Sparse Signal Recovery

Analysis and Synthesis Formulations with Prior Support Information

by

Brock Edward Hargreaves

B.Sc. (Hons.), The University of Calgary, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2014

Abstract

The synthesis model for signal recovery has been the model of choice for many years in compressive sensing. Various weighting schemes using prior support information to adjust the objective function associated with the synthesis model have been shown to improve the recovery of the signal in terms of accuracy. Generally, even with no prior knowledge of the support, iterative methods can build support estimates and incorporate that into the recovery which has also been shown to increase the speed and accuracy of the recovery. However when the original signal is sparse with respect to a redundant dictionary (rather than an orthonormal basis) there is a counterpart model to synthesis, namely the analysis model, which has been less popular but has recently attracted more attention. The analysis model is much less understood and thus there are fewer theorems available in both the context of non-weighted and weighted signal recovery.

In this thesis, we investigate weighting in both the analysis model and synthesis model in weighted ℓ_1-minimization. Theoretical guarantees on reconstruction and various weighting strategies for each model are discussed. We give conditions for weighted synthesis recovery with frames which do not require strict incoherency conditions, this is based on recent results of regular synthesis with frames using optimal dual ℓ_1 analysis. A novel weighting technique is introduced in the analysis case which outperforms its traditional counterparts in the case of seismic wavefield reconstruction. We also introduce a weighted split Bregman algorithm for analysis and optimal dual analysis. We then investigate these techniques on seismic data and synthetically created test data using a variety of frames.

Preface

This thesis is original, unpublished, independent work by the author, Brock Edward Hargreaves.

Table of Contents

List of Figures

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisors Dr. Özgür Yilmaz and Dr. Felix Herrmann for their patience, understanding, and for always pushing me to be a better mathematician and researcher. Without their continuous support I would certainly not be where I am today.

I would also like to thank all the members of the Seismic Laboratory of Imaging and Modelling (SLIM), along their support staff, for their expertise.

Dedication

To my mother and father for their continuous love and support throughout my university life. To my sister for joyfully boasting of my accomplishments, though her triumphs supersede my own. To my brother and his wife for their affection and support while far from home, for which I will be forever grateful.

Chapter 1

Introduction to Compressive Sensing

1.1 The Compressed Sensing Problem

Compressive sensing is a sampling paradigm introduced in two seminal papers Candés et al [10] and Donoho [14] which asserts that one can recover signals from fewer samples or measurements than traditional methods such as the Shannon-Nyquist sampling theorem [53, 58, 64], assuming certain prior knowledge of the structure of the signal.

Before posing the mathematical problem of compressive sensing, let us motivate one of the underlying principals of compressed sensing. Consider how a basic compact digital camera collects and stores an image. The camera's sensor is exposed to light passing through the lens and the image is then generally stored using a lossy compression method such as JPEG. A typical lossy compression algorithm might assume a sparse (or compressible) representation of the image in a particular basis or frame and store only those coefficients that contribute the most energy in the image. Many times this type of compression reduces the storage cost by orders of magnitude. Of course the computed coefficients come from the transform of a fully sampled image, so only after the fact can we perform lossy compression. However, one may pose the following question: If a signal is characterized by a small amount of transform coefficients, is it possible to change the acquisition design to reduce the number of samples collected? Among other things, compressive sensing attempts to answer this question by acquiring a compressed signal representation. Let us begin the mathematical framework behind compressive sensing by starting with the definition of sparsity.

Definition 1.1. A vector $x \in \mathbb{R}^N$ is called *k-sparse* if $||x||_0 \leq k$ where $||x||_0$ represents the number of non-zero coefficients of x. For $k \in \{1, 2, ..., N\}$,

define Σ_k^N as the set of all k-sparse vectors in $\mathbb{R}^N$:

$$\Sigma_k^N := \left\{x \in \mathbb{R}^N : ||x||_0 \leq k\right\} \tag{1.1}$$

We do not typically observe sparse vectors directly in the practical applications. Consider the following continuous valued signal $f(t)$ where t represents time.

$$f(t) = \begin{cases} 1 & \text{if } t = 0.5 \\ 0 & \text{if } t \in [0,1] \backslash \{t = 0.5\} \end{cases} \tag{1.2}$$

Suppose we had a measurement device that measures $f(t)$ every Δt seconds. In order to observe the only non-zero portion of this signal, we would require $0.5 \bmod \Delta t = 0$, i.e., we measure the precise moment for which the signal is non-zero. Such a condition on the measurement device is impractical since in general we will not know apriori where the non-zero element of the signal is. What happens more often is that a particular signal is sparse *with respect to* a basis (or frame).

Definition 1.2. A signal $f \in \mathbb{R}^d$ is said to be *k-sparse with respect to a matrix* $D \in \mathbb{R}^{d \times N}$, if $f = \sum_{j=1}^N d_j x_j$, for some $x \in \Sigma_k^N$, and $d_j \in \mathbb{R}^d$ are the column vectors[1] of D. Equivalently, $f = Dx$ for some $x \in \Sigma_k^N$.

Denote T_0 as the set of indices for which x is non zero where $|T_0| \leq k$. In this case, we would have the following decomposition of the signal:

$$f = \sum_{j \in T_0} d_j x_j$$

and the signal is a linear combination of only k out of the N vectors d_j. Though the vector f is typically dense, it is characterized by the vector x which is sparse. The collection $\{d_j\}_{j=1}^N$ is typically a complete system in $\mathbb{R}^d$, such as a basis (in which case $d = N$) or a frame ($N \geq d$). See Chapter 2 for a discussion on frames. For now, it is sufficient that we think of the d_j as the building blocks for the signal f.

Typically, in compressed sensing, measurements will be collected from f in the form of $y = \Phi f$ where $\Phi \in \mathbb{R}^{n \times d}$ and $n \ll d$. The matrix Φ is referred

[1] Note that as a slight abuse of notation d by itself is an integer representing the ambient dimension of the dictionary D, and the d_j are column vectors of D. Context should make it clear to which we are referring to.

to the sensing or measurement matrix.

The compressed sensing problem is then simple stated as the following [38]: Recover x from the knowledge of

$$y = \Phi D x$$

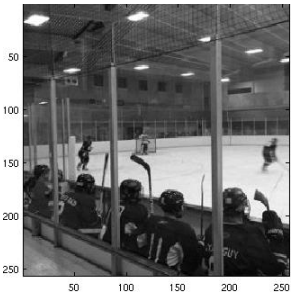
That is from a seemingly incomplete set of linear measurements, recover the sparse vector x.

A major question, which is still highly researched, is for which measurement matrices Φ will recovery of the sparse vector x be possible? Generally, what does the recovery process look like? Without any knowledge of where the k non-zero entries of the sparse vector x are, performing Gaussian elimination will be fruitless and yield infinitely many solutions. The seminal papers referred to at the beginning of this section were the first papers to begin answering these questions and indeed gave theoretical conditions for which the recovery is successful and the process of the recovery is computationally feasible. In this thesis we will focus on the recovery problem and in particular how we can incorporate prior knowledge of the locations of some of the non-zero entries, possibly inaccurately, in the recovery process. For more comprehensive introduction and surveys regarding compressive sensing, the reader is referred to [8, 19, 20, 22].

1.2 Applications of Sparse Approximation

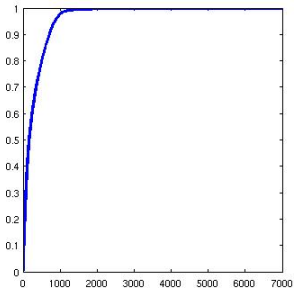
One might ask if the sparsity assumption is valid in practice. Indeed, signals in the real world (such as natural images) are typically difficult to model. Some theoretical work has been done modelling images as functions of bounded variation or belonging to an appropriate smoothness space such as certain Sobolev or Besov spaces. However these characterizations are not perfect, e.g., [28].

