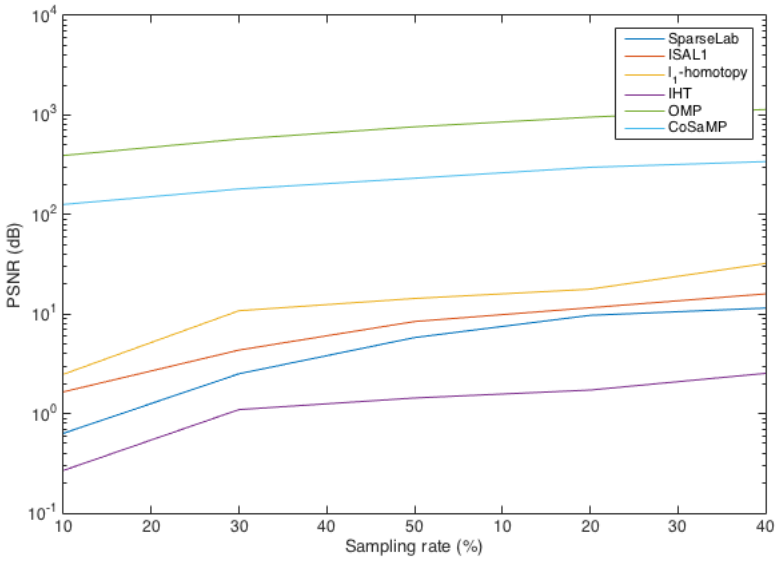


Figure 2.25: Average computation time versus sampling rate comparison for test images

2.3.1 Possibility of compressed sensing MRI

This section explores the possibility of applying compressed sensing framework to the MRI data. The MRI data has been one of the most interested in types of data among compressed sensing researchers. The primary reason for this is the nature of the MRI data and its acquisition scheme. As previously discussed in Section 2.1.1, the raw data is acquired from MRI hardware readily in the frequency domain, *i.e.*, the Fourier coefficients. It is one of the applications where the data is acquired directly in a sparse domain, thus, it has been mentioned as an example for an application of compressed sensing ever since the earliest works. Indeed, the first work on compressed sensing by Candès [28] aimed to retrieve the complete data from undersampled Fourier measurements, which fits into the MRI paradigm perfectly.

While the MRI raw data is a sparse signal in nature; to apply compressed sensing to MRI in practice, there are many MRI-specific issues arise. The biggest MRI specific issue involving the implementation of the sensing operation. Theoretically the sensing operation is defined as a linear operation of a sensing matrix to a target signal (Eq. (2.18) in Section 2.2.3). The challenge is to realise this equation in the physical hardware. The target signal in this case is a sparse signal, thus, instead of Eq. (2.18), the sensing operation reduces to

$$\mathbf{y} = \mathbf{\Phi x}. \tag{2.53}$$

In practice, the sensing matrix $\mathbf{\Phi}$ has to be implemented as a hardware or as a special pulse sequence (refer to Section 2.1.1).

The partial Fourier basis, a type of Subsampled Incoherent Basis

Matrices (Section 2.2.3), is one of the most popular sensing matrices for the MRI data. The partial Fourier matrix can be implemented relatively easily by randomly dropping some line from the scan trajectories, as first shown by Lustig *et al.* in [106, 105]. Instead of Cartesian trajectory, it is also shown that randomly drop lines in other scan trajectories—radial trajectory in particular—performs better as a sensing operator [104, 137, 161]. These works showed that not only compressed sensing can be incorporated into MRI, but they are also highly compatible.

There is also a real practical reason for attempting to apply compressed sensing to the MRI acquisition scheme. As discussed in Section 2.1.3, the MRI data has a trade-off between its spatial resolution and the acquisition time. To acquire data with higher resolution, the subject is required to remain in the scanner for a longer period of time. This has negative effects on both the subject's mental conditions and the limitation of data acquisition, particularly in fMRI experiments. A Recent study showed that compressed sensing, using random subsampled and iterative thresholding reconstruction, can accelerate the acquisition speed up to 4 times while retaining the diagnostically acceptable quality of the reconstruction results [92].

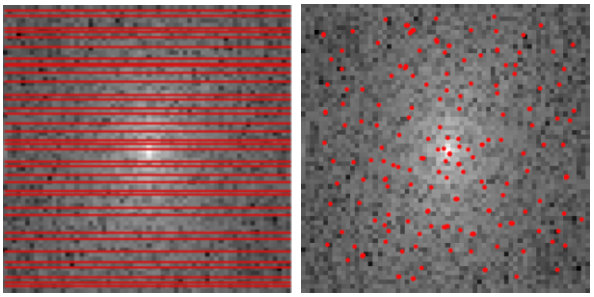
2.3.2 Compressed Sensing with Structural MRI

The majority of works attempted to apply compressed sensing to MRI, especially during the early days, focussed on the structural MRI. The terms structural MRI, anatomical MRI, and diagnostic MRI refer to the high-resolution scan that is used mainly for clinical diagnosis and medical study. The aim of the structural MRI is to provide the accurate anatomical details inside of the body. It has the characteristic similar

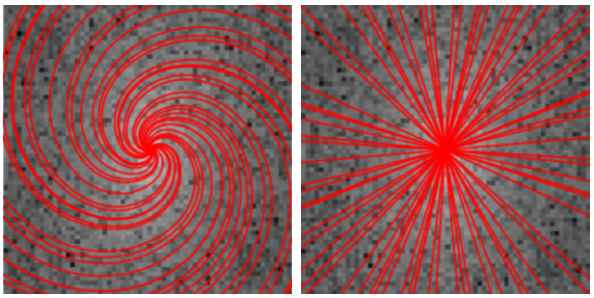
to other common medical imaging techniques, such as X-ray and CT, consisting of fine anatomical details of a single cross-sectional snapshot at a specific location of the body. High-resolution data is compulsory for structural MRI to reveal small anatomical details, thus, the aim of applying compressed sensing to the structural MRI is mainly to achieve the best reconstruction of data from the accelerated scan. Here the undersampled measurements are acquired as a mean to improve the acquisition speed.

There are many comprehensive works of literature on compressed sensing with structural MRI, both from the engineering point-of-view [105, 70, 117], and the clinical point-of-view [83, 81]. Clinical literature focuses primarily on the diagnosis effects of the CS-MRI data. On the other hand, from the engineering point-of-view, literature focuses more on the implementation issue and signal processing of such system. This thesis, too, looks at the CS-MRI from the engineering point-of-view. Roughly speaking, we can group these works of literature into two categories: the work on sensing operation and the reconstruction operation.

Let us first consider the works on the sensing operation. Because of the fact that MRI data is acquired in the sparse domain directly, most of the issue involving the sensing operation is the implementation of the sensing matrix. As previously mentioned, the first work on this issue is Lustig's Sparse MRI [105]. In this work, Lustig suggested several incoherent sensing strategy for MRI, all of them involve randomly dropping some trajectory lines during the scan, creating the random Subsampled Incoherent Matrix (Section 2.2.3). The randomly dropped lines can be done in many scan trajectories, such as a 2-dimensional Cartesian, a 3-dimensional Cartesian, and even a radial trajectory (Shown in



(a) 2D Cartesian (b) 3D Cartesian



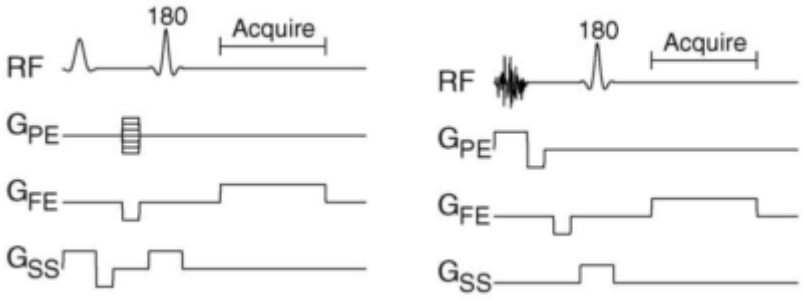
(c) Spiral (d) Radial

Figure 2.26: Randomly dropped trajectory in (a) 2-dimentional Cartesian, (b) 3-dimentional Cartesian (random points in 2D plain), (c) spiral trajectory, and (d) radial trajectory

Figure 2.26). He also suggested that randomly dropped lines in radial trajectory achieves the highest incoherence to the K-space data.

These trajectories studied by Lustig are the normal MRI trajectories widely employed by radiologists. However, it is also possible to have a random trajectory specifically tailored for compressed sensing, such as proposed in [81]. This kind of tailored trajectory works well with compressed sensing because it can capture most of the k-space energy in only a small number of sample.

Another approach to implement the sensing matrix aims to modify the RF excitation pulse instead of the readout trajectory. As mentioned in Section 2.1.1, MRI depends on the excitation using radio-frequency pulses. Rather than using a constant frequency pulse, this pulse can be used to encode the randomness into the acquisition. In [75], it proposed



(a) Conventional encoding (b) Random encoding

Figure 2.27: Random encoding using random magnitude RF pulse [75]

the use of random magnitude RF pulse as an implementation of the sensing matrix (shown in Figure 2.27).

Later, more research has followed in this direction. It finally developed into Spread Spectrum Compressed Sensing MRI, where the randomness is applied to the phase encoding via specialised RF pulse. For examples, it can be implemented by using linear frequency-swept RF pulses [135], random phase encoding gradient [174], or even a physical modification of the machine—as originally proposed—by using shim coils [131]. It is shown in [131] that the Spread Spectrum Compressed Sensing MRI has lower distortion compared to the variable-density scan, as proposed by Lustig.

While the issue regarding the implementation of the sensing operation is important, more research is focusing on the reconstruction operation. The reason for this is because, as we have discussed, the accuracy of the standard out-of-the-box compressed sensing reconstructions do not satisfy the extra-high accuracy requirement of medical imaging. This leads to an explosion of research on novel reconstruction algorithms and methods tailored for compressed sensing MRI application. One of the common approaches is the two-stage reconstruction method where the first stage

solves the compressed sensing reconstruction problem itself follows by the second stage where the reconstructed solution is optimised for better visual quality. Many works of literature using this approach employ different technique for the second stage; for examples, using the smooth l_0-norm optimisation for the second stage [134],using the weighted linear least square estimation [165], and using the relaxed l_1-norm minimisation [44]. The authors of [44] later developed the two-stage reconstruction approach further. This two stage reconstruction separates the MR data into low frequency and high frequency parts, and reconstruct each part separately, using the low-frequency reconstruction to guide the reconstruction of the high-frequency part [175]. This approach can also be seen in a similar approach proposed in [180], where the reconstruction is done using two different scales of discrete wavelet transform. It should be noted that the main goal of structural MRI is the visual quality for clinical diagnosis. The overall subjective meaning is more important than the accuracy of each voxel's value. As such, these mentioned two-stage approaches do not aim to provide an accurate reconstruction directly, but rather to provide the most meaningful reconstruction—or the reconstruction with the best visual quality.

Another big issue regarding the reconstruction operator of compressed sensing MRI is the computational complexity. It is very undesirable to have the acquisition system that takes ages to process the acquired data for viewing. This has a severe clinical impact since in many cases the diagnosis is needed to be done in urgency. Not to mentioned the frustration and difficulty of patients and practitioners involved. Indeed, if the reconstruction of the MRI data took too long, the technique would be abandoned in favour of other, more instantaneous imaging techniques

instead. This is exactly the issue of compressed sensing MRI, of which the processing time is dramatically increased as a trade-off for the higher temporal resolution. To tackle this issue, many works have been done to develop fast reconstruction methods for compressed sensing MRI. Chartrand is among the first who proposes the fast algorithms for the compressed sensing MRI in [33] by extending the two-stage reconstruction approach in [175] into a multi-stage reconstruction with the aim to reduce the complexity. The SENSE reconstruction algorithm, one of the popular reconstruction algorithm for compressed sensing MRI, also employs the same splitting concept of Chartrand [87]. Though this splitting approach work with general MRI techniques, more specialised approaches with better performance are developed for each specialised MRI techniques. For example, [116] proposes the faster variation of l_1-SPiRiT algorithm for compressed sensing Parallel MRI. While most reconstruction techniques focus on the 2-dimensional problem, following the conventional compressed sensing paradigm, [113] proposes the extension of the two-stage splitting approach by exploiting the feature of 3-dimensional space, which is the actual space that the MRI data is in. On the other hand, [181] proposes the fast reconstruction by reducing the problem's dimension using the principal component analysis.

The structural MRI generally consists of only a single high-resolution volume, thus, there is no temporal redundancy available. In other types of MRI—which will be discussed in the next part—that have temporal redundancy, the exploitation of this temporal redundancy has become the biggest improvement in reconstruction quality. Before we move on, we would like to mention few works of literature on structural MRI, which are out of categories that were discussed. Compressed Sensing

MRI based on Nonsubsampled Contourlet Transform, proposed in [136], employs a contourlet transform as a sparsifying basis, which is shown to outperform the wavelet basis and Fourier basis. However, because of the complexity of the contourlet transform, and also its non-rigorous mathematical formulation, this approach never catches on. [160] employs the Gaussian Mixture Model to the sparse representation of the structural MR data. This approach, however, finds its success in methods involves temporal redundancy which will be discussed in the next section.

Finally, while most of the aforementioned works are from the engineering field, there are some works of literature from the medical field verifying the successfulness of compressed sensing with the structural MRI data. In particular, [110] reports the evaluation from clinical point-of-view of compressed sensing MRI which shows a satisfactory result. Also, [2] reports a successful application of compressed sensing in Hyperpolarized 3He Lung Angiography.

2.3.3 Compressed sensing with Dynamic MRI

In this section, we will focus on the works of literature applying compressed sensing framework to the Dynamic MRI. Dynamic MRI is a special application of MRI aims specifically to capture the movement (dynamic) of body organs. It is especially useful as a diagnosis tool for the heart, lung, and joint monitoring, allowing the practitioners to study the working of these organs in real time (Figure 2.28), as opposed to just a snapshot of their appearance as with the structural MRI.

Because the data of Dynamic MRI is sequentially acquired over time, it is obvious that the Dynamic MRI data is rich with temporal redundancy. As we have seen from in the case of general compressed sensing,

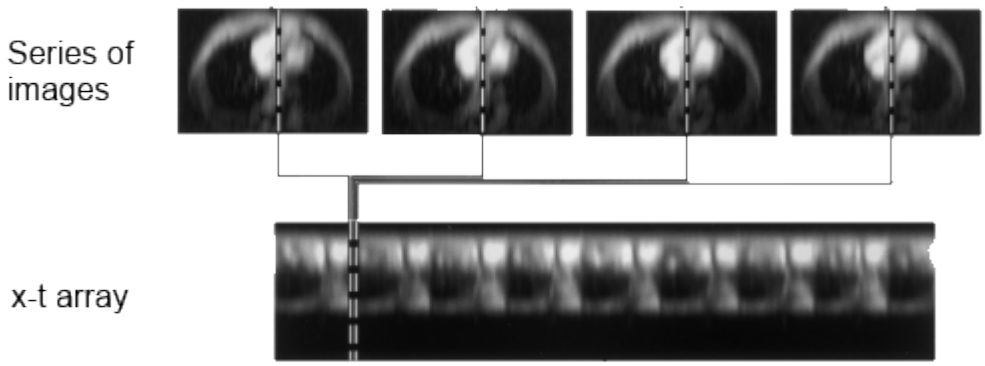


Figure 2.28: Example of dynamic MRI [158]

the use of the temporal redundancy as a side-information works well and is very desirable. Thus, it comes at no surprise that most works of literature involving compressed sensing and Dynamic MRI exploit this temporal redundancy in one way or another. The presence of the temporal redundancy in the data is the clear distinction between methods for Dynamic MRI and methods for the structural MRI.

The exploitation of the spatio-temporal redundancy in Dynamic MRI dates back beyond the introduction of compressed sensing. The methods exploiting such redundancy to increase the acquisition rate of Dynamic MRI, namely k-t Broad-use Linear Acquisition Speed-up Technique (k-t BLAST) and k-t Sensitivity Encoding (k-t SENSE), have been introduced in [158]. Both techniques essentially undersampled the k-space data in temporal direction and filled in the missing data using a trained interpolation method.

The k-t BLAST/SENSE methods have become the conventional mean of acquiring the Dynamic MRI data. The exploitation of the spatio-temporal redundancy is later incorporated into compressed sensing Dynamic MRI in most literature. The works in [69], [35], and [107] all employ the temporal redundancy to improve the compressed sensing re-

construction of the Dynamic MRI data. This concept of using both the redundancy and compressed sensing for dynamic MRI has matured into a state-of-the-art algorithm known as k-t FOCUSS (k-t Focal Underdetermined System Solver) [89]. k-t FOCUSS works by exploiting the sparsity of the frame difference. To put it simply, given two consecutive Dynamic MRI volumes V_1 and V_2, it is well-known that $V_2 - V_1$ is sparser than V_2 itself, thus solving for $\min \|V_2 - V_1\|_1$ yields better results than solving for $\min \|V_2\|_1$. It is shown that k-t FOCUSS outperforms the conventional k-t BLAST/SENSE techniques and is successfully applied to the *in vivo* cardiac Dynamic MRI experiment [88].

Apart from k-t FOCUSS, there are many attempts trying to improve the reconstruction quality and acquisition rate of Dynamic MRI further. Zhao improves the reconstruction quality by applying the two-stage reconstruction approach commonly employed in the structural MRI method into Dynamic MRI [179]. In [153], the authors improved the acquisition rate by adopting the spiral trajectories scan in place of the Cartesian trajectories commonly used by k-t FOCUSS.

The characteristic of Dynamic MRI can be viewed as the affine transformation of an object (including translation and rotation for joint, and skewing for the heart). This fact leads to the attempts to apply methods designed to cope with the motion; for example, by applying the motion compensation before exploiting the redundancy [14]. Also, there is a suggestion involving a trained sparsity basis tailored specifically for Dynamic MRI application. For example, a novel dictionary learning time sparsity basis for Dynamic MRI, proposed in [21], can outperform the fixed sparsity basis used by k-t FOCUSS.

Finally, even though Dynamic MRI can be viewed as the affine trans-

formation, more general methods exploiting the temporal redundancy can also be used. Weizman and Eldar, in particular, proposed the compressed sensing for the longitudinal MRI, a generalised format of MRI in the temporal domain. The redundancy can be exploited in both the temporal redundancy between 2 adjacent volumes (V_n and V_{n+1}) [166], and the spatial redundancy between 2 adjacent slices ($I_n and I_{n+1}$) [167]. These methods share similarity with k-t FOCUSS. This approach is called the temporal compressed sensing (TCS). Given a compressive measurements $\mathbf{y} = \mathbf{\Phi x}$, at any time instance t, the reconstructed k-space $\hat{\mathbf{x}}$ is

$$\min \|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\|_2 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}}^{(t)} = \mathbf{y}. \tag{2.54}$$

This approach is not limited to work only on the affine changing time-series data like dynamic MRI but also on the stationary, intensity-varying data of functional MRI, which leads us to the next section on functional MRI.

2.3.4 Compressed Sensing with Functional MRI

As the fMRI data consists of a time series of several volumes, fMRI can be benefited from the presence of temporal redundancy similar to the Dynamic MRI. However, the goal—and thus the characteristic—of fMRI is different from that of Dynamic MRI. In dynamic MRI, the goal is to study the moving parts of bodily organs. In essence, it is still anatomically focused like the structural MRI, with only the extension for monitoring the changing over time. On the other hand, fMRI primary aim is to monitor the brain activity, *i.e.*, its function. This difference in focus between functional and structural results in a different feature of interest. Whereas both the structural MRI and Dynamic MRI are concerned

with the image data as is, fMRI is interested in the brain activity the data inferring to. Whereas both the structural MRI and Dynamic MRI focus on how the organ looks like spatially, fMRI focuses on the changing intensity value of each voxel, specifically the changes due to the blood-oxygen-level dependent (BOLD) effect. These differences not only affect how the techniques are used but also affect how the compressed sensing be applied.

Most compressed sensing literature on the structural MRI, as discussed in Section 2.3.2, employ the two-stage reconstruction technique, where overall visual quality is preferred over the exact voxel intensity value. This technique is clearly not suitable for fMRI data, as the changes in voxel intensity lie at the heart of fMRI and is more important that the visual quality. Techniques for Dynamic MRI which are mostly based on k-t BLAST/SENSE also depend on the interpolation of intensity value. Unlike the linear translation of physical organs, the BOLD signal is not linear and cannot be interpolated easily. It is shown that by using k-t SENSE with the fMRI data, while the acquisition time can be successfully cut in half, suffered a lost in the final activity map [43]. Thus, the technique commonly employed in Dynamic MRI is also not suitable for fMRI.

It is notable that very few works have been done in the field of compressed sensing and fMRI compared to other types of MRI. The number of research is even strikingly small compared to compressed sensing field as a whole. The lack of literature in this area of study is an obvious gap that requires more research.

Among the current works of literature on this subject so far, there is no clear "theme" or the general trend of how the compressed sensing

should be applied. The first attempt to apply compressed sensing to fMRI is in [90], where the k-t FOCUSS technique is applied to the fMRI experiment in order to evaluate the effect on the BOLD signal acquisition and analysis. The aim of this work is to improve the temporal resolution of MRI and [90] successfully shows that using k-t FOCUSS in conjunction with Karhunen-Loeve transform (KLT) in temporal direction can yield an accurate BOLD signal detection.

On the other hand, [146] employs the use of low-rank and sparse decomposition to obtain the higher temporal resolution. In this approach, the entire time-series fMRI data is modelled as a large matrix $\mathbf{M}$, which is a superposition of a low-rank matrix $\mathbf{X}_L$ and a sparse matrix $\mathbf{X}_S$. The sparse matrix part can then be undersampled and reconstructed using the convex optimisation, *i.e.*, the reconstructed data is obtained by solving

$$\min \|\mathbf{X}_L\|_1 + \|\mathbf{\Psi}\mathbf{X}_S\|_1 \text{ subject to } \mathbf{\Phi}(\mathbf{X}_L + \mathbf{X}_S) = \mathbf{y}, \tag{2.55}$$

where $\mathbf{\Psi}$ is the sparsifying basis, $\mathbf{\Phi}$ is the undersampling operator and $\mathbf{y}$ is the undersampled data.

Both approaches aim to increase the temporal resolution of the fMRI data. However, the temporal resolution can also be improved by aiming directly for the brain activity information. As the brain activity is the true goal of fMRI, it is reasonable to aim for the more accurate activity detection rather than only the data with high temporal resolution. An example of this approach is the proposal in [31] to use a Bayesian approach to classifying the activation regions. Another example is [100], which goes a step forward to propose the use of a learned dictionary tailored for the undersampled acquisition of the BOLD signal. In this work, the notion of acquiring the brain image has been put aside for the BOLD

signal. The learned dictionary is decided so that the reconstructed sparse signal, which is not a common image representation, is compatible with the General Linear Model and the BOLD signal can be extracted.

Also, not all works of literature focus on the improvement of the temporal resolution. As there are many conventional fast acquisition techniques available that improve the temporal resolution by sacrificing the spatial resolution (see Section 2.1.3), it is desirable to have a technique that can improve the spatial resolution of such fast acquisition techniques. For example, in [63], the authors aimed to obtain high spatial resolution fMRI data. In this work, randomly undersampled measurements from a specially decided trajectory are obtained and reconstructed using the convex optimisation. It is reported to achieve 6 times improvement in spatial resolution with low distortion.

Despite the lack of the research trend in this field, we can see the general requirement for the desirable compressed sensing fMRI. The compressed sensing fMRI should

1. yield an accurate—or, if possible, even more detailed—BOLD signal and activity detection,

2. achieve a good ratio of the spatial resolution/temporal resolution trade-off which outperforms the conventional fast acquisition techniques such as EPI, and

3. be efficiently implemented, without and with little hardware alteration.

Apart from these requirements, the ability to obtain the reconstructed data quickly or in real time is also desirable. As with the case of dynamic MRI, the real-time acquisition is important in many clinical conditions.

The research in this thesis will aim to adhere to the requirement of compressed sensing fMRI stated here.

2.4 Summary

This chapter has surveyed the background and literature of fMRI and compressed sensing. The fMRI data captures the activities of the brain in the form of temporal intensity variation of the voxel magnitude. Its acquisition speed and resolution are dictated by the physical constraints of the system.

To improve the trade-off ratio between the acquisition speed and resolution, one method is to use compressed sensing. Compressed sensing allows a full-length signal to be acquired from undersampled measurements, using an incoherent sensing operation and a non-linear reconstruction operation based on mathematical optimisation. The chapter has presented several variations of the sensing and reconstruction operations and identified the benefits and disadvantages of them. There are many attempts to combine compressed sensing can be applied to MRI acquisition scheme. Most of these works focus on a specific application of MRI; mainly the structural MRI and Dynamic MRI. On the other hands, the literature combining compressed sensing to fMRI presents only in small number.

From this review, we have identified the requirements for the compressed sensing fMRI: high accuracy, high spatial/temporal resolution trade-off, and easy to implement. The reconstruction operation that can satisfy these requirements is introduced in the next chapter.

Chapter 3

Compressive Sensing Reconstruction From A Correlated Reference

Compressed sensing aims to obtain the full-length signal from compressive measurements by minimising the l_1-norm objective function. However, this is not always necessary. This chapter proposes a novel compressed sensing reconstruction method that minimises the difference between a signal and an arbitrary reference. As a proof of concept, the reconstruction method, called Referenced Compressed Sensing (Referenced CS), is shown to reconstruct images in the spatial domain and works especially well with video sequences.

3.1 Motivation

One major characteristic of natural signals,*i.e.*, signals acquired using a physical system from real-world sources, is the information redundancy. Unlike artificially created signals, in which abrupt changes can

Figure 3.1: Example of a natural image

be observed, variation in natural signals exhibits smooth and continuous changes. It results in a high amount of redundancy in any given signal. Exploiting the signal redundancy is also a main technique in data compression, in which a set of redundant signals is represented using a much smaller representation (such as entropy coding).

Let us explore such redundancy in more details. Consider an image in Figure 3.1. We can observe many redundancies between each part of the image. For examples, each row of the image resembles each other, as in Figure 3.2. The image can be split into several small blocks. This block-based analysis is one of the most common image processing techniques. Figure 3.3 shows the examples of these blocks of the image. We can notice that each block resembles its neighbours to some degree.

All these redundancies in the image's pixel data are referred to as the *spatial redundancy*. The spatial redundancy finds many applications in intra-coding technique, where the spatial redundancy of a single video frame is exploited.

Objectively, we can compute the distance between each part of the image. This distance, specifically the Euclidean distance, is computed

Figure 3.2: Example of spatial redundancy in image rows

Figure 3.3: Example of spatial redundancy in image blocks

Table 3.1: Average Euclidean distance between $\mathbf{I}^{(j)}$ and $\mathbf{I}^{(j-l)}$

Image	l	Distance	
		rows	columns
Lena	1	155.71	224.45
	2	251.84	356.25
	3	310.69	425.93
	5	398.41	514.96
	10	483.02	669.98
Peppers	1	190.34	228.95
	2	314.58	378.54
	3	398.93	471.75
	5	508.46	569.79
	10	611.93	681.92
Cameraman	1	183.83	288.33
	2	294.62	327.44
	3	344.16	387.06
	5	423.73	453.87
	10	522.30	553.55

as the l_2-norm between two signals. We say that the signal is more redundant or more *correlated* as the distance gets smaller. Let a signal $\mathbf{I}^{(j)}$ be a part of the image in Figure 3.1. For examples, each row in Figure 3.2 can be a signal $\mathbf{I}^{(j)}$, where j denotes the index of the row. The distance between two parts $\mathbf{I}^{(j)}$ and $\mathbf{I}^{(k)}$, for any j, k within the boundary of the image is

$$\|\mathbf{I}^{(j)} - \mathbf{I}^{(k)}\|_2 = \sqrt{\sum_i |(I_i^{(j)} - I_i^{(k)})^2|}, \tag{3.1}$$

where $I_i^{(j)}$ denotes each pixel of $\mathbf{I}^{(j)}$.

Table 3.1 shows that, indeed, the distance between neighbouring parts, *i.e.*, between $\mathbf{I}^{(j)}$ and $\mathbf{I}^{(j-1)}$ is smaller than those further away. It also shows that this spatial redundancy between neighbours is observable both between rows and columns.

A more obvious type of redundancy is the *temporal redundancy*, which

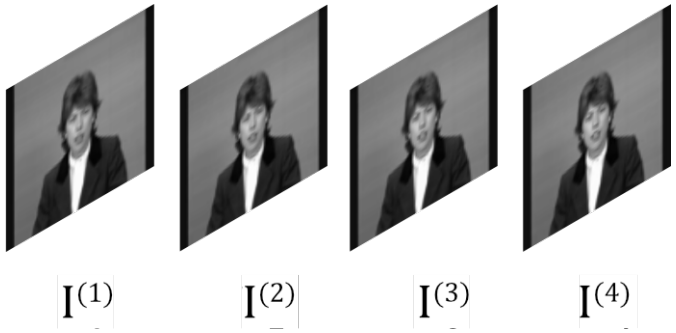


Figure 3.4: Example of temporal redundancy in a video sequence

is the redundant information between each time instance of a temporal sequence (such as a video sequence). Consider a video sequence in Figure 3.4, it is clear that the difference between each frame is very small compared to the whole frame. By exploiting this observation, the temporal redundancy leads to a dramatic result in the inter-frame compression, where only the information regarding the difference is kept most of the time.

Again, we can employ the same analysis we did for the spatial redundancy. Let $\mathbf{I}^{(t)}$ denotes each frame of the video sequence in Figure 3.4 at time instance t. We can compute the Euclidean distance between any two frames $\mathbf{I}^{(t)}$ and $\mathbf{I}^{(q)}$ as

$$\|\mathbf{I}^{(t)} - \mathbf{I}^{(q)}\|_2 = \sqrt{\sum_i |(I_i^{(t)} - I_i^{(q)})^2|}, \tag{3.2}$$

where $I_i^{(t)}$ denotes each pixel of $\mathbf{I}^{(t)}$.

Table 3.2 shows that the relationship between the temporal redundancy and the neighbour distance follows the same pattern as of spatial redundancy. The adjacent neighbours (*e.g.*$\mathbf{I}^{(t)}$ and $\mathbf{I}^{(t-1)}$) are indeed closer than the further ones.

Suppose that we want to perform a compressive sensing reconstruction of the frame $\mathbf{I}^{(t)}$. The signal $\mathbf{I}^{(t)}$, having n elements in total, can

Table 3.2: Average Euclidean distance between $\mathbf{I}^{(t)}$ and $\mathbf{I}^{(t-l)}$

Sequence	l	Distance
Claire	1	352.89
	2	621.35
	3	842.22
	5	1141.06
	10	1576.83
Coast guard	1	1482.23
	2	2285.91
	3	2697.42
	5	3193.51
	10	3912.04
News	1	837.07
	2	1333.93
	3	1706.59
	5	2210.63
	10	2837.14

Table 3.3: Average l_1-norm of $\mathbf{I}^{(t)}$ and average l_1-norm of $\mathbf{I}^{(t)} - \mathbf{I}^{(t-1)}$

Sample sequence	$\|\mathbf{I}^{(t)}\|_1$	$\|\mathbf{I}^{(t)} - \mathbf{I}^{(t-1)}\|_1$
Akiyo	2.37×10^6	1.90×10^4
Claire	2.44×10^6	2.12×10^4
Coast guard	2.88×10^6	1.37×10^5
Hall	3.33×10^6	4.39×10^4
News	2.13×10^6	3.78×10^4

be viewed as a point in n-dimensional space. The traditional l_1-norm objective function, $\|\hat{\mathbf{I}}^{(t)}\|_1$, can be viewed as different between the reconstructed signal $\hat{\mathbf{I}}^{(t)}$ and the origin. In other words, we can express the l_1-norm objective function as

$$\|\hat{\mathbf{I}}^{(t)}\|_1 = \|\hat{\mathbf{I}}^{(t)} - \mathbf{0}\|_1, \tag{3.3}$$

with $\mathbf{0}$ denotes an all-zero vector of length n. We can see in Table 3.3 that, for any frame $\mathbf{I}^{(t)}$, the l_1-norm of the difference between $\mathbf{I}^{(t)}$ and its neighbour $\mathbf{I}^{(t-1)}$ is much smaller than the l_1-norm of $\mathbf{I}^{(t)}$ itself.

Geometrically, we can say that $\mathbf{I}^{(t-1)}$ is closer to $\mathbf{I}^{(t)}$ than the ori-

gin. The difference between two signal is a much precise estimation of the location of $\mathbf{I}^{(t)}$ than the l_1-norm, thus this prior knowledge can be served as *a priori* information for the reconstruction algorithm. This is a motivation of many compressive sensing reconstruction algorithms based on prior knowledge. Most notably, the Temporal Compressive Sensing [167, 166].

3.2 The Proposed Method

Here we propose a novel compressive sensing reconstruction method that exploits the mentioned redundancies in signals. This method, named *Referenced Compressive Sensing* (Referenced CS), is a broad framework on the changing the basic l_1-minimisation problem into a minimisation of error between a signal and another signal. First, let us define the related symbols. A large signal, such as images and video sequences, can be viewed as a collection C consists of several smaller sub-signals.

Definition 5. *Any* $\mathbf{x}^{(j)} \in \mathbb{R}^n, j = 1, 2, ..., k$ *is a sub-signal in the collection* C*, where* n *is the length of each signal* $\mathbf{x}^{(j)}$ *and* k *is the total number of* $\mathbf{x}$ *in* C*.*

From *Definition 5*, we can define a notion of a *correlated reference*.

Definition 6. *For any signal* $\mathbf{x} \in \mathbb{R}^n$*, a correlated reference* $\mathbf{r}$ *of* $\mathbf{x}$ *is a signal such that* $\mathbf{r} \in \mathbb{R}^n$ *and*

$$\|\mathbf{r} - \mathbf{x}\|_2 \leq \epsilon, \tag{3.4}$$

for a small $0 < \epsilon \ll \|\mathbf{x}\|_2$.

Definition 6 shows that the correlated reference $\mathbf{r}$ can, in fact, be

any arbitrary signal close to $\mathbf{x}$. This is one of the novel features of the Referenced CS that the correlated reference is not required to be a member of the same collection $\mathbf{x}$ belongs, *i.e.*, the condition $\mathbf{r} \in C$ is not necessary. The distance between the reference signal $\mathbf{r}$ and the signal $\mathbf{x}$ in *Definition 6*, denoted $\delta = \|\mathbf{r} - \mathbf{x}\|_2$, has a very important role in Referenced CS. With this, we propose the following proposition:

Proposition 1. *Given a sensing operator* $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$*, a compressive measurement* $\mathbf{y} \in \mathbb{R}^m, \mathbf{y} = \mathbf{\Phi x}$*, and a correlated reference* $\mathbf{r} \in \mathbb{R}^n$*, an orthogonal projection* $\hat{\mathbf{x}}_P$ *from* $\mathbf{r}$ *onto the feasible subspace* $\hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}} = \{\hat{\mathbf{x}} | \mathbf{y} = \mathbf{\Phi x}\}$ *satisfies*

$$\sup_{\hat{\mathbf{x}}_P} \|\hat{\mathbf{x}}_P - \mathbf{x}\|_2 \leq \delta. \tag{3.5}$$

Proof. Consider the case of $\mathbf{x} \in \mathbb{R}^2$ in Figure 3.5. Let a reference $\mathbf{r}$ be a member of a set of references R such that $R = \{\mathbf{r} | \|\mathbf{r} - \mathbf{x}\|_2 \leq \delta\}$. Also define an orthogonal projection $\hat{\mathbf{x}}_P$, which is the projection of the reference $\mathbf{r} \in R$ onto the feasible subspace $\hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}}$.

Let $L = \|\hat{\mathbf{x}}_P - \mathbf{r}\|_2$ be the length of the projection. One finds that

$$L \leq \delta \sin\theta, \tag{3.6}$$

where θ is the angle between the vectors $\mathbf{r} - \mathbf{x}$ and $\hat{\mathbf{x}}_P - \mathbf{x}$. Because of the relationship

$$\|\hat{\mathbf{x}}_P - \mathbf{x}\|_2^2 = \delta^2 - L^2, \tag{3.7}$$

we can see that $\|\hat{\mathbf{x}}_P - \mathbf{x}\|_2$ is maximised when $L = 0$, *i.e.*, $\theta = 0$. This implies that, for any $\mathbf{r} \in R$,

$$\sup_{\hat{\mathbf{x}}_P} \|\hat{\mathbf{x}}_P - \mathbf{x}\|_2 \leq \delta. \tag{3.8}$$

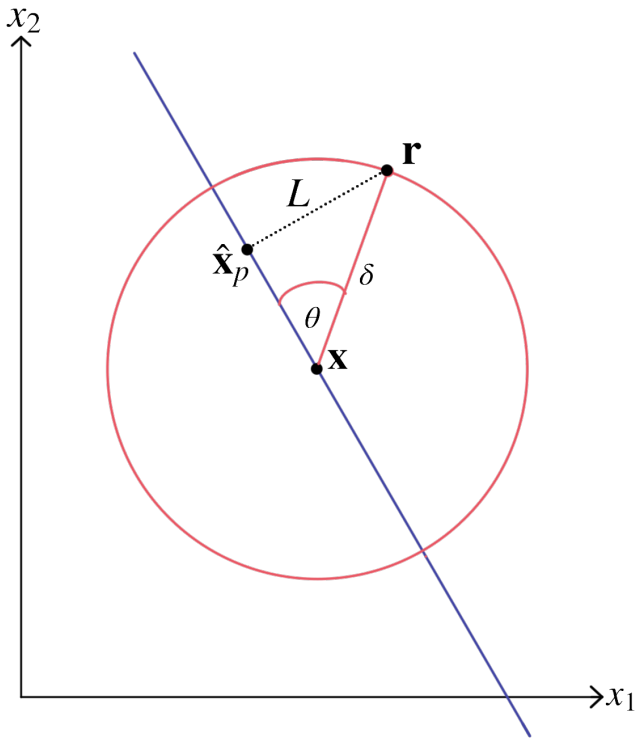


Figure 3.5: Geometric example of *Proposition 1* where $\mathbf{x} \in \mathbb{R}^2$.

□

Proposition 1 leads to the main finding for our proposed reconstruction method. We can find the solution $\hat{\mathbf{x}}_1$ using l_1-norm minimisation instead of using the orthogonal projection. In this case, unlike $\hat{\mathbf{x}}_P$, the sparsity of $\hat{\mathbf{x}}_1$ is promoted.