Nonetheless, it has been empirically shown that natural images as well as images gathered from medical and seismic experiments have sparse approximations with respect to a variety of different transforms. It is this fundamental observation which compressive sensing takes advantage of. Let us look at a very simple type of wavelet transformation on an image, namely the Haar wavelet [65]:



(a) People playing hockey

(b) Haar wavelet decomposition



(c) Cumulative Energy: The first 7000 terms out of 65536

Figure 1.1: Haar wavelet coefficients computed via the Discrete Wavelets Toolbox by Patrick J. Van Fleet [65]

This image has 256 by 256 pixels, and thus has 65536 coefficients. In Figure 1.1b we see that over 99% of the images energy is stored in only 3% of the wavelet coefficients which is about 2000 terms. We could compress this by only recording those 2000 terms. This is an example of lossy compression via transform coding, where a transform takes the signal to a different domain and only a selection of the coefficients are kept. A famous example of such compression is the JPEG compression scheme, which utilizes the discrete cosine transform (DCT). This method was later updated into JPEG2000 which uses the frame transform (wavelets) and tends to be more effective. Wavelets are based on multiresolutional analysis, a tool used in applied harmonic analysis which has grown into a stable research field over the last two decades. For wavelet based compression schemes see [65].

Sparse approximation is also used in the context of imaging while conducting mineral, oil, and gas exploration. In this thesis we are most concerned with the interpolation of seismic data, commonly referred to as seismic wavefield reconstruction. Seismic receivers are placed on the surface and then record seismic events from reflections of waves in the earth. We restrict ourself to the scenario of missing random receivers on an otherwise regularly sampled grid; this is portrayed by the example below:

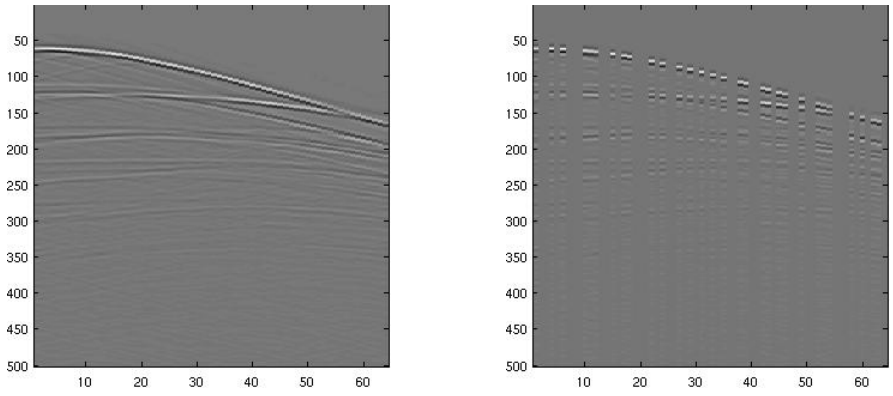


Figure 1.2: Left: Seismic shot record, Right: Seismic shot record with missing traces

One can take advantage of sparsity in the Curvelet domain to recover the missing traces, which we will elaborate on in Chapter 6. There is now a vast amount of literature that explores how to take advantage of this sparsity in the context of seismic data and image interpolation in a variety of domains

including Curvelet, Radon, and Fourier. Some exposure of these techniques can be found in the following: Herrmann et al [33], Zwartjes and Sacchi [70], Shahidi et al [57] and Trad et al [61].

Another common challenge in any seismic workflow is dealing with the massive amount of data that is being collected and processed. In seismic exploration it is not uncommon to deal with data in up to 5 dimensions, and the complexity of seismic imaging problems does not scale linearly with size or dimension. This is referred to as the curse of dimensionality [33]. Compressive sensing can help alleviate these problems two-fold; by requiring fewer measurements thus driving down the cost of acquisition, and by increasing the speed of which an image is processed. There are now a collection of seismic imaging algorithms which take advantage of the sparse approximation framework: "Efficient least-squares imaging with sparsity promotion and compressive sensing" by Herrmann and Li [32], "Randomized marine acquisition with compressive sampling matrices" by Mansour, Wason, Lin, and Herrmann [49], "Robust inversion, dimensionality reduction, and randomized sampling" by Aravkin, Friedlander, Herrmann, and van Leeuwen [1], "Robust inversion via semistochastic dimensionality reduction" by Aravkin, Friedlander, and van Leeuwen [2], and "Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization" by Mansour, Herrmann, and Yilmaz [48] are among them.

1.3 Thesis Outline

In Chapter 2 we review bases in sparse approximation theory and motivate the necessity for frames, followed by a brief introduction of frame theory in finite dimensions. In Chapter 3 we give a review of the literature in the sparse recovery problem with respect to orthonormal basis and frames. This includes a variety of theorems on uniqueness, equivalence, and recovery conditions for the analysis and synthesis formulations. Chapter 4 is dedicated to recent work in how partial support information can be incorporated into the analysis and synthesis formulations of the sparse recovery problem. Recovery conditions for the weighted formulations are given, much akin to the recovery conditions of Chapter 3. Chapters 4 through 6 are where the main contributions of this thesis lie. In Chapter 4 a condition for the weighted synthesis formulation is introduced which does not require strict incoherent conditions, along with introducing a novel weighting method for the analysis formulation. Chapter 5 reviews a split Bregman method for solving the

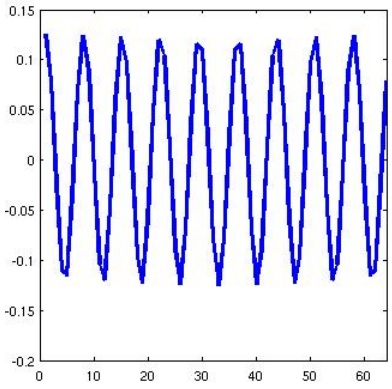
optimal dual analysis formulation, along with the introduction of a weighted split Bregman method for solving the weighted optimal dual analysis formulation. Chapter 6 applies the formulations discussed in this thesis to the seismic interpolation problem and compares their results to a seismic experiment conducted in the Gulf of Mexico.

Chapter 2

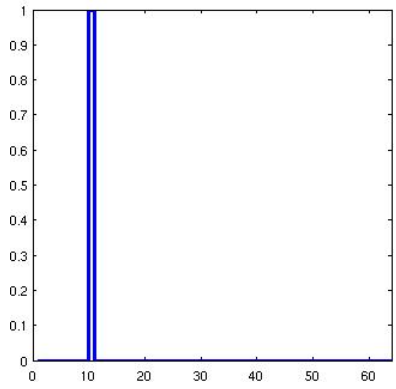
Frames and Bases in Sparse Approximation

2.1 Motivation

In the context of sparse approximation, many signals do not have sufficiently sparse coefficient sequences with respect to a basis. Often we can overcome this inefficiency by considering frames. Consider signals that are sums of sines and spikes. For example, consider the Fourier basis $\{d_k = e^{\frac{-i(k-1)\pi}{m}}\}$ and the standard basis $\{e_k\}$ for $k = 1 \dots m$ where the standard basis is the familiar one, i.e., $e_1 = [1, 0, 0, ..., 0]$. As a toy example, suppose we have two signals f_1 and f_2 which are sparse respectively to d_k and e_k with $m = 64$, as shown in Figure 2.1:



(a) The real part of a signal f_1 created by one Fourier mode



(b) A signal f_2 created by a single non-zero element

Figure 2.1: Two sparse signals

Let us create a third signal f_3, shown in Figure 2.2, which is the sum of the signals f_1 and f_2 in Figure 2.1. The signal f_3 cannot be sparsely represented in either the standard basis or the Fourier basis alone. However

if we consider f_3 as a linear combination of vectors in $\{d_k\}_{k=1}^m \cup \{e_k\}_{k=1}^m$ we can indeed give the exact sparse representation. It is in fact true that the union of two bases is a frame; next we formalize the definition of a frame.

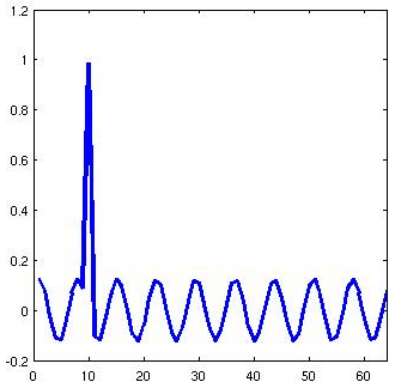


Figure 2.2: The sum of the two sparse signals

2.2 Finite Frames for $\mathbb{R}^d$

This section is for reference and gives a brief review of basic frame theory. For a comprehensive coverage, see [13].

Definition 2.1. A set of vectors $\{d_k\}_{k=1}^N$ in $\mathbb{R}^d$ is a *finite frame* for $\mathbb{R}^d$ if there exist constants $0 < A \leq B < \infty$ such that

$$\forall f \in \mathbb{R}^d, \quad A||f||_2^2 \leq \sum_{k=1}^{N} |< f, d_k >|^2 \leq B||f||_2^2 \tag{2.1}$$

where A and B are called frame bounds and $\{d_k\}_{k=1}^N$ are referred to as the frame elements. The frame bounds are not unique. The optimal lower frame bound is the supremum over all lower frame bounds and the optimal upper frame bound is the infimum over all upper frame bounds. This can be reformulated in matrix form:

$$A||f||_2^2 \leq f^*(DD^*)f \leq B||f||_2^2 \tag{2.2}$$

where $\{d_k\}_{k=1}^N$ are the columns of the matrix $D \in \mathbb{R}^{d \times N}$. That is, $A \leq \sigma_{min}^2(D)$ and $B \geq \sigma_{max}^2(D)$ where $\sigma_{min}(D)$ is the smallest (in magnitude) non-zero singular value of D and $\sigma_{max}(D)$ is the largest (in magnitude) singular value of D. Throughout the rest of this thesis if we refer to D

as a frame, then we consider the columns of D as the frame elements. It is obvious that this definition holds true if and only if DD^* is invertible. Clearly, this is equivalent to saying that D is full rank, i.e., $rank(D) = d$. In the specific case when D is a basis, this implies that D is also a frame.

Definition 2.2. A frame $\{d_k\}_{k\in I}$ is called a *tight frame* if we can choose $A = B$ and is called Parseval (unit norm tight frame) if $A = B = 1$.

In the case of a tight frame, the frame preserves angles between two vectors. In this sense, the frame bounds can be thought of how well the frame preserves the geometry of the vector space. Applied and computational scientists will often prefer to work with tight frames as they tend to have fast computational implementations which is important for large scale problems, typically this translates into fast matrix vector products, which are a key component in a wide variety of imaging algorithms. An important example of a frame for $\mathbb{R}^d$ is a *harmonic frame* as defined below.

Definition 2.3 (Harmonic Frame)**.** A harmonic frame [39] for $\mathbb{R}^d$ is generated by the columns of a Fourier matrix and the definition $H_N^d = \{h_k^N\}_{k=1}^N, N \geq d$ depends on whether d is odd or even. If d is even let

$$h_k^N = \sqrt{\frac{2}{d}}\left[cos\frac{2\pi k}{N}, sin\frac{2\pi k}{N}, cos\frac{4\pi k}{N}, sin\frac{4\pi k}{N}, cos\frac{6\pi k}{N}, sin\frac{6\pi k}{N}, \ldots, cos\frac{2\pi\frac{d}{2}k}{N}, sin\frac{2\pi\frac{d}{2}k}{N}\right]^T$$

for $k = 1, 2, \ldots, N$. If d is odd let

$$h_k^N = \sqrt{\frac{2}{d}}\left[\frac{1}{\sqrt{2}}, cos\frac{2\pi k}{N}, sin\frac{2\pi k}{N}, cos\frac{4\pi k}{N}, sin\frac{4\pi k}{N}, cos\frac{6\pi k}{N}, sin\frac{6\pi k}{N}, \ldots, cos\frac{2\pi\frac{d-1}{2}k}{N}, sin\frac{2\pi\frac{d-1}{2}k}{N}\right]^T$$

It was shown by Zimmermann [69] that H_N^d as defined above, is a unit-norm tight frame for $\mathbb{R}^d$ with frame bound $A = \frac{N}{d}$. We will now introduce the notion of the synthesis and analysis operators, associated with a frame. This will allow us to formalize the notion of synthesis and analysis coefficients.

Definition 2.4. For $\mathbb{R}^d$, equipped with a frame $\{d_k\}_{k=1}^N$, the *synthesis* operator is the linear mapping defined as follows:

$$T : \mathbb{R}^N \to \mathbb{R}^d \ , \ T\{c_k\}_{k=1}^N = \sum_{k=1}^N c_k d_k$$

We have already seen the usage of the synthesis operator in the introduction. The matrix representation of the synthesis operator in the standard basis is the matrix $D \in \mathbb{R}^{d \times N}$ whose columns are $\{d_k\}_{k=1}^N$. Indeed, when we build a signal $f = Dx$, the entries of x are fulfilling the role of c_k in the above definition. We will refer to these coefficients as the synthesis coefficients. As in the introduction, these synthesis coefficients are the weights of the building blocks of the signals of interest. On the other hand, given a signal f, can we decompose it into its constituent building blocks? This is the role of the analysis operator:

Definition 2.5. The *analysis* operator is defined as the adjoint of the synthesis operator and is given by:

$$T^* : \mathbb{R}^d \to \mathbb{R}^N \ , \ T^* f = \{\langle f, d_k \rangle\}_{k=1}^N$$

Using matrix notation, $T^* f = D^* f$. Composing T with its adjoint T^*, we obtain the *frame operator*:

$$S : \mathbb{R}^d \to \mathbb{R}^d, \ Sf = TT^* f = \sum_{k=1}^N \langle f, d_k \rangle d_k \tag{2.3}$$

Again, using matrix notation, $S = DD^*$.

Theorem 2.2.1 (Theorem 1.1.5 [13])**.** *Let $\{d_k\}_{k=1}^N$ be a frame for $\mathbb{R}^d$ with frame operator S, then:*

1. *S is invertible and self-adjoint.*
2. *Every $f \in \mathbb{R}^d$ can be represented as*

$$f = \sum_{k=1}^N \langle f, S^{-1} d_k \rangle d_k = \sum_{k=1}^N \langle f, d_k \rangle S^{-1} d_k \tag{2.4}$$

An important sequence of vectors, $\{S^{-1} d_k\}_{k=1}^m$ also form a frame, and are referred to as the canonical dual frame of $\{d_k\}_{k=1}^m$. In matrix notation, we write the canonical dual as $\bar{D} = (DD^*)^{-1} D$ where as usual the columns

of D are d_k. In linear algebra $\bar{D}^*$ is commonly referred to as the pseudo-inverse and is a right inverse of D, where (*) is the Hermitian conjugate. An important theorem regarding the canonical dual frame:

Theorem 2.2.2 (Proposition 5.1.4 [29]). *If $\{d_k\}_{k=1}^N$ is a frame for $\mathbb{R}^d$ and $f = \sum_{k=1}^N c_k d_k$ for some coefficients c_k, then*

$$\sum_{k=1}^{N} |c_k|^2 \geq \sum_{k=1}^{N} |\langle f, S^{-1} d_k \rangle|^2$$

with equality only if $c_k = \langle f, S^{-1} d_k \rangle$ for all k.

This theorem says that among all coefficients c_k which satisfy $f = \sum_{k=1}^N c_k d_k$, the canonical coefficients have the smallest ℓ_2-norm. This is important, for example, in the context of recovering signals when there is noise involved. There are many theorems related to the denoising problem and frames the reader is referred to works of Dragotti and Vetterli [17], Donoho and Johnstone [16], and Hennenfent and Herrmann [30]. For proofs and finer details of this section, the reader is referred to the works of Christensen [13], Casazza et al [12], Gröchenig [29], and Kovacevic and Chebira [37].