Proposition 2. *Given a sensing operator* $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$*, a compressive measurement* $\mathbf{y} \in \mathbb{R}^m, \mathbf{y} = \mathbf{\Phi x}$*, and a correlated reference* $\mathbf{r} \in \mathbb{R}^n$*, a least* l_1*-norm reconstruction* $\hat{\mathbf{x}}_1$*, which is the solution of*

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}} - \mathbf{r}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}, \tag{3.9}$$

satisfies

$$\|\hat{\mathbf{x}}_1 - \mathbf{x}\|_2 \leq 2\delta. \tag{3.10}$$

Proof. Consider the case of $\mathbf{x} \in \mathbb{R}^2$ in Figure 3.6. Let a reference $\mathbf{r}$ be a member of a set of references R such that $R = \{\mathbf{r} | \|\mathbf{r} - \mathbf{x}\|_2 \leq \delta\}$. The least l_1-norm solution $\hat{\mathbf{x}}_1$ is a point on the feasible subspace $\hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}} = \{\hat{\mathbf{x}} | \mathbf{y} = \mathbf{\Phi}\mathbf{x}\}$ such that the norm $\|\hat{\mathbf{x}}_1 - \mathbf{r}\|_1$ is minimised. Define the angle between vector $\hat{\mathbf{x}}_1 - \mathbf{r}$ and $\hat{\mathbf{x}}_P - \mathbf{r}$ as ρ. Both $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_P$ are on the feasible subspace $\hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}}$, because $\hat{\mathbf{x}}_P$ is an orthogonal projection, it is clear that

$$\|\hat{\mathbf{x}}_P - \mathbf{r}\|_2 \leq \|\hat{\mathbf{x}}_1 - \mathbf{r}\|_2. \tag{3.11}$$

From Figure 3.6, it can be seen that

$$\|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_P\|_2 = \|\hat{\mathbf{x}}_P - \mathbf{r}\|_2 \tan\rho. \tag{3.12}$$

From Proposition 1, we know that $\|\hat{\mathbf{x}}_P - \mathbf{r}\|_2 \leq \delta$. Also, from Figure 3.6, it can be seen that $\rho \leq \frac{\pi}{4}$ for $\hat{\mathbf{x}}_1$ to be the smallest l_1-norm solution, thus,

$$\|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_P\|_2 \leq \delta. \tag{3.13}$$

This implies that, for any $\mathbf{r} \in R$,

$$\sup_{\hat{\mathbf{x}}_1} \|\hat{\mathbf{x}}_1 - \mathbf{x}\|_2 = \|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_P\|_2 + \|\hat{\mathbf{x}}_P - \hat{\mathbf{x}}\|_2, \tag{3.14}$$

$$\leq 2\delta. \tag{3.15}$$

□

Proposition 2 shows that the error of the least l_1-norm reconstruction $\hat{\mathbf{x}}_1$ depends on the distance δ from the reference $\mathbf{r}$ to $\mathbf{x}$. Moreover, if the

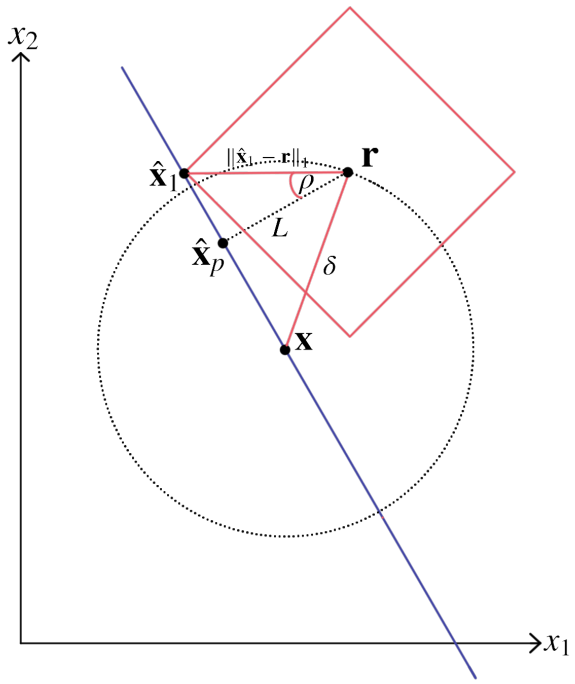


Figure 3.6: Geometric example of *Proposition 2* where $\mathbf{x} \in \mathbb{R}^2$.

sensing operation $\mathfrak{S}(\mathbf{x}) = \mathbf{\Phi x}$ satisfies the conditions of the compressed sensing matrix (Section 2.2.3), then the *Lemma 1*—asserted by Donoho in [52]—holds.

Lemma 1. *[52] Any* $\hat{\mathbf{x}}_c \in \hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}}$ *that is a solution of a general* l_p*-norm minimisation problem*

$$\min_{\mathbf{\Phi}\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_p \text{ subject to } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}} \tag{3.16}$$

satisfies

$$\sup \|\mathbf{x} - \hat{\mathbf{x}}_c\|_2 \leq 2\mathfrak{R}_c(\mathbf{\Phi}, \hat{\mathbf{X}}_{\mathbf{\Phi},\mathbf{y}}), \tag{3.17}$$

where $\mathfrak{R}_c(\mathbf{\Phi}, \hat{\mathbf{X}}_{\Phi,\mathbf{y}})$ *is the optimal result obtained from the central algorithm [128, 155].*

Lemma 1 is asserted by employing the concept of central algorithm, which is the theoretical algorithm that yields the closest possible optimal solution according to the fields of Optimal Recovery (OR) [128] and Information-Based Complexity (IBC) [155]. Unfortunately, as Donoho suggested, the central algorithm is intractable in practice. Nonetheless, this concept of the central algorithm from OR/IBC is the main tool Donoho used to derive the original compressed sensing reconstruction method.

Lemma 1 shows that any least l_1-norm solution must be within the factor of two from the theoretical optimal solution. Eq. (3.15) together with *Lemma 1* suggest that the Referenced CS solution from Eq. (3.9) must also be close to the optimal solution, with the exact solution distance depending on the reference distance δ.

3.2.1 Comparison with Temporal Compressive Sensing

Referenced Compressive Sensing find the best reconstruction by solving the optimisation problem in Eq. (3.9). This can be seen as a minimisation of error between 2 signals, in this case, the reconstructed signal $\hat{\mathbf{x}}$ and the correlated reference $\mathbf{r}$. This approach is very similar to many reconstruction algorithms that employ side-information, particularly the TCS[167].

The TCS is a reconstruction method specifically designed for temporal signal reconstruction, as its name suggests. This means for the TCS, the collection C is a temporal signal such as a video sequence and each sub-signal $\mathbf{x}^{(t)} \in C$ is a progression in the temporal axis at each time instance t. The TCS reconstructs each sub-signal $\hat{\mathbf{x}}^{(t)}$ from compressed measurements by solving the following optimisation problem:

$$\min_{\hat{\mathbf{x}}^{(t)}} \|\boldsymbol{\Psi}\hat{\mathbf{x}}^{(t)}\|_1 + \|\hat{\mathbf{x}}^{(t)} - \mathbf{x}^{(t-1)}\|_1 \text{ subject to } \boldsymbol{\Phi}\mathbf{x}^{(t)} = \mathbf{y}. \tag{3.18}$$

The two main differences between Referenced CS and the TCS are as follows:

1. The first term of the TCS objective function promotes the sparsity of $\mathbf{x}^{(t)}$. In Referenced CS, there is no explicit notion of maximising the sparsity of $\mathbf{x}$ itself.

2. The TCS requires another signal from within the same collection C as *a priori*, the correlated reference $\mathbf{r}$ of Referenced CS is arbitrary.

While Referenced CS is indeed inspired by the TCS and is primarily designed to reconstruct temporal signals, it is not limited only to such setting.

One can also see that Referenced CS can be extended to promote the sparsity of the signal in the same fashion as the TCS. Specifically, one can promote both the sparsity and reference closeness by solving the optimisation problem:

$$\min_{\hat{\mathbf{x}}^{(t)}} \|\mathbf{\Psi}\hat{\mathbf{x}}^{(t)}\|_1 + \|\hat{\mathbf{x}}^{(t)} - \mathbf{r}^{(t)}\|_1 \text{ subject to } \mathbf{\Phi}\mathbf{x}^{(t)} = \mathbf{y}. \tag{3.19}$$

Eq. (3.19) can be converted into the TCS in Eq. (3.18) easily by choosing the immediate neighbour of $\mathbf{x}^{(t)}$ in C, *i.e.*, $\mathbf{x}^{(t-1)}$, as the reference $\mathbf{r}^{(t)}$.

3.3 Relationship between Correlated Reference, Sampling Rate, and Reconstruction Error

The relationship between the correlated reference $\mathbf{r}$, the sampling rate, and the reconstruction error is explored experimentally. As discussed, the reference $\mathbf{r}$ is characterised by the distance from signal $\mathbf{x}$, *i.e.*, $\delta = \|\mathbf{r} - \mathbf{x}\|_2$. The sampling rate s directly affects the sensing operator. Specifically, the sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ has $m = sn$ rows. Lastly, our goal is to minimise the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x}) = \|\hat{\mathbf{x}} - \mathbf{x}\|_2$.

Let us first consider the relationship between δ and $E(\hat{\mathbf{x}}, \mathbf{x})$. Here we employ the Monte Carlo method over the sensing-reconstruction operation. We consider a pool of signal $P = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^{256}\}$ containing 1000 linearly independent random vectors. Each element $x_i \in \mathbf{x}$ ranges between 0 and 255, with the number of non-zero elements fixed to 50. Figure 3.7 shows an example of such vector $\mathbf{x}$ from the pool P.

A set of 20 randomly chosen vectors $P_I \subset P$ is used as an input

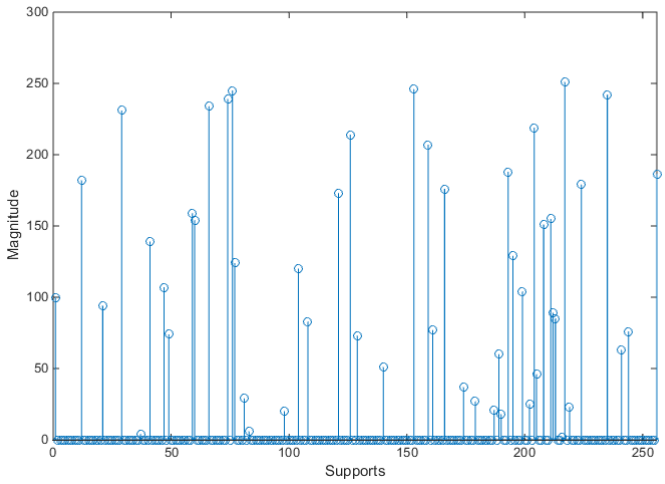


Figure 3.7: Example of a vector $\mathbf{x}$ in the random pool P

signal. For each signal $\mathbf{x} \in P_I$, we apply the sensing and reconstruction operations on it repeatedly, each time using another vector in P as a reference $\mathbf{r}$. In other words, we perform

$$\hat{\mathbf{x}} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}, \mathbf{r}) | \forall \mathbf{x} \in P_I, \forall \mathbf{r} \in P, \mathbf{x} \neq \mathbf{r}, \tag{3.20}$$

where $\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}, \mathbf{r})$ denotes the reconstruction operation. Figure 3.8 shows the scatter plot of each reconstructed signal between δ against $E(\hat{\mathbf{x}}, \mathbf{x})$, along with their linear regression trend. The reconstruction is done using Referenced CS (Eq. (3.9)). It clearly shows that the lower δ indeed leads to the lower reconstruction error. Figure 3.9 shows the same scatter plot, this time the reconstruction also employs the sparsity term (Eq. (3.19)). It can be seen that, given that it is known in advance that $\mathbf{x}$ is sparse, adding the sparsity term greatly improves the reconstruction error. Also, in both figures, a number of reconstruction results are observed to be reconstructed exactly due to the low sparsity of $\mathbf{x}$. It should be noted that the average reconstruction error is far lower than the limit imposed by *Proposition 2*.

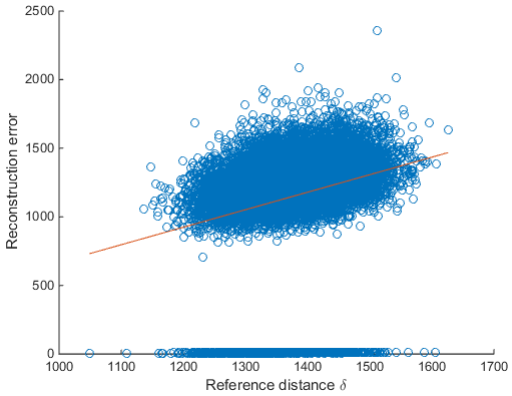


Figure 3.8: A scatter plot shows the relationship between the reference distance δ and the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$of Referenced CS, along with their linear regression.

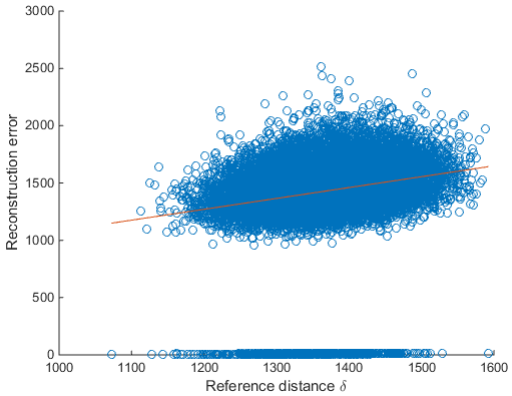


Figure 3.9: A scatter plot shows the relationship between the reference distance δ and the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$of Referenced CS with sparsity term, along with their linear regression.

The random pool P used in this experiment has the exaggeratedly high reference distance δ compared to other settings. This is because both components of each vector, sparse supports, and magnitudes, are completely random. Thus, each signal-reference pair contains no obvious correlation. This is intentional in order to study the case where the reference is completely arbitrary to the target signal. This strategy is, of course, not optimal and may result in a very large reconstruction error.

Now, let us consider the relationship between the sampling rate s and the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$, with and without the presence of the reference $\mathbf{r}$. Again, we employ the Monte Carlo method here using the same pool of random vectors P. This time, every signal $\mathbf{x} \in P$ is used as an input. We apply the sensing and reconstruction operations on each signal repeatedly, each time using a different sampling rate s in the range from 0.1 to 0.9. The reference $\mathbf{r}$ is created by adding random white noise on the input signal. Specifically, for each $\mathbf{x}$, we create a reference signal

$$\mathbf{r} = \mathbf{x} + k\mathbf{n}, \tag{3.21}$$

where $\mathbf{n}$ is a uniform white noise, such that

$$\|\mathbf{r} - \mathbf{x}\|_2 = k\|\mathbf{x}\|_2. \tag{3.22}$$

In this experiment, three values of k are chosen, resulting in three distances: $\delta_L = 0.25\|\mathbf{x}\|_2$, $\delta_M = 0.5\|\mathbf{x}\|_2$, and $\delta_H = 0.75\|\mathbf{x}\|_2$.

Figure 3.10 and Figure 3.11 show the average plot between the sampling rate s and the error $E(\hat{\mathbf{x}}, \mathbf{x})$. In Figure 3.10, the reconstruction is done using Referenced CS without the sparsity term (Eq. (3.9)) As one would expect, the error goes up as the sampling rate decreases. However,

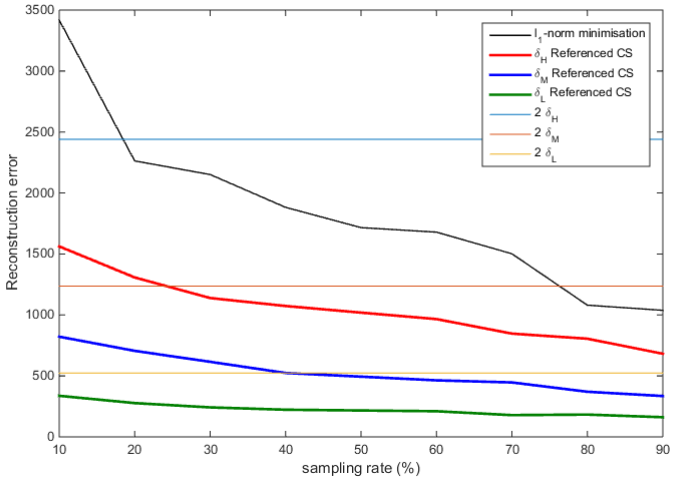


Figure 3.10: Relationship between the sampling rate s, the reference distance δ, and the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$of Referenced CS

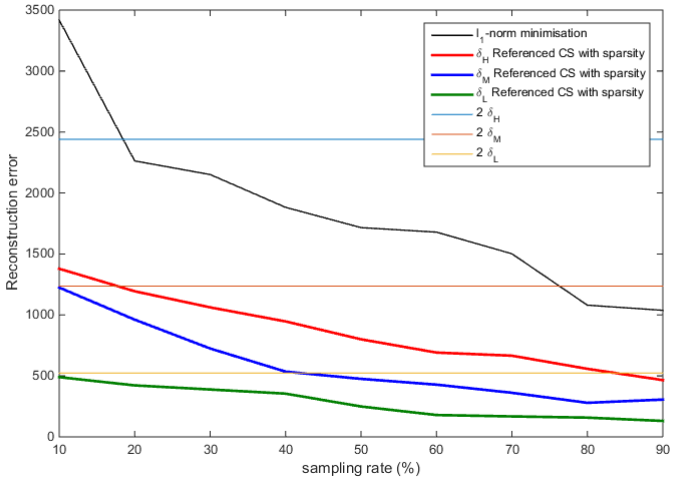


Figure 3.11: Relationship between the sampling rate s, the reference distance δ, and the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$of Referenced CS with sparsity term

the error also reduces as the distance δ decreases. The small δ–where $\mathbf{r}$ is close to $\mathbf{x}$–clearly results in the lower reconstruction error.

In addition, the difference between the reconstruction error between the low and high sampling rate also reduces along with δ. Let $E_{\delta,s}$ denotes the average reconstruction error when $\mathbf{x}$ is sampled at rate s and is reconstructed using the reference with distance δ. $E_{l_1,s}$ denotes the reconstruction using l_1-norm minimisation, which error difference between two rates $E_{l_1,0.1} - E_{l_1,0.9} = 2379.7$. Using Referenced CS reduces the difference dramatically. For examples, with large reference distance $\delta_H = 0.75\|\mathbf{x}\|_2$, the difference is $E_{\delta_H,0.1} - E_{\delta_H,0.9} = 880.18$. With small reference distance $\delta_L = 0.25\|\mathbf{x}\|_2$, the difference reduces to $E_{\delta_L,0.1} - E_{\delta_L,0.9} = 175.58$. In other words, the lower sampling rate can be reconstructed more accurately if the reference distance δ is low.

Figure 3.11 shows the average reconstruction error, akin to Figure 3.10, but using Referenced CS with the sparsity term (Eq. (3.19)). The same observations as of Figure 3.10 can be made with this figure. Compare Figure 3.11 with Figure 3.10, we can notice that Referenced CS with the sparsity term (Figure 3.11) yields lower error than without the sparsity term when the sampling rate is high, *i.e.*, more than 40 %. When the sampling rate is small—especially at 10–30%—Referenced CS without the sparsity term (Figure 3.10) yields lower error. Moreover, the effect of the sparsity term is more significant when the reference distance δ is higher. Figure 3.10 and Figure 3.11 also show lines indicating the value twice the reference distance (2δ). This is the limit imposed by *Proposition 2*, which is derived for Eq. (3.9). It shows that Referenced CS reconstruction strictly obeys *Proposition 2*. However, by adding the sparsity term as in Eq. (3.19), *Proposition 2* is not held. Thus, Refer-

enced CS is recommended to be employed without the sparsity term, as stated in Eq. (3.9).

3.4 Implementation of Referenced CS

Referenced CS presented in Section 3.2 is a non-linear optimisation problem, much like its l_1-norm minimisation predecessor. As shown in Section 2.2.4, unlike its linear counterpart, the non-linear optimisation problem requires more sophisticated reconstruction algorithm to solve. In this section, a few modifications to the popular l_1-norm minimisation algorithms are demonstrated to make them work with the Referenced CS problem.

3.4.1 Linear programming

The most primitive strategy to solve the l_1-norm minimisation problem is to cast the problem into a linear programming problem, then solve it using standard linear programming solvers such as the Simplex method or the Interior-point method. As discussed in Section 2.2.4, most early compressive sensing reconstruction algorithms work in this fashion.

The l_1-norm minimisation problem,

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}, \tag{3.23}$$

can be written in a parametric form as

$$\begin{aligned} &\min_{\mathbf{x}} \mu\|\mathbf{x}\|_1 + \|\epsilon\|_1 \\ &\text{subject to } \mathbf{\Phi}\mathbf{x} + \epsilon = \mathbf{y}. \end{aligned} \tag{3.24}$$

By expressing $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$ and $\epsilon = \epsilon^+ - \epsilon^-$, where $\mathbf{x}^+, \mathbf{x}^-, \epsilon^+, \epsilon^-$ are non-negative, Eq. (3.24) can be cast into a linear programming as

$$\begin{aligned} \min_{\mathbf{x}^+, \mathbf{x}^-, \epsilon^+, \epsilon^-} \; & \mu \mathbf{1}^{\mathrm{T}}(\mathbf{x}^+ + \mathbf{x}^-) + \mathbf{1}^{\mathrm{T}}(\epsilon^+ + \epsilon^-) \\ \text{subject to } & \mathbf{\Phi}(\mathbf{x}^+ - \mathbf{x}^-) + (\epsilon^+ - \epsilon^-) = \mathbf{y}, \\ & \mathbf{x}^+, \mathbf{x}^-, \epsilon^+, \epsilon^- \geq 0. \end{aligned} \tag{3.25}$$

The Referenced CS replaces the objective function from $\|\mathbf{x}\|_1$ to $\|\mathbf{x} - \mathbf{r}\|_1$. We can cast the Referenced CS problem into a linear programming problem in the same fashion by letting $\mathbf{z} = \mathbf{x} - \mathbf{r}$ and $\mathbf{z} = \mathbf{z}^+ - \mathbf{z}^-$, it follows that Referenced CS can be written in the parametric form as

$$\begin{aligned} \min_{\mathbf{z}} \; & \mu \|\mathbf{z}\|_1 + \|\epsilon\|_1 \\ \text{subject to } & \mathbf{\Phi}\mathbf{z} + \mathbf{\Phi}\mathbf{r} + \epsilon = \mathbf{y}. \end{aligned} \tag{3.26}$$

Thus it can be cast into a linear programming problem as

$$\begin{aligned} \min_{\mathbf{z}^+, \mathbf{z}^-, \epsilon^+, \epsilon^-} \; & \mu \mathbf{1}^{\mathrm{T}}(\mathbf{z}^+ + \mathbf{z}^-) + \mathbf{1}^{\mathrm{T}}(\epsilon^+ + \epsilon^-) \\ \text{subject to } & \mathbf{\Phi}(\mathbf{z}^+ - \mathbf{\Phi}^-) + \mathbf{A}(\mathbf{r}^+ - \mathbf{r}^-) + (\epsilon^+ - \epsilon^-) = \mathbf{y}, \\ & \mathbf{z}^+, \mathbf{z}^-, \mathbf{r}^+, \mathbf{r}^-, \epsilon^+, \epsilon^- \geq 0. \end{aligned} \tag{3.27}$$

The linear programming problem in Eq. (3.27) solves for the solution of Referenced CS in *Proposition 2*. This problem can be solved by any linear programming solvers. This is an easy to use option since there are many solvers available. The standard methods for solving linear programming work fine here, including the simplex method and the interior-point method. L1-Magic toolbox [23] employs the simplex method while

Sparselab [51] employs the interior-point method. These toolboxes, however, hard-coded their solver to solve the l_1-norm minimisation problem specifically and thus are not easy to modify. Many generic linear programming solvers are available, both open source such as GNU Linear Programming Kit (GLPK) or commercial software such as Matlab optimisation toolbox and CVX toolbox [74, 73]. However, generic solvers tend to have a much lower performance and result in a much longer runtime compared to dedicated l_1-norm minimisation solvers.

3.4.2 Projected subgradient Method

Instead of solving the problem using the linear programming, it can be solved using convex optimisation methods directly. One of the most efficient means to solve l_1-norm minimisation problem is the Projected Subgradient method. This method is an extension of the subgradient method, such as the steepest gradient and the conjugate gradient method, commonly used to solve fully-determined linear systems. The projected subgradient, however, is designed to solve a convex optimisation problem. This method performs the basic two steps. Step 1, the signal is updated toward the optimal solution directly using a subgradient – a scaled-down gradient by a dynamic step size λ. The step size in this method, however, is not determined to be an optimal value as in the normal subgradient method. Rather, it is set to a large value initially and decreases over time. This results in the updated signal sometimes get outside the set of feasible solutions defined by the constraints. In step 2, the updated signal is projected back onto the feasible set. In this method, as the step size λ decreases, the signal converges gradually to the optimal solution, thus, the optimal solution $\hat{\mathbf{x}}$ is obtained when $\lambda \to 0$.

One good implementation of this method is the Infeasible-point Subgradient Algorithm (ISA), implementing in ISAL1 solver [102]. Its working follows the steps mentioned above, as it projects the signal toward the optimal objective function using a sub-gradient, follows by the back-projection operation φ using Conjugate Gradient method. *Algorithm 4* summarises ISA, with the objective function $f = \|\mathbf{x}\|_1$ in case of l_1-norm minimisation problem. The complete implementation details, such as the setting of each parameter, can be found in [103].

Algorithm 4 Infeasible Subgradient Algorithm (ISA) [103]

Input: a starting point $\mathbf{x}^{(0)}$, objective function f, estimated optimal objective φ
Output: an approximate solution to 3.9

initialise $k := 0$, $\mathbf{x}^{(1)} := \mathbf{x}^{(0)}$, subgradient $\mathbf{h}^{(1)} := 0$, step size $\lambda^{(1)} := 0$
repeat
    choose subgradient $h^{(k)} \in \partial f(\mathbf{x}^{(k)})$
    if $f(\mathbf{x}^{(k)}) \leq \varphi$ or $\mathbf{h}^{(k)} = 0$ **then**
        stop (optimal if $\mathbf{h}_k = 0$)
    end if
    compute step size $\lambda^{(k)}$
    compute the next iterate $\mathbf{x}^{(k+1)} := \mathcal{P}(\mathbf{x} - \lambda^{(k)}\mathbf{h}^{(k)})$
    increase $k := k + 1$
until a stopping criterion is satisfied

The ISA algorithm is efficient and accurate. To solve Referenced CS using the ISA algorithm, it can be done easily by replace the objective function term, *i.e.*, replacing $f = \|\mathbf{x}\|_1$ with $f = \|\mathbf{x} - \mathbf{r}\|_1$. If fact, many simple convex objective function (those involve only matrix arithmetic) can be solved by ISA as easily. More complex objective function, such as the total variation-norm minimisation, can also be solved with ISA. However, the modification is no longer trivial. Unless stated otherwise, all results of Referenced CS in this study are obtained using the implementation of the ISA algorithm.

3.4.3 Greedy methods

Other widely used reconstruction algorithms for l_1-norm minimisation problem are the greedy methods. As discussed in Section 2.2.4, greedy methods such as Matching Pursuit, Subspace Pursuit, and CoSaMP do not reconstruct the signal by solving the optimisation problem mathematically. Rather, each method follows its heuristic function during the reconstruction. The heuristic may not directly contribute to the actual objective function.

For example, in the OMP, the algorithm starts from the vector consisting of all zero—the sparsest possible solution. It then tries to fit this vector to the set of feasible solutions as dictated by the constraints. It can be said that OMP works inversely to the convex optimisation-based methods, where the solution starts off within the feasible set and moves to the sparsest solution afterward.

The summary of the OMP is shown in *Algorithm 1* (Section 2.2.4). Even though the OMP does not have an explicit objective function in the algorithm, we can modify OMP to work with Referenced CS easily. As the OMP minimises $\|\mathbf{x}\|_1$ by setting the initial solution $\mathbf{x}^{(0)} = \mathbf{0}$, where $\mathbf{0}$ is an all-zero vector—the smallest l_1-norm possible—we can minimises $\|\mathbf{x} - \mathbf{r}\|_1$ by setting the initial solution $\mathbf{x}^{(0)} = \mathbf{r}$. Thereby makes $\mathbf{x}^{(0)}$ to be closest to the reference $\mathbf{r}$, *i.e.*, $\|\mathbf{x}^{(0)} - \mathbf{r}\|_1 = 0$. It is also possible to modify other greedy methods, such as CoSaMP, in a similar manner to work with Referenced CS.

3.5 Experimental Results

This section shows the experimental results of the Referenced CS reconstruction in various scenarios. There are three scenarios considered here: the row-based reconstruction of images, the block-based reconstruction of images, and the video sequence reconstruction. The first two scenarios exploit the spatial redundancy within an image while the last one exploits the temporal redundancy between video frames.

3.5.1 Row-based Image Reconstruction

The first scenario exploits the spatial redundancy between each row of an image. In this scenario, we show that Referenced CS can outperform the l_1-norm minimisation and can reconstruct images in a row-by-row fashion. This scenario is a proof-of-concept to demonstrate the fact that Referenced CS, by default, does not employ the sparsity information and can be applied to both sparse and dense signal reconstruction.

Following the convention used previously, we can regard an image as a collection of sub-signals $\mathbf{I}^{(j)}$, $j = 0, ..., n$, each $\mathbf{I}^{(j)}$ represents each row of the image, where n is the number of row presents. The compressed measurements $\mathbf{y}^{(j)}$ are obtained for each $\mathbf{I}^{(j)}$ using a random Gaussian sensing matrix $\mathbf{\Phi}$. The sampling rates s used here are set to 0.25, 0.5, and 0.75.

A set of test images used in this experiment is shown in Figure 3.12.

Firstly, let us consider the reconstruction results when the sensing and reconstruction operations are applied to the sparse domain. In this case, the sensing operation is

$$\mathbf{y}^{(j)} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{I}^{(j)}, \tag{3.28}$$

(a) Aerial (b) Baboon (c) Barbara (d) Boat

(e) Cameraman (f) Goldhill (g) Lena (h) Peppers

Figure 3.12: Test images

where $\mathbf{\Psi}$ is the discrete cosine transform basis and $\mathbf{\Phi}$ is the random Gaussian sensing matrix. The reconstruction is obtained via the reconstruction operation

$$\hat{\mathbf{x}}^{(j)} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}^{(j)}, \mathbf{r}^{(j)}), \tag{3.29}$$

$$\hat{\mathbf{I}}^{(j)} = \mathbf{\Psi}^{-1}\hat{\mathbf{x}}^{(j)}, \tag{3.30}$$

where $\mathbf{r}^{(j)} = \hat{\mathbf{x}}^{(j-1)}$, and the operation $\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}, \mathbf{r})$ is either Referenced CS (Eq. (3.9)) or the l_1-norm minimisation. Table 3.4 compares the Structural Similarity Measure (SSIM) between the results from Referenced CS and the l_1-norm minimisation. It can be seen that Referenced CS outperforms the l_1-norm minimisation. At the sampling rate of 50%, Referenced CS yields on average 32% higher SSIM index than the l_1-norm minimisation.

Even more interesting is Table 3.5 which shows the comparison be-

Table 3.4: SSIM comparison between l_1-norm minimisation results and Referenced CS results on sparse domain

Test image	Sampling rate s	l_1-min SSIM	RefCS SSIM
Aerial	0.1	0.067	0.075
	0.2	0.108	0.110
	0.3	0.138	0.178
	0.4	0.165	0.229
	0.5	0.195	0.283
Baboon	0.1	0.056	0.035
	0.2	0.079	0.109
	0.3	0.089	0.129
	0.4	0.126	0.150
	0.5	0.191	0.213
Barbara	0.1	0.045	0.064
	0.2	0.088	0.122
	0.3	0.125	0.156
	0.4	0.156	0.153
	0.5	0.201	0.260
Boat	0.1	0.076	0.068
	0.2	0.111	0.107
	0.3	0.132	0.155
	0.4	0.146	0.204
	0.5	0.199	0.235
Cameraman	0.1	0.062	0.067
	0.2	0.091	0.109
	0.3	0.113	0.155
	0.4	0.141	0.191
	0.5	0.123	0.217
Goldhill	0.1	0.060	0.055
	0.2	0.076	0.107
	0.3	0.142	0.125
	0.4	0.142	0.159
	0.5	0.151	0.186
Lena	0.1	0.068	0.044
	0.2	0.152	0.115
	0.3	0.150	0.168
	0.4	0.175	0.189
	0.5	0.211	0.241
Peppers	0.1	0.071	0.059
	0.2	0.101	0.119
	0.3	0.129	0.172
	0.4	0.184	0.235
	0.5	0.233	0.283

Table 3.5: Average norms and distances for image row reconstruction in sparse domain

Test image	$\|\mathbf{r}-\mathbf{x}\|_2$	$\|\mathbf{x}\|_2$	$E(\hat{\mathbf{x}},\mathbf{x})$	$\|\mathbf{r}-\mathbf{x}\|_1$	$\|\mathbf{x}\|_1$
Aerial	597.49	1547.70	607.11	10171.00	5218.80
Baboon	579.01	1505.50	577.45	9339.10	5112.30
Barbara	601.69	1355.00	595.16	8263.50	5027.81
Boat	616.71	1529.30	1125.80	9183.22	5239.19
Cameraman	655.66	1484.43	625.59	8493.30	5010.40
Goldhill	793.87	1324.31	1313.70	8015.61	6068.80
Lena	592.37	1245.80	586.95	7531.60	4987.53
Peppers	600.24	1333.42	583.77	7805.90	4923.22

tween the reference distance $\delta = \|\mathbf{r}-\mathbf{x}\|_2$, the origin distance $\|\mathbf{x}\|_2$, and the reconstruction error $E(\hat{\mathbf{x}},\mathbf{x})= \|\hat{\mathbf{x}}-\mathbf{x}\|_2$. It can be seen that the reconstruction errors of Referenced CS are indeed less than twice the size of δ, as predicted in *Proposition 2*. Table 3.5 also shows that the l_1 distance of $\|\mathbf{r}-\mathbf{x}\|_1$ in this scenario is smaller than the l_1-norm of $\mathbf{x}$ itself, making it a better objective function.

This effect is very clear when the reconstruction is applied to the non-sparse signal. Let us consider another scenario when the sensing and reconstruction operations are applied instead to the dense domain. The sensing operation in this case is

$$\mathbf{y}^{(j)} = \mathbf{\Phi}\mathbf{I}^{(j)}, \tag{3.31}$$

without any sparsifying bases. This time the reconstruction operation yields the reconstruction directly, *i.e.*,

$$\hat{\mathbf{I}}^{(j)} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}^{(j)}, \mathbf{r}^{(j)}). \tag{3.32}$$

In other words, this time the sensing and reconstruction are done to the non-sparse signal directly in its spatial domain. The same reconstruction

operator in Eq. (3.29) is employed here. Table 3.6 compares the SSIM between the Referenced CS reconstruction and the l_1-norm minimisation reconstruction. It comes at no surprise that the l_1-norm minimisation fails completely to reconstruct any image. Referenced CS, however, successfully reconstructs more structure of the image, albeit with lower accuracy than the sparse version. The difference of SSIM index between two algorithms is larger compared to the sparse reconstruction, with 86% higher SSIM index on average for Referenced CS. Table 3.7 shows the reference distance δ, $\|\mathbf{x}\|_2$, and the error $E(\hat{\mathbf{x}}, \mathbf{x})$. This time, we can see that the reference is much closer to $\mathbf{x}$ than the origin. Again, Table 3.7 shows that the norm $\|\mathbf{r} - \mathbf{x}\|_1$ is much smaller than $\|\mathbf{x}\|_1$.

3.5.2 Block-based Image Reconstruction

This scenario is similar to the previous one as it also deals with the image reconstruction. In this case, the spatial redundancy exploited comes from each small block in the image instead of its rows. Again, we regard an image as a collection of sub-signals $\mathbf{I}^{(j,k)}$. Let $\mathbf{I}^{(j,k)}$ denotes a block at the row j and column k of the image, and $\mathbf{i}^{(j,k)}$ is a vectorised form of a block $\mathbf{I}^{(j,k)}$. As shown previously in Section 3.1, these blocks are highly correlated to their neighbours, thus, they are good candidates for being used as references.

The sensing operation is applied to each block individually, *i.e.*, the compressed measurements $\mathbf{y}^{(j,k)}$ are obtained from

$$\mathbf{y}^{(j,k)} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{i}^{(j,k)}. \tag{3.33}$$

Table 3.6: SSIM comparison between l_1-norm minimisation results and Referenced CS results on spatial domain

Test image	Sampling rate s	l_1-min SSIM	RefCS SSIM
Aerial	0.1	0.066	0.074
	0.2	0.091	0.115
	0.3	0.101	0.155
	0.4	0.118	0.229
	0.5	0.169	0.266
Baboon	0.1	0.065	0.053
	0.2	0.069	0.078
	0.3	0.062	0.132
	0.4	0.097	0.148
	0.5	0.109	0.229
Barbara	0.1	0.082	0.053
	0.2	0.078	0.093
	0.3	0.104	0.163
	0.4	0.104	0.203
	0.5	0.106	0.233
Boat	0.1	0.044	0.055
	0.2	0.040	0.099
	0.3	0.084	0.124
	0.4	0.095	0.158
	0.5	0.125	0.176
Cameraman	0.1	0.057	0.053
	0.2	0.091	0.094
	0.3	0.102	0.155
	0.4	0.093	0.158
	0.5	0.106	0.177
Goldhill	0.1	0.051	0.058
	0.2	0.065	0.105
	0.3	0.080	0.101
	0.4	0.092	0.141
	0.5	0.098	0.157
Lena	0.1	0.037	0.058
	0.2	0.105	0.134
	0.3	0.116	0.141
	0.4	0.127	0.242
	0.5	0.170	0.263
Peppers	0.1	0.066	0.049
	0.2	0.087	0.136
	0.3	0.128	0.183
	0.4	0.129	0.161
	0.5	0.087	0.245

Table 3.7: Average norms and distances for image row reconstruction in spatial domain

Test image	$\|\mathbf{r}-\mathbf{x}\|_2$	$\|\mathbf{x}\|_2$	$E(\hat{\mathbf{x}},\mathbf{x})$	$\|\mathbf{r}-\mathbf{x}\|_1$	$\|\mathbf{x}\|_1$
Aerial	606.18	1547.70	620.39	16907	5554.20
Baboon	536.79	1508.50	541.79	16533	4880.50
Barbara	640.88	1355.05	644.35	14394	5992.71
Boat	787.01	1529.31	796.64	16603	7808.70
Cameraman	717.19	1484.42	1208.80	15197	6622.40
Goldhill	792.99	1324.31	786.33	14364	7727.21
Lena	678.65	1245.80	677.99	12805	6459.12
Peppers	618.62	1333.42	626.17	13348	5649.41

The reconstruction

$$\hat{\mathbf{x}}^{(j,k)} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}^{(j,k)}, \mathbf{r}^{(j,k)}), \tag{3.34}$$

$$\hat{\mathbf{i}}^{(j,k)} = \mathbf{\Psi}^{-1}\hat{\mathbf{x}}^{(j,k)}, \tag{3.35}$$

can be obtained using the Referenced CS or the l_1-norm minimisation as the reconstruction operation. The previous block on the same row is used as a reference, *i.e.*, $\mathbf{r}^{(j,k)} = \hat{\mathbf{x}}^{(j,k-1)}$. The special case is the first column ($k = 1$), which are always reconstructed using the l_1-norm minimisation. Figure 3.13 demonstrates the block-based reconstruction strategy.

Figure 3.14 shows the reconstruction results from using this strategy. This figure, together Table 3.8, shows that the results of Referenced CS contain less noise than the results from l_1-norm minimisation. The same analysis we did in Section 3.5.1 can be employed here. On average, the reconstruction from 50% sample using Referenced CS has 18% higher SSIM index. Table 3.9 shows the average reference distance δ, the origin distance $\|\mathbf{x}_{i,j}\|_2$, and the reconstruction error $E(\hat{\mathbf{x}},\mathbf{x})$.

Table 3.8: SSIM comparison between l_1-norm minimisation results and Referenced CS results on sparse domain

Test image	Sampling rate s	l_1-min SSIM	RefCS SSIM
Aerial	0.1	0.157	0.168
	0.2	0.265	0.329
	0.3	0.349	0.439
	0.4	0.439	0.552
	0.5	0.524	0.639
Baboon	0.1	0.171	0.167
	0.2	0.306	0.373
	0.3	0.425	0.507
	0.4	0.522	0.625
	0.5	0.610	0.715
Barbara	0.1	0.111	0.226
	0.2	0.328	0.416
	0.3	0.454	0.544
	0.4	0.557	0.664
	0.5	0.655	0.755
Boat	0.1	0.122	0.185
	0.2	0.249	0.339
	0.3	0.398	0.464
	0.4	0.519	0.610
	0.5	0.613	0.709
Cameraman	0.1	0.119	0.177
	0.2	0.199	0.266
	0.3	0.279	0.371
	0.4	0.353	0.457
	0.5	0.426	0.546
Goldhill	0.1	0.178	0.245
	0.2	0.302	0.359
	0.3	0.422	0.522
	0.4	0.549	0.629
	0.5	0.644	0.745
Lena	0.1	0.177	0.224
	0.2	0.337	0.401
	0.3	0.446	0.528
	0.4	0.543	0.635
	0.5	0.629	0.711
Peppers	0.1	0.124	0.214
	0.2	0.283	0.383
	0.3	0.415	0.504
	0.4	0.501	0.617
	0.5	0.597	0.702

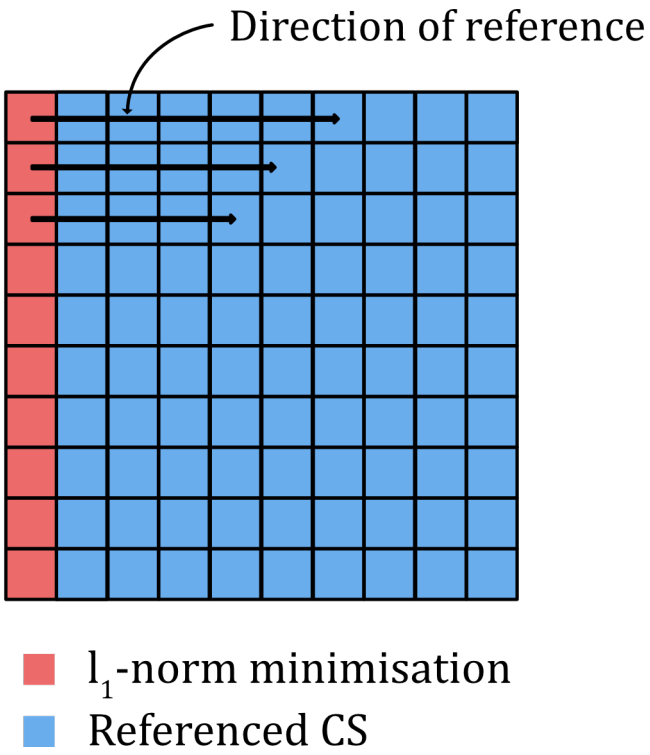
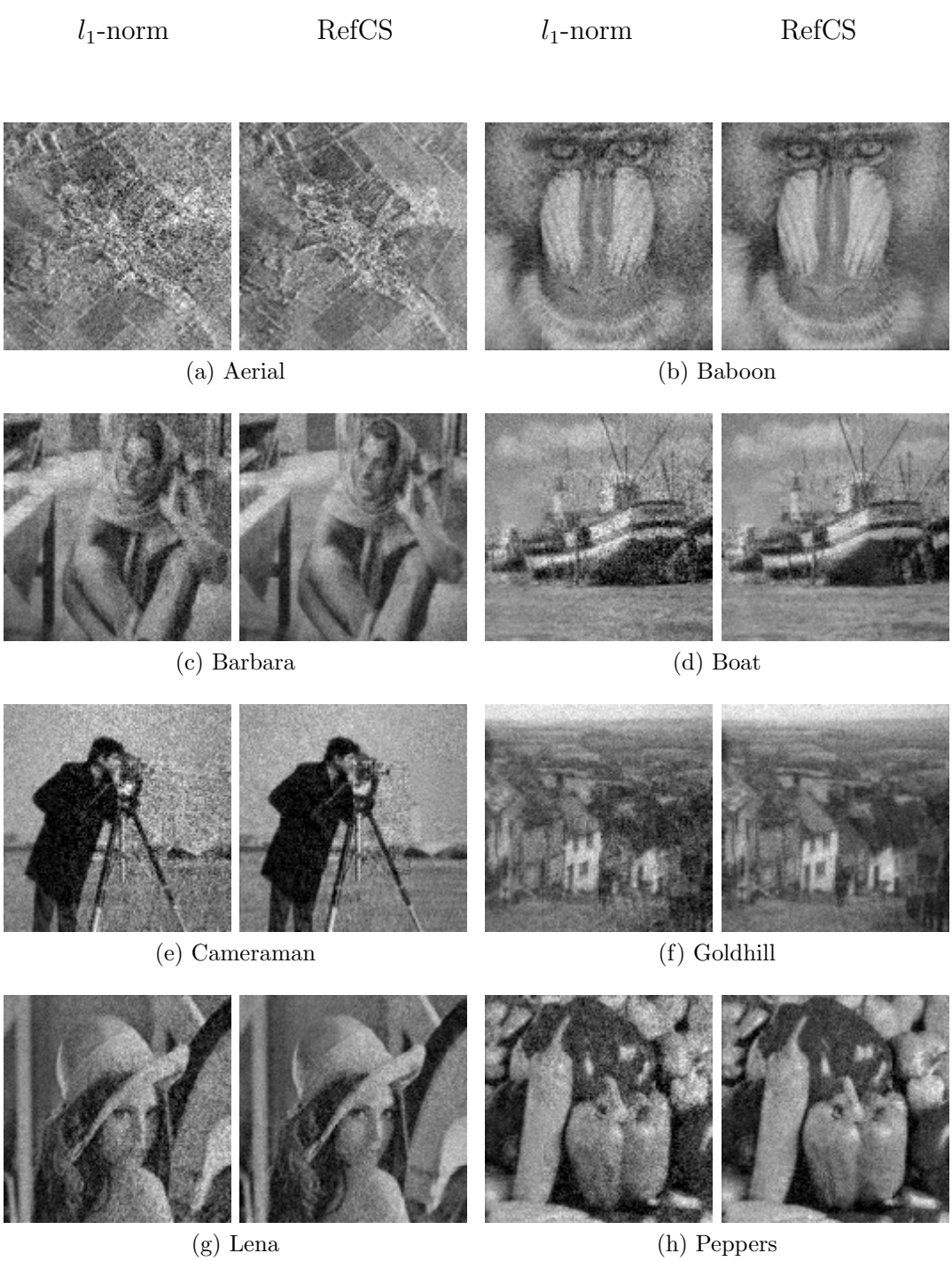


Figure 3.13: Block-based image reconstruction strategy

Table 3.9: Average norms and distances for image block reconstruction in sparse domain

Test image	$\|\mathbf{r}-\mathbf{x}\|_2$	$\|\mathbf{x}\|_2$	$E(\hat{\mathbf{x}},\mathbf{x})$	$\|\mathbf{r}-\mathbf{x}\|_1$	$\|\mathbf{x}\|_1$
Aerial	2.13×10^3	4.36×10^3	854.17	2.55×10^4	2.12×10^4
Baboon	2.11×10^3	4.24×10^3	527.10	1.76×10^4	1.98×10^4
Barbara	2.34×10^3	3.81×10^3	516.02	1.71×10^4	2.18×10^4
Boat	2.26×10^3	4.33×10^3	548.99	1.78×10^4	2.07×10^4
Cameraman	2.93×10^3	4.14×10^3	684.81	1.92×10^4	2.66×10^4
Goldhill	1.96×10^3	3.77×10^3	481.30	1.56×10^4	1.85×10^4
Lena	2.37×10^3	3.48×10^3	519.35	1.76×10^4	2.17×10^4
Peppers	2.73×10^3	3.73×10^3	623.44	2.05×10^4	2.61×10^4



(a) Aerial (b) Baboon

(c) Barbara (d) Boat

(e) Cameraman (f) Goldhill

(g) Lena (h) Peppers

Figure 3.14: Results of the reconstruction in sparse domain, using l_1-norm minimisation and Referenced CS. The reconstruction is done from 50% samples.

3.5.3 Video Sequence Reconstruction

The last scenario to experiment is the reconstruction of video sequences by exploiting their temporal redundancy. This is the most intuitive scenario for Referenced CS because, as discussed in Section 3.1, each frame in a video sequence usually contains a high amount of temporal redundancy to its neighbours. In this scenario, each signal $\mathbf{I}^{(t)}$ is a frame of the video sequence at time t. The compressed measurements $\mathbf{y}^{(t)}$ are obtained for each frame by the sensing operation:

$$\mathbf{y}^{(t)} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{i}^{(t)}, \tag{3.36}$$

where $\mathbf{i}^{(t)}$ is a vectorised form of $\mathbf{I}^{(t)}$. As usual, Referenced CS and the l_1-norm minimisation are used as the reconstruction operation. The reconstructed frame $\hat{\mathbf{i}}^{(t)}$ is obtained by

$$\hat{\mathbf{x}}^{(t)} = \mathfrak{R}(\mathbf{\Phi}, \mathbf{y}^{(t)}, \mathbf{r}^{(t)}), \tag{3.37}$$

$$\hat{\mathbf{i}}^{(t)} = \mathbf{\Psi}^{-1}\hat{\mathbf{x}}^{(t)}, \tag{3.38}$$

where $\mathbf{r}^{(t)} = \hat{\mathbf{x}}^{(t-1)}$. The first frame of the sequence is always reconstructed using the l_1-norm minimisation. Figure 3.15 summarises the video sequence reconstruction strategy used in this experiment.

In this experiment, the sensing-reconstruction is done on 14 standard test sequences, shown in Figure 3.16. These sequences can be seen online at `http://svc.group.shef.ac.uk/shefcswh.html`. These sequences can be grouped into 3 categories: low activity sequences, medium activity sequences, and high activity sequences. Figure 3.17 shows the examples of the reconstructed sequences using the l_1-norm minimisation, while Figure 3.18 shows the reconstructed sequences using Referenced

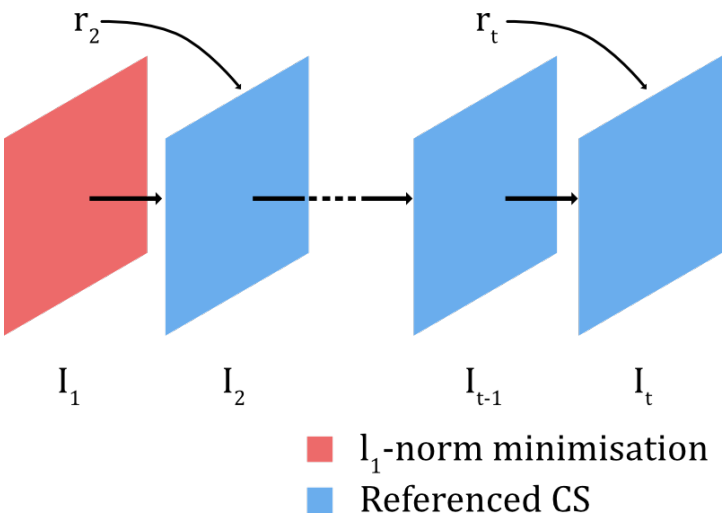


Figure 3.15: Referenced CS strategy for video sequence reconstruction

CS. The visual quality of Referenced CS results is superior to those of l_1-norm minimisation. In particular, Referenced CS results show the reduction of the noise-like reconstruction error which can be observed easily in l_1-norm minimisation results.

It can be noticed in Figure 3.18 that the visual quality of Referenced CS results is proportional to the amount of activity in the sequences. This is because the distance between 2 frames is higher in the high activity sequences that consist of lots of movements, such as the Skate sequence, than that of the low activity ones. In effect, the high activity sequences have larger reference distance δ. This amount of activity does not affect the result of the l_1-norm minimisation, though, as each frame is reconstructed independently from each other.

Figure 3.19 shows the objective comparison of the reconstruction accuracy between both reconstruction methods using PSNR, which shows that Referenced CS results in at least 40% higher in PSNR than the l_1-norm minimisation. The improvement of Referenced CS is largely the suppression of the reconstruction error. More detailed analysis can be seen in Table 3.10, which compares the average reference distance δ,

(a) Akiyo (b) Clair (c) Container

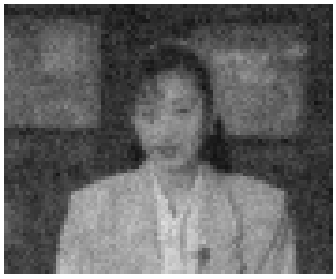
(d) Road (e) MissUSA (f) Car

(g) Coastguard (h) Office (i) Mother

(j) News (k) Salesman (l) Foreman

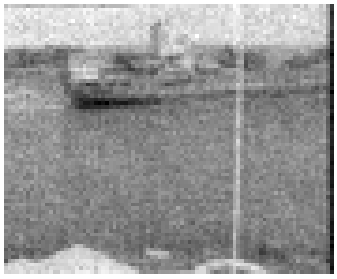
(m) Skate (n) Silence

Figure 3.16: Example frames from the set of test video sequences. (a)–(e) are low activity sequences, (f)–(k) are medium, and (l)–(n) are high activity sequences.



(a) Akiyo

(b) Clair

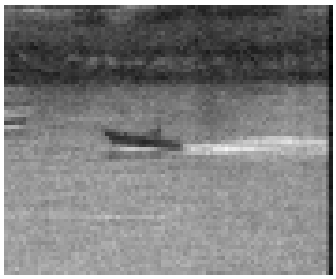


(c) Container

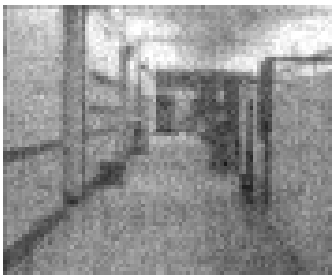
(d) Road

(e) MissUSA

(f) Car

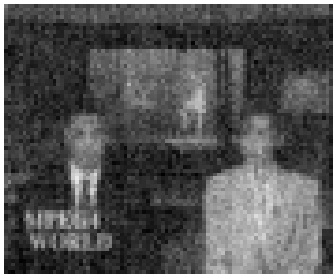


(g) Coastguard



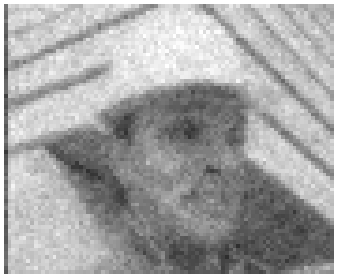
(h) Office

(i) Mother

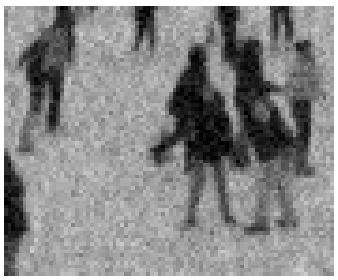


(j) News

(k) Salesman



(l) Foreman



(m) Skate

(n) Silence

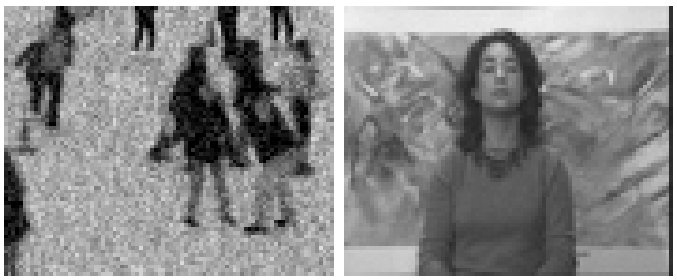
Figure 3.17: Examples of reconstructed video sequences using l_1-norm minimisation from 50% of samples

(a) Akiyo (b) Clair (c) Container

(d) Road (e) MissUSA (f) Car

(g) Coastguard (h) Office (i) Mother

(j) News (k) Salesman (l) Foreman



(m) Skate (n) Silence

Figure 3.18: Examples of reconstructed video sequences using Referenced CS from 50% of samples

Table 3.10: Average norms and distances for video sequence reconstruction

Test image	$\|\mathbf{r}-\mathbf{x}\|_2$	$\|\mathbf{x}\|_2$	$E(\hat{\mathbf{x}},\mathbf{x})$	$\|\mathbf{r}-\mathbf{x}\|_1$	$\|\mathbf{x}\|_1$
Low activity sequence					
Akiyo	9.09×10^2	8.26×10^3	8.00×10^2	5.53×10^4	7.76×10^4
Clair	9.26×10^2	8.79×10^3	7.83×10^2	5.76×10^4	8.95×10^4
Container	11.48×10^2	12.05×10^3	10.31×10^2	6.92×10^4	9.57×10^4
Road	6.79×10^2	14.22×10^3	5.45×10^2	4.22×10^4	6.51×10^4
MissUSA	5.78×10^2	5.49×10^3	4.84×10^2	3.58×10^4	5.24×10^4
Medium activity sequence					
Car	12.19×10^2	9.20×10^3	10.35×10^2	7.42×10^4	10.29×10^4
Coastguard	8.56×10^2	10.07×10^3	7.18×10^2	5.17×10^4	7.32×10^4
Office	11.93×10^2	11.87×10^3	10.53×10^2	7.18×10^4	9.82×10^4
Mother	7.67×10^2	9.63×10^3	6.51×10^2	4.69×10^4	7.16×10^4
News	13.93×10^2	7.16×10^3	12.19×10^2	8.22×10^4	10.14×10^4
Salesman	9.60×10^2	6.29×10^3	8.52×10^2	5.71×10^4	7.28×10^4
High activity sequence					
Foreman	12.81×10^2	13.43×10^3	9.94×10^2	7.69×10^4	10.76×10^4
Skate	26.60×10^2	11.14×10^3	12.87×10^2	12.33×10^4	12.25×10^4
Silence	10.13×10^2	10.81×10^3	8.57×10^2	6.14×10^4	8.96×10^4

origin distance $\|\mathbf{x}^{(t)}\|_2$, and the reconstruction error $E(\hat{\mathbf{x}},\mathbf{x})$of each sequence. It also compares two objective function, $\|\mathbf{x}^{(t)}\|_1$ and $\|\mathbf{r}^{(t)}-\mathbf{x}^{(t)}\|_1$. Table 3.10 shows that, again, the reference distance is much smaller than the sparsity and, thus, is more desirable. The reconstruction error $E(\hat{\mathbf{x}},\mathbf{x})= \|\hat{\mathbf{x}}-\mathbf{x}\|$ is also shown to be less than twice the reference distance δ, satisfying *Proposition 2*.

3.6 Summary

In this chapter, Referenced Compressed Sensing, a novel reconstruction approach for compressed sensing, has been introduced. Referenced CS exploits the redundancy between the signal and an arbitrary correlated reference. It is possible to exploit both the spatial redundancy, with more than 18% improvement in SSIM index for the image reconstruction, and

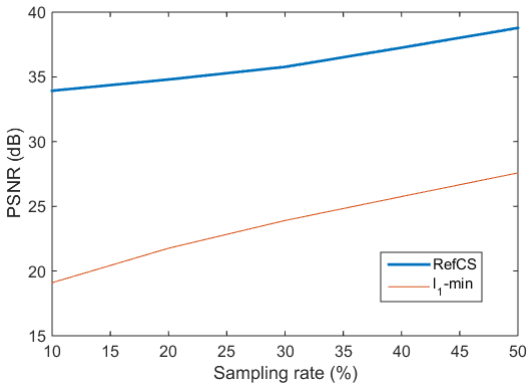


Figure 3.19: Average PSNR across 14 test sequences versus sampling rate s.

the temporal redundancy which results in much less reconstruction error, with at least 50% improvement in PSNR. Referenced CS can be implemented easily from currently available l_1-norm minimisation algorithms. The results in this chapter have shown the proof-of-concept of several possible scenarios where Referenced CS could be employed. However, to employ Referenced CS in practical applications, there are two issues to be considered. Finally, it is shown that the performance of Referenced CS depends largely on the distance between the correlated reference and the signal, thus, it is important to have the reference with the highest correlation. Another issue with Referenced CS is its high computational complexity due to the use of the iterative reconstruction algorithm to minimise the l_1-norm objective function. This complexity prevents the use of Referenced CS in many applications, including the fMRI acquisition, and must be addressed. Both issues that are to be discussed in the next chapter.

Chapter 4

Extensions on Referenced Compressed Sensing

A few complimentary works for Referenced CS (Chapter 3) are presented in this chapter. Firstly, this chapter proposes an estimation strategy for the correlated reference based on Running Gaussian Average (RGA). The effect of the estimation and its specific learning parameter are also studied. Secondly, this chapter presents the low-complexity variation of the Referenced CS reconstruction by substituting the l_1-norm objective function with the l_2-norm. This technique, referred to as Referenced CS with the Least Squares (Referenced CS/LS), demonstrates the fascinating outcome that, by exploiting the correlated reference, it is possible to perform the compressed sensing reconstruction without using iterative algorithms.

4.1 Correlated Reference Estimation using Running Gaussian Average

The first thing to look at is the estimation of the correlated reference $\mathbf{r}$. In Chapter 3, we have discussed the compressed sensing reconstruction that exploits the redundancy between a signal $\mathbf{x}$ and a correlated reference—an arbitrary signal close to the target signal. It also shows that the performance of the proposed reconstruction algorithm depends heavily on the reference distance $\delta = \|\mathbf{r} - \mathbf{x}\|_2$ (defined in *Definition 6*), thus, it is clear that the task of finding a good reference is important.

In this section, a reference estimation strategy designed for the Referenced CS reconstruction of temporal signals is proposed. The strategy uses the concepts of Running Gaussian Average to create a dynamic reference with a high amount of temporal redundancy suitable for Referenced CS.

4.1.1 Motivation and Method

Consider a simple Referenced CS reconstruction of a video sequence in Figure 3.15, where the reference $\mathbf{r}$ is the reconstructed result of the previous frame, *i.e.*, $\mathbf{r}^{(t)} = \hat{\mathbf{x}}^{(t-1)}$. It can be seen that the quality of references gets worse as time progresses. Specifically, the reference distance δ becomes larger over time due to the reconstruction error and content changes. These distorted references results in distorted reconstructions, which in turn result in an even more inaccuracy of reference. This phenomenon is called the error propagation of reconstruction error, where the error from older frames carries onto the next frame. Unless one can guarantee perfect, error-free reconstruction results, using previous results

as references always generate this error propagation issue. Unfortunately, the compressed sensing reconstruction operation is prone to reconstruction error by many factors, such as the limitation of objective function, algorithm performance, and even the implementation and round-up error could result in reconstruction error. Thus, it is clear that procedure to obtain reliable references despite this shortcoming is needed.

First, consider a task of reconstructing a video sequence. Compressive measurements $\mathbf{y}^{(t)} = \mathbf{\Phi\Psi i}^{(t)}$, where $\mathbf{i}^{(t)}$ is a vectorised form of a frame $\mathbf{I}^{(t)}$, can be observed for every time instance t. Here, we model each reconstructed frame $\hat{\mathbf{i}}^{(t)} = \mathbf{\Psi}^{-1}\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}^{(t)}, \mathbf{r}^{(t)})$ as a combination of the lossless signal $\mathbf{i}^{(t)}$ and the reconstruction error $\mathbf{e}^{(t)}$, *i.e.*,

$$\hat{\mathbf{i}}^{(t)} = \mathbf{i}^{(t)} + \mathbf{e}^{(t)}, \tag{4.1}$$

where the error $\mathbf{e}^{(t)}$ is assumed to be a vector of random variable drawn from a random process of some unspecified distribution.

Here we propose the correlated reference estimator with the aim to negate the effect of the error propagation. This is the RGA-based reference estimator, which performs a running average over the entire collection of the reconstructed signals. RGA is inspired by the Gaussian Mixture Model in background estimation problem [39]. While Gaussian Mixture models the value of each pixel as a combination of several Gaussian distributions, each representing the possible value of background and foreground objects [12, 13], RGA in its simplified form only models each pixel using one Gaussian distribution.

The correlated reference $\mathbf{r}^{(t)}$ used in Referenced CS reconstruction is obtained from $\mathbf{r}^{(t)} = \mathbf{\Psi i}_{\mathbf{r}}^{(t-1)}$, where $\mathbf{i}_{\mathbf{r}}$ denotes a vectorised reference image in the spatial domain. Each pixel of the reference image $\mathbf{i}_{\mathbf{r}}^{(t)}$ is

drawn from the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{i_r}}^{(t)}, \boldsymbol{\varsigma}_{\mathbf{i_r}}^{(t)})$, where $\boldsymbol{\mu}_{\mathbf{i_r}}^{(t)}$ and $\boldsymbol{\varsigma}_{\mathbf{i_r}}^{(t)}$ are the vectors containing the mean and the variance of each pixel in $\mathbf{i}_{\mathbf{r}}^{(t)}$ respectively.

We define the update rule of $\mathbf{i}_{\mathbf{r}}^{(t)}$, the reference image at frame t, in terms of $\boldsymbol{\mu}_{\mathbf{i_r}}^{(t)}$ and $\boldsymbol{\varsigma}_{\mathbf{i_r}}^{(t)}$ as

$$\boldsymbol{\mu}_{\mathbf{i_r}}^{(t)} = (1-\alpha)\boldsymbol{\mu}_{\mathbf{i_r}}^{(t-1)} + \alpha\hat{\mathbf{i}}^{(t)}, \tag{4.2}$$

$$\boldsymbol{\varsigma}_{\mathbf{i_r}}^{(t)} = (1-\alpha)\boldsymbol{\varsigma}_{\mathbf{i_r}}^{(t-1)} + \alpha\left[(\hat{\mathbf{i}}^{(t)} - \boldsymbol{\mu}_{\mathbf{i_r}}^{(t)})^{\mathrm{T}}(\hat{\mathbf{i}}^{(t)} - \boldsymbol{\mu}_{\mathbf{i_r}}^{(t)})\right], \tag{4.3}$$

where $\hat{i}^{(j,t)}$ is a pixel of the reconstructed image $\hat{\mathbf{i}}^{(t)} = \boldsymbol{\Psi}^{-1}\hat{\mathbf{x}}^{(t)}$. In other words, both the mean and variance of the reference is updated on the fly with the latest reconstruction result $\hat{\mathbf{i}}^{(t)}$. The parameter α is called a learning parameter, which defines the rate of which the reference is updated given the reconstruction results. Large values of α gives the reference a faster response to the changes of contents inside a sequence, but reduce the performance in terms of reducing error propagation. Small values of α, on the other hands, can suppress more of the error propagation, but response slowly to the changes of contents. This could potentially lead to outdated reference and, thus, results in the large reference distance.

In practice, the references used in Referenced CS are obtained as vectors of mean values.

4.1.2 Relationship between Learning Parameter and Reconstruction Error

To study the relationship between the learning parameter and the reconstruction accuracy, we employ the Monte Carlo method. A sensing

operation and the Referenced CS reconstruction, described in Chapter 3, are applied to a sequence of signals C, which is constructed as follows. Each signal $\mathbf{x} \in C$ is a sparse signal with K non-zero elements, *i.e.*, K-sparse, for a small K. Both the locations (supports) and magnitude of the non-zero elements of $\mathbf{x}^{(1)}$, the first instance of the sequence C, are drawn from uniformly random process. To generate other instances $\mathbf{x}^{(t)} \in C, t > 1$, while maintaining their likeliness to $\mathbf{x}^{(1)}$, we employ the following procedures for each instance $\mathbf{x}^{(t)}$:

1. A new support is randomly added to the support set of $\mathbf{x}^{(t)}$. Its magnitude is drawn randomly.

2. An existing support of $\mathbf{x}^{(t)}$ is randomly removed. Its magnitude is reset to zero.

3. Each element of $\mathbf{x}^{(t)}$ is multiplied by a random gain $0.9 \leq \beta \leq 1.1$.

4. Each support, along with its magnitude, of $\mathbf{x}^{(t)}$ is randomly shifted.*

Each sequence C is then compressively sampled and reconstructed using Referenced CS,

$$\min \|\hat{\mathbf{x}}^{(t)} - \mathbf{r}^{(t)}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}}^{(t)} = \mathbf{y}^{(t)}, \tag{4.4}$$

and $\mathbf{y}^{(t)} = \mathbf{\Phi}\mathbf{x}^{(t)}$. The reference $\mathbf{r}^{(t)}$ used for the reconstruction is estimated using the RGA estimator in Eq. (4.2) using a various value of the learning parameter α.

Two sets, each contains the total of 1000 sequences of C, are employed in the Monte Carlo simulation to map the relationship between the learning parameter α and PSNR of the reconstruction results. The first set of C is created without using the 4th procedure (*), *i.e.*, no shift

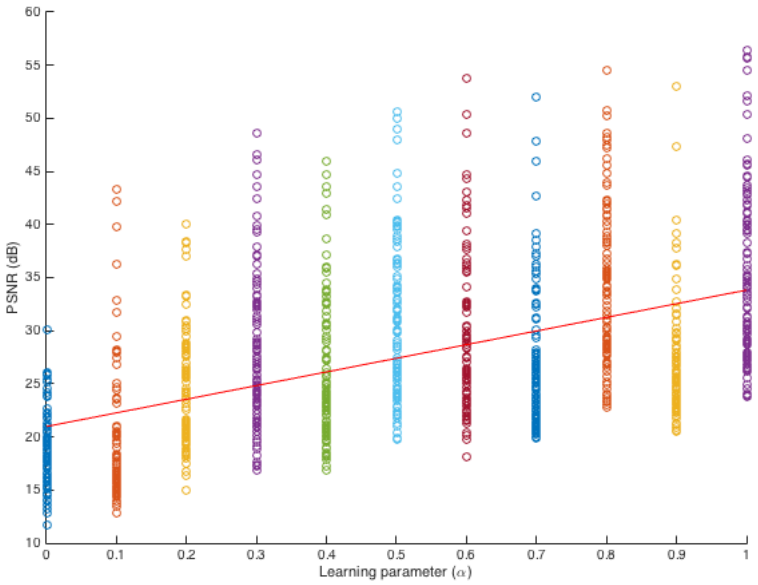


Figure 4.1: PSNR versus learning parameter of test set C without support shifts.

in the locations of the non-zero elements. Figure 4.1 shows the scatter plot between α and PSNR of this set. The regression line shows that when the sparse supports are stationary, a larger value of α provides the results with high PSNR with the highest probability. The use of the naive reference ($\alpha = 1$) also provides a very good accuracy, thus, the use of RGA estimator is trivial.

The second set of C is created with the random support shift procedure (*). It can be seen in Figure 4.2 that the situation is in reverse when the sparse supports are no longer stationary. In this case, the use of small values of α gives better reconstruction accuracy than the larger ones. The middle range of α, therefore, provides a middle ground for both signals with stationary and non-stationary supports.

It should be noted that natural signals acquired in real-world applications have, in general, non-stationary supports. An example case study of this is shown in Figure 4.3, where two frames from a video sequence are

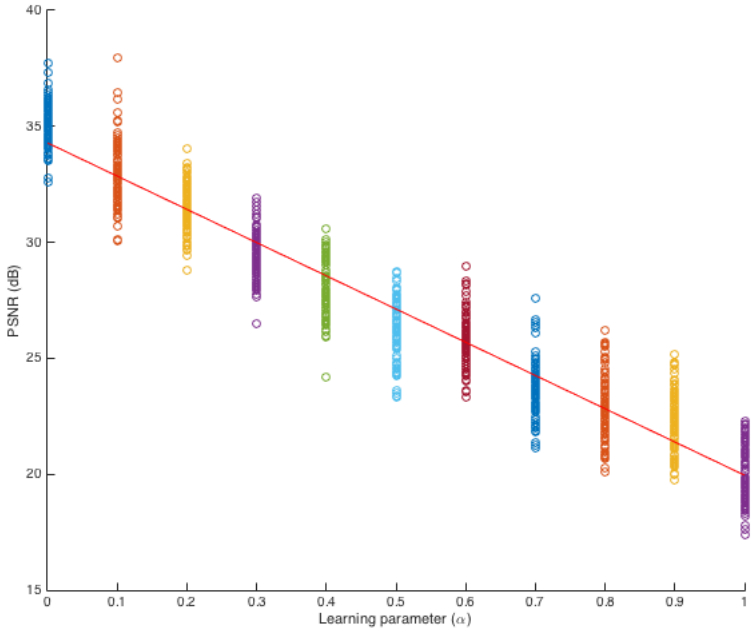


Figure 4.2: PSNR versus learning parameter of test set C with support shifts.

represented using discrete Fourier transform. While the general positions and shape of the sparse coefficients seem similar, many of the supports are not in the same position. Figure 4.3 shows the position of approximately sparse support—those with the coefficient value larger than 5% of the maximum value—and the difference in position between two frames. Surprisingly, as many as 100 supports out of 480, *i.e.*, 20%, change their position between these two visually similar frames. As natural signals consist of both stationary and non-stationary supports, the middle range of α is preferable than any one extreme.

4.1.3 Optimal dynamic learning parameter

Instead of using any fixed value for the learning parameter α, it is possible to use a dynamic rate. Recall that, given a collection of spatio-temporal

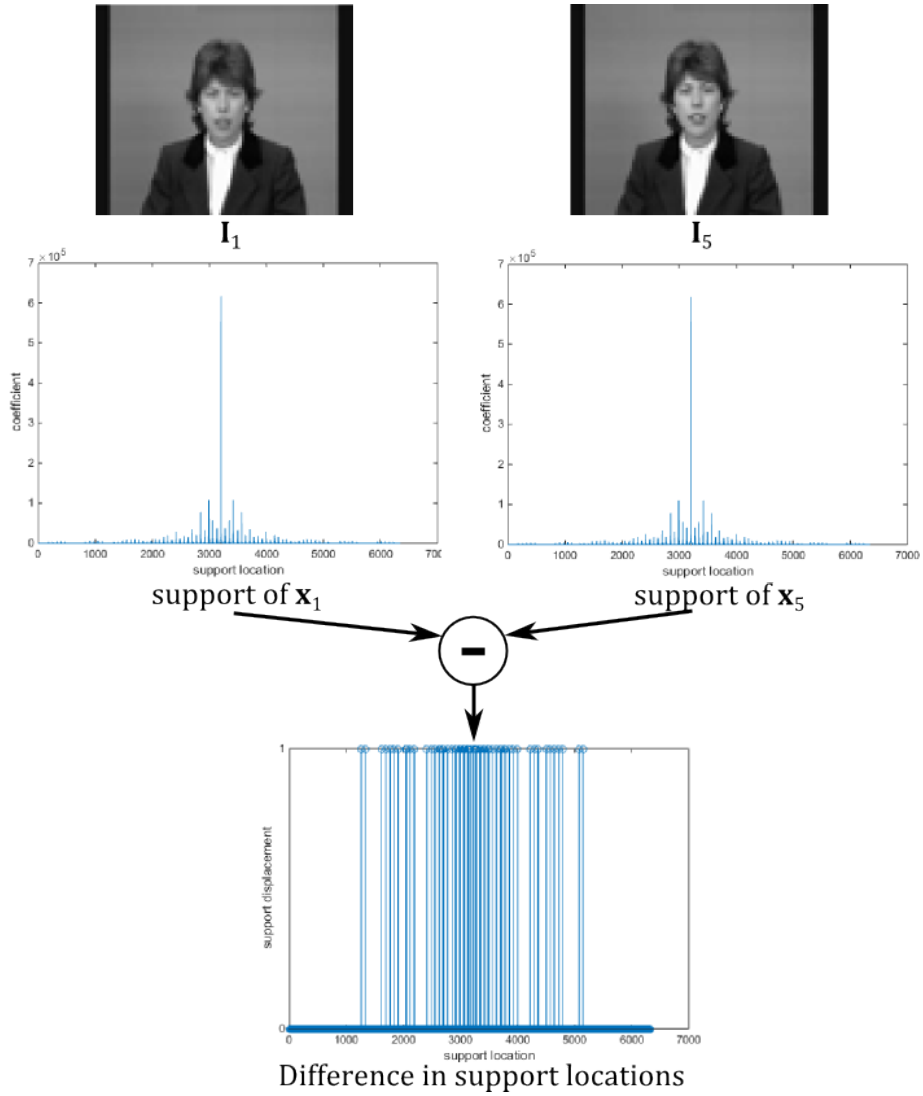


Figure 4.3: Displacement in support locations between two video frames

signal C, we can express a reconstructed signal $\hat{\mathbf{i}}^{(t)}$ of $\mathbf{i}^{(t)} \in C$ as

$$\hat{\mathbf{i}}^{(t)} = \mathbf{i}^{(t)} + \mathbf{e}^{(t)}, \tag{4.5}$$

where $\mathbf{e}^{(t)}$ is the reconstruction error that is assumed to be a vector of random variable drawn from a random process of some unspecified distribution. Suppose we require the reference $\mathbf{r}^{(t+1)} = \mathbf{\Psi}\mathbf{i}_{\mathbf{r}}^{(t+1)}$ to be the sparse domain of the average of the first t instances of $\hat{\mathbf{i}}$, *i.e.*,

$$\mathbf{i}_{\mathbf{r}}^{(t+1)} = \frac{1}{t}\sum_{k=1}^{t}\hat{\mathbf{i}}^{(k)}. \tag{4.6}$$

Since

$$\mathbf{i}_{\mathbf{r}}^{(t)} = \frac{1}{t-1}\sum_{k=1}^{t-1}\hat{\mathbf{i}}^{(k)}, \tag{4.7}$$

we can derive that

$$\mathbf{i}_{\mathbf{r}}^{(t+1)} = \frac{1}{t}\hat{\mathbf{i}}^{(t)} + \frac{t-1}{t}\mathbf{i}_{\mathbf{r}}^{(t)}. \tag{4.8}$$

Compare Eq. (4.8) to Eq. (4.2), thus, we can see that by setting the learning parameter $\alpha = \frac{1}{t}$ for any value of t makes $\mathbf{i}_{\mathbf{r}}^{(t)}$ to be the average of the first $t-1$ instances of $\hat{\mathbf{i}}$. This has the advantage of computing the average of all $t-1$ instances without the need to store all the reconstructed results within the memory.

The learning parameter $\alpha = \frac{1}{t}$ is optimal in the sense that the reference $\mathbf{r}$ completely disregards the presence of reconstruction noise in the reconstructed sequence. By expressing Eq. (4.7) in terms of Eq. (4.5),

we can see that

$$\mathbf{i}_{\mathbf{r}}^{(t)} = \frac{1}{t-1}[\hat{\mathbf{i}}^{(1)} + \cdots + \hat{\mathbf{i}}_{(t-1)}] \tag{4.9}$$

$$= \frac{1}{t-1}[(\mathbf{i}^{(1)} + \mathbf{e}^{(1)}) + \cdots + (\mathbf{i}^{(t-1)} + \mathbf{e}^{(t-1)})] \tag{4.10}$$

$$= \frac{1}{t-1}[(\mathbf{i}^{(1)} + \cdots + \mathbf{i}^{(t-1)}) + (\mathbf{e}^{(1)} + \cdots + \mathbf{e}^{(t-1)})]. \tag{4.11}$$

Since each $\mathbf{e}$ term is drawn from a random process, by Central Limit Theorem, when t is large, the distribution of the summation of $\mathbf{e}$ becomes a normal distribution. Thus, we obtain

$$\mathbf{i}_{\mathbf{r}}^{(t)} \equiv \frac{\sum_{k=1}^{t-1} \mathbf{i}^{(k)}}{t-1} + \frac{(t-1)\mathcal{N}(0,\sigma^2)}{t-1} \tag{4.12}$$

$$\equiv \bar{\mathbf{i}}^{(t-1)} + \mathcal{N}(0,\sigma^2), \tag{4.13}$$

where $\bar{\mathbf{i}}^{(t-1)}$ is the average of the first $t-1$ instances.