We should note that there are generally, but not always, other frames which lead to decompositions as above. Any frame $\{v_k\}_{k=1}^N$ for $\mathbb{R}^d$ satisfying

$$\forall f \in \mathbb{R}^d \ , \ f = \sum_{k=1}^{N} \langle f, d_k \rangle v_k \tag{2.5}$$

is called a dual frame to $\{d_k\}_{k=1}^N$. If $\{v_k\}_{k=1}^N$ is not the canonical dual frame, then it is referred to as an alternative dual frame.

Chapter 3

The Sparse Recovery Problem

3.1 Sparsity with respect to Orthonormal Bases

3.1.1 The Synthesis Formulation

Suppose a signal $f \in \mathbb{R}^d$ has a sparse synthesis representation with respect to an orthonormal basis $D \in \mathbb{R}^{d \times d}$. That is

$$f = Dx \tag{3.1}$$

where $x \in \Sigma_k^d$ and the columns of D form an orthonormal basis of $\mathbb{R}^d$. Furthermore, we only have access to a set of linear non-adaptive measurements y via the measurement matrix $\Phi \in \mathbb{R}^{m \times d}$ with $m \ll d$.

$$y = \Phi f = \Phi D x \tag{3.2}$$

Solving the sparse recovery problem is to find the sparsest vector of coefficients that satisfies the measurements. This can be posed as the following optimization problem:

$$\hat{x} = \underset{\tilde{x} \in \mathbb{R}^d}{\arg\min} ||\tilde{x}||_0 \text{ subject to } \Phi D \tilde{x} = y \tag{3.3}$$

where $||\tilde{x}||_0$ is defined as the number of non-zero elements of $\tilde{x}$. Under this recovery, a reconstruction of our signal would be

$$\hat{f} = D\hat{x} \tag{3.4}$$

Often in practice there is noise in the measurements and the constraint equations adapts for this:

$$\hat{x} = \underset{\tilde{x} \in \mathbb{R}^d}{\arg\min} ||\tilde{x}||_0 \text{ subject to } ||\Phi D \tilde{x} - y||_2 \leq \epsilon \tag{3.5}$$

where ϵ is an estimate of the level of noise in the measurements. It is well known that this combinatorial problem (3.5) is NP-hard and thus alternative optimization problems must be considered. One such alternative is the convex relaxation of (3.5), given by

$$\hat{x} = \underset{\tilde{x}\in\mathbb{R}^d}{\arg\min} ||\tilde{x}||_1 \text{ subject to } ||\Phi D\tilde{x} - y||_2 \leq \epsilon \tag{3.6}$$

where our objective function now considers the ℓ_1 norm rather than the ℓ_0-"norm". As before, the signal is then reconstructed

$$\hat{f} = D\hat{x}$$

3.1.2 The Analysis Formulation

Recall the synthesis based formulation of the sparse recovery problem has a sparsity assumption on the synthesis coefficients of the signal. The analysis based formulation of the sparse recovery problem proposes that the analysis coefficients be sparse. In practice, signals are never exactly sparse, and dictionaries are often designed to make Ωf for some classes of f as sparse as possible where Ω is some appropriate analysis operator. Common examples of analysis operators include: the shift invariant wavelet transform [47]; the finite difference operator, which concatenates the horizontal and vertical derivatives of an image and is closely connected to total variation [56]; and the Curvelet transform [59]. In this case, the following optimization problem is considered:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d}{\arg\min} ||\Omega\tilde{f}||_0 \text{ subject to } y = \Phi\tilde{f} \tag{3.7}$$

An important recently theoretical study of the analysis and synthesis formulations comes when considering the cosparse analysis model of Nam et al [51], in which both the synthesis and analysis models are given as a union of subspace model. In the compressive sensing literature when comparing analysis and synthesis, the analysis operator Ω has often been chosen as the canonical dual of some frame D and (3.7) becomes:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d}{\arg\min} ||\bar{D}^*\tilde{f}||_0 \text{ subject to } y = \Phi\tilde{f} \tag{3.8}$$

where $\bar{D} \equiv (DD^*)^{-1}D$ denotes the canonical dual frame of D. This formulation will be our focus in this thesis. As in the synthesis case, we can consider

the following convex relaxation along with denoising:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d}{\arg\min} ||\bar{D}^*\tilde{f}||_1 \text{ subject to } ||\Phi\tilde{f} - y||_2 \leq \epsilon \tag{3.9}$$

Notice that in this formulation there is no reconstruction after the optimization problem is solved. In words, the analysis formulation attempts to directly find a signal whose measurements fit that data and whose associated analysis coefficient sequence is sparse.

3.2 Equivalence of Synthesis and Analysis for Orthonormal Bases

An important, and still ongoing, research area is when are the analysis and synthesis formulations equivalent and when does one outperform the other in terms of recovery? Let us start with the case when D is an orthonormal basis.

Theorem 3.2.1. *If D is an orthonormal basis, the analysis and synthesis formulations are equivalent.*

Proof. In this case $DD^* = D^*D = I$, i.e., $D^{-1} = D^*$ and $\bar{D} = D$. Starting with the synthesis formulation:

$$\hat{f} = D\hat{x} = D \cdot \underset{x\in\mathbb{R}^N}{\arg\min} ||x||_1 \text{ subject to } ||y - \Phi Dx||_2 \leq \epsilon \tag{3.10}$$

We have $x = D^{-1}f$ so that

$$\begin{aligned}\hat{f} &= D \cdot \underset{D^{-1}f\in\mathbb{R}^N}{\arg\min} ||D^{-1}f||_1 \text{ subject to } ||y - \Phi f||_2 \leq \epsilon \\ &= \underset{f\in\mathbb{R}^N}{\arg\min} ||\bar{D}^*f||_1 \text{ subject to } ||y - \Phi f||_2 \leq \epsilon\end{aligned}$$

and we arrive at the analysis formulation. Conversely, suppose the analysis formulation:

$$\begin{aligned}\hat{f} &= \underset{f\in\mathbb{R}^N}{\arg\min} ||\bar{D}^*f||_1 \text{ subject to } ||y - \Phi f||_2 \leq \epsilon \\ &= \underset{Dx\in\mathbb{R}^N}{\arg\min} ||\bar{D}^*Dx||_1 \text{ subject to } ||y - \Phi Dx||_2 \leq \epsilon\end{aligned}$$

which simplifies to the synthesis formulation:

$$f = D \cdot \underset{x \in \mathbb{R}^N}{\arg\min} ||x||_1 \text{ subject to } ||y - \Phi D x||_2 \leq \epsilon$$

□

3.3 Synthesis and Analysis Formulations for Redundant Frames

3.3.1 General Frames and Duals

What changes when we consider the case when $D \in \mathbb{R}^{d \times N}$ is a frame? The recovery process in the synthesis formulation stays the same,

$$\hat{x} = \underset{\tilde{x} \in \mathbb{R}^N}{\arg\min} ||\tilde{x}||_1 \text{ subject to } ||\Phi D \tilde{x} - y||_2 \leq \epsilon \tag{3.11}$$

where the optimization still takes place over an N-dimensional space, however we then take the obtained sparse coefficient vector $\hat{x}$ and map it from the high N-dimensional space to a lower d-dimensional space via the reconstruction $\hat{f} = D\hat{x}$. This is somewhat of a blessing in disguise if our only goal is to recover the signal f. Note that the mapping is certainly not one-to-one, thus it is possible to have a recovery $\hat{x}$ of the sparse coefficient vector x where the error $||\hat{x} - x||_2$ is large but still have $\hat{x}$ map to the correct signal f and thus having the error $||f - \hat{f}||_2$ be small. For example, if we add anything in the nullspace of D to our recovered coefficient sequence, we still obtain the same reconstruction.