While the main focus of this approach is to provide a reference that is more robust to reconstruction error, quantitatively the general reconstruction accuracy is about on par with the middle range fixed learning parameters. By using Monte Carlo on the same test sets used in Section 4.1.2, the average PSNR when using the dynamic learning parameter is 28.23 dB for the stationary supports set and 29.72 dB for the non-stationary set.

4.1.4 Experimental results

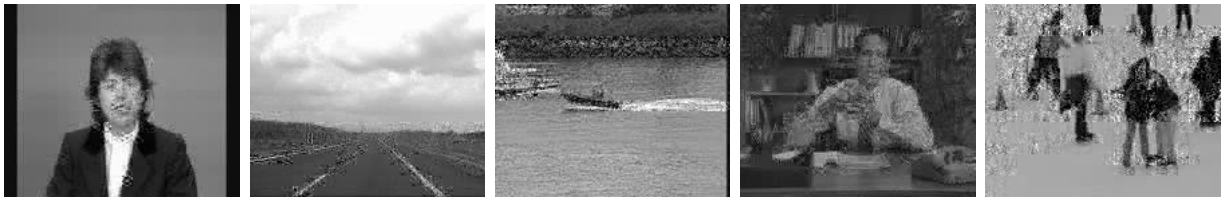
The experiment in this section aims to demonstrate the effect of the learning parameter α in the reconstruction of real video sequences. The dataset used here is the same set used in Section 3.5.3(Figure 3.16), consisting of 14 sequences grouped into 3 categories of low, medium, and

high activity. The sampling operation and reconstruction operation are applied to each sequence. The reconstruction in this experiment is done strictly using the Referenced CS. Several strategies for choosing the correlated reference $\mathbf{r}$ are employed and compared between each other. The simplest reference, referred to as a *naive reference*, is the case when $\alpha = 1$. In other words, only the immediate reconstructed frame is used as a reference for the next frame ($\mathbf{r}^{(t)}_{naive} = \hat{\mathbf{x}}^{(t-1)}$). This is exactly what was done in Section 3.5.3. The Running Gaussian references are estimated using fixed learning parameters $\alpha = 0.1$, 0.3, and 0.5—denoted as $\mathbf{r}^{(t)}_{0.1}$, $\mathbf{r}^{(t)}_{0.3}$, and $\mathbf{r}^{(t)}_{0.5}$ respectively—as well as the optimal reference $\mathbf{r}^{(t)}_{opt}$ using the dynamic parameter $\alpha = \frac{1}{t}$. Moreover, the reconstructed results using these references are also compared with the results using lossless references. The lossless reference is a controlled benchmark, obtained directly from the input without the application of the sensing and reconstruction operations, *i.e.*, $\mathbf{r}^{(t)}_{lossless} = \mathbf{x}^{(t-1)}$. Such a reference is, of course, unavailable in practice and is shown here only for the comparison purpose.

Figure 4.4 shows the examples of the reconstructed sequences. Each row in Figure 4.4 shows the results obtained using a different type of references. The sequences in the first and seconds columns are the examples of low activity sequences. The third and fourth columns are medium activity sequences while the last column shows the examples of high activity sequences. Subjectively it is clear that, in all sequences, the reconstructions using references with optimal learning parameter result in better visual quality than other settings. Also, it can be noticed that the results obtained using $\mathbf{r}^{(t)}_{0.1}$ have resulted in less error than those obtained using $\mathbf{r}^{(t)}_{0.5}$. Moreover, the results using naive references demonstrate the highest amount of reconstruction error. This follows the discussion in

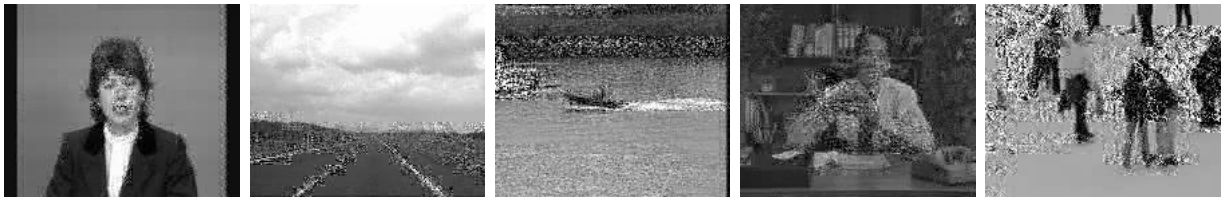
(a) Lossless reference

(b) Optimal reference



(c) Reference with $\alpha = 0.1$

(d) Reference with $\alpha = 0.3$



(e) Reference with $\alpha = 0.5$

(f) Naive reference

Figure 4.4: Examples of reconstructed sequences using various types of references

Section 4.1.2 that when the locations of signal sparse supports are not stationary, as in the case of most natural signals, the small values of α provides the most robust reconstruction results.

The above observation is verified by Table 4.1, which shows the PSNR of each reconstructed sequence using each reference setting. This table confirms that despite the tendency of higher reconstruction quality as α gets smaller, the optimal reference using the dynamic parameter outperforms all fixed learning parameters in natural sequences reconstruction. The dynamic parameter can be seen as a progression from large α ($\alpha = 1$) to smaller ones ($\alpha \to 0$ as t increases). This implies that the estimator is more confident during the initial part of the reconstructed sequence, where the level of reconstruction error is still small. Table 4.1 also shows that the effects of the learning parameter are more prominent when the activity level in the sequence is higher. As such, the difference in reconstruction quality using references with different learning parameters can be observed more easily in high activity sequences than in the lower ones.

4.2 Improvement in Referenced CS Complexity

One of the major issues that prevent a practical deployment of compressed sensing is the high computational complexity of the iterative reconstruction operation. High complexity makes the reconstruction of large data be a time-consuming task, especially if the data is very large. This issue heavily affects the use of CS to obtain the fMRI data because, as discussed in Section 2.1.3, fMRI requires a very large dataset compared to other imaging techniques. This chapter attempts to tackle this

Table 4.1: Peak Signal-to-Noise Ratio of reconstructed video sequences using various types of references

	Lossless	Naive	$\alpha = 0.5$	$\alpha = 0.3$	$\alpha = 0.1$	optimal
Low activity sequences						
Akiyo	46.16	34.79	34.96	35.20	35.43	**35.59**
Clair	47.76	35.17	35.48	36.69	36.94	**36.94**
Container	44.64	25.75	25.78	25.78	25.89	**32.87**
Road	40.13	34.06	34.58	35.19	35.93	**37.16**
MissUSA	46.05	37.18	37.42	38.05	38.87	**39.24**
Medium activity sequences						
Car	37.53	31.53	31.36	32.37	32.50	**33.56**
Coastguard	35.37	29.66	29.96	30.62	31.24	**32.11**
Office	41.08	32.20	32.70	32.81	33.08	**33.56**
Mother	44.78	34.63	34.73	35.75	35.79	**36.09**
News	42.89	32.31	32.45	32.55	33.06	**33.11**
Salesman	41.98	31.55	31.84	32.57	32.74	**33.17**
High activity sequences						
Foreman	35.88	29.54	29.73	30.55	31.05	**32.04**
Skate	36.56	29.60	29.76	30.36	30.82	**32.61**
Silence	41.30	31.85	32.06	32.76	32.83	**33.18**

issue by introducing a method to greatly reduce the complexity of the CS reconstruction. This method involves a relaxation of the objective function from the l_1-norm to the l_2-norm, thus allows the reconstruction to be done using a linear method: the least squares approximation.

4.2.1 Complexity of Iterative Reconstruction Algorithms

As discussed in Section 3.2, the non-smooth nature of the l_1-norm objective function requires the reconstruction algorithms to work iteratively. Moreover, the complexity of the algorithms is proportional to the size of the signal in a non-linear fashion. Thus, reconstructing several small signals are faster than a single large signal.

The reconstruction algorithms modified for Referenced CS in Section 3.4 have vary degree of complexity. The linear programming algo-

Table 4.2: Comparison of computational time per frame using different reconstruction algorithms

Algorithm	small (88×72 pixel)	large (176×144 pixel)
SparseLab	22.67	906.84
ISAL1	17.34	693.61
l_1-homotopy	42.87	1714.83
IHT	1.12	44.87
OMP	217.16	8686.41
CoSaMP	198.62	7944.87

rithms have high complexity. The simplex method has the worst case complexity of $O(n^2 2^n)$ for the signal of n elements. The worst case complexity of the Interior-point Method (used in SparseLab toolbox) is $O(n^{3.5})$ [94]. The worst running time of the OMP is $O(Kmn)$, to reconstruct K-sparse signal of length n from the compressive measurements of m elements [156]. The worst case complexity of Infeasible Point Subgradient Method (ISA) has not been assessed, however, the author did an intensive simulation test to find out that ISA's complexity is in between those of linear-programming methods' and OMP's [103].

In Section 2.2.5, Figure 2.25 plots the computational time against the sampling rate, comparing between each algorithm when reconstructing the set of test images. Table 4.2 shows this comparison in details, showing the time it takes for each algorithm to reconstruct a frame of the video sequences from Figure 3.16 using various sampling rates.

Table 4.2 also shows the computational time per frame for the same set of sequences with twice the resolution. It can be seen clearly that when the length n of signal is four times larger, the complexity increases dramatically. If we increase the resolution even higher, soon the reconstruction algorithms will become intractable in practice.

This observation is especially true for the fMRI data. The typical data from the EPI technique generally have the resolution of 64-by-64

pixels per slice, with around 30-64 slices per volume. This makes the total number of element per volume to be around 1.23×10^5 to 2.62×10^5. Though this number is not too large in the general imaging sense, the reconstruction time is already quite high. As a single fMRI experiment could consist of several hundreds of volumes, easy calculation shows that using CS reconstruction is indeed intractable for a practical fMRI experiment.

4.2.2 Referenced CS with Least Squares Method

One of the main promises of Referenced CS is the fact that it moves away from the notion of sparsity. Recall the same symbols from Section 3.2. As shown in *Proposition 2* that the reconstruction error $E(\hat{\mathbf{x}}, \mathbf{x})$ depends only on the distance $\delta = \|\mathbf{r} - \mathbf{x}\|_2$ between the reference $\mathbf{r}$ and the signal $\mathbf{x}$. This proposition is valid for any l_p-norm, $0 < p \leq \infty$. As we no longer need to maximise the sparsity using the l_1-norm, it is desirable to relax the optimisation problem back to the l_2-norm objective function. The l_2-norm objective function has found its use in a great many fields of science and engineering because it is smooth and differentiable. This allows the l_2-norm to be solved using a straightforward linear method, such as the least squares approximation method, rather than complex iterative algorithms. Linearity makes the least squares method much simpler than iterative algorithms used to solve the l_1-norm minimisation problem.

Here, we assert that the least squares should be able to perform reasonably well with the incorporation of the correlated reference. Particularly since the degree of sparsity of most natural signals does not reach the exact reconstruction level—and with the bound defined by *Proposi-*

tion 2—the reconstruction error from the least squares method should be similar to the error of the l_1-norm solution in terms of magnitude. The sources of error are different, however. The error of l_1-norm minimisation comes from the fact that the l_1-norm solution is too sparse than the natural signals. The error of the least squares, on the other hand, is due to the fact that the l_2-norm solution's sparsity is too small.

We propose to use the least squares method to solve for the solution of Referenced CS, referred to as Referenced CS with the Least Squares (Reference CS/LS), following the *Proposition 3*.

Proposition 3. *Let* $\mathbf{r} \in \mathbb{R}^n$ *be a correlated reference of a signal* $\mathbf{x} \in \mathbb{R}^n$*, the reconstructed signal* $\hat{\mathbf{x}} \in \mathbb{R}^n$ *can be obtained from the compressive measurements* $\mathbf{y} \in \mathbb{R}^m, \mathbf{y} = \mathbf{\Phi x}$ *by*

$$\hat{\mathbf{x}} = \mathbf{r} + \mathbf{\Phi}^{\mathrm{T}}(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}})^{-1}(\mathbf{y} - \mathbf{\Phi r}). \tag{4.14}$$

Proof. Following Eq. (3.9), we define a Referenced CS l_2-norm minimisation problem as

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}} - \mathbf{r}\|_2 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}. \tag{4.15}$$

Define a Lagrangian function as

$$\mathcal{L}(\hat{\mathbf{x}}) = \|\hat{\mathbf{x}} - \mathbf{r}\|_2^2 + \lambda^{\mathrm{T}}(\mathbf{\Phi}\hat{\mathbf{x}} - \mathbf{y}), \tag{4.16}$$

where λ is the Lagrange multiplier. Set a derivative of $\mathcal{L}(\hat{\mathbf{x}})$ to zero, *i.e.*,

$$\frac{\partial}{\partial \hat{\mathbf{x}}}\mathcal{L}(\hat{\mathbf{x}}) = 2\hat{\mathbf{x}} - 2\mathbf{r} + \mathbf{\Phi}^{\mathrm{T}}\lambda = 0, \tag{4.17}$$

to obtain

$$\hat{\mathbf{x}} = \mathbf{r} - \frac{1}{2}\mathbf{\Phi}^{\mathrm{T}}\lambda. \tag{4.18}$$

To solve for the Lagrange multiplier λ, substitute Eq. (4.18) into $\mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}$ to obtain

$$\mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{\Phi}(\mathbf{r} - \frac{1}{2}\mathbf{\Phi}^{\mathrm{T}}\lambda) \tag{4.19}$$

$$= \mathbf{\Phi}\mathbf{r} - \frac{1}{2}\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}\lambda. \tag{4.20}$$

From Eq. (4.20), we can get

$$\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}\lambda = -2(\mathbf{y} - \mathbf{\Phi}\mathbf{r}), \tag{4.21}$$

and finally,

$$\lambda = -2(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}})^{-1}(\mathbf{y} - \mathbf{\Phi}\mathbf{r}). \tag{4.22}$$

Substitute Eq. (4.22) back into Eq. (4.18) to obtain

$$\hat{\mathbf{x}} = \mathbf{r} + \mathbf{\Phi}^{\mathrm{T}}(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}})^{-1}(\mathbf{y} - \mathbf{\Phi}\mathbf{r}). \tag{4.23}$$

□

4.2.3 Experiment on Referenced CS using Least Squares

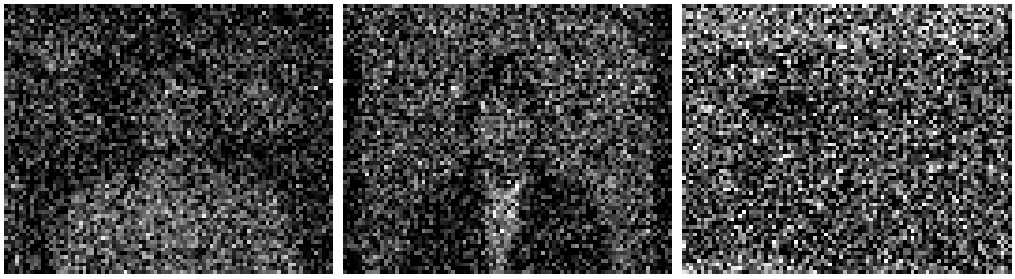
To verify the performance of Referenced CS/LS, proposed in *Proposition 3*, here the same experiment performed in Section 3.5.3 is repeated. The sampling and reconstruction operations are applied to the same set of 14 test sequences in Figure 3.16. In this experiment, the reconstruction results come from using the following reconstruction operation: 1) Referenced CS/LS (using the least squares approximation), 2) Referenced CS

(using l_1-norm objective function), 3) the l_1-norm minimisation, and 4) the conventional least squares method. Both the l_1-norm minimisation and the least squares method are for the benchmarking purpose.

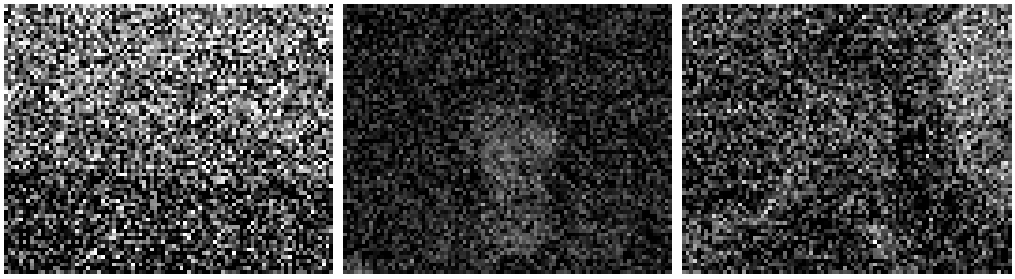
Figure 4.5 shows the example frames from the reconstructed sequences using the conventional least squares method while Figure 4.6 shows the examples of the Referenced CS/LS results. The examples of reconstructed sequences using the l_1-minimisation and Referenced CS can be found in Figure 3.17 and Figure 3.18 respectively. While it comes at no surprise that the results of the conventional least squares method have extremely poor quality, the results of Referenced CS/LS, on the other hand, show very good visual quality, comparable to those obtained from the iterative algorithms. The same observation made with Referenced CS can be made here also. The reconstruction error is present mainly as additive reconstruction error. The reconstruction accuracy largely depends on the reference distance δ, thus, the quality is higher in the low activity sequences than in high-quality ones.

Figure 4.7 shows the average PSNR versus sampling rate of each algorithm. It can be seen that as the sampling rate s gets higher, the reconstruction accuracy of Referenced CS/LS improves rapidly. With a sampling rate higher than 35%, the results of Referenced CS/LS—despite having a much lower complexity—outperforms the conventional l_1-norm minimisation method. At 50%, Referenced CS/LS yields 19% higher PSNR than l_1-norm minimisation and 3 times higher than the conventional least squares method.

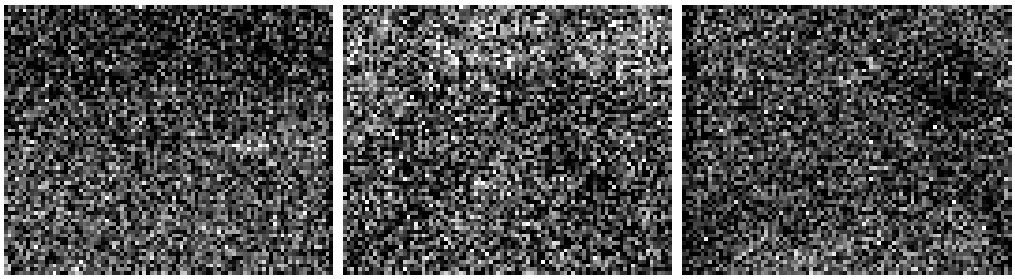
Table 4.3 shows the average time required to reconstruct a frame of the test sequences. It shows clearly than the computational time required by the least squares-based method, including Referenced CS/LS, is sev-



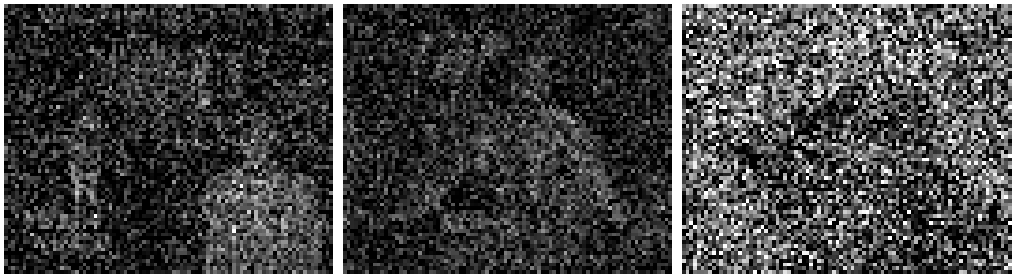
(a) Akiyo (b) Clair (c) Container



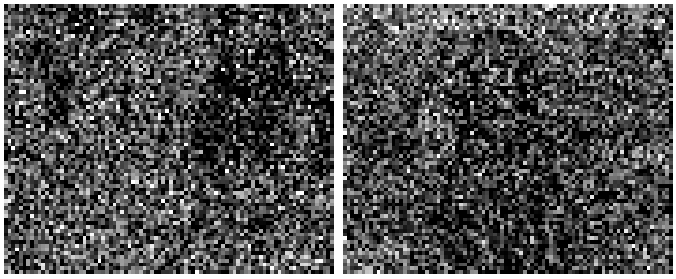
(d) Road (e) MissUSA (f) Car



(g) Coastguard (h) Office (i) Mother



(j) News (k) Salesman (l) Foreman



(m) Skate (n) Silence

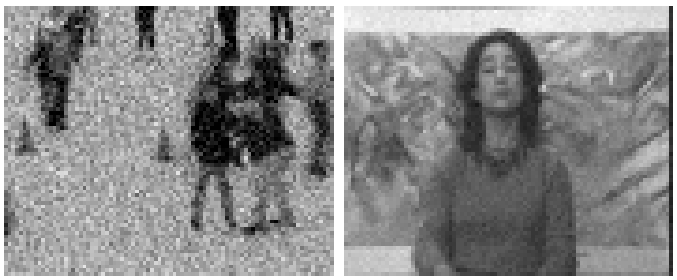
Figure 4.5: Examples of reconstructed video sequences using the least squares method from 50% of samples

(a) Akiyo (b) Clair (c) Container

(d) Road (e) MissUSA (f) Car

(g) Coastguard (h) Office (i) Mother

(j) News (k) Salesman (l) Foreman



(m) Skate (n) Silence

Figure 4.6: Examples of reconstructed video sequences using Referenced CS/LS from 50% of samples

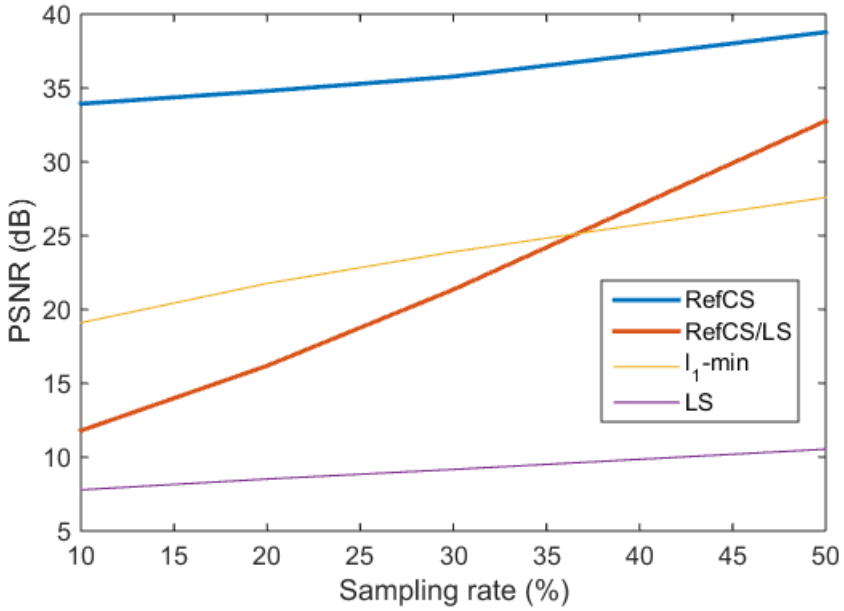


Figure 4.7: Average PSNR across 14 test sequences versus sampling rate. The reconstruction is done using Referenced CS (RefCS), Referenced CS/LS (RefCS/LS), l_1-norm minimisation (l_1-min), and the least squares method (LS).

Table 4.3: Average reconstruction time per frame in seconds

Method	sampling rate s (%)				
	10	20	30	40	50
RefCS	0.90	1.14	1.59	1.82	2.21
RefCS/LS	3.80×10^{-4}	7.57×10^{-4}	11.19×10^{-4}	14.72×10^{-4}	17.89×10^{-4}
l_1-min	0.86	1.12	1.41	1.79	2.18
LS	1.86×10^{-4}	3.39×10^{-4}	5.18×10^{-4}	6.93×10^{-4}	9.09×10^{-4}

eral degree less than that required by the iterative algorithms. Because Referenced CS/LS uses a linear operation, the computational time now depends on the performance of the mathematical software package used to complete such operation rather than the complexity of the algorithm. On average, at the sampling rate of 50%, Table 4.3 shows that Referenced CS with the Least Squares can reconstruct the result at least 1200 times faster than the l_1-norm minimisation via convex optimisation.

4.3 Summary

This chapter addressed two main issues of the Referenced Compressed Sensing. The first issue addressed is the estimation of the correlated reference. In this chapter, the novel correlated reference estimator based on Running Gaussian Average has been proposed, where the correlated reference is obtained from a Running Gaussian Average of the reconstructed frames. The effect of the learning parameter of the Running Gaussian Average has been studied, and it is shown that the reference that can suppress error propagation can be obtained by using the dynamic optimal learning parameter rather than a fixed value.

The second issue addressed in this chapter is the computational complexity of Referenced CS. The low complexity version, referred to as Referenced CS with the Least Squares, is proposed. It is shown that, by replacing the l_1-norm objective function to the l_2-norm, the reconstruction can be more than 1200 times faster. Moreover, it is shown that by exploiting the redundancy of the correlated reference, the reconstruction result using the least squares approximation is, surprisingly, comparable to the results of iterative algorithms.

Chapter 5

Compressed Sensing for Functional MRI

In previous chapters, the novel compressed sensing reconstruction technique, Referenced CS, has been developed. This technique is to be applied to the fMRI data in this chapter. The aim is to obtain the fMRI data from compressive measurements such that it has a high spatial resolution and maintains the accuracy of its activity map. Firstly, we will describe the methods and settings for applying the compressed sensing to the fMRI data. Each relevant parameter is to be analysed in terms of visual quality to identify the best setting for Referenced Compressed Sensing fMRI. The reconstruction using this novel technique will then be compared to the results from other state-of-the-art reconstruction algorithms as well as the currently employed fast acquisition technique—the EPI. Finally, we introduce another analysis approach, the baseline independent analysis of the data, to show that the error resulting from the compressed sensing technique can be excluded from the analysis.

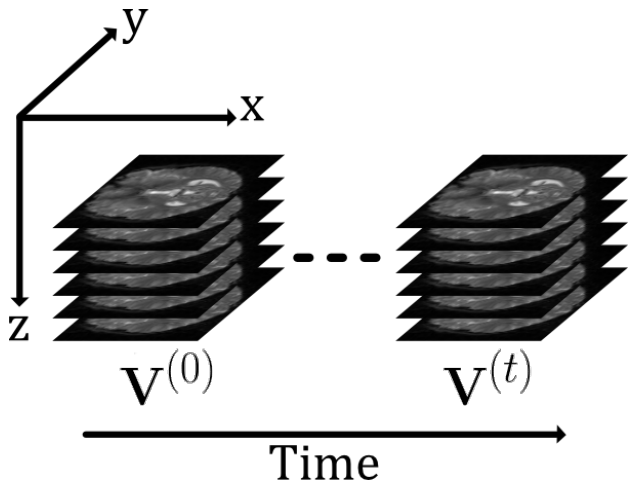


Figure 5.1: MRI data is a temporal collection of several 3-dimensional MR volumes

5.1 Applying Compressed Sensing to fMRI

As discussed in Section 1.1, the fMRI data can be viewed as a temporal collection of several MRI volumes. As such, Referenced CS, which is designed specifically for the temporal data reconstruction, is applicable to the fMRI data.

Figure 5.1 shows the characteristic of the fMRI data, which is a collection of MRI volumes $\mathbf{V}$. Each volume itself is a collection of slices $\mathbf{I} \in \mathbb{R}^{m \times n}$. Each slice $\mathbf{I}$ is obtained from the K-space $\mathbf{x}$, which is the frequency domain of $\mathbf{I}$. In other words, $\mathbf{i} = \mathbf{\Psi}^{-1}\mathbf{x}$, where $\mathbf{i} \in \mathbb{R}^{mn}$ is the vectorised representation of $\mathbf{I}$ and $\mathbf{\Psi}^{-1}$ is the inverse Fourier basis.

Instead of acquiring a complete K-space from the scanner, the compressed sensing fMRI aims to reconstruct the K-space $\hat{\mathbf{x}}$ from the compressive measurements

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{5.1}$$

where $\mathbf{\Phi}$ denotes the incoherent sensing matrix (described in Section 2.2.3). There exist many approaches to implement the sensing operation into the practical MRI scheme, as discussed in Section 2.3.

However, our focus is the reconstruction operation. The reconstructed K-space $\hat{\mathbf{x}}$ can be obtained using various reconstruction methods, including the conventional l_1-norm minimisation, the proposed Referenced CS method, and even the low complexity variation of Referenced CS such as Referenced CS/LS.

The straightforward extension of compressed sensing to fMRI is to treat each slice $\mathbf{i}$ as an independent signal. First we reconstruct the full resolution K-space $\hat{\mathbf{x}}$ from the measurement $\mathbf{y}$. For each K-space $\mathbf{x}$ of a slice $\mathbf{i}$, with its compressive measurements obtained from the sampling operation $\mathbf{y} = \mathbf{\Phi x}$, the reconstruction operation is defined as

$$\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}) = \min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}. \tag{5.2}$$

Eq. (5.2) can be solved using any available l_1-norm minimisation solvers. Afterward, the reconstructed slice $\hat{\mathbf{i}}$ is obtained from $\hat{\mathbf{x}}$ by applying the basis, *i.e.*,

$$\hat{\mathbf{i}} = \mathbf{\Psi}^{-1}\hat{\mathbf{x}} \tag{5.3}$$

$$= \mathbf{\Psi}^{-1}\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}). \tag{5.4}$$

On the other hand, using the proposed Referenced CS, each slice is reconstructed based on its own reference. Thus, the collection of reference slices themselves can be considered to be a volume, called the Reference Volume ($\mathbf{V_r}$). Given the compressive measurement $\mathbf{y}$, and the corresponding reference slice $\mathbf{r}$, the reconstruction operation is defined as

$$\mathfrak{R}(\mathbf{\Phi}, \mathbf{y}, \mathbf{r}) = \min \|\hat{\mathbf{x}} - \mathbf{r}\|_1 \text{ subject to } \mathbf{\Phi}\hat{\mathbf{x}} = \mathbf{y}. \tag{5.5}$$

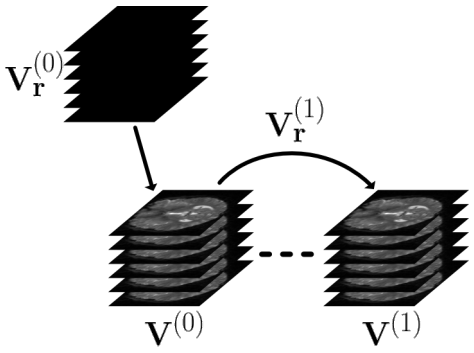
Eq. (5.5) can be solved using the modified solvers presented in Section 3.4. Again the reconstructed slice $\hat{\mathbf{i}}$ can then be obtained from $\hat{\mathbf{i}} = \Psi^{-1}\hat{\mathbf{x}}$.

The complexity of the reconstruction operation can be reduced by using Referenced CS/LS. As discussed in Section 4.2.2, by relaxing the l_1-norm objective function into the l_2-norm, it is possible to obtain a reconstruction result in much less time at the cost of the reconstruction accuracy. With the proper choice of $\mathbf{V_r}$, however, it is possible to obtain an acceptable reconstruction result much faster than the conventional l_1-norm minimisation could.

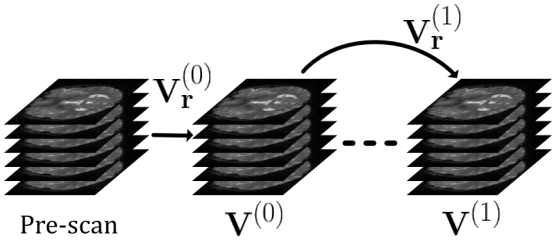
By *Definition 6*, the reference volume $\mathbf{V_r}$ can be any arbitrary 3-dimensional signal. However, as the reconstruction accuracy of Referenced CS depends heavily on the reference distance $\|\mathbf{V} - \mathbf{V_r}\|_2$, it is desirable to have the reference volume that is close to the target K-space. There are several variations on the Referenced CS setting, each with its strength and weakness. The choices of Initial Reference Volume, reference reset strategy, and update strategy all affect the performance of Referenced CS.

5.1.1 Initial Reference Volume

Figure 5.2 shows two strategies for the Initial Reference Volume $\mathbf{V_r}^{(0)}$, *i.e.*, the reference for reconstructing the first volume $\mathbf{V}^{(0)}$ of the data. A simple and straightforward choice for the Initial Reference Volume $\mathbf{V_r}^{(0)}$ is a zero volume, denoted $\mathbf{0}$, which is the volume that consists of all zeros (Figure 5.2a). This in effect is equivalent to the reconstruction of the first volume using the conventional l_1-norm minimisation. This choice of $\mathbf{V_r}^{(0)}$ is easy to implement and requires no alteration to the readout sequence. However, this method leads to the reconstruction error constitutes to the



(a) Using zero volume as R_0

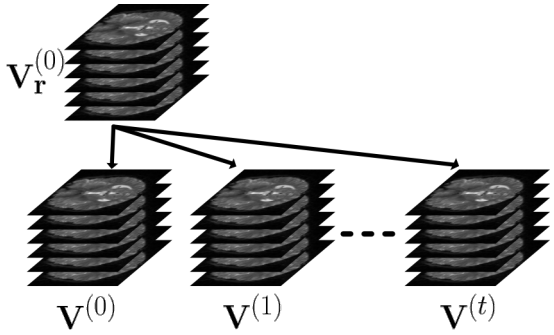


(b) Using a lossless pre-scan volume as R_0

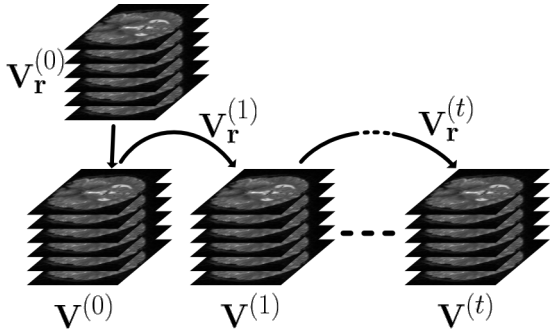
Figure 5.2: Strategies for the Initial Reference Volume

l_1-norm minimisation method, which will propagate to the next reference volume. Despite the fact that we can reduce the effect of reconstruction error propagation by using RGA estimation, it is more desirable to have a good $\mathbf{V}_\mathbf{r}^{(0)}$ in the first place.

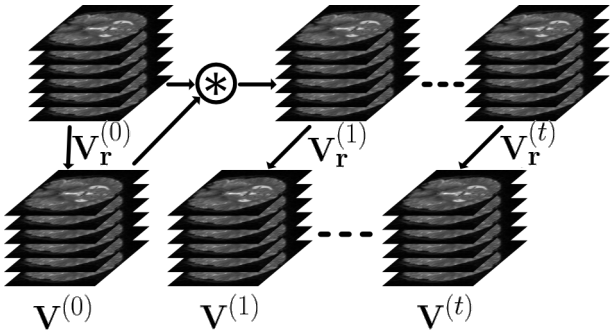
A good Initial Reference Volume can be obtained by doing a complete, lossless acquisition of the K-space. It could be done either by acquiring the first volume uncompressed—sampling at 100%—or by using a pre-scan volume as a reference volume, *i.e.*, $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{V}^{(-1)}$, where $\mathbf{V}^{(-1)}$ denotes a pre-scan volume (Figure 5.2b). While this strategy might sound unattractive in other applications, it is natural in the MRI acquisition where a pre-scan is needed most of the time to compensate and calibrate the hardware [77].



(a) No update



(b) Naive reference



(c) Running Gaussian reference

Figure 5.3: Strategies for the Initial Reference Volume

5.1.2 Reference Volume Update Strategy

Figure 5.3 shows several ways of how to use the reference volume to reconstruct each volume. The first choice is to use the Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$ to reconstruct each and every successive volume after it (Figure 5.3a). Using the reference volume in this fashion does not make the reference to be up-to-date to the data. However, because the Initial Reference Volume is used purely, its quality will dictate all the reconstruction quality. If $\mathbf{V}_\mathbf{r}^{(0)}$ is obtained from the uncompressed acquisition, as suggested previously, then it is in the best, reconstruction-error-free condition that can yield a very good result. However, if it is obtained from the l_1-norm minimisation, its reconstruction error will also greatly damage the reconstruction results.

Another choice is to use the latest reconstructed volume as the reference volume for the next reconstruction, *i.e.*, let $\mathbf{V}_\mathbf{r}^{(t)} = \hat{\mathbf{V}}^{(t-1)}$ where t denotes the volume index. This approach is referred to as the naive reference (Figure 5.3b), similar to what was used in Section 3.5.3. As discussed in Chapter 4, this approach is very sensitive to reconstruction error in the long run, thus, the error will propagate from one volume to its subsequent volumes rapidly.

Instead of these two extremes, the reference volume can be estimated using the RGA estimator (Figure 5.3c). Section 4.1.1 shows that the running Gaussian reference works better than the naive reconstruction while provides a way to negate the propagation of reconstruction error. The learning parameter α required by the estimator can be fixed to any value, or set to the optimal value as discussed in Section 4.1.3.

5.1.3 Reference Volume Reset

Similar to the case of natural video sequences, it is desirable to reset the Initial Reference Volume from time to time. Despite the fact that the RGA estimator can provide the changing reference volume with minimal error propagation as time passes, the reference distance will become too great due to the aggregated error or the changing of the signal itself. For examples, the movement of the subject in the machine or the changing of the subject's haemodynamic system could result in a totally uncorrelated volume compared to the $\mathbf{V}_{\mathbf{r}}^{(0)}$.

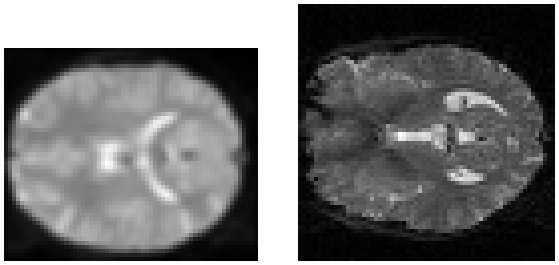
Having a reset for the reference volume will provide a great improvement in the reconstruction accuracy at the cost of the design complexity. For example, a new Initial Reference Volume can be acquired at every period T using the lossless sampling. All successive volumes during the same period T are acquired compressively and reconstructed using Referenced CS instead. In essence, this strategy provides a trade-off between the robustness and accuracy of the compressed sensed fMRI data and the length and the complexity of the experiment.

These strategies for Referenced CS will be evaluated in terms of reconstruction error in details later on in Section 5.4.

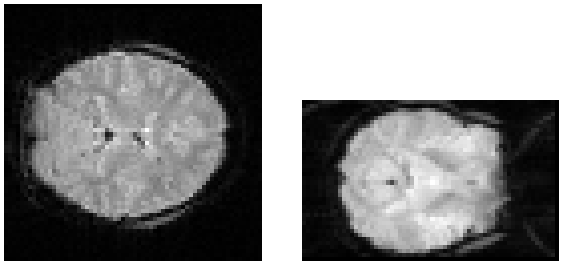
5.2 fMRI Datasets

Throughout this study, the following fMRI data are used during the study:

- **Dataset 1**: Attention to visual motion dataset [19],
- **Dataset 2**: Auditory fMRI dataset [19],



(a) Dataset 1 (b) Dataset 2



(c) Dataset 3 (d) Dataset 4

Figure 5.4: Examples of the datasets

- **Dataset 3**: Mixed-gambles task fMRI dataset (obtained from the OpenfMRI database with accession number ds000005 [154]),
- **Dataset 4**: Visual object recognition dataset (obtained from the OpenfMRI database with accession number ds000105 [124]).

The example of these datasets are shown in Figure 5.4.

The sensing and reconstruction operations are applied to these original datasets using various settings to create the final test data for this study. Further on, each data will be identified not only by the source data but also by their CS acquisition setting, including the sampling rate, the strategy for Initial Reference Volume, its update, and its reset. To simplify, each dataset is labelled in the following format: *<the source number>–<sampling rate>–<acquisition flag>*. For example, data DS1-30-110 refers to Dataset 1 which is compressively sampled at 30% using pre-scan Initial Reference Volume, the each reference is estimated naive reference, and no reset to the reference is employed. Table 5.1 shows all

the available data that has been created.

The first digit of the acquisition flag identifies the choice of the Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$—0: $\mathbf{V}_\mathbf{r}^{(0)}$ is obtained using the traditional l_1-norm minimisation, 1: $\mathbf{V}_\mathbf{r}^{(0)}$ is obtained using the lossless pre-scan. The second digit indicates the reference update strategy— 0 means no update employed ($\mathbf{V}_\mathbf{r}^{(t)} = \mathbf{V}_\mathbf{r}^{(0)}$), 1 being the naive reference ($\mathbf{V}_\mathbf{r}^{(t)} = \hat{\mathbf{V}}^{(t-1)}$), and 2 being the Running Gaussian-based reference. The third digit of the flag indicates the Reference reset strategy; 0: no reference reset at all, 1 means that the new Initial Reference Volume is reacquired at every arbitrary T volumes, resetting the reference periodically. 2 and 3 are similar to 1, but the reset is done at every $T/2$ and $T/4$ respectively. The choice of T used in in this study is chosen based on the TR of each dataset, such that the period $T \times$ TR is equal to one minute. In the case of the RGA-based reference update, there will also be the forth flag indicating the value of the learning parameter α.

Due to the limitation of time, it is not possible to explore every combination of parameters. Section 5.3 shows that, in general, the reconstruction operation could take up to more than 30 hours. To workaround this limitation, in this study, the effect of each parameter will be evaluated individually. Only one parameter is treated as an independent variable at a time while all other parameters are fixed. Once the effect of the acquisition parameters has been evaluated, the proposed Referenced CS will be compared to the l_1-norm minimisation obtained from the state-of-the-art algorithm.

Firstly, the complexity of compressed sensing fMRI will be evaluated, followed by the visual quality and the accuracy of the activity map.

Table 5.1: Compressive sensing fMRI dataset

Data ID	Source	sampling rate (%)	$\mathbf{V}_{\mathbf{r}}^{(0)}$	Update strategy	Reset strategy
DS1-30-000	DS1	30	l_1-min	no update	no reset
DS1-30-100	DS1	30	Pre-scan	no update	no reset
DS1-30-101	DS1	30	Pre-scan	no update	at T
DS1-30-102	DS1	30	Pre-scan	no update	at $T/2$
DS1-30-104	DS1	30	Pre-scan	no update	at $T/4$
DS1-30-112	DS1	30	Pre-scan	naive	at $T/2$
DS1-30-1221	DS1	30	Pre-scan	RGA,$\alpha = 0.1$	at $T/2$
DS1-30-1223	DS1	30	Pre-scan	RGA,$\alpha = 0.3$	at $T/2$
DS1-30-1225	DS1	30	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS1-30-122o	DS1	30	Pre-scan	RGA,optimal α	at $T/2$
DS1-10-1225	DS1	10	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS1-50-1225	DS1	50	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS2-30-000	DS2	30	l_1-min	no update	no reset
DS2-30-100	DS2	30	Pre-scan	no update	no reset
DS2-30-101	DS2	30	Pre-scan	no update	at T
DS2-30-102	DS2	30	Pre-scan	no update	at $T/2$
DS2-30-104	DS2	30	Pre-scan	no update	at $T/4$
DS2-30-112	DS2	30	Pre-scan	naive	at $T/2$
DS2-30-1221	DS2	30	Pre-scan	RGA,$\alpha = 0.1$	at $T/2$
DS2-30-1223	DS2	30	Pre-scan	RGA,$\alpha = 0.3$	at $T/2$
DS2-30-1225	DS2	30	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS2-30-122o	DS2	30	Pre-scan	RGA,optimal α	at $T/2$
DS2-10-1225	DS2	10	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS2-50-1225	DS2	50	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$

Table 5.2: Compressive sensing fMRI dataset (continue)

Data ID	Source	sampling rate (%)	$\mathbf{V}_\mathbf{r}^{(0)}$	Update strategy	Reset strategy
DS3-30-000	DS3	30	l_1-min	no update	no reset
DS3-30-100	DS3	30	Pre-scan	no update	no reset
DS3-30-101	DS3	30	Pre-scan	no update	at T
DS3-30-102	DS3	30	Pre-scan	no update	at $T/2$
DS3-30-104	DS3	30	Pre-scan	no update	at $T/4$
DS3-30-112	DS3	30	Pre-scan	naive	at $T/2$
DS3-30-1221	DS3	30	Pre-scan	RGA,$\alpha = 0.1$	at $T/2$
DS3-30-1223	DS3	30	Pre-scan	RGA,$\alpha = 0.3$	at $T/2$
DS3-30-1225	DS3	30	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS3-30-122o	DS3	30	Pre-scan	RGA,optimal α	at $T/2$
DS3-10-1225	DS3	10	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS4-30-000	DS4	30	l_1-min	no update	no reset
DS4-30-100	DS4	30	Pre-scan	no update	no reset
DS4-30-101	DS4	30	Pre-scan	no update	at T
DS4-30-102	DS4	30	Pre-scan	no update	at $T/2$
DS4-30-104	DS4	30	Pre-scan	no update	at $T/4$
DS4-30-112	DS4	30	Pre-scan	naive	at $T/2$
DS4-30-1221	DS4	30	Pre-scan	RGA,$\alpha = 0.1$	at $T/2$
DS4-30-1223	DS4	30	Pre-scan	RGA,$\alpha = 0.3$	at $T/2$
DS4-30-1225	DS4	30	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS4-30-122o	DS4	30	Pre-scan	RGA,optimal α	at $T/2$
DS4-10-1225	DS4	10	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$
DS4-50-1225	DS4	50	Pre-scan	RGA,$\alpha = 0.5$	at $T/2$

5.3 Acquisition Time and Reconstruction Time

The first thing to be evaluated is the time involved in the compressed sensing fMRI. There are two types of time to be considered; the acquisition time is the time required to obtain to raw measurements from the MRI scanner, and the reconstruction time which is the time required to reconstruct the MRI volumes from the measurements.

5.3.1 Acquisition Time

The acquisition time depends mainly on the scanning technique and pulse sequence used for the scan. The high-resolution scan used in the structural MRI requires several signal readouts to obtain a single K-space. Even though this process can be accelerated up to 4 times using Fast Spin Echo (Section 2.1.3), it still takes several TRs cycle per K-space. On the other hand, EPI—commonly used in the fMRI experiments—can obtain an entire K-space within one TR at the cost of the loss of spatial resolution.

The use of compressed sensing in fMRI allows the high-resolution scan, with the same resolution as in the structural MR data, to be acquired within the same time required for the EPI data. Table 5.3 shows the comparison of the acquisition time required for each acquisition method. Here the high-resolution data is supposed to have the resolution of $n \times n \times k$, where n, k are an arbitrary integer. The acquisition time in Table 5.3 is presented in terms of the number of TR period required to obtain the data. Because faster acquisition means more data can be acquired in a given period of time, Table 5.3 clearly shows the

Table 5.3: Acquisition time required to obtain MR data

Acquisition method	Time per slice (# of TR)	Time per volume (# of TR)	Resolution
High resolution (Conventional)	n	$n \times k$	$n \times n \times k$
High resolution (Fast Spin Echo)	$\frac{n}{4}$	$\frac{n}{4} \times k$	$n \times n \times k$
Echo Planar Imaging	1	k	$\frac{n}{4} \times \frac{n}{4} \times k$
CS-MRI	1	k	$n \times n \times k$

benefit of applying the compressed sensing to the fMRI data: to improve the ratio between the spatial resolution and the temporal resolution.

5.3.2 Reconstruction times

The reconstruction time refers to the time required to reconstruct the MRI volume $\mathbf{V}$ from the raw measurements. In the conventional MRI, the reconstruction time is nearly instantaneous, as the only step required is to apply the Fourier basis to the K-space. In the compressed sensing MRI, on the other hand, the reconstruction time is much higher due to the time required to solve the optimisation problem. Thus, the reconstruction time depends largely on the complexity of the reconstruction algorithm.

Table 5.4 shows the average time required to compute a volume of MRI data from the datasets used in this study. The computation is done using Matlab2014b on a machine running Windows 7 64-bit, with Intel Core i5 3.3GHz processor and 4GB of memory. The clear separation can be noticed between the algorithms based on the l_1-norm and the l_2-norm minimisation. The time required for both the l_1-norm minimisation and Referenced CS methods is several degree longer than Referenced CS/LS and the least squares method, both employ the l_2-norm minimisation. For

Table 5.4: Average time required to reconstruct each volume in datasets

Methods	sampling rate	$\mathcal{T}_{\mathcal{R}}$ (seconds)			
		DS1	DS2	DS3	DS4
l_1-min	0.1	218.91	374.28	206.66	194.34
	0.3	391.42	602.31	380.46	310.78
	0.5	641.68	874.00	662.47	513.71
RefCS	0.1	214.32	373.63	208.13	194.93
	0.3	383.95	579.32	385.86	310.87
	0.5	636.43	872.67	660.37	496.87
Least squares	0.1	0.27	0.41	0.22	0.28
	0.3	0.44	0.86	0.49	0.44
	0.5	0.66	1.27	0.68	0.59
RefCS/LS	0.1	0.41	0.53	0.35	0.27
	0.3	0.92	1.22	0.77	0.72
	0.5	1.21	1.71	1.15	1.08

example, the computational time of Referenced CS is at least 4.56×10^4 percent longer than that of Referenced CS/LS.

In the context of this specific study, Table 5.5 shows the total time required to compute each dataset in its completeness. It can be seen that the computational complexity of the l_1-based methods is impractical in real fMRI experiments. The amount of time requires to reconstruct a complete series of MRI volumes would be too large for any clinical purposes. Even in the situation where the data is not required ungently, such as in the research environment, the time it takes to process the data would slow down the research more than the benefit of the accelerated acquisition offered by the compressed sensing fMRI. It is clear that the method with lower complexity, such as Referenced CS/LS, is more desirable in real applications. The difference in reconstruction quality between these methods is to be discussed in the next section.

Table 5.5: Time required to reconstruct each dataset

Methods	sampling rate	Time (seconds)			
		DS1	DS2	DS3	DS4
l_1-min	0.1	37871.75	35556.24	48595.28	23320.24
	0.3	67715.31	57219.39	89409.16	37293.82
	0.5	111010.71	83030.31	146518.39	61645.29
RefCS	0.1	37078.33	35494.29	48911.35	23391.1
	0.3	66423.21	55035.74	90677.98	37624.81
	0.5	110103.01	82903.69	145976.49	59624.8
Least squares	0.1	46.78	45.09	60.061	25.01
	0.3	105.29	90.91	132.24	56.54
	0.5	161.37	144.21	196.20	80.59
RefCS/LS	0.1	70.66	49.83	82.11	32.23
	0.3	159.69	115.44	180.25	86.82
	0.5	209.27	162.6	270.76	129.81

5.4 Reconstruction Quality of Compressed Sensing fMRI

In this section, the simulated compressed sensing fMRI data are evaluated in terms of visual quality. The purpose of this evaluation is not only to compare the reconstruction from Referenced CS with the state-of-the-art algorithm, but it is also to study the effect of each acquisition strategy and setting for Referenced CS in the fMRI environment. The various strategies for Referenced CS discussed in Section 5.1 are to be evaluated first, followed by the comparisons with the state-of-the-art algorithm.

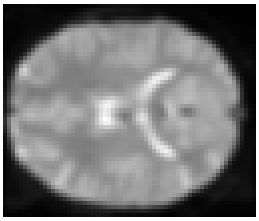
5.4.1 Study on the Initial Reference Volume

In this first study, the goal is to determine the best strategy for choosing the Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$. As discussed in previous section, the naive approach for choosing $\mathbf{V}_\mathbf{r}^{(0)}$, the reference for the very first volume to reconstruct, is to set $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$. This essentially turns the Referenced CS problem into a normal l_1-norm minimisation problem.

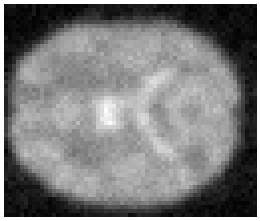
In contrast, it is a common practice in the MRI experiment that a pre-scan is to be performed before the actual experiment. This serves many reasons, ranging from the machine calibration to the stabilisation of the subject's physiology. The ultimate fact is, there usually is redundant information readily at hands prior to the experiment, which we can employ as a initial reference volume.

In the strict application viewpoint, the use of pre-scan initial reference might seem to defeat the goal of the compressive sensing fMRI, since the complete information is required. It is worth noting that our goal here is to improve the spatial-temporal resolution trade-off rather than to reduce the resolution of the acquisition hardware. However, it is true that by employing the pre-scan information as a reference, the implementation complexity is increased—also increases is the complexity of experiment design.

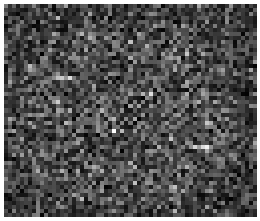
Figure 5.5–Figure 5.8 show the examples of the reconstructed data using different types of the initial reference volume. In this test, apart from the choice of $\mathbf{V}_{\mathbf{r}}^{(0)}$, all other parameters are fixed, namely, the $\mathbf{V}_{\mathbf{r}}^{(0)}$ reset is disabled (flag 0), no reference update over time (flag 0), and the sampling rate is fixed at 30%. These examples show that when using $\mathbf{V}_{\mathbf{r}}^{(0)} = \mathbf{0}$, essentially reconstructs the first volume using the l_1-norm minimisation, the noise-like reconstruction error of the l_1-norm minimisation is propagated over to all the subsequent reconstructions of Referenced CS. This effect is more severe in the case of Referenced CS/LS, as by using $\mathbf{V}_{\mathbf{r}}^{(0)} = \mathbf{0}$, the reconstruction problem effectively turns into a regular least squares problem. It is, however, possible to use Referenced CS exclusively to reconstruct the first volume using $\mathbf{V}_{\mathbf{r}}^{(0)} = \mathbf{0}$, followed by Referenced CS/LS for all other volumes. On the other



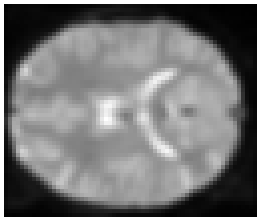
(a) Lossless scan



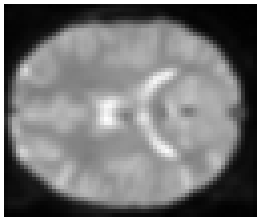
(b) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$



(c) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$

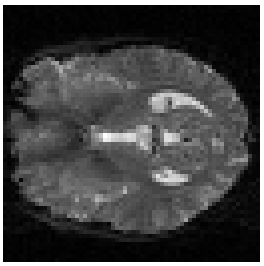


(d) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

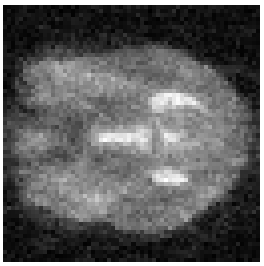


(e) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

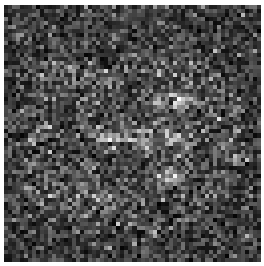
Figure 5.5: Examples of Dataset 1 reconstruction using Referenced CS and Referenced CS/LS employing different strategies for Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$. DS1-30-000 ($\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$) and DS1-30-100 ($\mathbf{V}_\mathbf{r}^{(0)}$ from pre-scan data) are compressively sampled at 30.



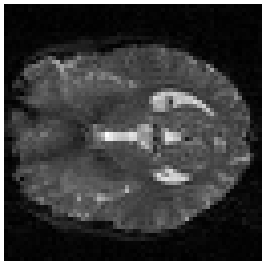
(a) Lossless scan



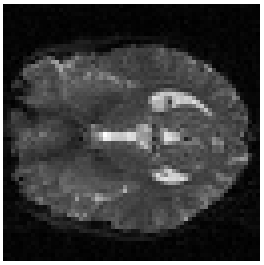
(b) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$



(c) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$

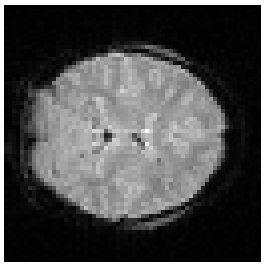


(d) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

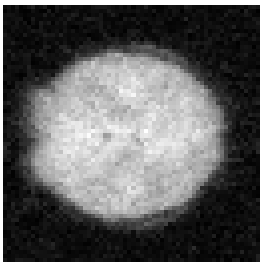


(e) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

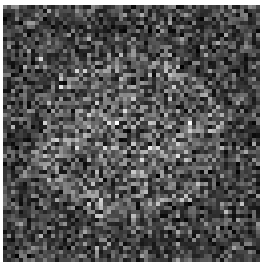
Figure 5.6: Examples of Dataset 2 reconstruction using Referenced CS and Referenced CS/LS employing different strategies for Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$. DS2-30-000 ($\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$) and DS2-30-100 ($\mathbf{V}_\mathbf{r}^{(0)}$ from pre-scan data) are compressively sampled at 30.



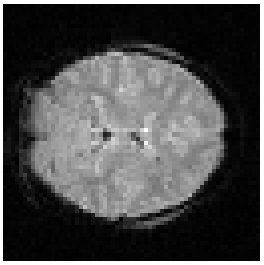
(a) Lossless scan



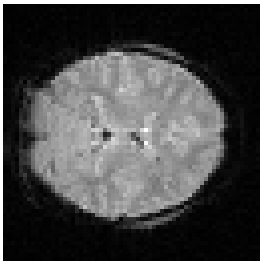
(b) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$



(c) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$

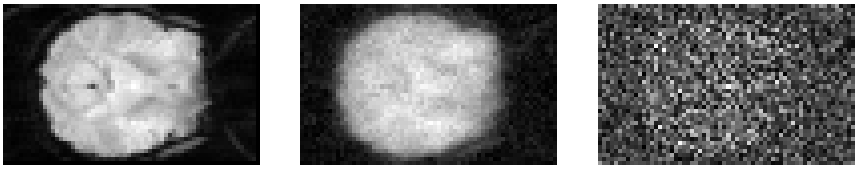


(d) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data



(e) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

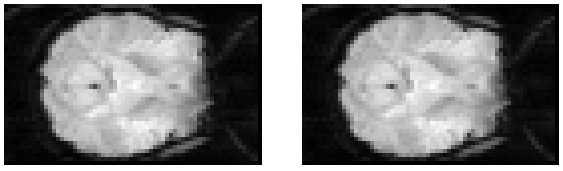
Figure 5.7: Examples of Dataset 3 reconstruction using Referenced CS and Referenced CS/LS employing different strategies for Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$. DS3-30-000 ($\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$) and DS3-30-100 ($\mathbf{V}_\mathbf{r}^{(0)}$ from pre-scan data) are compressively sampled at 30.



(a) Lossless scan

(b) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$

(c) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$



(d) RefCS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

(e) RefCS/LS reconstruction using $\mathbf{V}_\mathbf{r}^{(0)}$ from Pre-scan data

Figure 5.8: Examples of Dataset 4 reconstruction using Referenced CS and Referenced CS/LS employing different strategies for Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$. DS4-30-000 ($\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$) and DS4-30-100 ($\mathbf{V}_\mathbf{r}^{(0)}$ from pre-scan data) are compressively sampled at 30.

Table 5.6: Average PSNR of reconstruction using different Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$

Dataset	$\mathbf{V}_\mathbf{r}^{(0)} = \mathbf{0}$		pre-scan	
	RefCS	RefCS/LS	RefCS	RefCS/LS
DS1	38.68	22.28	**59.47**	56.41
DS2	30.90	21.99	**39.31**	37.90
DS3	37.29	25.39	48.01	**48.09**
DS4	30.31	17.57	44.04	**44.13**

hand, when the pre-scan $\mathbf{V}_\mathbf{r}^{(0)}$ is used, both the problem of initial error propagation and the failure of Referenced CS/LS to reconstruct the first volume are avoided. As a result, the subsequent reconstruction volumes show a much less reconstruction error.

Table 5.6 shows the visual quality metric using the PSNR. Also Figure 5.9–Figure 5.12 show the variation of the PSNR across each dataset. These results indeed verify the observation we got from the examples. In general, the quality of the Initial Reference Volume acts as the upper

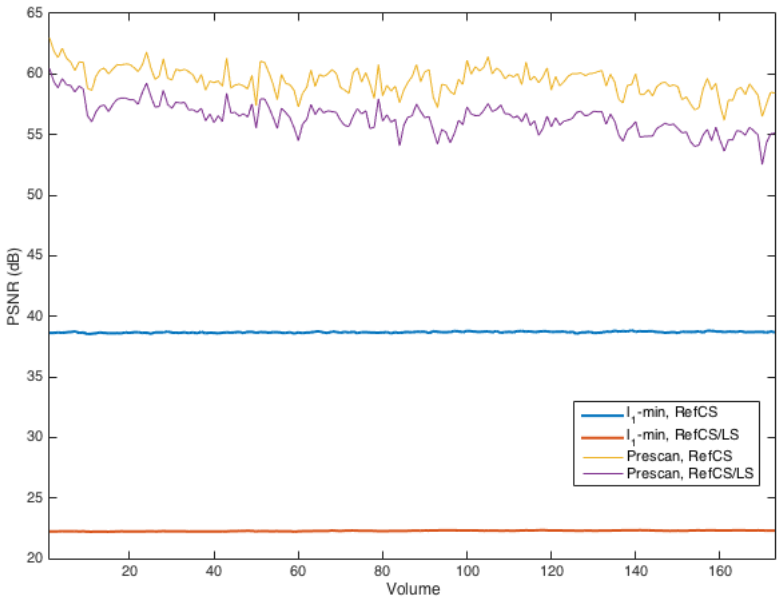


Figure 5.9: PSNR variation across Dataset 1 showing the difference in the choice of $\mathbf{V}_\mathbf{r}^{(0)}$

bound for all the subsequent reconstruction. Once $\mathbf{V}_\mathbf{r}^{(0)}$ already suffers from poor reconstruction quality, the entire series of data would suffer also. Thus, it is mandatory to employ the $\mathbf{V}_\mathbf{r}^{(0)}$ with the best quality possible, which is the lossless pre-scan volume in the setting of the fMRI experiment.

5.4.2 Study on the Initial Reference Reset

In this study, we compare several reset strategies for the reference volume $\mathbf{V}_\mathbf{r}$. The strategies employed here are no reset (flag 0), reset a new $\mathbf{V}_\mathbf{r}^{(0)}$ at every period T (flag 1), at every $T/2$ (flag 2), at every $T/4$ (flag 4). Specifically, in this experiment, T is chosen to be 20 for DS1, DS2, and DS3, and 16 for DS4, in order to make a reset period $T \times$ TR to be one minute, corresponding to the TR parameter of each dataset. All other parameters are fixed throughout this study. The choice of Initial

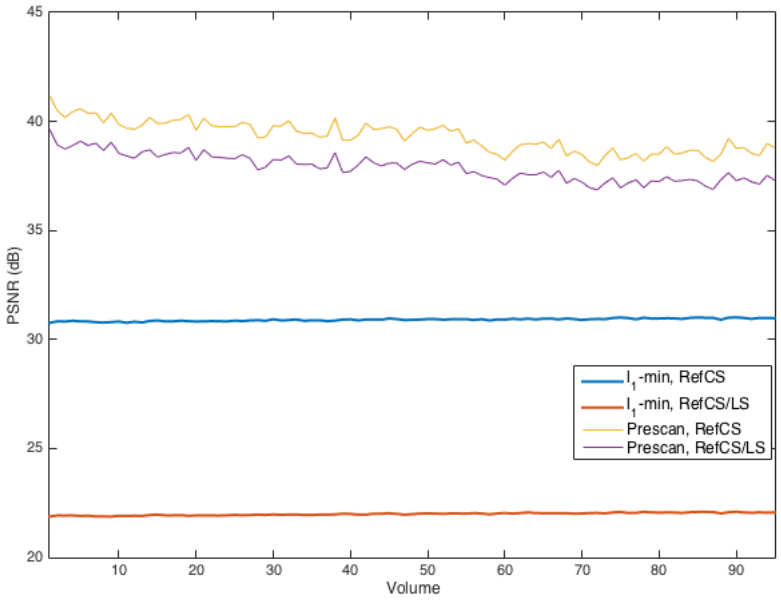


Figure 5.10: PSNR variation across Dataset 2 showing the difference in the choice of $\mathbf{V}_{\mathbf{r}}^{(0)}$

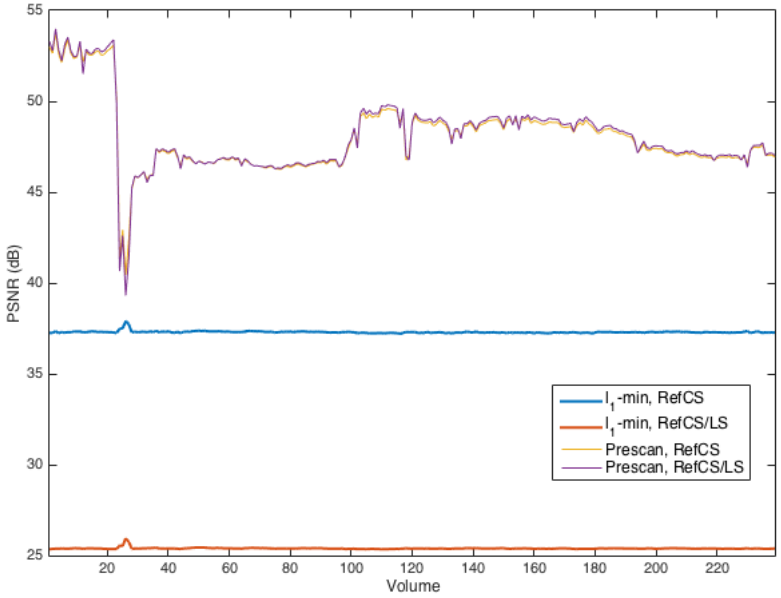


Figure 5.11: PSNR variation across Dataset 3 showing the difference in the choice of $\mathbf{V}_{\mathbf{r}}^{(0)}$

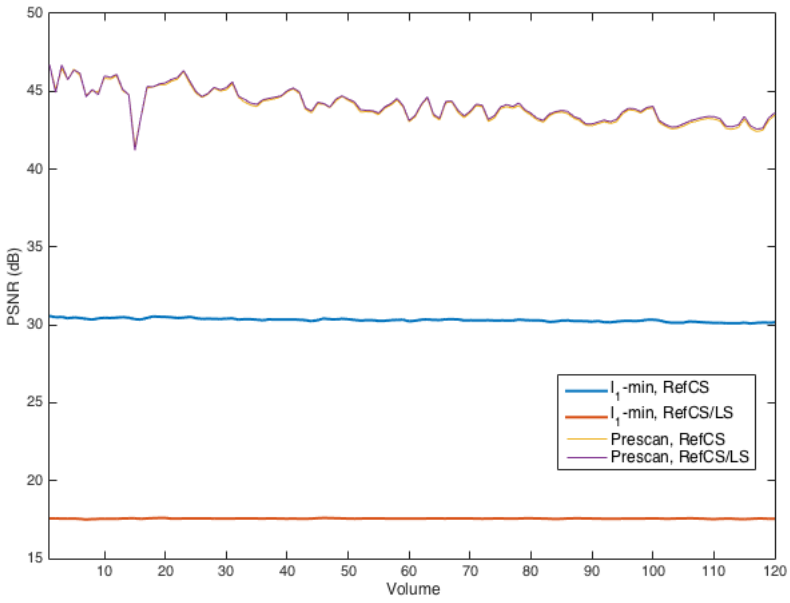
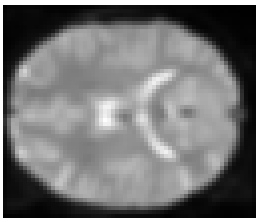


Figure 5.12: PSNR variation across Dataset 4 showing the difference in the choice of $\mathbf{V}_{\mathbf{r}}^{(0)}$

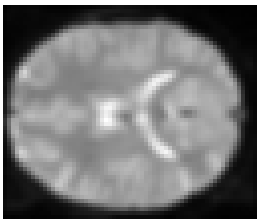
Reference Volume is set to be pre-scan (flag 1), with no the reference update (flag 0), and the sampling rate at 30%.

Figure 5.13–Figure 5.16 show the examples of the reconstructed data using different reset strategies. Table 5.7 and Figure 5.17–Figure 5.20 show the visual quality metric using PSNR.

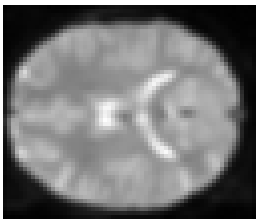
The benefit of the reference reset can be seen clearly in Dataset 2 (Figure 5.18), Dataset 3 (Figure 5.19), and Dataset 4 (Figure 5.20). In these datasets, there exist the sudden changes in the data due to the physical—not related to the reconstruction operation—phenomena. The source of these changes could be the movement of the subject in the scanner or a sudden change of the haemodynamic process. In these datasets, with the reset strategy for the reference, the quality of the reconstruction can be recovered after every reset, providing a more robust and adaptable system. On the other hands, without the reference reset, once such a change is experienced, the subsequent reconstructions suffer to the loss



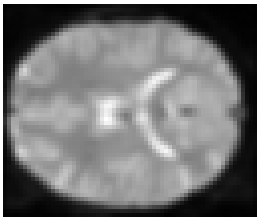
(a) Lossless scan



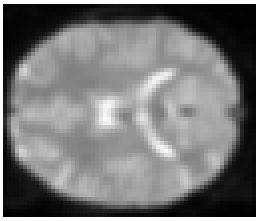
(b) RefCS with no reset



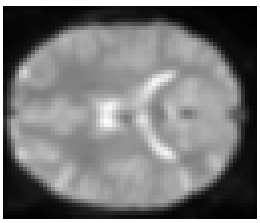
(c) RefCS/LS with no reset



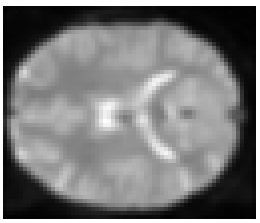
(d) RefCS with reset at T



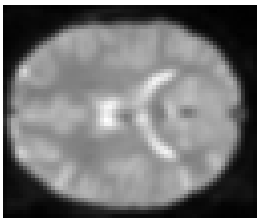
(e) RefCS/LS with reset at T



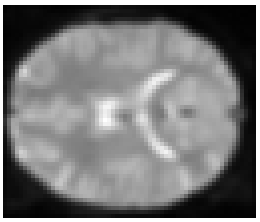
(f) RefCS with reset at $T/2$



(g) RefCS/LS with reset at $T/2$

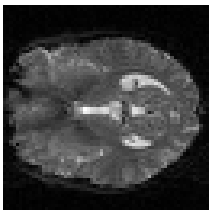


(h) RefCS with reset at $T/4$

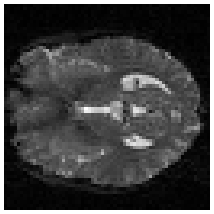


(i) RefCS/LS with reset at $T/4$

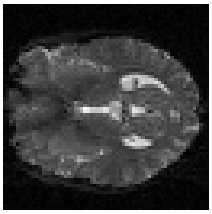
Figure 5.13: Examples of Dataset 1 reconstruction using Referenced CS and Referenced CS/LS employing different reset strategies for the Initial Reference Volume $\mathbf{V}_{\mathbf{r}}^{(0)}$. DS1-30-30-100 (no reset), DS1-30-101 (reset at T), DS1-30-102 (reset at $T/2$), and DS1-30-103 (reset at $T/4$) are compressively sampled at 30%.



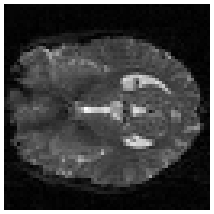
(a) Lossless scan



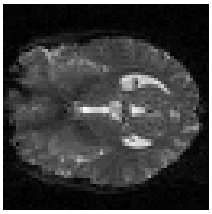
(b) RefCS with no reset



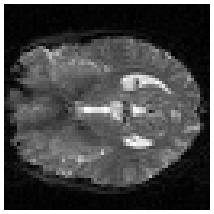
(c) RefCS/LS with no reset



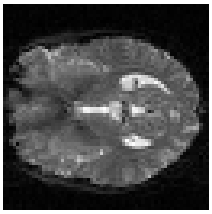
(d) RefCS with reset at T



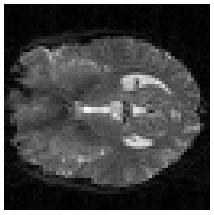
(e) RefCS/LS with reset at T



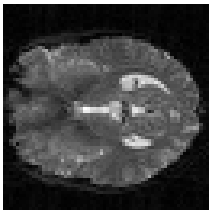
(f) RefCS with reset at $T/2$



(g) RefCS/LS with reset at $T/2$

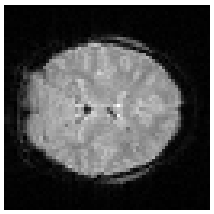


(h) RefCS with reset at $T/4$

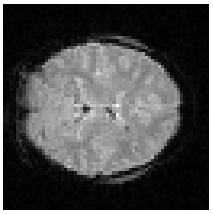


(i) RefCS/LS with reset at $T/4$

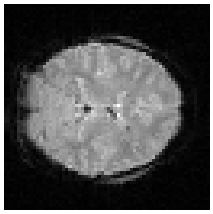
Figure 5.14: Examples of Dataset 2 reconstruction using Referenced CS and Referenced CS/LS employing different reset strategies for the Initial Reference Volume $\mathbf{V}_{\mathbf{r}}^{(0)}$. DS2-30-100 (no reset), DS2-30-101 (reset at T), DS2-30-102 (reset at $T/2$), and DS2-30-103 (reset at $T/4$) are compressively sampled at 30%.



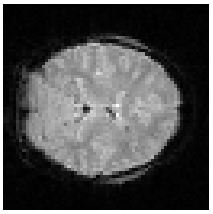
(a) Lossless scan



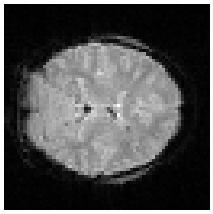
(b) RefCS with no reset



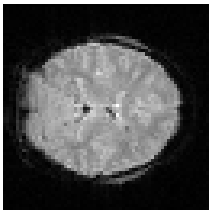
(c) RefCS/LS with no reset



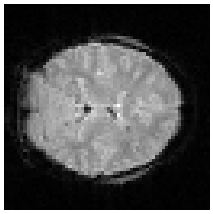
(d) RefCS with reset at T



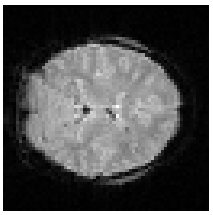
(e) RefCS/LS with reset at T



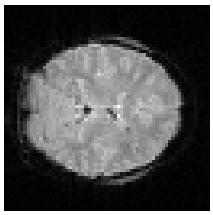
(f) RefCS with reset at $T/2$



(g) RefCS/LS with reset at $T/2$

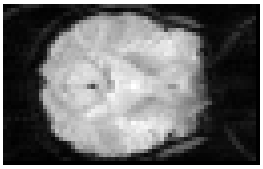


(h) RefCS with reset at $T/4$

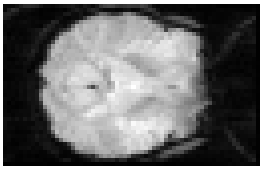


(i) RefCS/LS with reset at $T/4$

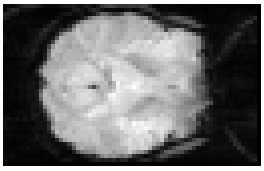
Figure 5.15: Examples of Dataset 3 reconstruction using Referenced CS and Referenced CS/LS employing different reset strategies for the Initial Reference Volume $\mathbf{V}_{\mathbf{r}}^{(0)}$. DS3-30-100 (no reset), DS3-30-101 (reset at T), DS3-30-102 (reset at $T/2$), and DS3-30-103 (reset at $T/4$) are compressively sampled at 30%.



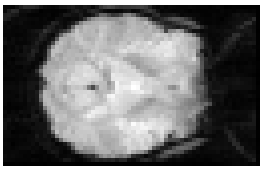
(a) Lossless scan



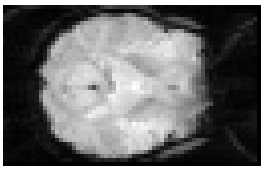
(b) RefCS with no reset



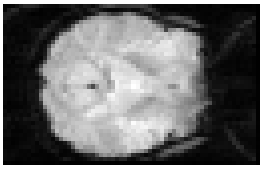
(c) RefCS/LS with no reset



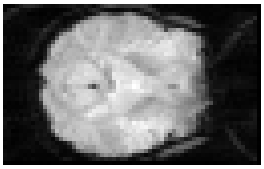
(d) RefCS with reset at T



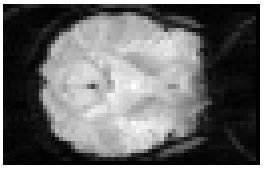
(e) RefCS/LS with reset at T



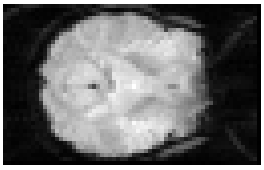
(f) RefCS with reset at $T/2$



(g) RefCS/LS with reset at $T/2$



(h) RefCS with reset at $T/4$



(i) RefCS/LS with reset at $T/4$

Figure 5.16: Examples of Dataset 4 reconstruction using Referenced CS and Referenced CS/LS employing different reset strategies for the Initial Reference Volume $\mathbf{V}_{\mathbf{r}}^{(0)}$. DS4-30-100 (no reset), DS4-30-101 (reset at T), DS4-30-102 (reset at $T/2$), and DS4-30-103 (reset at $T/4$) are compressively sampled at 30%.