The major change in the analysis formulation for frames as seen in (3.12) is that the space in which optimization takes place is now d-dimensional instead of N-dimensional[2]:

$$\hat{f} = \underset{f \in \mathbb{R}^d}{\arg\min} ||\bar{D}^* f||_1 \text{ subject to } ||\Phi f - y||_2 \leq \epsilon \tag{3.12}$$

The approach of using a change of variables to prove equivalency breaks down when D is a redundant frame, due to the fact that D is no longer

[2]One should not be quick to assume that since the space of optimization is smaller then computationally the analysis formulation is easier to solve than the synthesis one. In fact there are far more solvers available to solve synthesis problems than analysis. The issue can essentially be boiled down to the coupling of the matrix $\bar{D}^*$ inside the ℓ_1-norm with f, which makes it harder to develop fast solvers for the analysis formulation.

invertible. Let us illustrate this non-equivalence with a simple example. Consider the signal $f = Dx$ with D and x defined as

$$D = \begin{bmatrix} 1 & 2 & -1 & -2 & 0 \\ 0 & 1 & 2 & -4 & 3 \\ -4 & -2 & 4 & 1 & 4 \\ 2 & 3 & -3 & 1 & -4 \end{bmatrix}$$

$$x = \begin{bmatrix} 0 & 0 & 0 & -2 & 0 \end{bmatrix}^T$$

so that

$$f = \begin{bmatrix} 4 & 8 & -2 & -2 \end{bmatrix}^T$$

We also construct the canonical dual frame $\bar{D} = (DD^*)^{-1}D$ for solving the analysis formulation:

$$\bar{D} = \begin{bmatrix} -0.5691 & 0.1436 & -1.2336 & -0.0241 & 0.7423 \\ 0.3103 & 0.1034 & 0.7555 & -0.1192 & -0.3637 \\ -0.2654 & 0.2449 & -0.0241 & 0.1243 & 0.1001 \\ 0.0548 & 0.3516 & 0.5504 & 0.1038 & -0.3458 \end{bmatrix}$$

We have that D has column rank 4 and thus the columns of D form a frame for $\mathbb{R}^4$, noting also that canonical dual frame $\bar{D}$ satisfies $D\bar{D}^* = I$. Suppose now we observe Gaussian measurements $y = Af$ by multiplying f by a matrix whose entries are sampled from a standard normal Gaussian distribution A whose rows have been normalized:

$$A = \begin{bmatrix} 0.9447 & 0.0668 & -0.0059 & 0.3209 \\ 0.4910 & -0.5374 & 0.6784 & -0.0998 \\ -0.4415 & -0.5692 & -0.3935 & 0.5712 \end{bmatrix}$$

Using the software SPGℓ_1 by van den Berg and Friedlander [66] we solve the synthesis formulation (3.11), with noise parameter set to 0, obtaining the approximation:

$$\hat{f} = \begin{bmatrix} 3.9964 & 7.9944 & -1.9948 & -2.0007 \end{bmatrix}^T$$

Using the software NESTA by Bobin, Becker, and Candés [5] we solve the analysis formulation (3.12), with noise and smoothing parameters set to 0, obtaining the approximation:

$$\hat{f} = \begin{bmatrix} 3.9703 & 8.0509 & -1.9267 & -1.9217 \end{bmatrix}^T$$

so we observe the solutions of the analysis and the synthesis recoveries do

not coincide. It should be noted that these solutions are quite similar and one might consider numerical accuracy to be the fault. However, these solutions were calculated to machine precision and their differences are a consequence of non-equivalence of the analysis and synthesis formulation. Though not included in this thesis, one can replace the pseudo-inverse with an optimal dual analysis operator (described in the following section) and see the solutions coincide.

Some of the first theoretical work in analysis and synthesis formulations was conducted in Elad et al [18] where the authors investigated the geometry of the signals associated with the synthesis and analysis models. More recently Candés and Needell [9] gave the first recovery condition which imposes no incoherence restriction on the dictionary. These mentioned papers use the canonical dual as the analysis operator. It should not be surprising that there might exist other duals which give a sparser representation. Indeed, this has been investigated; [18, 45, 51] all consider analysis operators which are not the canonical dual. Li et al [45] generalize the analysis formulation by proposing the general dual ℓ_1-analysis formulation:

$$\hat{f} = \underset{f\in\mathbb{R}^d}{\arg\min} ||\tilde{D}^* f||_1 \text{ s.t. } ||y - \Phi f||_2 \leq \epsilon \tag{3.13}$$

where columns of the analysis operator $\tilde{D}$ form a general (and any) dual frame of D. We could possibly choose the canonical dual, though there might be more sparsifying dual, i.e., choose $\tilde{D}$ such that $||\tilde{D}^* f||_1$ is smaller than $||\bar{D}^* f||_1$.

3.3.2 Optimal Dual ℓ_1-Analysis

For a given frame D, there are in fact infinitely many dual frames to D. Among those duals, we hope that some of them provide an optimal sparsification. One possible solution is the *optimal-dual-based-ℓ_1 analysis* [45] which optimizes not only over the signal space but also over the set of all duals:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d, D\tilde{D}^*=I}{\arg\min} ||\tilde{D}^* \tilde{f}||_1 \text{ s.t. } ||y - \Phi \tilde{f}||_2 \leq \epsilon \tag{3.14}$$

Note that the class of all dual frames for D is given by [40]

$$\begin{aligned}\tilde{D} &= (DD^*)^{-1}D + W^*(I_d - D^*(DD^*)^{-1}D) \\ &= \bar{D} + W^*P\end{aligned}$$

where $\bar{D}$ is the canonical dual frame of D, P is the orthogonal projection onto the null space of D, and $W \in \mathbb{R}^{N\times d}$ is an arbitrary matrix. Plugging the above formula into (3.14) we obtain

$$(\hat{f}, \hat{g}) = \underset{\tilde{f}\in\mathbb{R}^d, g\in\mathbb{R}^N}{\arg\min} \; ||\bar{D}^*\tilde{f} + Pg||_1 \;\text{ s.t. } ||y - \Phi\tilde{f}||_2 \leq \epsilon \tag{3.15}$$

The solution of (3.15) corresponds to the solution of (3.13) with some optimal dual frame, say $\tilde{D}_0$, as the analysis operator. As a byproduct of the optimization, one can construct the optimal dual $\tilde{D}_0$. The construction the authors [44] propose is the following:

$$\tilde{D}_0 = \bar{D}^* + \frac{\hat{f} \otimes \hat{g}}{||\hat{f}||_2^2} \tag{3.16}$$

The optimality here is that in the sense that $||\tilde{D}_0^*\tilde{f}||_1$ achieves the smallest $||\tilde{D}^*\tilde{f}||_1$ among all dual frames $\tilde{D}$ of D and feasible signals $\tilde{f}$ satisfying the constraint in (3.15). Further, and quite remarkably, it can then be shown that the formulation (3.15) is equivalent to the synthesis formulation (3.11), leading to the following theorem; the proof of which is short and we give now.

Theorem 3.3.1 (Liu, Li, Mi, Lei, and Yu [44])**.** *ℓ_1-synthesis and optimal-dual-based ℓ_1-analysis are equivalent.*

Proof. Start with the optimal-dual-based ℓ_1-analysis formulation (3.15). Let $\tilde{x} = \bar{D}^*\tilde{f} + Pg$, then we have

$$\begin{aligned} D\tilde{x} &= D\bar{D}^*\tilde{f} + DPg \\ &= I\tilde{f} + 0 \\ &= \tilde{f} \end{aligned}$$

Since the columns of $[\bar{D}^*, P]$ span the whole N-dimensional space and both $\tilde{f}$ and g are free, then any $\tilde{x} \in \mathbb{R}^N$ can be put in this form. Putting these two facts into (3.15), we obtain the ℓ_1-synthesis formulation (3.11). Conversely, starting with the ℓ_1-synthesis formulation, we have that for any $\tilde{x} \in \mathbb{R}^N$, the following decomposition always holds

$$\begin{aligned} \tilde{x} = \tilde{x}_R + \tilde{x}_N &= D^*(DD^*)^{-1}D\tilde{x} + P\tilde{x} \\ &= \bar{D}^*D\tilde{x} + P\tilde{x} \end{aligned}$$

where $\tilde{x}_R$ and $\tilde{x}_N$ are the components of $\tilde{x}$ belonging to the row space and null space of D, respectively. By definition of ℓ_1-synthesis, we have $\tilde{f} = D\tilde{x} \in \mathbb{R}^d$. Let $g = \tilde{x} \in \mathbb{R}^N$, we can arrive at the optimal dual-based ℓ_1-analysis formulation. □

In other words, their exists appropriate analysis operators (namely optimal duals) for which the analysis and synthesis formulations are equivalent for the case of frames.