Table 5.7: Average PSNR of reconstruction using different reset strategies for Initial Reference Volume $\mathbf{V}_{\mathbf{r}}^{(0)}$

		DS1	DS2	DS3	DS4
No reset	RefCS	59.47	39.31	48.01	44.04
	RefCS/LS	56.41	37.90	48.09	44.13
Reset at T	RefCS	**60.32**	42.36	52.06	46.14
	RefCS/LS	57.57	42.00	52.29	46.28
Reset at $T/2$	RefCS	60.29	42.99	52.40	46.31
	RefCS/LS	57.63	42.86	52.61	46.49
Reset at $T/4$	RefCS	60.25	**43.38**	52.64	46.25
	RefCS/LS	58.41	43.33	**52.83**	**46.42**

of quality. In this particularly important because, in the real fMRI experiments, the first few volumes acquired are especially unstable due to the physical limitation of the system.

5.4.3 Study on the Reference Update Strategy

In this study, we are going to compare between several update strategies for the reference volume. Even though we have demonstrated in Chapter 4 that the RGA with optimal learning parameter can provide a good reference that negate the propagation of the reconstruction error in natural video sequences, its effect on the fMRI data is still uncertain. Because the fMRI data can be viewed as a collection of temporally-sequenced data, just like video sequences, it is trivial that the same reconstruction error propagation problem exists in the fMRI data. The difference, however, is the different characteristics of the video sequence and fMRI. In video sequences, the changes are typically spatially object displacements, *i.e.*, objects moving around the frame. In contrast, objects in the fMRI data are mostly stationary, and movements are undesirable. Thus, the changes in the fMRI data are typically magnitude-wise, very similar to the video sequences with low-activity level.

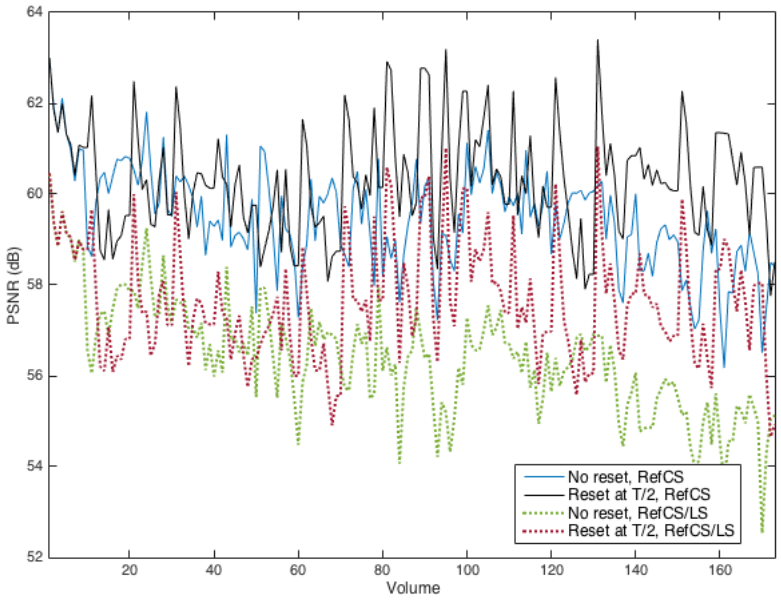


Figure 5.17: PSNR variation across Dataset 1 with different reset strategies

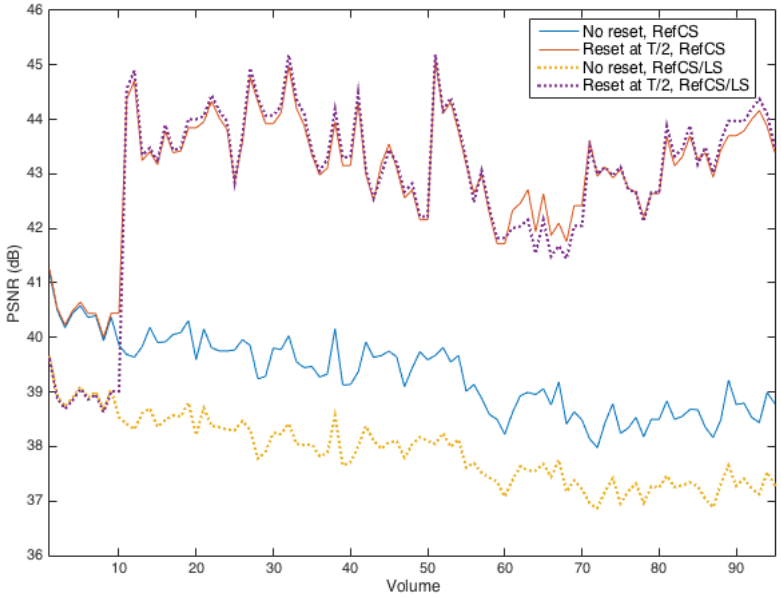


Figure 5.18: PSNR variation across Dataset 2 with different reset strategies

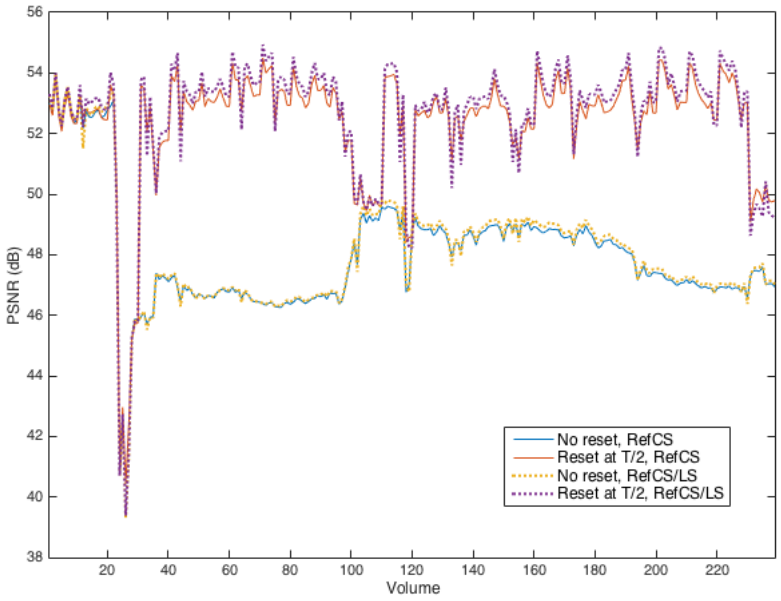


Figure 5.19: PSNR variation across Dataset 3 with different reset strategies

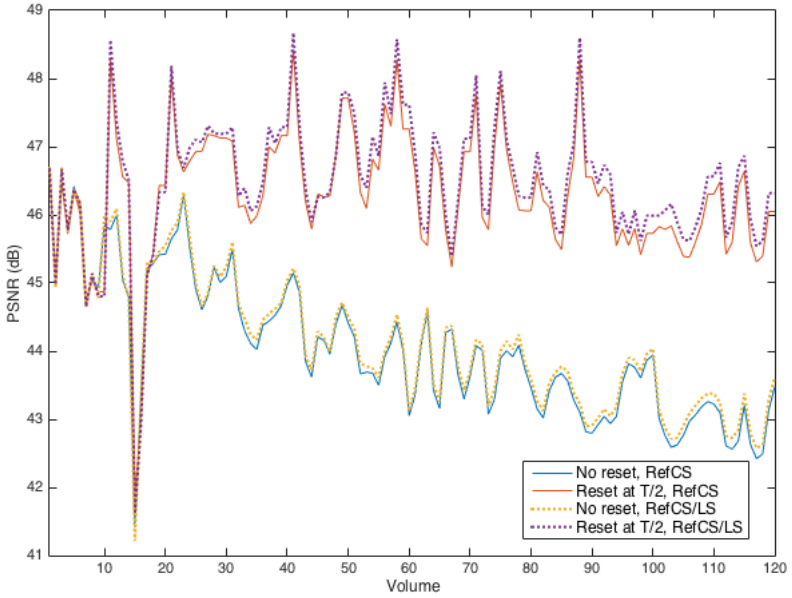


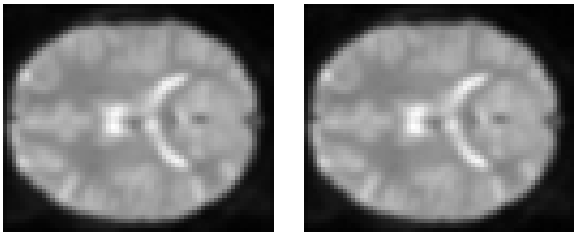
Figure 5.20: PSNR variation across Dataset 4 with different reset strategies

Table 5.8: Average PSNR of reconstruction using different reference update strategies

		DS1	DS2	DS3	DS4
No update	RefCS	60.2930	42.9933	52.4020	46.3077
	RefCS/LS	57.6256	42.8647	52.6056	46.4855
Naive update	RefCS	59.0277	42.0219	51.3442	45.0733
	RefCS/LS	57.6163	42.8528	52.6070	46.4753
RGA with $\alpha = 0.1$	RefCS	60.3007	42.9970	52.4304	46.3288
	RefCS/LS	57.5834	42.8516	52.6058	46.4726
RGA with $\alpha = 0.3$	RefCS	60.1664	42.8420	52.4402	46.3275
	RefCS/LS	57.6106	42.8589	52.6062	46.4772
RGA with optimal α	RefCS	**60.4532**	**43.0626**	52.4727	46.3965
	RefCS/LS	57.6224	42.8563	**52.6074**	**46.4889**

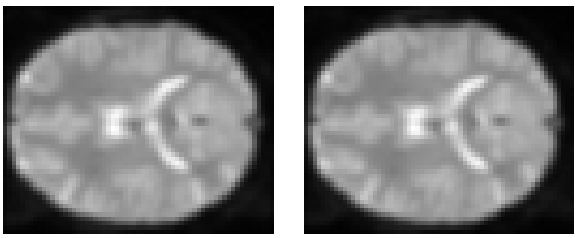
To study the reference update, here we compare the visual quality and the activity map of reconstructed data obtained using different update strategies. Figure shows the examples of the reconstructed data using no update ($\mathbf{V}_{\mathbf{r}}^{(t)} = \mathbf{V}_{\mathbf{r}}^{(0)}$) (flag 0), naive reference ($\mathbf{V}_{\mathbf{r}}^{(t)} = \hat{\mathbf{V}}^{(t-1)}$) (flag 1), and RGA-based reference (flag 2) with $\alpha = 0.1, 0.3$, and the optimal α presented in Section 4.1.3. Again, all other parameters are fixed; namely, the choice of $\mathbf{V}_{\mathbf{r}}^{(0)}$ is pre-scan, the reference is reset at every period T, and the sampling rate is 30% throughout. Examples of the results using different update strategies are shown in Figure 5.21–Figure 5.24.

Table 5.8 and Figure 5.25–Figure 5.28 show the quality metric using the PSNR. These results verify that the RGA-based reference with optimal learning parameter indeed yields the best reconstruction quality. The more interesting observation is that, in fact, any low-value fixed learning parameter, such as $\alpha = 0.1$ or $\alpha = 0$ (no update), also yield good results compared to the naive reference. The naive reference (where $\alpha = 1$), also known as the TCS, shows the worst performance among all the possible strategies. This is following the result presented in Section 4.1.2 that the small value of α gives a better performance than the large value of α.



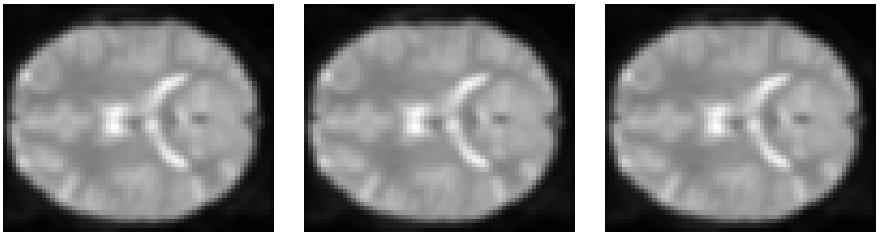
(a) RefCS with no update

(b) RefCS with naive update



(c) RefCS/LS with no update

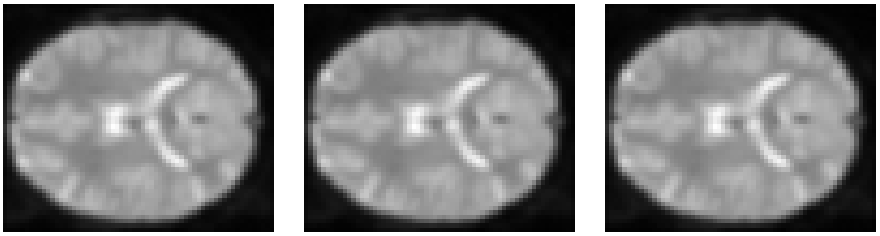
(d) RefCS/LS with naive update



(e) RefCS with RGA, $\alpha = 0.1$

(f) RefCS with RGA, $\alpha = 0.3$

(g) RefCS with RGA, optimal α

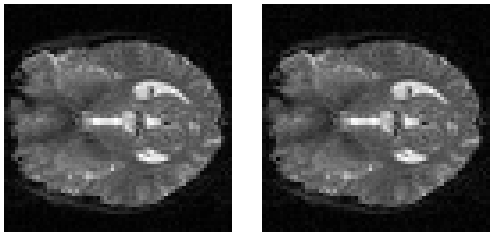


(h) RefCS/LS with RGA, $\alpha = 0.1$

(i) RefCS/LS with RGA, $\alpha = 0.3$

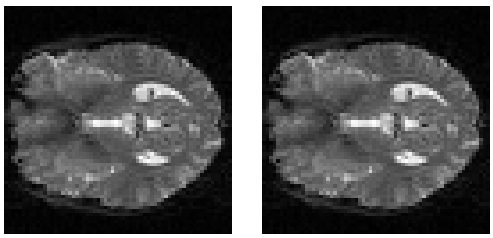
(j) RefCS/LS with RGA, optimal α

Figure 5.21: Examples of Dataset 1 reconstruction using Referenced CS and Referenced CS/LS employing different reference update strategies. DS1-30-102 (no update), DS1-30-112 (naive update), DS1-30-122 (RGA update with $\alpha = 0.1, 0.3, 0.5$) are compressively sampled at 30%.



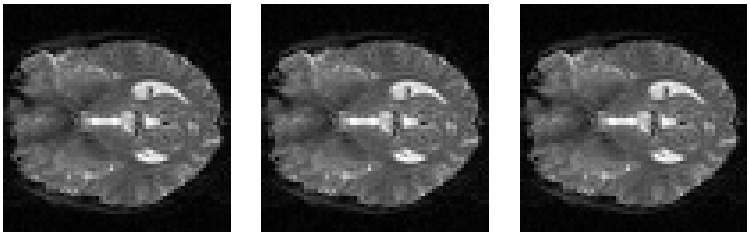
(a) RefCS with no update

(b) RefCS with naive update



(c) RefCS/LS with no update

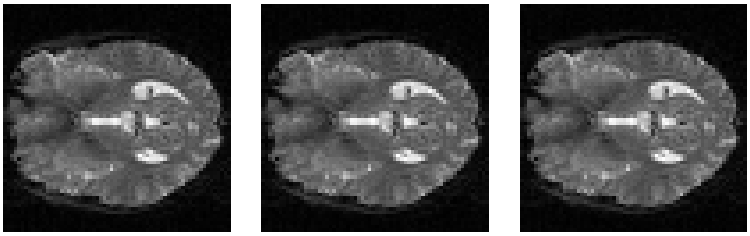
(d) RefCS/LS with naive update



(e) RefCS with RGA, $\alpha = 0.1$

(f) RefCS with RGA, $\alpha = 0.3$

(g) RefCS with RGA, optimal α

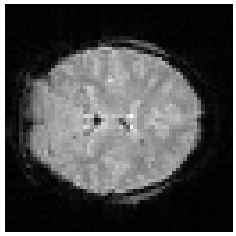


(h) RefCS/LS with RGA, $\alpha = 0.1$

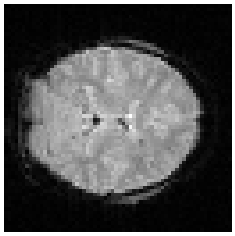
(i) RefCS/LS with RGA, $\alpha = 0.3$

(j) RefCS/LS with RGA, optimal α

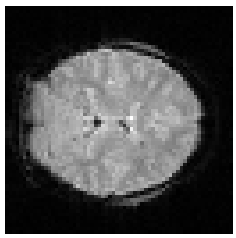
Figure 5.22: Examples of Dataset 2 reconstruction using Referenced CS and Referenced CS/LS employing different reference update strategies. DS2-30-102 (no update), DS2-30-112 (naive update), DS2-30-122 (RGA update with $\alpha = 0.1, 0.3, 0.5$) are compressively sampled at 30%.



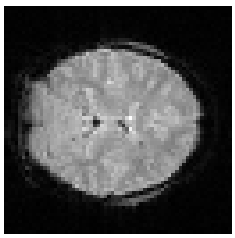
(a) RefCS with no update



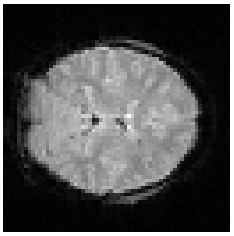
(b) RefCS with naive update



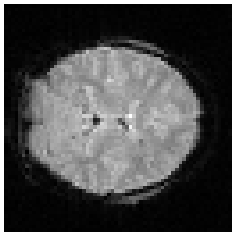
(c) RefCS/LS with no update



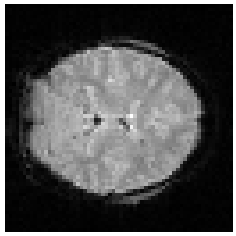
(d) RefCS/LS with naive update



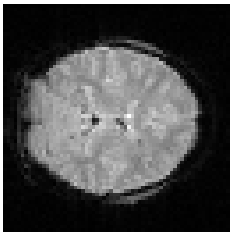
(e) RefCS with RGA, $\alpha = 0.1$



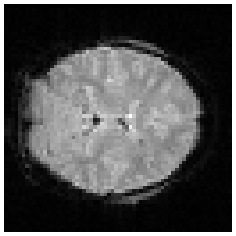
(f) RefCS with RGA, $\alpha = 0.3$



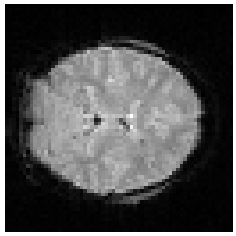
(g) RefCS with RGA, optimal α



(h) RefCS/LS with RGA, $\alpha = 0.1$



(i) RefCS/LS with RGA, $\alpha = 0.3$

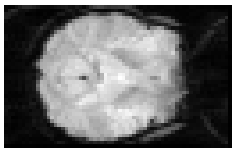


(j) RefCS/LS with RGA, optimal α

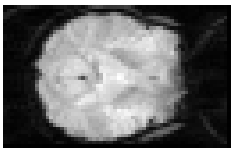
Figure 5.23: Examples of Dataset 3 reconstruction using Referenced CS and Referenced CS/LS employing different reference update strategies. DS3-30-102 (no update), DS3-30-112 (naive update), DS3-30-122 (RGA update with $alpha = 0.1, 0.3, 0.5$) are compressively sampled at 30%.

(a) RefCS with no update

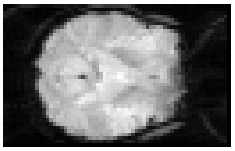
(b) RefCS with naive update



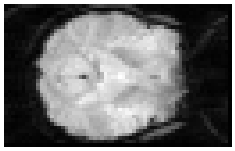
(c) RefCS/LS with no update



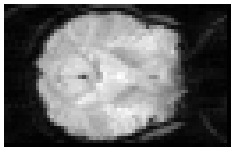
(d) RefCS/LS with naive update



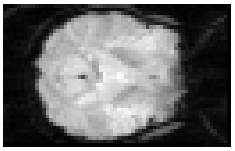
(e) RefCS with RGA, $\alpha = 0.1$



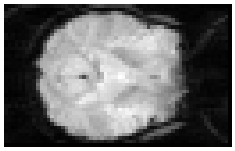
(f) RefCS with RGA, $\alpha = 0.3$



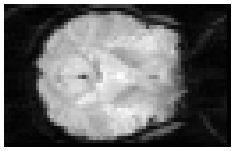
(g) RefCS with RGA, optimal α



(h) RefCS/LS with RGA, $\alpha = 0.1$



(i) RefCS/LS with RGA, $\alpha = 0.3$



(j) RefCS/LS with RGA, optimal α

Figure 5.24: Examples of Dataset 4 reconstruction using Referenced CS and Referenced CS/LS employing different reference update strategies. DS4-30-102 (no update), DS4-30-112 (naive update), DS4-30-122 (RGA update with $alpha = 0.1, 0.3, 0.5$) are compressively sampled at 30%.

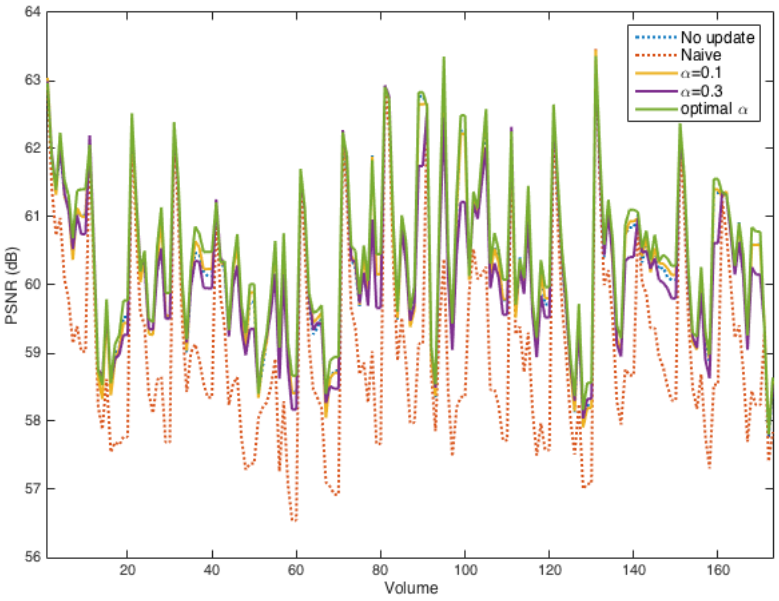


Figure 5.25: PSNR variation across Dataset 1 with different reference update strategies

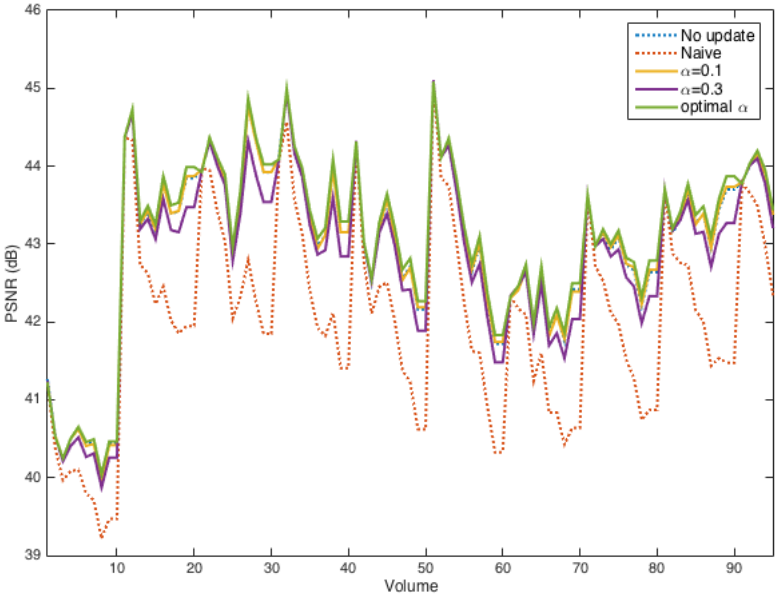


Figure 5.26: PSNR variation across Dataset 2 with different reference update strategies

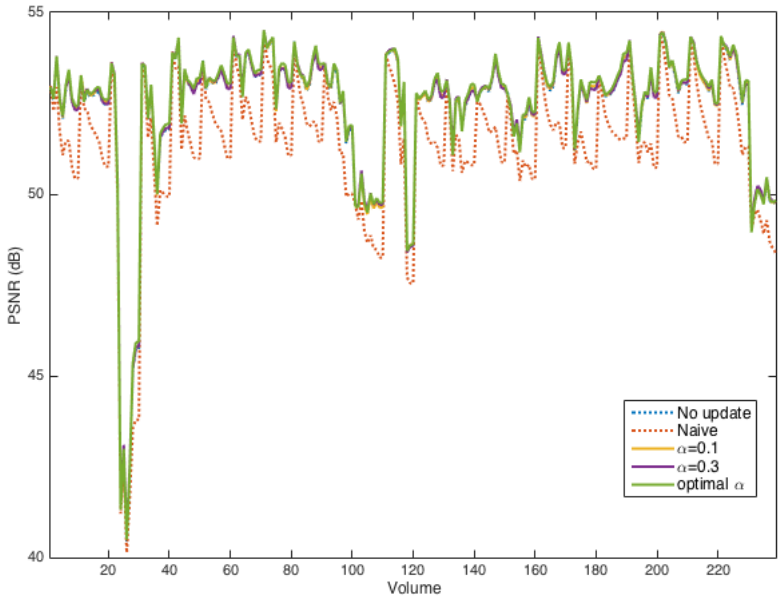


Figure 5.27: PSNR variation across Dataset 3 with different reference update strategies

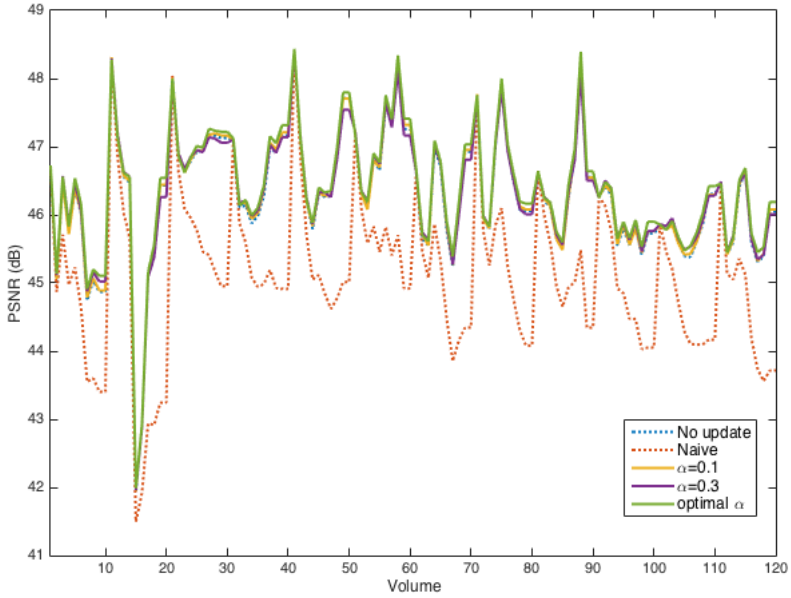


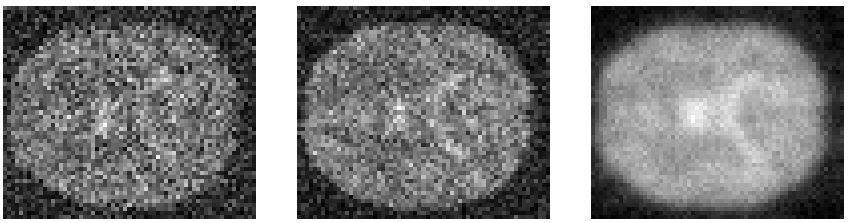
Figure 5.28: PSNR variation across Dataset 4 with different reference update strategies

5.4.4 Comparison Between Reconstruction Strategies

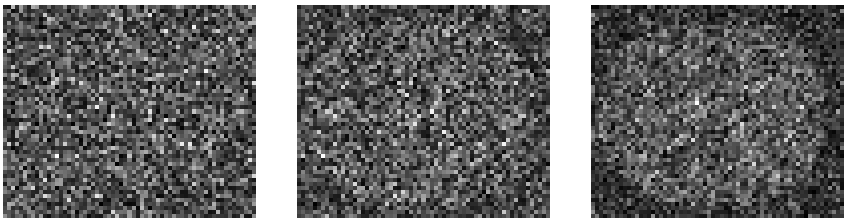
Now that we have identified the optimal parameters setting for the proposed Referenced CS method, it is now possible to compare the performance between Referenced CS and the traditional l_1-norm minimisation reconstruction. Specifically, both visual quality and the activity map are to be compared with the reconstructed data using a) the l_1-norm minimisation, b) the l_2-norm minimisation (the least squares), c) Referenced CS, and d) Referenced CS/LS. The parameters for Referenced CS are fixed throughout this study as the optimal parameters obtained from previous studies. Not only several reconstruction strategies are employed, here the reconstruction is also performed using several sampling rates, *i.e.*, sampling rate $s = 0.1$ (10%), 0.3 (30%) and 0.5 (50%) are used.

Figure 5.29–Figure 5.32 display the examples of reconstructed data using the l_1-norm minimisation against Referenced CS at several sampling rates. Subjectively, it is clear that the reconstruction quality of the Referenced CS reconstruction results is far superior to the l_1-norm minimisation results as the absence of reconstruction error is obvious. Interestingly, while the results of the least squares are completely abysmal, the results of Referenced CS/LS have good reconstruction quality, despite the fact that it employs the similar linear reconstruction as the least squares.

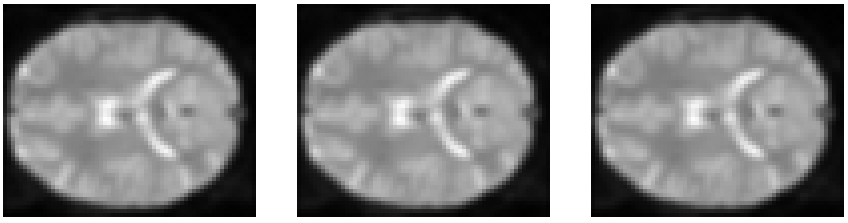
Objectively, Table 5.9 and Figure 5.33–Figure 5.36 verify the result of subjective observation using the PSNR. Not only that Referenced CS yields 2–3 times higher PSNR than the l_1-norm minimisation, they also show that, in general, the performance of Referenced CS/LS is comparable to that of Referenced CS. It should be noted that these dramatic



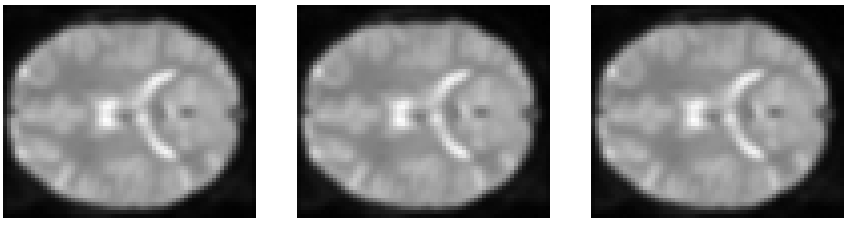
(a) l_1-min at $s = 0.1$ (b) l_1-min at $s = 0.3$ (c) l_1-min at $s = 0.5$



(d) Least squares at $s = 0.1$ (e) Least squares at $s = 0.3$ (f) Least squares at $s = 0.5$

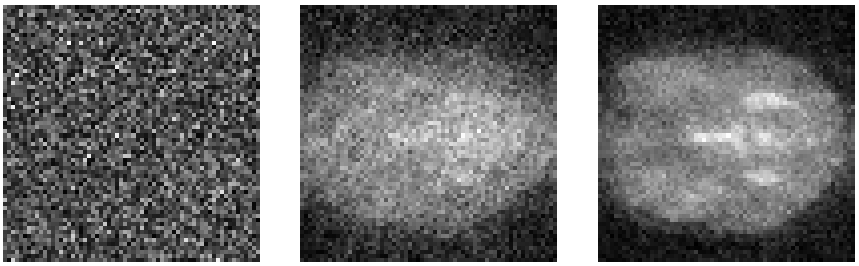


(g) RefCS at $s = 0.1$ (h) RefCS at $s = 0.3$ (i) RefCS/LS at $s = 0.5$

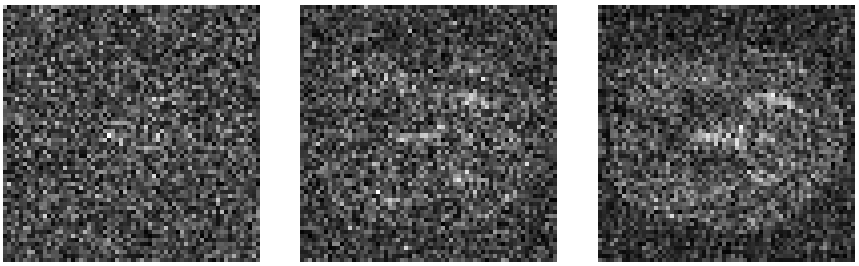


(j) RefCS/LS at $s = 0.1$ (k) RefCS/LS at $s = 0.3$ (l) RefCS/LS at $s = 0.5$

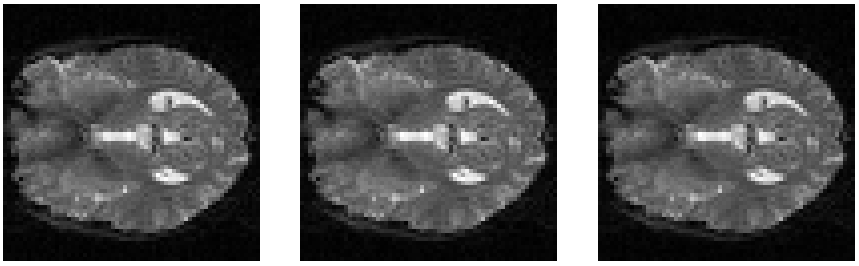
Figure 5.29: Examples of Dataset 1 reconstruction using l_1-norm minimisation, Least Squares, Referenced CS and Referenced CS/LS. The reconstruction is done on the compressively sampled data using the sampling rate of 10% ($s = 0.1$), 30% ($s = 0.3$), and 50% ($s = 0.5$).



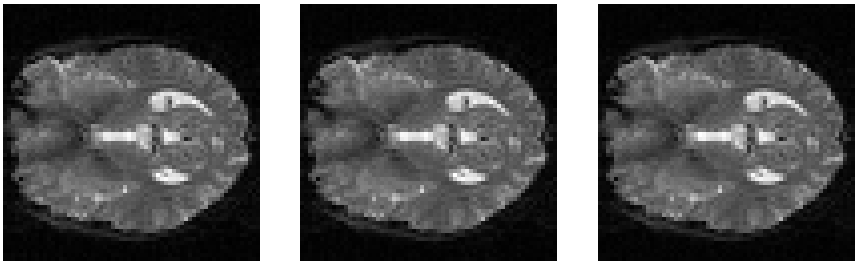
(a) l_1-min at $s = 0.1$ (b) l_1-min at $s = 0.3$ (c) l_1-min at $s = 0.5$



(d) Least squares at $s = 0.1$ (e) Least squares at $s = 0.3$ (f) Least squares at $s = 0.5$

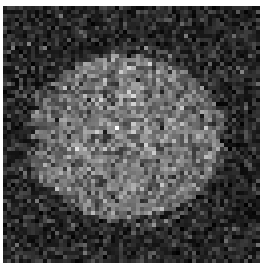


(g) RefCS at $s = 0.1$ (h) RefCS at $s = 0.3$ (i) RefCS/LS at $s = 0.5$

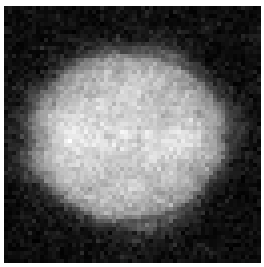


(j) RefCS/LS at $s = 0.1$ (k) RefCS/LS at $s = 0.3$ (l) RefCS/LS at $s = 0.5$

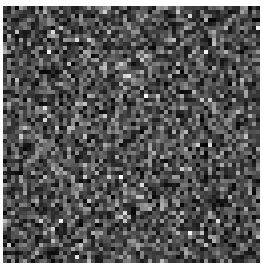
Figure 5.30: Examples of Dataset 2 reconstruction using l_1-norm minimisation, Least Squares, Referenced CS and Referenced CS/LS. The reconstruction is done on the compressively sampled data using the sampling rate of 10% ($s = 0.1$), 30% ($s = 0.3$), and 50% ($s = 0.5$).



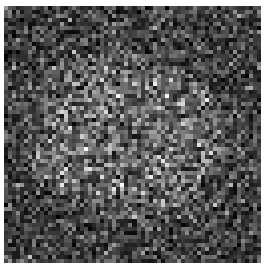
(a) l_1-min at $s = 0.1$



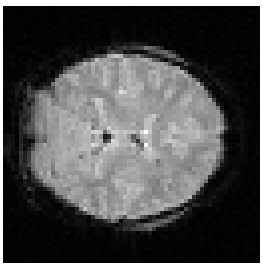
(b) l_1-min at $s = 0.3$



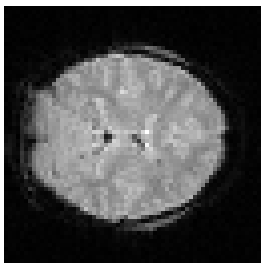
(c) Least squares at $s = 0.1$



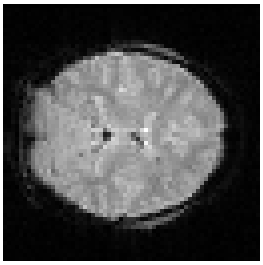
(d) Least squares at $s = 0.3$



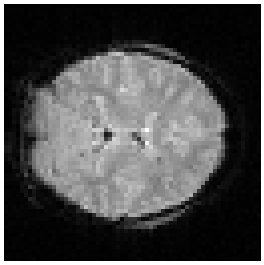
(e) RefCS at $s = 0.1$



(f) RefCS at $s = 0.3$

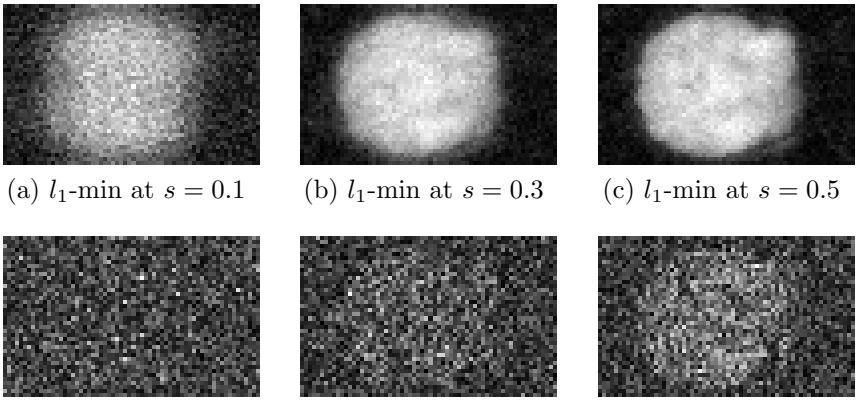


(g) RefCS/LS at $s = 0.1$



(h) RefCS/LS at $s = 0.3$

Figure 5.31: Examples of Dataset 3 reconstruction using l_1-norm minimisation, Least Squares, Referenced CS and Referenced CS/LS. The reconstruction is done on the compressively sampled data using the sampling rate of 10% ($s = 0.1$), and 30% ($s = 0.3$).



(a) l_1-min at $s = 0.1$ (b) l_1-min at $s = 0.3$ (c) l_1-min at $s = 0.5$

(d) Least squares at $s = 0.1$ (e) Least squares at $s = 0.3$ (f) Least squares at $s = 0.5$

(g) RefCS at $s = 0.1$ (h) RefCS at $s = 0.3$ (i) RefCS/LS at $s = 0.5$

(j) RefCS/LS at $s = 0.1$ (k) RefCS/LS at $s = 0.3$ (l) RefCS/LS at $s = 0.5$

Figure 5.32: Examples of Dataset 4 reconstruction using l_1-norm minimisation, Least Squares, Referenced CS and Referenced CS/LS. The reconstruction is done on the compressively sampled data using the sampling rate of 10% ($s = 0.1$), 30% ($s = 0.3$), and 50% ($s = 0.5$).

Table 5.9: Average PSNR of reconstruction using different reconstruction algorithms

sampling rate s	Algorithm	DS1	DS2	DS3	DS4
0.1	l_1-min	24.97	16.22	26.35	21.79
	Least squares	20.67	20.53	24.09	16.03
	RefCS	56.35	41.19	50.86	44.69
	RefCS/LS	56.05	41.27	51.03	44.92
0.3	l_1-min	25.56	24.94	34.86	26.03
	Least squares	22.28	22.04	25.59	17.62
	RefCS	60.14	42.85	52.25	46.11
	RefCS/LS	57.60	42.86	52.60	46.48
0.5	l_1-min	32.87	25.24	36.12	30.21
	Least squares	23.96	23.61	26.11	19.18
	RefCS	65.23	44.78	55.32	48.01
	RefCS/LS	59.59	44.84	55.89	48.45

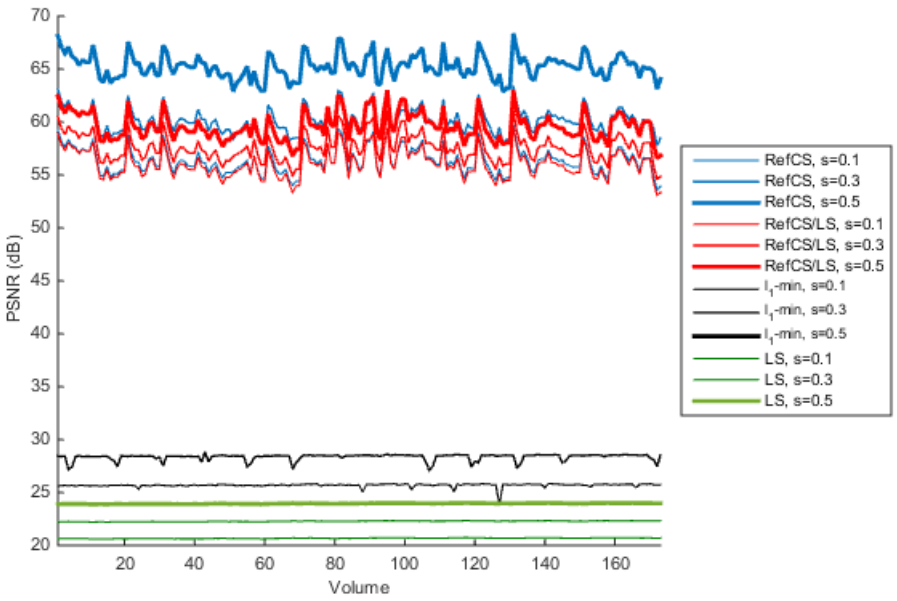


Figure 5.33: PSNR variation across Dataset 1 using different reconstruction strategies and sampling rate s

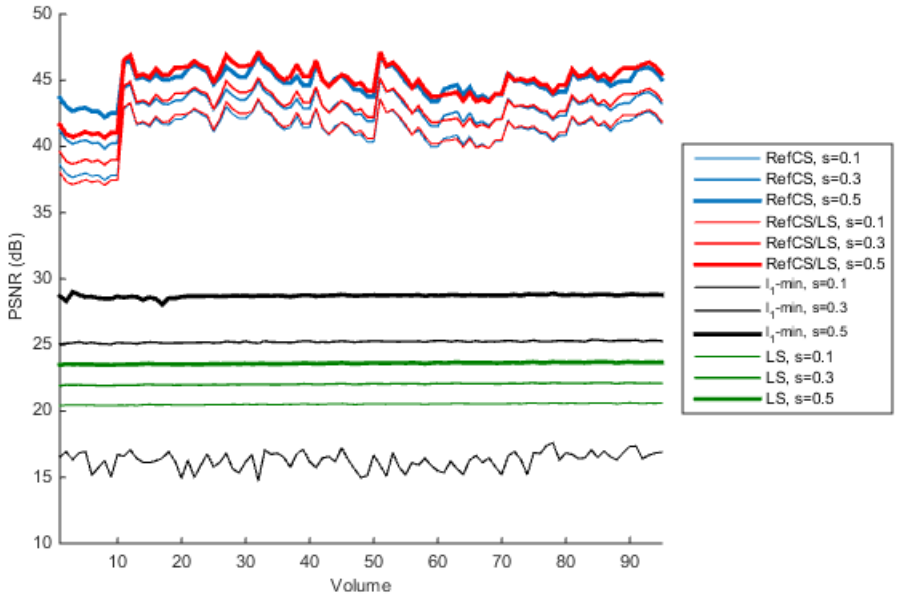


Figure 5.34: PSNR variation across Dataset 2 using different reconstruction strategies and sampling rate s

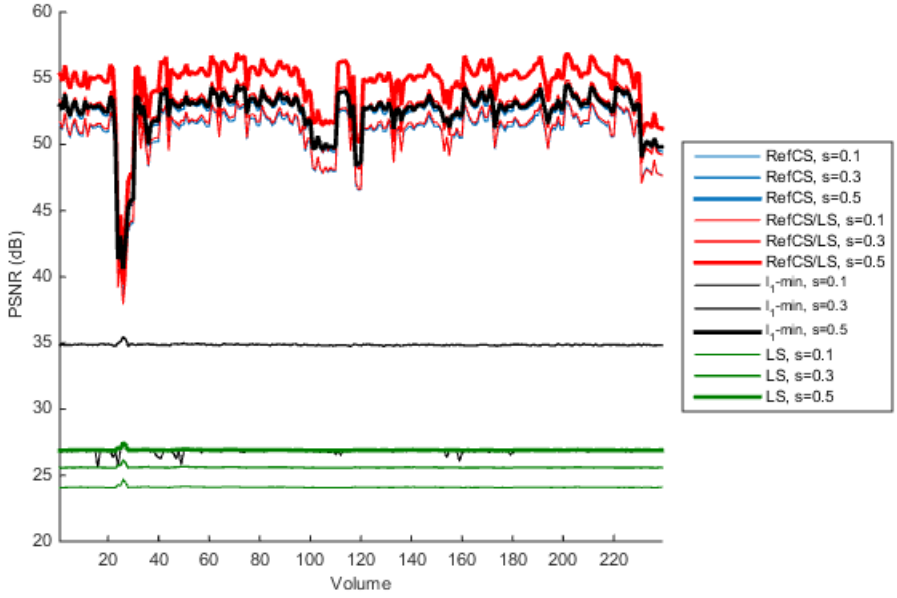


Figure 5.35: PSNR variation across Dataset 3 using different reconstruction strategies and sampling rate s

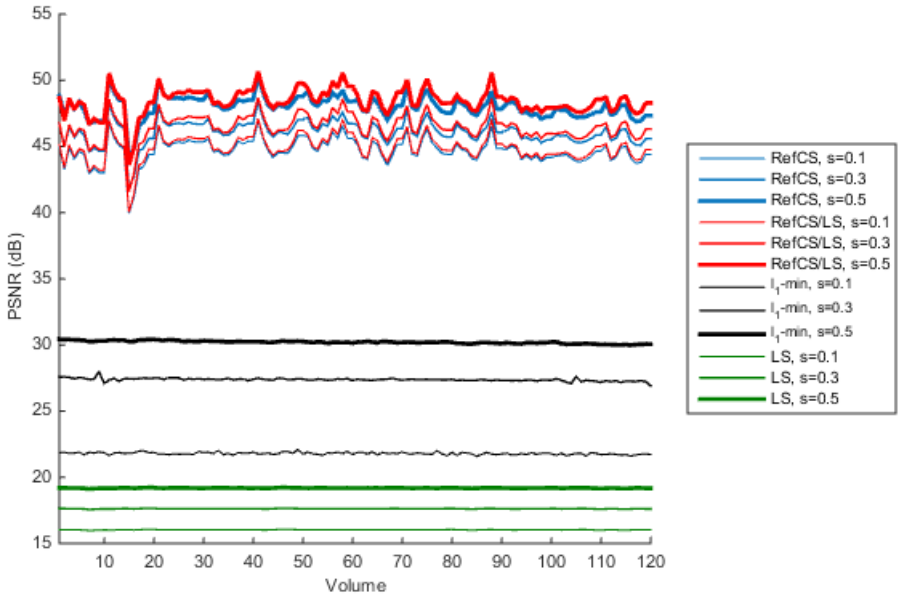


Figure 5.36: PSNR variation across Dataset 4 using different reconstruction strategies and sampling rate s

results require an accurate Initial Reference Volume $\mathbf{V}_\mathbf{r}^{(0)}$, such as the lossless pre-scan $\mathbf{V}_\mathbf{r}^{(0)}$.

5.4.5 Comparison with Conventional Fast Acquisition Techniques

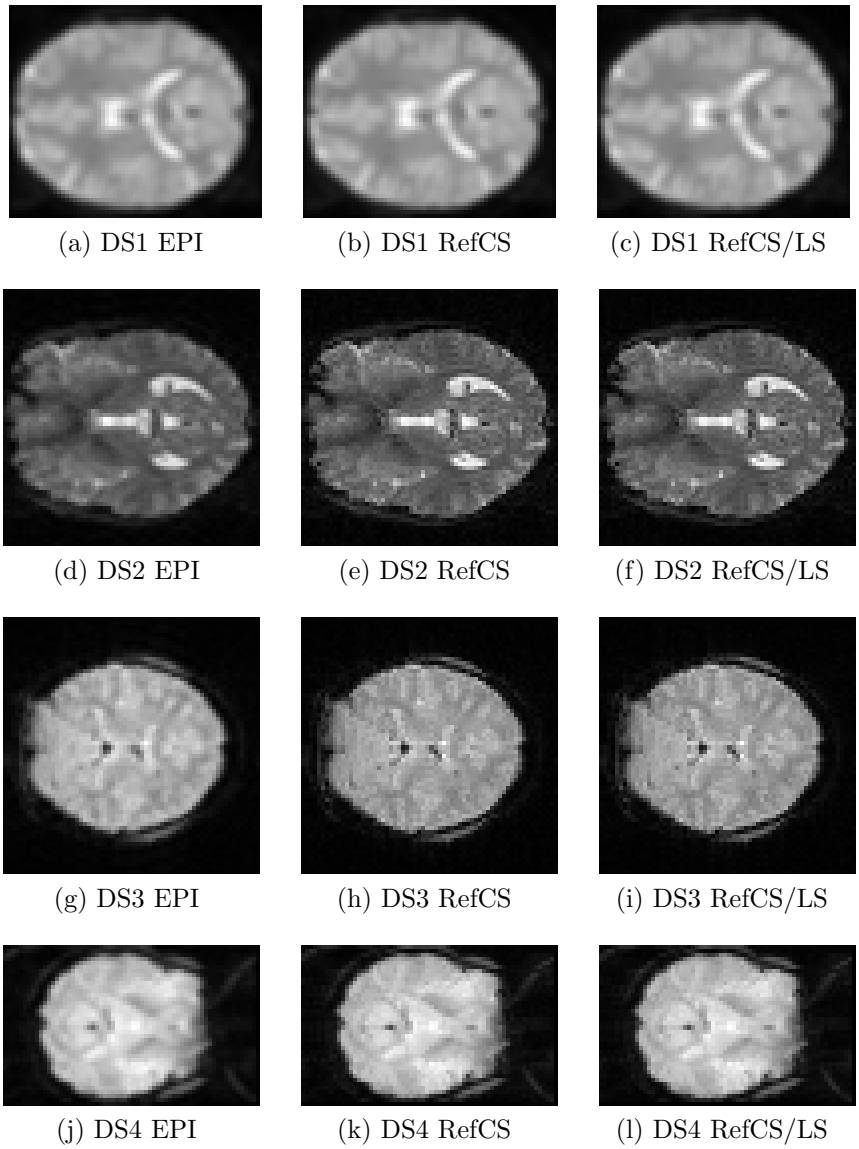
One of the biggest questions on the usefulness of the compressed sensing fMRI is how it is compared to the fast acquisition techniques currently employed—such as EPI. One clear distinction between the compressive sensing fMRI and EPI is their different goals. As discussed in Section 2.1.3, EPI aims to improve the acquisition rate by sacrificing the spatial resolution. In contrast, the goal of the compressed sensing fMRI is to reconstruct the full resolution data, albeit with some reconstruction error, from the same amount of samples acquired by EPI. Thus, this approach can be viewed as either the improved spatial resolution of

the undersampled acquisition or the improved temporal resolution of the full-size acquisition.

To verify this aim, this study compares the error of the compressed sensing reconstructed data to the error of the low-resolution EPI data, using the lossless data as the ground truth. Unfortunately, because we have no opportunity to perform the experiment with live subjects in a physical scanner, the EPI data used in this study are procedurally generated from the test datasets.

Figure 5.37 compares the examples of data from the compressed sensing reconstruction against the low-resolution data obtained using EPI, where the number of K-Space sample in both cases are fixed at 30%. Unlike in compressed sensing, in the case of EPI, the sampling rate is the ratio between the area of acquired K-space to the area of total K-space. From the results in Figure 5.37, it can be seen that the Referenced CS results have better fine details. The sharpen edges in the Referenced CS results can be observed quite clearly in Dataset 2 and Dataset 3. This is to be expected because EPI data only contains the low frequency coefficients of the K-Space, as opposed to the Referenced CS data that contains the entire fully-reconstructed K-Space.

Objectively, Figure 5.38–Figure 5.41 show the PSNR across the entire datasets. These figures show that, apart from some outliers, most of the errors in Dataset 2 and Dataset 3 is due to sudden subject movements, in which case Referenced CS yields higher PSNR than EPI. In this particular case of 30% sample, on average, Referenced CS and Referenced CS/LS yield 10.3% and 9.6% more PSNR than EPI respectively.



(a) DS1 EPI (b) DS1 RefCS (c) DS1 RefCS/LS

(d) DS2 EPI (e) DS2 RefCS (f) DS2 RefCS/LS

(g) DS3 EPI (h) DS3 RefCS (i) DS3 RefCS/LS

(j) DS4 EPI (k) DS4 RefCS (l) DS4 RefCS/LS

Figure 5.37: Examples of EPI result compared with the results of Referenced CS and Referenced CS/LS. All results shown are obtained from 30% of sample.

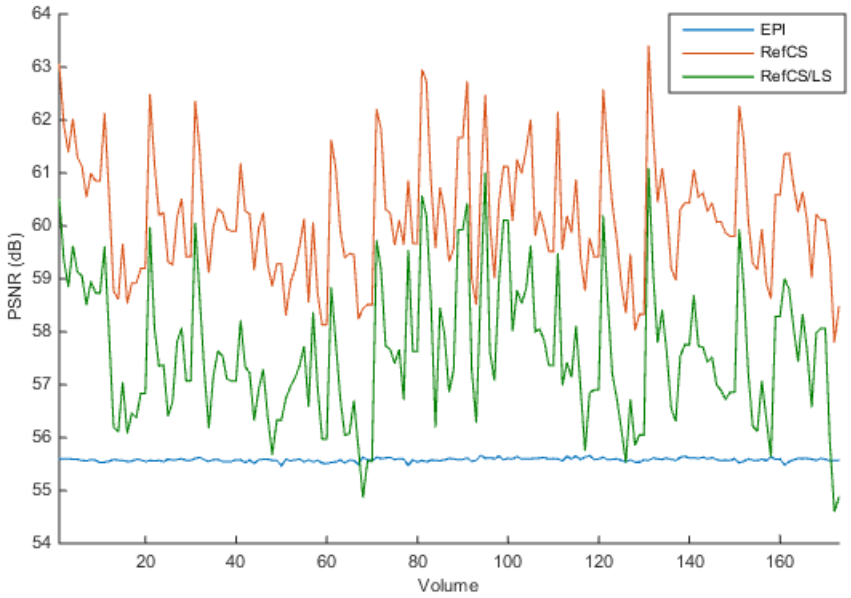


Figure 5.38: PSNR variation across Dataset 1 using EPI, Referenced CS, and Referenced CS/LS with sampling rate at 30%.

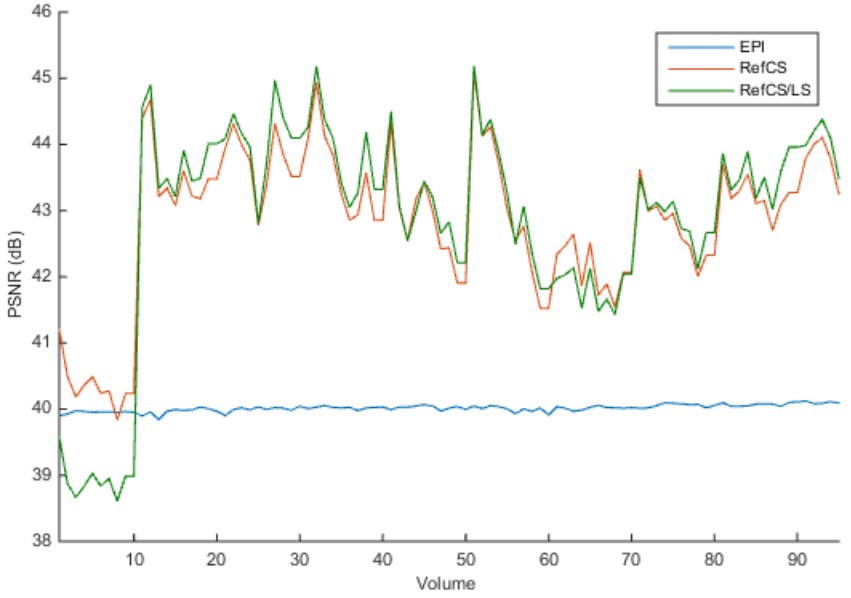


Figure 5.39: PSNR variation across Dataset 2 using EPI, Referenced CS, and Referenced CS/LS with sampling rate at 30%.

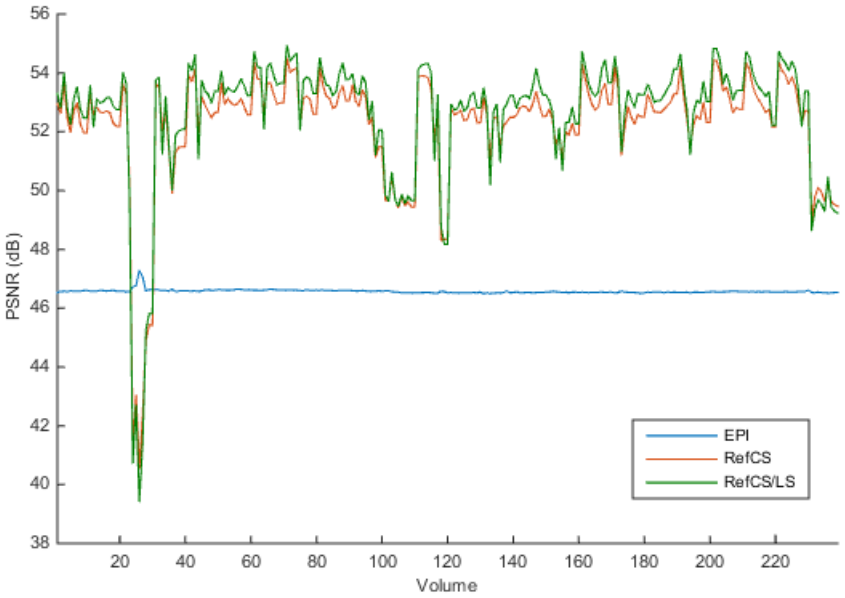


Figure 5.40: PSNR variation across Dataset 3 using EPI, Referenced CS, and Referenced CS/LS with sampling rate at 30%.

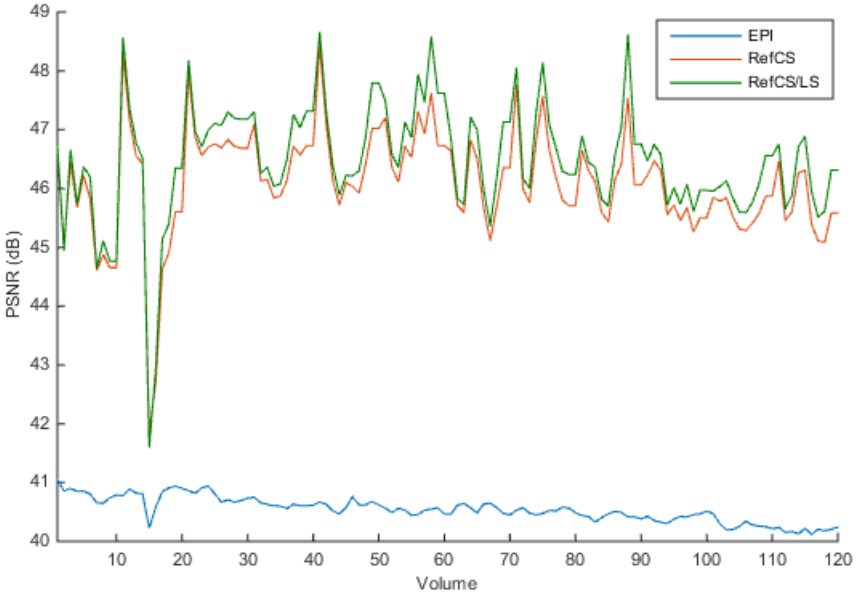
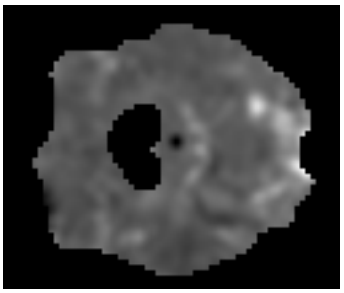


Figure 5.41: PSNR variation across Dataset 4 using EPI, Referenced CS, and Referenced CS/LS with sampling rate at 30%.

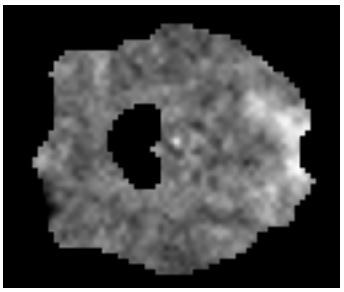
5.5 Accuracy and Preservation of Clinical Features

In this section, the reconstructed fMRI data is evaluated in terms of the accuracy and preservation of clinical features, namely, the brain activity map. One of the most common practices to obtain the brain activity map is through the use of an analysis toolbox, such as SPM. In this study, SPM12 is used as a primary analysis toolbox. Figure 5.42 and Figure 5.43 show the activity maps of Dataset 1 and Dataset2, respectively, obtained using Referenced CS and Referenced CS/LS compared against the map of the lossless data. The reconstructed data using the l_1-norm minimisation and the least squares, unfortunately, fail completely to create their resulting activity maps, and the attempt to do so are rejected by the toolbox. These figures show that the activity maps obtained from Referenced CS and Referenced CS/LS contain the noise-like fluctuation due to the reconstruction error. Despite this, it can be seen that the peak spots (positive activity regions) are situated in the correct location. The effect of these fluctuations is, in fact, small (less than 10%) compared to the activity regions. In practice, a threshold is applied to these activity maps according to the discretion of the experimenter.

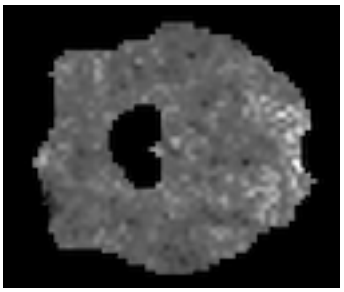
To evaluate the accuracy of the clinical features based on the maps obtained using a toolbox has a shortcoming, *i.e.*, the resulting maps depend heavily on the parameters and setting of the toolbox. These settings can be varied according to the situation and discretion of the experimenter. Also, it is possible that there exist some bias between the settings and the acquisition method; for example, the setting which yields the best result for the low-resolution data but disregard high-frequency contents are



(a) Lossless

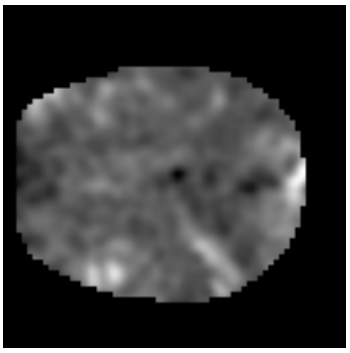


(b) Referenced CS

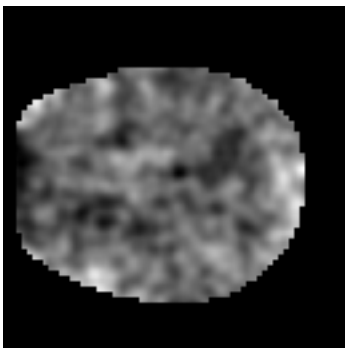


(c) Referenced CS/LS

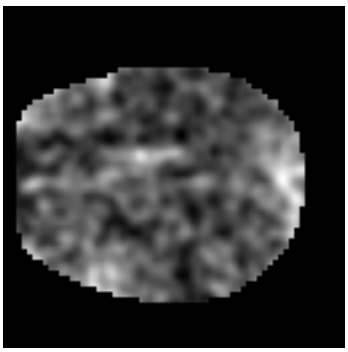
Figure 5.42: The activity map of Dataset 1 obtained from SPM toolbox. The data is reconstructed from 30% sample.



(a) Lossless



(b) Referenced CS



(c) Referenced CS/LS

Figure 5.43: The activity map of Dataset 2 obtained from SPM toolbox. The data is reconstructed from 30% sample.

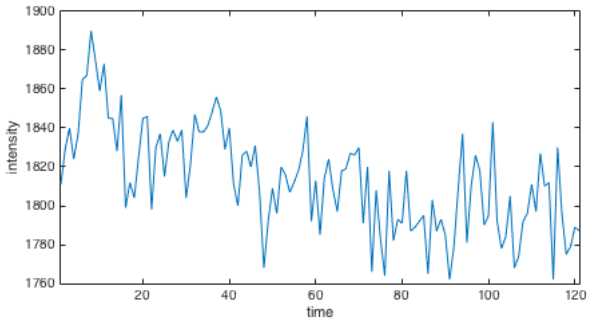
not suitable for the higher resolution acquisition method. Unfortunately, the exact relationship between the setting and the acquisition method are not included in this study due to the time constraint and requires further study on the topic.

To avoid the above complication, the metric used in this study is the normalised cross-correlation (NCC) of the voxel temporal signals. A voxel temporal signal is a one-dimensional signal of the changing magnitude of a voxel along the temporal axis. The voxel temporal magnitude signal $\mathbf{v}^{(i,j,k)}$ at the position i, j of slice k is the magnitude of a voxel $v^{(i,j,k,t)} \in \mathbf{V}^{(t)}$, for all t. Figure 5.44 shows the example of a voxel temporal signal from the active region of Dataset 2. The normalised cross-correlation $\mathcal{C}$ of the voxel temporal signals $\mathbf{v}_1$ and $\mathbf{v}_2$ is defined as

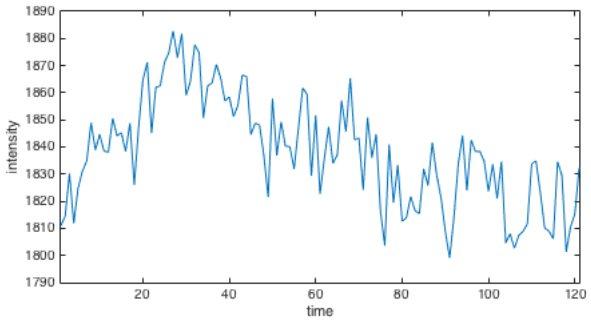
$$\mathcal{C}(\mathbf{v}_1, \mathbf{v}_2) = \frac{1}{2} \frac{(\mathbf{v}_1 - \mu_1 \mathbf{1})^{\mathrm{T}} (\mathbf{v}_2 - \mu_2 \mathbf{1})}{\sigma_1 \sigma_2}, \tag{5.6}$$

where n is the length of the signal $\mathbf{v}_1$ and $\mathbf{v}_2$, μ_1, μ_2 and σ_1, σ_2 are the mean and the standard deviation of $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. The vector $\mathbf{1}$ is a vector of all 1s with length n.

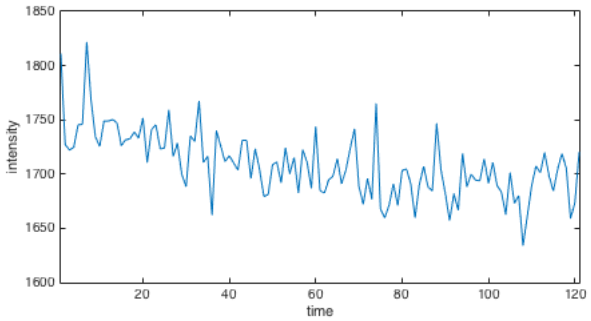
Table 5.10 shows that, in 3 out of 4 datasets, the results of Referenced CS/LS are comparable to the results of Referenced CS in terms of preserving the temporal variation of the voxel magnitude. The results of the l_1-norm minimisation, on the other hand, largely fail to preserve the voxel temporal signals especially at low sampling rates. Despite this, the correlation between these reconstructed data and the lossless data is admittedly very low. By observation, it is found that one of the biggest sources of error is due to the loss of dynamic range in the low-frequency component (Figure 5.45c). This error in the low frequency variation is referred to as a baseline error. The next section addresses this issue



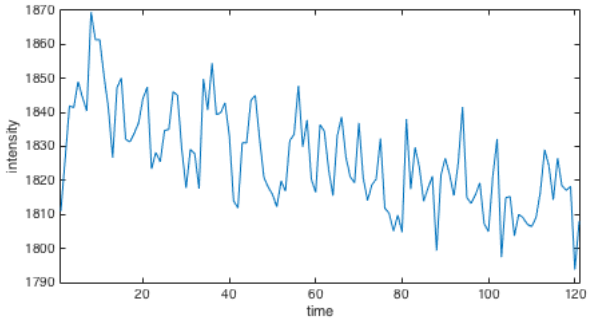
(a) Lossless



(b) l_1-min



(c) RefCS



(d) RefCS/LS

Figure 5.44: Examples of voxel temporal signal of an activated voxel. This example shows the reconstructed data using sampling rate of 30%

Table 5.10: Normalised Cross-correlation coefficient of fMRI data using different reconstruction methods

Method	Sampling rate (%)	NCC coefficient (%)			
		DS1	DS2	DS3	DS4
RefCS	10	31.10	29.37	32.28	28.35
	30	54.19	45.09	45.02	43.47
	50	70.48	59.99	58.14	58.96
RefCS/LS	10	28.68	28.35	32.84	29.90
	30	40.51	45.29	47.57	46.15
	50	52.14	61.25	61.53	61.72
l_1-min	10	14.52	11.52	8.57	8.24
	30	26.13	24.09	19.36	22.09
	50	35.24	36.77	46.10	33.75
LS	10	1.23	3.09	3.86	8.22
	30	1.61	10.75	13.72	13.17
	50	5.37	13.35	27.11	23.62

specifically and proposes several techniques to improve the correlation between the reconstructed and lossless data further.

5.6 Baseline-independent Analysis of Compressed sensing fMRI

The previous section has demonstrated that compressed sensing can be used to obtain the fMRI data from the compressive measurements. Not only it can reconstruct the visually resemble the fMRI data, but it is also possible to perform clinical analysis—specifically to obtain the brain activity map—on these reconstructed data.

The problem of the l_1-norm minimisation reconstruction error which affects the accuracy of analysis results has improved by the use of the proposed Referenced CS. However, while the issue of error propagation has been addressed by the introduction of RGA reference, it is found that the fMRI data reconstructed using Referenced CS suffers the loss

of dynamic range. The magnitude of a single activated voxel over time, shown in Figure 5.44, clearly demonstrates this phenomenon. In the fMRI data, the main feature of interest, the haemodynamic response function (HRF), lies in the high frequency variation of the signal, which appears as the shape and the local relationship. While these features, referred to as a voxel's profile, are similar in all the shown signals, the difference between each signal lies on its low frequency variation—the baseline.

This section attempts to tackle this issue by using the baseline independent (BI) analysis on these data. Because the loss of dynamic range in the Referenced CS data is due to the use of the temporal information which slows down the variation of the low frequency components, BI analysis offers a way to remove this effect from the clinical analysis results.

5.6.1 Methods and Formulations

In this work, assume that the fMRI data is composed of 2 components: profile and baseline. The profile is primarily defined by the haemodynamic response function, which directly corresponds to the stimuli presented to the brain. The profile presents itself as the high frequency variation of fMRI data. The baseline, on the other hand, is governed by other physical factors unrelated to the stimuli and presents itself as the low frequency variation. The magnitude of the profile is relatively small compared to the baseline function.

Let $\mathbf{v}$ denotes the signal of a voxel's intensity over time. The signal $\mathbf{v}$ can be modelled as

$$\mathbf{v} = \mathbf{v}_b + \mathbf{v}_p + \mathbf{e}, \tag{5.7}$$

where $\mathbf{v}_b$ denotes the baseline signal, $\mathbf{v}_p$ denotes the profile signal, and $\mathbf{e}$ denotes the acquisition error. Also let $v^{(t)}, v_b^{(t)}, v_p^{(t)}$, and $e^{(t)}$ denote a point on each signal $\mathbf{v}, \mathbf{v}_b, \mathbf{v}_p$, and $\mathbf{e}$ at time t respectively. Because only the profile $\mathbf{v}_p$ is interested in, it is desirable to minimise the effects of the baseline $\mathbf{v}_b$ and the error $\mathbf{e}$.

To remove the effect of the baseline, the first task is to estimate the baseline signal $\mathbf{v}_b$. Unlike the profile, the baseline can be estimated relatively easily. Here we consider several strategies for the baseline estimation. Once the baseline is estimated, the baseline independent voxel intensity signal $\mathbf{v}'$ can be obtained by

$$\mathbf{v}' = \mathbf{v} - \mathbf{v}_b. \tag{5.8}$$

Low-pass filtering

A straightforward way to estimate the baseline is to view it as a regular low-passed signal of $\mathbf{v}$. This strategy is especially captivating if it is assumed that the distribution of $\mathbf{v}_p$ and $\mathbf{e}$ is identical and independent. Under this assumption, the baseline can be estimated using various low-pass filtering methods.

The most direct method is to apply a window function to $\mathbf{v}$. Any low-pass windows—such as Gaussian, Hamming, or Blackman—could be used. Because window functions are applied globally, this method of estimation does not perform well with signals containing multi-scale features, as it risks incorporating large scale profiles into the estimation. The more localised low-pass filtering can be achieved using the spatial low-pass filtering such as the moving average filter. The k-point weighted

moving average can be computed as

$$b^{(t)} = \sum_{i=t-\frac{k-1}{2}}^{t+\frac{k-1}{2}} w^{(i)} v^{(i)}, \tag{5.9}$$

where k is an odd integer and $w^{(i)}$ is the weight function.

Curve estimation

A more sophisticate way of estimating the baseline is to fit the baseline function $\mathbf{v}_b$ to the voxel signal $\mathbf{v}$ such that the error between $\mathbf{v}_b$ and $\mathbf{v}$ is minimised. Specifically, assume that the baseline function $\mathbf{v}_b$ is a degree n polynomial in the form of

$$\mathbf{v}_b^{(t)} = a_n (x^{(t)})^n + a_{n-1}(x^{(t)})^{n-1} + \cdots + a_1 (x^{(t)})^1 + a_0, \tag{5.10}$$

where $\mathbf{A} = \{a_0, ..., a_n\}$ are the coefficients of $\mathbf{v}_b$, the baseline function is obtained from

$$\mathbf{v}_b = \arg\min_{\mathbf{A}} \|\mathbf{v}_b - \mathbf{v}\|_2. \tag{5.11}$$

The solution of Eq. (5.11) can be solved using any optimisation techniques.

The baseline estimated this way is a more "whole picture" approach than the previous low-pass filtering approach. It works well to extract the slowly changing nature of the baseline, without compromising the high frequency information of the profile.

5.6.2 Simulation Results

Here, the proposed BI fMRI data are created by applying the baseline estimation methods to the reconstructed fMRI data. The reconstructed

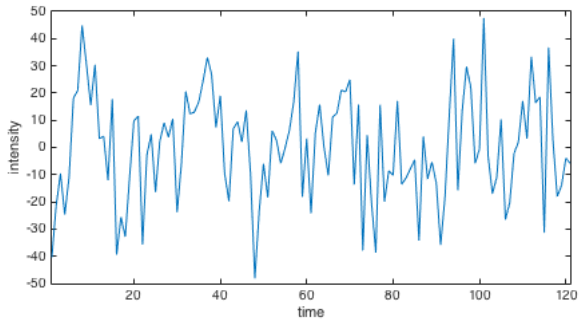
Table 5.11: Average NCC coefficients in percentage (%) across all 4 datasets between reconstructed and lossless BI data using global filtering estimation.

Method	Sampling rate (%)	Baseline-dependent	Baseline-independent	
			Gaussian	Blackman
RefCS	0.1	30.27	30.69	31.16
	0.3	46.94	47.63	48.71
	0.5	61.89	62.78	64.14
RefCS/LS	0.1	29.94	34.31	34.49
	0.3	44.88	52.09	52.32
	0.5	59.16	66.81	67.05
l_1-min	0.1	10.71	10.66	10.53
	0.3	22.92	22.50	22.32
	0.5	37.97	36.91	36.71
LS	0.1	4.10	4.22	4.17
	0.3	9.81	10.04	9.98
	0.5	17.36	17.53	17.46

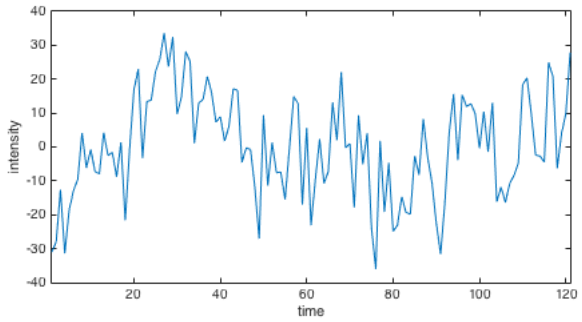
data is the same set of data used in Section 5.5. The estimation of the baseline are done using global filtering (Gaussian Window and Blackman Window), Moving Average (5-point (5-pt MA) and 10-point (10-pt MA)), and curve fitting (degree 1,2, and 3).

Figure 5.45 shows the examples of the baseline-independent activated voxel, the same voxel shown in Figure 5.44. It can be seen that the low frequency variation of the baseline is removed and that the voxel temporal signal now contains only the profile, very small in magnitude compared to the magnitude of the baseline-dependent version. It can also be noticed that the dynamic range of all data is now in a much closer range. Table 5.11–Table 5.13 show that the correlation between the BI reconstructed data and BI lossless data improves greatly in the case of Referenced CS and Referenced CS/LS data, compared to baseline-dependent version. The use of BI analysis, however, does not benefit the regular l_1-norm minimisation and the least squares results.