3.4 Uniqueness of ℓ_0 Recovery for Synthesis

Before moving to recovery conditions via ℓ_1-minimization we will consider the problem of uniqueness of both the analysis and synthesis formulations. As before, let us consider the issue of whether D is a basis or a frame. Consider a signal $f = Dx$ where the columns of D form a basis, then this linear system has a unique solution. Now consider a signal $f = Dx$ where D is a frame, there are infinitely many solutions to this equation. That is, even before sampling we have a non-uniqueness problem which did not exist when D was a basis. A definition defined early in compressive sensing is the notion of the spark of a matrix, which will help us better understand uniqueness.

Definition 3.1 (Spark)**.** The spark of a given matrix Φ is the smallest number of columns of Φ that are linearly dependent.

The following theorem associates sparse signal recovery and the spark of a matrix in the orthonormal basis case:

Theorem 3.4.1 (Donoho, Elad, [15])**.** *Let Φ be an $m \times d$ matrix and let $k \in \mathbb{N}$. Then the following conditions are equivalent:*

1. *If a solution x to* (3.3) *satisfies $||x||_0 \leq k$ then it is the unique solution*
2. *$spark(\Phi) > 2k$*

Thus when dealing with exactly sparse vectors, spark provides a complete characterization of sparse signal recovery. However, even for modestly sized problems, this is computationally infeasible to check. This definition has been extended to handle sparsity with respect to a frame D.

Definition 3.2 (D-Spark)**.** Given a matrix Φ we define D-Spark(Φ) as the smallest possible number of columns from D, marked as T , such that range(D_T)$\cap$ Null(Φ) $\neq 0$

Coming back to the issues introduced when using D as a frame, notice that D-Spark(Φ) $\leq$ Spark(Φ) with equality if and only if D is an invertible matrix, thus quantifying the uniqueness problem associated with frames. As in the orthonormal case, there is an associated theorem with sparse recovery.

Theorem 3.4.2 (Giryes, Elad, [26]). *Let $y = \Phi f$ where f has a k-sparse representation x under D. If $k <$D-Spark(Φ)$/2$ then $f = D\hat{x}$ for $\hat{x}$ the minimizer of* (3.3).

The theorems above give us conditions for uniqueness in the synthesis setting. Clearly, if we recover the correct sparse vector x then we would indeed recover the correct signal f via $f = D\hat{x}$. In practice, the convex relaxation of both the synthesis and analysis formulations are preferred. The question is then when do the convex relaxations of the analysis and synthesis formulations give the same solution as their non-convex counterparts? We investigate this in the following section. A small note on the case when D is redundant: When one is only interested in recovering the signal $f = Dx$, than it is not necessary to first find it's sparsest representation. As discussed previously, adding any vector to x in the null space of D (which will most likely giving a less sparse solution) does not change it's reconstruction. Thus a large amount of research is focused on bounding the error $||f - \hat{f}||_2$ of the recovery of the convex relaxations.

3.5 Recovery Guarantees for ℓ_1-Minimization

In practice, the convex relaxation of the sparse recovery problem is considered. In this section we summarize existing recovery conditions for both the convex relaxations of analysis and synthesis formulations.

3.5.1 Synthesis

In order to quantify the errors in the reconstruction we will require two definitions. Namely, the k-term approximation error of a vector and the restricted isometry property(RIP) of a matrix.

Definition 3.3. The best *k-term approximation error* of a vector $x \in \mathbb{C}^N$ in ℓ_p is defined as

$$\sigma_k(x)_p = \inf_{z \in \Sigma_k} ||x - z||_p \tag{3.17}$$

Definition 3.4. *The Restricted Isometry Property(RIP)*: For an $m \times n$ measurement matrix Φ, the *s-restricted isometry constant* δ_s of Φ is the smallest quantity such that:

$$(1 - \delta_s)||x||_2^2 \leq ||\Phi x||_2^2 \leq (1 + \delta_s)||x||_2^2 \tag{3.18}$$

holds for all s-sparse signals x.

Constructing measurement matrices which satisfy RIP is a difficult and ongoing research area. This is beyond the scope of this thesis and we will rather focus on using this condition to show and prove error estimates for signal recovery. The following was the first theorem using the definition of RIP for sparse signal recovery.

Theorem 3.5.1 (Candés, Tao, Romberg [10])**.** *The recovery* (3.11) *satisfies the following:*

$$||\hat{x} - x||_2 \leq C_0 \frac{\sigma_s(x)_1}{\sqrt{s}} + C_1 \epsilon \tag{3.19}$$

provided that the 2s-restricted isometry constant of the matrix composition ΦD *obeys* $\delta_{2s} < \sqrt{2} - 1$.

This result was improved to $\delta_{2s} \leq \frac{3}{4+\sqrt{6}} \approx 0.4652$ in 2010 by Foucart [21] and further still $\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246$. in 2013 by Foucart and Rauhut [22] and a sharp bound of $\delta_{2s} < \frac{\sqrt{2}}{2}$ has now been found by Cai and Zhang in 2013 [7].

Notice that if x is s-sparse and there is no noise then Theorem 3.5.1 guarantees a perfect recovery. At this point we should note that when Φ is a Gaussian and D is a frame it is well known that even if Φ has a small RIP constant, ΦD may no longer satisfy the assumptions of the above theorem. Thus when using frames as a sparsity inducing transform, traditional methods of analyzing the residual of the recovered signal break down.

3.5.2 Analysis

The definition of RIP has been extended for usage in the analysis formulation of the sparse recovery problem.

Definition 3.5 (D-RIP)**.** Let Σ_s be the union of all subspace spanned by all subsets of s columns of D. We say that the measurement matrix Φ obeys the restricted isometry property adapted D(abbreviated D-RIP) with constant δ_s if

$$(1-\delta_s)||v||_2^2 \leq ||\Phi v||_2^2 \leq (1+\delta_s)||v||_2^2 \tag{3.20}$$

holds for all $v \in \Sigma_s$.

Notice that Σ_s is the image under D of all s-sparse vectors. Again, we will not focus on which matrices satisfy D-RIP but rather what D-RIP conditions do we require to ensure accurate signal reconstructions. The following is the first theorem using D-RIP to give an estimate of the error in the signal reconstruction for the analysis formulation.

Theorem 3.5.2 (Candés, Needell [9])**.** *Let D be an arbitrary tight frame and let Φ be a measurement matrix satisfying D-RIP with $\delta_{2s} < .08$. Then the solution $\hat{f}$ to* (3.9) *satisfies*

$$||\hat{f} - f||_2 \leq C_0\epsilon + C_1 \frac{||D^*f - (D^*f)_s||_1}{\sqrt{s}}$$

for some well behaved constants C_0 and C_1.

Again, we see a perfect recovery if D^*f is s-sparse and there is no noise. We should mention an important recent result by Baker [4] in the case of tight frames though, which provides a sharp upper bound of $\delta_{2s} < \frac{\sqrt{2}}{2}$, a result analogous to its synthesis counterpart.

These ideas have now been generalized for arbitrary frames, not requiring the condition of tightness however not claiming optimality, and we investigate this in the following section.