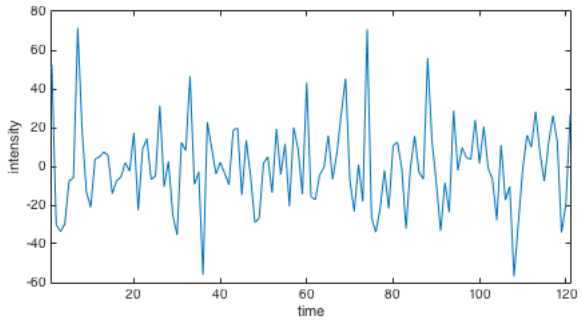
Table 5.11 shows that by using the global filtering to estimate the



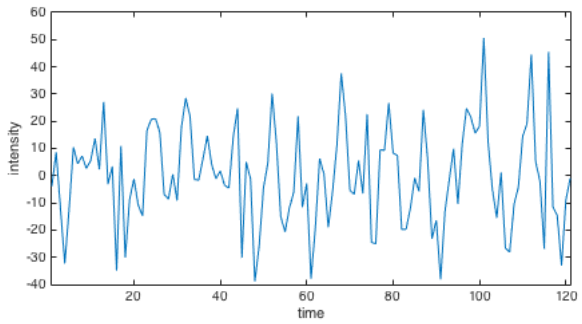
(a) Lossless



(b) l_1-min



(c) RefCS



(d) RefCS/LS

Figure 5.45: Examples of baseline-independent voxel temporal signal of an activated voxel. This example shows the reconstructed data using sampling rate of 30%. The baseline is estimated using Degree 2 Curve Fitting.

Table 5.12: Average NCC coefficients in percentage (%) across all 4 datasets between reconstructed and lossless BI data using moving average estimation.

Method	Sampling rate (%)	Baseline-dependent	Baseline-independent	
			5-pt MA	10-pt MA
RefCS	0.1	30.27	39.71	38.92
	0.3	46.94	61.16	59.76
	0.5	61.89	78.03	76.41
RefCS/LS	0.1	29.94	34.84	33.29
	0.3	44.88	52.41	50.40
	0.5	59.16	67.13	65.16
l_1-min	0.1	10.71	11.16	11.43
	0.3	22.92	22.56	21.92
	0.5	37.97	36.23	35.48
LS	0.1	4.10	3.98	4.12
	0.3	9.81	9.9	10.00
	0.5	17.36	17.26	17.33

Table 5.13: Average NCC coefficients in percentage (%) across all 4 datasets between reconstructed and lossless BI data using curve fitting estimation.

Method	Sampling rate (%)	Baseline-dependent	Baseline-independent		
			Degree 1 CF	Degree 2 CF	Degree 3 CF
RefCS	0.1	30.27	34.13	35.33	36.18
	0.3	46.94	52.99	55.10	56.39
	0.5	61.89	69.02	71.2	72.74
RefCS/LS	0.1	29.94	33.95	33.83	33.74
	0.3	44.88	51.27	51.91	53.45
	0.5	59.16	66.02	65.71	65.68
l_1-min	0.1	10.71	10.85	10.83	10.83
	0.3	22.92	22.40	22.17	22.10
	0.5	37.97	36.22	35.98	35.86
LS	0.1	4.10	4.75	5.03	5.31
	0.3	9.81	10.39	10.81	11.15
	0.5	17.36	17.79	18.1	18.42

baseline to create the BI data, the improvement of the NCC for the Referenced CS results is mere $\approx 1\%$ while the improvement of Referenced CS/LS is at least 14 %. The moving average (Table 5.12), on the other hand, shows the improvement of Referenced CS at least 25%, while the improvement of Referenced CS/LS is at about 16%. Finally, the curve fitting estimation (Table 5.13) shows the improvement of approximately 12% for both Referenced CS and Referenced CS/LS.

5.7 Summary

In this section, it has been shown that Referenced Compressed Sensing can be applied to the fMRI data successfully. The various possible settings for the compressed sensing fMRI environment are shown and evaluated. The reconstructed fMRI data using Referenced CS and Referenced CS/LS—despite much less complexity—are shown to have much better reconstruction quality than the data obtained using the l_1-norm minimisation method. It is also shown that it is possible to reconstruct the brain activity map from the reconstructed data. However, to use the method for medical purposes, the high amount of reconstruction accuracy is essential. Despite being able to construct the brain activity map from the Referenced CS data, the NCC coefficients of the reconstructed results are yet to reach an acceptable margin. The baseline-independent analysis of the data is also proposed to remedy the loss of dynamic range issue suffered by Referenced CS reconstructed data, which improves the NCC coefficient further. However, further improvements are needed before the compressed sensing fMRI can be employed for any clinical purposes.

Chapter 6

Conclusions

One of the biggest challenges of fMRI is the trade-off between the spatial and temporal resolution of the data. The current fast acquisition technique employed in the fMRI experiments sacrifices the spatial resolution in order to reduce the acquisition time required to acquire a large number of brain image volumes. One of the approaches to improve this trade-off is to incorporate the compressed sensing framework to the fMRI acquisition scheme. The main research question of this study, the possibility of applying compressed sensing to the fMRI acquisition scheme, has been thoroughly studied throughout this thesis.

In Chapter 2, it is shown that while the current acquisition technique has reached the maximum possible speed allowed by the physical limitation of the MRI theory, it is possible to use the same amount of sampled data to reconstruct the higher resolution data via compressed sensing. For compressed sensing to success, it requires two parts: the sensing operation and the reconstruction operation. This study focuses on the reconstruction operation, as it is shown that despite there are many works of literature present, there is clearly the lack of works on the compressed sensing fMRI. As most state-of-the-art MRI reconstruction

algorithms aim to reconstruct the data with high visual quality, which is ideal for diagnosis, they do not focus on the accurate reconstruction of each voxel which the fMRI studies heavily rely upon.

In Chapter 3, the novel compressed sensing reconstruction method has been proposed. The method, named Referenced Compressed Sensing, exploits either the spatial or temporal redundancy between a signal and an arbitrary reference. This method moves away from the general concept of maximising sparsity, allowing novel reconstruction scenario unsuitable for the traditional compressed sensing to be performed. It works especially well with spatio-temporal signals such as video sequences. In this chapter, it is shown that, on average, the use of Referenced CS yields at least 50% higher in terms of the PSNR when reconstructing video sequences.

The performance of Referenced CS depends heavily on the quality of the reference. When reconstructing video sequences, it is possible that the reconstruction error of one frame will propagate into the reference and get carried over to the next frame. To remedy this error propagation, the reference estimation method based on the Running Gaussian Average is proposed in Chapter 4. Also, one of the challenges of using compressed sensing in practice is the computational complexity of its iterative reconstruction algorithms. Chapter 4 addressed this issue by introducing the low complexity version of Referenced CS. The method, referred to as Referenced CS/LS, relaxes the objective function of the Referenced CS from the l_1-norm to the l_2-norm, essentially converts the problem into the linear least squares approximation. While the traditional least squares approximation fails to work with compressed sensing completely, it is shown that, by exploiting the correlated reference, the Referenced

CS/LS can yield the result comparable to the l_1-norm minimisation with at least 1200 times faster in computation time.

Finally, the proposed Referenced CS is applied to the fMRI data in Chapter 5. There are many possible settings for a practical Referenced CS-fMRI acquisition scheme. Chapter 5 explores these settings, including the Initial Reference Volume, the reset strategy, and the update strategy. The results of Referenced CS outperform the results of the l_1-norm minimisation vastly in terms of visual quality, with at least twice the value of the PSNR. It also shows that, surprisingly, the visual quality of Referenced CS/LS results are comparable to Referenced CS, despite its much faster computational time. The reconstructed fMRI data has much higher resolution than the conventional fast acquisition technique, resulting in more detailed image data. On the other hand, while it is possible to create the brain activity map from the reconstructed data with accurate activity regions, the resulting map contains some amount of noise due to the reconstruction error. One source of error is the loss of dynamic range due to the use of Referenced CS. To tackle this issue, Chapter 5 also suggested the use of baseline-independent analysis of the fMRI data, where the error in the low frequency baseline is negated.

In conclusion, by employing compressed sensing to the fMRI acquisition scheme, it is possible to obtain high-resolution data with high visual quality from a small amount of measurements. However, further development is still needed to guarantee the complete accuracy of the activity map created from the compressed sensing fMRI data.

6.1 Future works

This thesis has shown the proof-of-concept for the compressed sensing fMRI. During this study, many possibilities for further investigation arise.

For any technique to be employed in the actual clinical use, it is important that the technique is studied intensively and thoroughly. This, too, applies to the proposed method in this thesis. It is beneficial to evaluate the sensing operation and the proposed reconstruction operation in the *in vivo* experiments, *i.e.*, experiments involving live subjects.

A further research could be done in regards of the evaluation metric. Global quality metrics commonly employed to evaluate the quality of signals, such as Mean Squares Error and the PSNR, do not necessarily match the subjective quality observed by human. This is the main motivation behind many subjective-driven quality metric, such as the SSIM. As the activity map is subjected to experimenters and neurologists preferences and procedures, the only suitable way to evaluate the quality of the activity map is by using subjective expert opinion. The same concept of SSIM could be done for the fMRI data evaluation, *i.e.*, an objective metric that matches or represents the subjective quality as interested by experts. Such a metric will focus more on high-level features, such as the statistical changes in the signals, rather than just their absolute magnitude.

The baseline-independent analysis and the baseline estimations can be improved further upon. It is possible to use more advanced approximation techniques to separate the baseline from the features of interest. The extraction of baseline signal from the features of interest is a problem of source separation and, as such, many techniques can be employed. Some of these techniques, such as the Principal Components Analysis and

Independent Component Analysis, are of special interest because these techniques have already found their use in many conventional fMRI experiments.

The most important parameter that affects the performance of Referenced CS is the distance between the correlated reference $\mathbf{r}$ and the sparse signal $\mathbf{x}$. As such, a method that can blindly predict the reference distance can benefits Reference CS greatly. For example, one can use such prediction to determine whether to reconstruct the signal using Referenced CS or the conventional CS, which can be useful in many scenarios such as when a video sequence has high amount of activity or consists of scene changes. In other words, suppose that the reference distance prediction is lower than a threshold, which means that the reference is considerably different from the signal, the reconstruction can switch back to the conventional CS reconstruction for better reconstruction accuracy. This prediction is also useful to determine the efficient reference volume reset in the compressed sensing fMRI system that is more adaptive to changing conditions than the fixed period resets studied in this thesis. One potential technique to predict the reference distance could be discovered by studying the relationship between the compressed measurements $\mathbf{y} = \mathbf{\Phi x}$ and $\mathbf{y_r} = \mathbf{\Phi r}$ in more details.

The reconstruction based on Referenced CS can be greatly benefited from many advanced concepts in image and video coding. As shown in Section 3.5.2, Referenced CS can also work with the spatial redundancy in the signal. The intra coding concepts employed in latest video coding technique can be used as a way to exploit such spatial redundancy. Similarly, the temporal redundancy can be enhanced by using the motion estimation techniques. In general, it is possible for many techniques em-

ployed in the latest video coding standard, such as High Efficiency Video Coding, to be incorporated into the framework of Referenced CS.

Finally, the computational complexity is the biggest challenge of the successful compressed sensing fMRI application. While Referenced CS/LS proposed in this study provides a remedy to this issue, but as shown in Section 4.2.3 and Section 5.5, the reconstruction based on the l_1-norm can often yield superior reconstruction accuracy. A fast algorithm for l_1-norm-based iterative reconstruction methods will be of great benefit to both this specific application and compressed sensing application in the wider field. One possible way to improve the complexity of the iterative algorithms is to reduce the size of the signals. As shown in Section 4.2.1, the computational time is proportional to the length of the signal. Thus, by reducing the signal into a set of smaller signal, the iterative algorithms can be accelerated.

Bibliography

[1] S. Adams, P. R. Baum, T. Stuckensen, K. Bitter, and G. Hör. Prospective comparison of 18F-FDG PET with conventional imaging modalities (CT, MRI, US) in lymph node staging of head and neck cancer. *European Journal of Nuclear Medicine*, 25(9):1255–1260, 1998.

[2] S. Ajraoui, K. J. Lee, M. H. Deppe, S. R. Parnell, J. Parra-Robles, and J. M. Wild. Compressed sensing in hyperpolarized 3He Lung MRI. *Magnetic Resonance in Medicine*, 63(4):1059–1069, 2010.

[3] M. Akçakaya, J. Park, and V. Tarokh. Low Density Frames for Compressive Sensing. In *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 3642–3645, March 2010.

[4] M. Akçakaya, J. Park, and V. Tarokh. A Coding Theory Approach to Noisy Compressive Sensing Using Low Density Frames. *IEEE Transactions on Signal Processing*, 59(11):5369–5379, November 2011.

[5] A. Amini and F. Marvasti. Deterministic Construction of Binary, Bipolar, and Ternary Compressed Sensing Matrices. *IEEE Transactions on Information Theory*, 57(4):2360–2370, April 2011.

[6] E. Arias-Castro, E. J. Candès, and M. A. Davenport. On the Fundamental Limits of Adaptive Sensing. *IEEE Transactions on Information Theory*, 59(1):472–481, January 2013.

[7] M. S. Asif and J. Romberg. Fast and Accurate Algorithms for Re-Weighted l_1-Norm Minimization. *IEEE Transactions on Signal Processing*, 61(23):5905–5916, December 2013.

[8] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak. Toeplitz-Structured Compressed Sensing Matrices. In *Proceedings of 2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 294–298, August 2007.

[9] P. A. Bandettini and R. Cox. Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, 43(4):540–548, 2000.

[10] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A Simple Proof of the Restricted Isometry Property for Random Matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[11] D. Baron, S. Sarvotham, and R. G. Baraniuk. Bayesian Compressive Sensing Via Belief Propagation. *IEEE Transactions on Signal Processing*, 58(1):269–280, January 2010.

[12] H. Bhaskar, L. Mihaylova, and A. Achim. Video Foreground Detection Based on Symmetric Alpha-Stable Mixture Models. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(8):1133–1138, August 2010.

[13] H. Bhaskar, L. Mihaylova, and S. Maskell. Articulated human body parts detection based on cluster background subtraction and foreground matching. *Neurocomputing*, 100:58–73, 2013.

[14] C. Bilen, I. Selesnick, Y. Wang, R. Otazo, and D. K. Sodickson. A motion compensating prior for dynamic MRI reconstruction using combination of compressed sensing and parallel imaging. In *Proceedings of 2011 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–6, December 2011.

[15] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[16] T. Blumensath and M. E. Davies. Iterative Thresholding for Sparse Approximations. *Journal of Fourier Analysis and Applications*, 14(5):629–654, 2008.

[17] P. T. Boufounos and R. G. Baraniuk. 1-Bit compressive sensing. In *Proceedings of 2008 42nd Annual Conference on Information Sciences and Systems (CISS)*, pages 16–21, March 2008.

[18] A. Bourquard and M. Unser. Binary Compressed Imaging. *IEEE Transactions on Image Processing*, 22(3):1042–1055, March 2013.

[19] C. B uchel and K. J. Friston. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7(8):768–778, 1997.

[20] N. Burak Karahanoglu and H. Erdogan. A* Orthogonal Matching Pursuit: Best-First Search for Compressed Sensing Signal Recovery. *ArXiv e-prints*, September 2010.

[21] J. Caballero, A. N. Price, D. Rueckert, and H. J. V. Dictionary Learning and Time Sparsity for Dynamic MR Data Reconstruction. *IEEE Transactions on Medical Imaging*, 33(4):979–994, April 2014.

[22] E. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9–10):589–592, May 2008.

[23] E. Candes and J. Romberg. l1-magic: Recovery of sparse signals via convex programming, 2005. http://statweb.stanford.edu/ candes/l1magic/.

[24] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, April 2007.

[25] E. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, December 2005.

[26] E. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007.

[27] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, June 2007.

[28] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete fre-

quency information. *IEEE Transactions on Information Theory*, 52(2):489–509, February 2006.

[29] E. J. Candès and M. B. Wakin. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[30] N. R. Carlson. *Physiology of Behavior*. MyPsychKit Series. Allyn & Bacon, 2010.

[31] A. Carmi, T. N. Sainath, P. Gurfil, D. Kanevsky, D. Nahamoo, and B. Ramabhadran. The Use of isometric transformations and bayesian estimation in compressive sensing for fMRI classification. In *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 493–496, March 2010.

[32] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12):121105, 2008.

[33] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *Proceedings of 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI)*, pages 262–265, June 2009.

[34] K. M. Cheman. *Optimization techniques for solving basis pursuit problems*. PhD thesis, North Carolina State University, 2006.

[35] C. Chen, Y. Li, L. Axel, and J. Huang. Real time dynamic MRI by exploiting spatial and temporal sparsity. *Magnetic Resonance Imaging*, 3(4):473–482, 2016.

[36] S. Chen and D. Donoho. Basis pursuit. In *Proceedings of 1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 41–44, October 1994.

[37] P. Ciuciu, J. B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali. Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging*, 22(10):1235–1251, October 2003.

[38] R. Coifman, Y. Meyer, and F. Geshwind. Noiselets. *Applied and Computational Harmonic Analysis*, 10(1):27–44, January 2001.

[39] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino. Background subtraction for automated multisensor surveillance: A comprehensive review. *EURASIP Journal on Advances in Signal Processing*, 2010(1):343057, 2010.

[40] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image restoration by sparse 3D transform-domain collaborative filtering. In *Proceedings of SPIE Image Processing: Algorithms and Systems VI*, volume 6812, pages 681207–681212, 2008.

[41] W. Dai and O. Milenkovic. Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, May 2009.

[42] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf. Constrained adaptive sensing. *ArXiv e-prints*, June 2015.

[43] J. A. de Zwart, P. van Gelderen, P. Kellman, and J. H. Duyn. Application of sensitivity-encoded echo-planar imaging for blood oxygen level-dependent functional brain imaging. *Magnetic Resonance in Medicine*, 48(6):1011–1020, 2002.

[44] J. Deng, Z. Yang, C. Zhang, and W. Lu. Orthonormal expansion l1-minimization for compressed sensing in MRI. In *Proceedings of 2011 18th IEEE International Conference on Image Processing (ICIP)*, pages 2297–2300, September 2011.

[45] Z. Devir and M. Lindenbaum. Blind Adaptive Sampling of Images. *IEEE Transactions on Image Processing*, 21(4):1478–1487, April 2012.

[46] U. Dias and M. E. Rane. Comparative analysis of sensing matrices for compressed sensed thermal images. In *Proceedings of 2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing*, pages 265–270, March 2013.

[47] A. Divekar and O. K. Ersoy. Probabilistic matching pursuit for compressive sensing. *Department of Electrical and Computer Engineering Technical Reports, Purdue University*, 2010.

[48] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran. Sparsity adaptive matching pursuit algorithm for practical compressed sensing. In *Proceedings of 2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 581–587, October 2008.

[49] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran. Fast and Efficient Compressive Sensing Using Structurally Random Matrices. *IEEE Transactions on Signal Processing*, 60(1):139–154, January 2012.

[50] T. T. Do, T. D. Tran, and L. Gan. Fast compressive sampling with structurally random matrices. In *Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3369–3372. IEEE, 2008.

[51] D. Donoho, I. Drori, V. Stodden, Y. Tsaig, and M. Shahram. SparseLab, May 2007. http://sparselab.stanford.edu/.

[52] D. L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[53] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.

[54] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, January 2006.

[55] M. F. Duarte and R. G. Baraniuk. Kronecker Compressive Sensing. *IEEE Transactions on Image Processing*, 21(2):494–504, February 2012.

[56] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008.

[57] M. F. Duarte and Y. C. Eldar. Structured Compressed Sensing: From Theory to Applications. *IEEE Transactions on Signal Process*, 59(9):4053–4085, 2011.

[58] K. Egiazarian, A. Foi, and V. Katkovnik. Compressed Sensing Image Reconstruction Via Recursive Spatially Adaptive Filtering. In *Proceedings of 2007 IEEE International Conference on Image Processing (ICIP)*, volume 1, pages I–549–I–552, September 2007.

[59] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[60] M. Elad, B. Matalon, and M. Zibulevsky. Image Denoising with Shrinkage and Redundant Representations. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006*, volume 2, pages 1924–1931, 2006.

[61] Y. C. Eldar. Generalized sure for exponential families: Applications to regularization. *IEEE Transactions on Signal Processing*, 57(2):471–481, February 2009.

[62] M. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and Zooming Using Sparse Representations. *The Computer Journal*, 52(1):64–79, 2009.

[63] Z. Fang, N. Van Le, M. Choy, and J. H. Lee. High spatial resolution compressed sensing (HSPARSE) functional MRI. *Magnetic Resonance in Medicine*, 2015.

[64] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak. Majorization-Minimization Algorithms for Wavelet-Based Im-

age Restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, December 2007.

[65] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of l_1-balls for $0 < p \leq 1$. *Journal of Complexity*, 26(6):629–640, 2010.

[66] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.

[67] K. Friston, P. Fletcher, O. Josephs, A. Holmes, M. Rugg, and R. Turner. Event-Related fMRI: Characterizing Differential Responses. *NeuroImage*, 7(1):30–40, 1998.

[68] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994.

[69] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine*, 59(2):365–373, 2008.

[70] S. Geethanath, R. Reddy, A. S. Konar, S. Imam, R. Sundaresan, R. Babu D. R., and R. Venkatesan. Compressed Sensing MRI: A Review. *Critical Reviews in Biomedical Engineering*, 41(3):183–204, 2013.

[71] G. H. Glover. Deconvolution of Impulse Response in Event-Related BOLD fMRI. *NeuroImage*, 9(4):416–429, 1999.

[72] S. Gold, B. Christian, S. Arndt, G. Zeien, T. Cizadlo, D. Johnson, M. Flaum, and N. Andreasen. Functional MRI statistical software packages: A comparative analysis. *Human Brain Mapping*, 6(2):73–84, 1998.

[73] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[74] M. Grant and S. Boyd. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[75] J. P. Haldar, D. Hernando, and Z. P. Liang. Compressed-Sensing MRI With Random Encoding. *IEEE Transactions on Medical Imaging*, 30(4):893–903, April 2011.

[76] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65:413–497, April 1993.

[77] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti. *MRI: The Basics*. The Basics Series. Lippincott Williams & Wilkins, 2010.

[78] J. D. Haupt, W. U. Bajwa, G. M. Raz, and R. D. Nowak. Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation. *IEEE Transactions on Information Theory*, 56(11):5862–5875, November 2010.

[79] M. A. Herman and T. Strohmer. High-Resolution Radar via Compressed Sensing. *IEEE Transactions on Signal Processing*, 57(6):2275–2284, June 2009.

[80] R. J. Huster, S. Debener, T. Eichele, and C. S. Herrmann. Methods for Simultaneous EEG-fMRI: An Introductory Review. *The Journal of Neuroscience*, 32:6053–6060, May 2012.

[81] E. S. Ilievska and Z. A. Ivanovski. Customized k-space trajectory for compressed sensing MRI. In *Proceedings of 2011 19th Telecommunications Forum (TELFOR)*, pages 631–634, November 2011.

[82] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. Robust 1-Bit Compressive Sensing via Binary Stable Embeddings of Sparse Vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, April 2013.

[83] O. N. Jaspan, R. Fleysher, and M. L. Lipton. Compressed sensing MRI: a review of the clinical literature. *The British Journal of Radiology*, 88(1056):20150487, 2015. PMID: 26402216.

[84] M. Jenkinson, C. Beckmann, T. Behrens, M. Woolrich, and S. Smith. FSL. *NeuroImage*, 62:782–790, 2012.

[85] P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional Magnetic Resonance Imaging: An Introduction to Methods*. Oxford medical publications. OUP Oxford, 2001.

[86] S. Ji, Y. Xue, and L. Carin. Bayesian Compressive Sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, June 2008.

[87] M. Jiang, J. Jin, F. Liu, Y. Yu, L. Xia, Y. Wang, and S. Crozier. Sparsity-constrained SENSE reconstruction: An efficient imple-

mentation using a fast composite splitting algorithm. *Magnetic Resonance Imaging*, 31(7):1218–1227, 2013.

[88] H. Jung, J. Park, J. Yoo, and J. C. Ye. Radial k-t FOCUSS for high-resolution cardiac cine MRI. *Magnetic Resonance in Medicine*, 63(1):68–78, 2010.

[89] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine*, 61(1):103–116, 2009.

[90] H. Jung and J. C. Ye. Performance evaluation of accelerated functional MRI acquisition using compressed sensing. In *Proceedings of 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, pages 702–705, June 2009.

[91] D. S. Kalogerias and A. P. Petropulu. RIP bounds for naively subsampled Scrambled Fourier sensing matrices. In *Proceedings of 2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, March 2014.

[92] M. Kayvanrad, A. Lin, R. Joshi, J. Chiu, and T. Peters. Diagnostic quality assessment of compressed sensing accelerated magnetic resonance neuroimaging. *Journal of Magnetic Resonance Imaging*, 2016.

[93] K. Kim and G. Shevlyakov. Why Gaussianity? *IEEE Signal Processing Magazine*, 25(2):102–113, March 2008.

[94] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale l1-Regularized Least

Squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, December 2007.

[95] F. Krahmer and R. Ward. Stable and Robust Sampling Strategies for Compressive Imaging. *IEEE Transactions on Image Processing*, 23(2):612–622, February 2014.

[96] K. Li, L. Gan, and C. Ling. Convolutional Compressed Sensing Using Deterministic Sequences. *IEEE Transactions on Signal Processing*, 61(3):740–752, February 2013.

[97] S. Li, F. Gao, G. Ge, and S. Zhang. Deterministic Construction of Compressed Sensing Matrices via Algebraic Curves. *IEEE Transactions on Information Theory*, 58(8):5035–5041, August 2012.

[98] S. Li and G. Ge. Deterministic Construction of Sparse Sensing Matrices via Finite Geometry. *IEEE Transactions on Signal Processing*, 62(11):2850–2859, June 2014.

[99] S. Li and G. Ge. Deterministic Sensing Matrices Arising From Near Orthogonal Systems. *IEEE Transactions on Information Theory*, 60(4):2291–2302, April 2014.

[100] S. Li and H. Qi. Compressed dictionary learning for detecting activations in fMRI using double sparsity. In *Proceedings of 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 434–437, December 2014.

[101] J. Listerud, S. Einstein, E. Outwater, and H. Y. Kressel. First principles of fast spin echo. *Magnetic resonance quarterly*, 8(4):199–244, December 1992.

[102] D. A. Lorenz, M. E. Pfetsch, and A. M. Tillmann. An infeasible-point subgradient method using adaptive approximate projections. *Computational Optimization and Applications*, 57(2):271–306, 2013.

[103] D. A. Lorenz, M. E. Pfetsch, and A. M. Tillmann. Solving Basis Pursuit: Heuristic Optimality Check and Solver Comparison. *ACM Transactions on Mathematical Software*, 41(2):8:1–8:29, February 2015.

[104] H. Luong, B. Goossens, J. Aelterman, L. Platiša, and W. Philips. Optimizing image quality in MRI: On the evaluation of k-space trajectories for under-sampled MR acquisition. In *Proceedings of 2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 25–26, July 2012.

[105] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[106] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008.

[107] A. Majumdar, R. K. Ward, and T. Aboulnasr. Compressed Sensing Based Real-Time Dynamic MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 31(12):2253–2266, December 2012.

[108] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[109] R. L. McNamee and W. F. Eddy. Visual analysis of variance: A tool for quantitative assessment of fMRI data processing and analysis. *Magnetic Resonance in Medicine*, 46(6):1202–1208, 2001.

[110] J. Milles, M. J. Versluis, A. G. Webb, and J. H. C. Reiber. Quantitative evaluation of Compressed Sensing in MRI: Application to 7T time-of-flight angiography. In *Proceedings of 2010 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB)*, pages 1–4, November 2010.

[111] Q. Mo and Y. Shen. A Remark on the Restricted Isometry Property in Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 58(6):3654–3656, June 2012.

[112] L. B. Montefusco, D. Lazzaro, and S. Papi. Nonlinear Filtering for Sparse Signal Recovery From Incomplete Measurements. *IEEE Transactions on Signal Processing*, 57(7):2494–2502, July 2009.

[113] L. B. Montefusco, D. Lazzaro, S. Papi, and C. Guerrini. A Fast Compressed Sensing Approach to 3D MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 30(5):1064–1075, May 2011.

[114] M. Moussallam, L. Daudet, and G. Richard. Matching Pursuits with random sequential subdictionaries. *Signal Processing*, 92(10):2532–2544, 2012.

[115] G. Muehllehner and J. S. Karp. Positron emission tomography. *Physics in Medicine and Biology*, 51(13):R117, 2006.

[116] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig. Fast l1-SPIRiT Compressed Sensing Parallel Imag-

ing MRI: Scalable Parallel Implementation and Clinically Feasible Runtime. *IEEE Transactions on Medical Imaging*, 31(6):1250–1262, June 2012.

[117] Y. Nan, Z. Yi, and C. Bingxia. Review of Compressed Sensing for Biomedical Imaging. In *Proceedings of 2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, pages 225–228, November 2015.

[118] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[119] D. Needell and R. Vershynin. Signal Recovery From Incomplete and Inaccurate Measurements Via Regularized Orthogonal Matching Pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):310–316, April 2010.

[120] E. Niedermeyer and F. L. da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[121] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

[122] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872, December 1990.

[123] S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13):5951–5955, 1992.

[124] A. J. O'toole, F. Jiang, H. Abdi, and J. V. Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *Cognitive Neuroscience, Journal of*, 17(4):580–590, 2005.

[125] G. E. Pfander and H. Rauhut. Sparsity in Time-Frequency Representations. *Journal of Fourier Analysis and Applications*, 16(2):233–260, 2009.

[126] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of Matrices Having a Sparse Representation. *IEEE Transactions on Signal Processing*, 56(11):5376–5388, Nov 2008.

[127] G. E. Pfander, H. Rauhut, and J. A. Tropp. The restricted isometry property for time–frequency structured random matrices. *Probability Theory and Related Fields*, 156(3):707–737, 2012.

[128] A. Pinkus. N-widths and optimal recovery. In *Proceeding of Symposia in Applied Mathematics*, volume 36, pages 51–66, 1986.

[129] L. C. Potter, E. Ertin, and M. Parker, J. T.and Cetin. Sparsity and Compressed Sensing in Radar Imaging. *Proceedings of the IEEE*, 98(6):1006–1020, June 2010.

[130] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry, and T. Li. Principles and Applications of Echo-planar Imaging: A

Review for the General Radiologist. *RadioGraphics*, 21(3):767–779, 2001.

[131] G. Puy, J. P. Marques, R. Gruetter, J. P. Thiran, D. Van De Ville, P. Vandergheynst, and Y. Wiaux. Spread Spectrum Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, 31(3):586–598, March 2012.

[132] K. Qiu and A. Dogandzic. Sparse Signal Reconstruction from Quantized Noisy Measurements via GEM Hard Thresholding. *IEEE Transactions on Signal Processing*, 60(5):2628–2634, May 2012.

[133] K. Qiu and A. Dogandzic. Sparse signal reconstruction via ecme hard thresholding. *IEEE Transactions on Signal Processing*, 60(9):4551–4569, September 2012.

[134] X. Qu, X. Cao, D. Guo, C. Hu, and Z. Chen. Compressed sensing MRI with combined sparsifying transforms and smoothed l0 norm minimization. In *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 626–629, March 2010.

[135] X. Qu, Y. Chen, X. Zhuang, Z. Yan, D. Guo, and Z. Chen. Spread Spectrum Compressed Sensing MRI Using Chirp Radio Frequency Pulses. In *Proceedings of 2016 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016.

[136] X. Qu, D. Guo, Z. Chen, and C. Cai. Compressed sensing MRI based on nonsubsampled contourlet transform. In *Proceedings of*

2008 IEEE International Symposium on IT in Medicine and Education (ITME), pages 693–696, December 2008.

[137] S. Ramani, D. S. Weller, J. F. Nielsen, and J. A. Fessler. Non-Cartesian MRI Reconstruction With Automatic Regularization Via Monte-Carlo SURE. *IEEE Transactions on Medical Imaging*, 32(8):1411–1422, August 2013.

[138] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed Sensing and Redundant Dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219, May 2008.

[139] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, 52(3):471–501, 2010.

[140] P. Ritter and A. Villringer. Simultaneous EEG-fMRI. *Neuroscience & Biobehavioral Reviews*, 30(6):823–838, 2006.

[141] Y. Rivenson and A. Stern. Compressed Imaging With a Separable Sensing Operator. *IEEE Signal Processing Letters*, 16(6):449–452, June 2009.

[142] J. Romberg. Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, March 2008.

[143] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025, 2008.

[144] V. Saligrama. Deterministic Designs with Deterministic Guarantees: Toeplitz Compressed Sensing Matrices, Sequence Design and

System ID. *IEEE Transactions on Information Theory*, PP(99):1, 2012.

[145] P. Schniter, L. C. Potter, and J. Ziniel. Fast bayesian matching pursuit. In *Proceedings of 2008 Information Theory and Applications Workshop*, pages 326–333, January 2008.

[146] V. Singh, A. H. Tewfik, and D. B. Ress. Under-sampled functional MRI using low-rank plus sparse matrix decomposition. In *Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 897–901, April 2015.

[147] R. Srinivasan. Methods to improve the spatial resolution of EEG. *International Journal of Bioelectromagnetism*, 1(1):102–111, 1999.

[148] H. Stefan, G. Pawlik, H. G. Böcher-Schwarz, H. J. Biersack, W. Burr, H. Penin, and W. D. Heiss. Functional and morphological abnormalities in temporal lobe epilepsy: a comparison of interictal and ictal EEG, CT, MRI, SPECT and PET. *Journal of Neurology*, 234(6):377–384, 1987.

[149] B. L. Sturm and M. G. Christensen. Cyclic matching pursuits with multiscale time-frequency dictionaries. In *Proceedings of 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 581–585, November 2010.

[150] B. L. Sturm, M. G. Christensen, and R. Gribonval. Cyclic pure greedy algorithms for recovering compressively sampled sparse signals. In *Proceedings of 2011 Conference Record of the Forty Fifth*

Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pages 1143–1147, Novemver 2011.

[151] A. S. Tehrani, A. A. G. Dimakis, and G. Caire. Optimal deterministic compressed sensing matrices. In *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5895–5899, May 2013.

[152] A. M. Tillmann and M. E. Pfetsch. The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, February 2014.

[153] A. Tolouee, J. Alirezaie, and P. Babyn. Accelerating dynamic MRI by compressed sensing reconstruction from undersampled k-t space with spiral trajectories. In *Proceedings of 2014 Middle East Conference on Biomedical Engineering (MECBME)*, pages 17–20, February 2014.

[154] S. M. Tom, C. R. Fox, C. Trepel, and R. A. Poldrack. The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, 315(5811):515–518, 2007.

[155] J. Traub and H. Wozniakowski. *A general theory of optimal algorithms*. ACM monograph series. Academic Press, 1980.

[156] J. A. Tropp and A. C. Gilbert. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, December 2007.

[157] J. A. Tropp and S. J. Wright. Computational Methods for Sparse Solution of Linear Inverse Problems. *Proceedings of the IEEE*, 98(6):948–958, June 2010.

[158] J. Tsao, P. Boesiger, and K. P. Pruessmann. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine*, 50(5):1031–1042, 2003.

[159] T. Tuma and P. Hurley. On the incoherence of noiselet and Haar bases. In *Proceedings of 2009 8th International Conference on Sampling Theory and Applications (SAMPTA)*, page General session, May 2009.

[160] G. Tzagkarakis and P. Tsakalides. Sparse representation of medical images via compressed sensing using Gaussian Scale Mixtures. In *Proceedings of 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 744–747, April 2010.

[161] G. Vaillant, C. Prieto, C. Kolbitsch, G. Penney, and T. Schaeffter. Retrospective Rigid Motion Correction in k-Space for Segmented Radial MRI. *IEEE Transactions on Medical Imaging*, 33(1):1–10, January 2014.

[162] D. Valsesia and E. Magli. Compressive signal processing with circulant sensing matrices. In *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1015–1019, May 2014.

[163] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[164] E. van den Berg and M. P. Friedlander. Sparse Optimization with Least-Squares Constraints. *SIAM Journal on Optimization*, 21(4):1201–1229, 2011.

[165] J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen. Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. *NeuroImage*, 81:335–346, 2013.

[166] L. Weizman, Y. C. Eldar, and D. Ben Bashat. Compressed sensing for longitudinal MRI: An adaptive-weighted approach. *Medical Physics*, 42(9):5195–5208, 2015.

[167] L. Weizman, O. Rahamim, R. Dekel, Y. C. Eldar, and D. Ben-Bashat. Exploiting similarity in adjacent slices for compressed sensing MRI. In *Proceedings of 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1549–1552, August 2014.

[168] J. Wen, Z. Chen, Y. Han, J. Villasenor, and S. Yang. A compressive sensing image compression algorithm using quantized DCT and noiselet information. In *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1294–1297, March 2010.

[169] A. M. Wink, H. Hoogduin, and J. B. Roerdink. Data-driven haemodynamic response function extraction using Fourier-wavelet regularised deconvolution. *BMC Medical Imaging*, 8(1):1–16, 2008.

[170] M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith. Fully Bayesian spatio-temporal modeling of FMRI data. *IEEE Transactions on Medical Imaging*, 23(2):213–231, February 2004.

[171] G. Xu and Z. Xu. Compressed Sensing Matrices From Fourier Matrices. *IEEE Transactions on Information Theory*, 61(1):469–478, January 2015.

[172] M. Yan, Y. Yang, and S. Osher. Robust 1-bit Compressive Sensing Using Adaptive Outlier Pursuit. *IEEE Transactions on Signal Processing*, 60(7):3868–3875, July 2012.

[173] J. Yang and Y. Zhang. Alternating Direction Algorithms for l_1-Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, 2011.

[174] Y. Yang, F. Liu, Z. Jin, and S. Crozier. Aliasing Artefact Suppression in Compressed Sensing MRI for Random Phase-Encode Undersampling. *IEEE Transactions on Biomedical Engineering*, 62(9):2215–2223, September 2015.

[175] Y. Yang, F. Liu, W. Xu, and S. Crozier. Compressed Sensing MRI via Two-stage Reconstruction. *IEEE Transactions on Biomedical Engineering*, 62(1):110–118, January 2015.

[176] P. Ye, J. L. Paredes, G. R. Arce, Y. Wu, C. Chen, and D. W. Prather. Compressive confocal microscopy. In *Proceedings of 2009*

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2009.

[177] R. Zahedi, L. W. Krakow, E. K. P. Chong, and A. Pezeshki. Adaptive Estimation of Time-Varying Sparse Signals. *IEEE Access*, 1:449–464, 2013.

[178] E. Zarahn, G. Aguirre, and M. D'Esposito. A Trial-Based Experimental Design for fMRI. *NeuroImage*, 6(2):122–138, 1997.

[179] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z. P. Liang. Image Reconstruction From Highly Undersampled (k,y)-Space Data With Joint Partial Separability and Sparsity Constraints. *IEEE Transactions on Medical Imaging*, 31(9):1809–1820, September 2012.

[180] Z. Zhu, K. Wahid, P. Babyn, and R. Yang. Compressed Sensing-based MRI Reconstruction Using Complex Double-density Dual-tree DWT. *Journal of Biomedical Imaging*, 2013:10:10–10:10, January 2013.

[181] F. Zong, M. N. d'Eurydice, and P. Galvosas. Fast reconstruction of highly undersampled MR images using one and two dimensional principal component analysis. *Magnetic Resonance Imaging*, 34(2):227–238, 2016.