3.5.3 General and Optimal Dual Based Analysis

Using (3.13) as a decoder, Li et al [45] have the following theorem regarding the error of the signal reconstruction:

Theorem 3.6 (Liu, Mi, Li [45])**.** *Let D be a general frame of $\mathbb{R}^n$ with frame bounds $0 < A \leq B < \infty$. Let $\tilde{D}$ be any dual frame of D with frame bounds $0 < \tilde{A} \leq \tilde{B} < \infty$, and let $\rho = s/b$. Suppose*

$$\left(1 - \sqrt{\rho B\tilde{B}}\right)^2 \cdot \delta_{s+a} + \rho B\tilde{B} \cdot \delta_b < 1 - 2\sqrt{\rho B\tilde{B}} \tag{3.21}$$

holds for some positive integers a and b satisfying $0 < b - a \leq 3a$. Then, the solution $\hat{f}$ to (3.13) *satisfies*

$$||\hat{f} - f||_2 \leq C_0\epsilon + C_1 \frac{||\tilde{D}f - (\tilde{D}^* f)_s||_1}{\sqrt{s}} \tag{3.22}$$

where C_0 and C_1 are some constants and $(\tilde{D}^* f)_s$ denotes the vector consisting of the largest s entries of $\tilde{D}^* f$ in magnitude.

As a direct consequence of the equivalence between optimal-dual-ℓ_1-analysis and synthesis, we can make a direct application of Theorem 3.6, using an optimal dual, to arrive at a theorem for solving the synthesis formulation.

Theorem 3.7 (Liu, Li, Mi, Lei, and Yu [44])**.** *In Theorem 3.6 replace $\tilde{D}$ with some optimal dual frame $\tilde{D}_0$, as defined in* (3.14)*. Then, the solution $\hat{f}$ to* (3.14) *or* (3.11) *satisfies* (3.22)*.*

Note that the optimal dual cannot be known apriori since it is signal dependent. Rather than use optimal-dual-ℓ_1-analysis as a decoder, the authors use the equivalence to give theoretical conditions for synthesis based recovery with frames (using traditional compressive sensing arguments which have known to break down since they rely on incoherency of the measurement matrix with the frame). That said, the conditions recovered are pointwise dependent on the signal and the associated optimal dual so that we cannot calculate the error without explicitly finding the optimal dual.

In conclusion, we've given recovery conditions for the synthesis, analysis, and optimal dual analysis formulations of the sparse recovery problem. Over the past decade, the compressive sensing community has been in the pursuit of weaker RIP and D-RIP conditions while simultaneously improving error bounds. Suppose we had some specific prior knowledge of the solution of the problem. As it turns out, this can be used to both help numerical methods in sparse recovery along with improve theoretical conditions which we investigate in the following chapter.

Chapter 4

Weighted Methods and Partial Support Information

Within the optimization framework, one can consider having prior support information about the signal you wish to recover. For example, we might know where some of the non-zero synthesis coefficients are located if we are trying to solve the synthesis formulation. It is natural to suspect that having extra information about a signal should allow for a faster convergence or more accurate reconstruction method. In this chapter we investigate weighting in both the analysis and synthesis formulations. Though we will primarily focus on the weighting schemes by Friedlander et al [23] and Candés et al [11], we should recognize that the recovery of compressively sampled signals using prior support information has been studied in the literature by Von Borries, Mioss, and Potes, [54], Vaswani and Lu [67] [46], Jacques [35], and Khajehnejad et al [36].

4.1 Weighted Synthesis

One approach that utilizes prior information in the recovery algorithm is to replace ℓ_1 minimization with weighted ℓ_1 minimization:

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} ||x||_{1,w} \text{ subject to } ||\Phi D x - y||_2 \leq \epsilon \tag{4.1}$$

where $||x||_{1,\omega} := \sum_i w_i |x_i|$ is the weighted ℓ_1 norm. There are various ways to choose the weights. Here we consider two possible ways, fixed weights and variable weights.

4.1.1 Fixed Weights

The first case we consider is a fixed weighting strategy [23]:

$$w_i = \begin{cases} 1 & \text{if } i \in \tilde{T}^c \\ \omega & \text{if } i \in \tilde{T} \end{cases} \tag{4.2}$$

where $\tilde{T}$ is a "support" estimate of x. The following theorem gives a recovery condition for such a weighting scheme:

Theorem 4.1.1 (Friedlander, Mansour, Saab, Yilmaz [23]). *Let x be in $\mathbb{R}^N$ and let x_k be its best k-term approximation, supported on T_0. Let $\tilde{T} \subset \{1, ..., N\}$ be an arbitrary set and define ρ and α such that $|\tilde{T}| = \rho k$ and $|\tilde{T} \cap T_0| = \alpha\rho k$. Suppose that there exists an $a \in \frac{1}{k}\mathbb{Z}$, with $a \geq (1-\alpha)\rho$ where $a \geq 1$, and the measurement matrix satisfies RIP with*

$$\delta_{ak} + \frac{a}{\gamma^2}\delta_{(a+a)k} < \frac{a}{\gamma^2} - 1 \tag{4.3}$$

where $\gamma = \omega + (1-\omega)\sqrt{1+\rho-2\alpha\rho}$ for some $0 \leq \omega \leq 1$. Then the solution $\hat{x}$ to (4.1) obeys

$$||\hat{x} - x||_2 \leq C_0'\epsilon + C_1' k^{-1/2}\left(\omega||x - x_k||_1 + (1-\omega)||x_{\tilde{T}^c \cap T_0^c}||_1\right) \tag{4.4}$$

Moreover, when $\alpha > 0.5$ (which indicates an accuracy of the support greater than 50 percent) and $\omega < 1$, (4.1) outperforms (3.11) in terms of accuracy, stability, and robustness guarantees [23]. Practically, this means that whenever one has some reasonably accurate prior support information then the weighted formulation is preferred over it's non weighted counterpart.

4.1.2 Variable Weights

The second weighting scheme popular in compressive sensing literature we consider is the variable weighting scheme of [11], where the goal is to design the weights to be inversely proportional to the magnitude of the non-zero entries:

$$w_i = \frac{1}{|x_i| + a} \tag{4.5}$$

where a is a stability parameter which will be explained shortly.

Let us begin with a heuristic reasoning behind such a choice of weights given in [11]. They specify that as a rough rule of thumb the weights should relate inversely to the true signal magnitudes and more generally that large weights could be used to discourage nonzero entries in the recovered signal while small weights could be used to encourage nonzero entries, a heuristic also reflected in the strategy of Friedlander et al [23]. Suppose the sparse

signal was known and then weights were set as $w_i = \frac{1}{|x_i|}$. The weights would then be infinite at all locations outside of the support of x, forcing the coordinates of the solution vector x at these locations to be zero. In this case, as long as the number of measurements taken is at least as large as the number number of non-zeros of the actual signal, we would have $x = \hat{x}$. In practice we do not know the signal exactly, otherwise we would not have to solve any problems. As well, it is not possible to use infinity as a weight and thus **a** is used as a stability parameter.

We will see in the following section that Needell [52] gave theoretical guarantees for the case of iterative reweighting.

4.1.3 Iterative Reweighting

Even without prior support information, it is still possible to incorporate weights to the recovery problem. The approach is to solve the non-weighted formulation, use that solution to compute weights, and then solve the associate weighted formulation. Iterative reweighting in the synthesis case was first introduced by Candés, Wakin, and Boyd [11] and has the following structure:

Algorithm 1 Iterative reweighting with ℓ_1-synthesis

Set $l = 0$ and $w_j^l = 1$, $j \in J$ (J indexes the dictionary) and $w_j = W_{j,j}$
while $||x^{(l)} - x^{(l+1)}|| \geq \sigma$ and $l <$ maximum iterations **do**
  $x^{(l)} = \arg\min ||W^{(l)}x||_{\ell_1}$ subject to $||\Phi Dx - y||_2 \leq \epsilon$
  Update weights: $w_j^{(l+1)}$ according solution $x^{(l)}$.
  l = l + 1
end while

As mentioned in the previous section, there exists theoretical results on the convergence of the reweighting strategy:

Theorem 4.1.2 (Needell [52])**.** *Assume Φ satisfies RIP with constant $\delta_{2s} < \sqrt{2} - 1$. Let x be an s-sparse vector with noisy measurements $u = \Phi x + e$ where $||e||_2 \leq \epsilon$. Assume the smallest non-zero coordinate μ of x satisfies $\mu \geq \frac{4\alpha\epsilon}{1-\rho}$. Then the limiting approximation of reweighted ℓ_1-minimization using weights defined as* (4.5) *satisfies*

$$||x - \hat{x}||_2 \leq C''\epsilon \tag{4.6}$$

where $\rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}$, $\alpha = \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}$, and $C'' = \frac{2\alpha}{1+\rho}$

It is still current research for choosing smart and robust rules for selecting the parameter a though its purpose is clearly to provide stability but also to ensure that a zero-valued component in x does not strictly prohibit a nonzero estimate at the next step. That is, if a component is wrongly identified as zero it is still possible to recover it in further iterations.

4.2 Weighted Analysis

Weighting in the analysis case was introduced at the same time as weighting in the synthesis case in [11]. The weighted analysis problem is posed as the following:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d}{\arg\min} ||\bar{D}^*\tilde{f}||_{1,w} \text{ subject to } ||\Phi\tilde{f} - y||_2 \leq \epsilon \tag{4.7}$$

We can again choose our weights based on some sort of prior information, applying either a variable weight strategy as in [11], or a fixed weight strategy as in [23]. Recall that in the analysis case, we hope that $\bar{D}^*f$ is sparse. Further more if we are assuming that f has a sparse representation with respect to a dictionary D, we hope that $\bar{D}^*f$ is a close approximation to that representation, noting that if D is a basis then it would in fact be the exact sparse representation. More often than not though, $\bar{D}^*f$ will not be sparse but rather compressible in the sense that the majority of the signal's energy is stored in a small amount of its coefficients. The weights then should be calculated based on those entries which contribute the most to the signals energy. This is the approach taken in Chapter 6 on seismic wavefield reconstruction.

4.2.1 Iterative Reweighting

Just as in synthesis formulation, we consider iterative reweighting in the analysis formulation. The authors of [11] examine iterative reweighting in the analysis case empirically using examples from MRI imaging and radar. However, we are unaware of any theorems which give guarantees analogous to Theorem 4.1.2.

As before we start by solving a non-weighted analysis formulation and then solve a sequence of weighted analysis formulations using previous solutions as a way of calculating the weights.

Algorithm 2 Iterative reweighting with ℓ_1-analysis

Set $l = 0$ and $w_j^l = 1$, $j \in J$ (J indexes the dictionary) and $w_j = W_{j,j}$
while $||f^{(l)} - f^{(l+1)}|| \geq \sigma$ and $l <$ maximum iterations **do**
  $f^{(l)} = \arg\min ||W^{(l)}\bar{D}^* f||_{\ell_1}$ subject to $||\Phi f - y||_2 \leq \epsilon$
  Update weights: $w_j^{(l+1)}$ according solution $f^{(l)}$.
  l = l + 1
end while

4.3 Main Theoretical Contributions

Weighted Methods for Recovering Signals that are Sparse with respect to Coherent Dictionaries

In the remainder of this chapter, we will describe the major theoretical contributions of this thesis. Specifically, in Section 4.3.1, we will introduce a novel weighting technique for the analysis formulation. This technique is heuristically motivated and is observed to perform well for certain classes of signals including seismic data, for more on this see Chapter 6. In Section 4.3.2, we will introduce and discuss weighted-general-dual-analysis recovery. In Section 4.3.3, we apply these to the "optimal duals" and establish the equivalence of the optimal-dual-weighted-analysis and weighted-synthesis algorithms (see Theorem 4.3.2). As a consequence we obtain recovery gauruntees for the weighted-synthesis algorithm of [23] in the case when the sparsity is respect to highly coherent dictionaries.

4.3.1 A Novel Weighting Technique for Analysis

The last method of weighting which we examine in this thesis introduces a weight matrix inside of the canonical dual. It is posed as the following optimization problem:

$$\hat{f} = \underset{\tilde{f} \in \mathbb{R}^d}{\arg\min} ||\overline{DW}^* \tilde{f}||_1 \text{ subject to } ||\Phi\tilde{f} - y||_2 \leq \epsilon \tag{4.8}$$

where $\overline{DW}$ is the canonical dual of DW.

However, we make the following choice in weights which is reverse to that of Friedlander et al [23]:

$$w_i = \begin{cases} 1 & \text{if } i \in \tilde{T} \\ \omega & \text{if } i \in \tilde{T}^c \end{cases} \tag{4.9}$$

where $0 < \omega < 1$ and $\tilde{T}$ is a support estimate of x. As we shall see in Chapter 6 on seismic wavefield reconstruction, this method sometimes outperforms the weighted synthesis and analysis formulations with fixed weights.

A heuristic reasoning behind this choice of weighting can be seen if we consider the case that D is an orthonormal basis. Start with a sparse representation of a signal f with respect to an orthornomal basis D and then multiply both sides by $\overline{DW}^*$ and expand the right hand side:

$$\begin{aligned} f &= Dx \\ \overline{DW}^* f &= \overline{DW}^* Dx \\ &= \left(((DW(DW)^*)^{-1} DW \right)^* Dx \\ &= WD^*(DW^2D^*)^{-1}Dx \\ &= WD^*(D^{-1})^* W^{-2} D^{-1} Dx \\ &= W^{-1}x \end{aligned}$$

The final line which states that $\overline{DW}^* f = W^{-1}x$ is in fact reminiscent of the variable weighting of [11] in that the weights should relate inversely to the true signal magnitudes. Notice that since we had reversed the choice of weights in our definition, their inverse now correspond to assigning a large weight on the off support and a small weight to the true support as in the synthesis weighting case.

4.3.2 Weighted General Dual Analysis

We now consider weighting in the general dual formulation. The following optimization problem can be considered:

$$\hat{f} = \underset{f \in \mathbb{R}^d}{\arg\min} \, ||\tilde{D}^* f||_{1,w} \;\; \text{s.t. } ||y - \Phi f||_2 \leq \epsilon \tag{4.10}$$

where $\tilde{D}$ is an arbitrary dual frame of D. We combine the techniques of Friedlander et al [23] and Li et al [45] to arrive at the following theorem.

Theorem 4.3.1 (Weighted general dual ℓ_1-analysis)**.** *Consider the signal $f = Dx$, where $x \in \mathbb{R}^N$, and $D \in \mathbb{R}^{d\times N}$ is an arbitrary frame of $\mathbb{R}^d$ with frame bounds $0 < A \leq B < \infty$ and we have noisy linear measurements:*

$$y = \Phi Dx + z \tag{4.11}$$

Let $\tilde{D}$ be an alternative dual frame of D with frame bounds $0 < \tilde{A} \leq \tilde{B} < \infty$. Let $\tilde{T} \subset \{1, ..., N\}$ be an arbitrary set. Let T_0 index the support of the best s-term approximation of $\tilde{D}^* f$ and thus $|T_0| = s$. Define ρ and α such that $|\tilde{T}| = \rho s$ and $|\tilde{T} \cap T_0| = \alpha\rho s$. Suppose Φ has D-RIP with constant δ that satisfies

$$\left(1 - \gamma\sqrt{\frac{sB\tilde{B}}{b}}\right)^2 \delta_{s+a} + \frac{sB\tilde{B}}{b}\delta_b < 1 - 2\gamma\sqrt{\frac{sB\tilde{B}}{b}} \tag{4.12}$$

for some positive integers a and b satisfying $0 < b - a \leq 3a$ where $\gamma = \omega + (1-\omega)\sqrt{1 + \rho - 2\alpha\rho}$ and $0 \leq \omega \leq 1$ with our weights chosen as follows:

$$w_i = \begin{cases} 1 & \text{if } i \in \tilde{T}^c \\ \omega & \text{if } i \in \tilde{T} \end{cases} \tag{4.13}$$

and $\tilde{T}$ is a support estimate of $\tilde{D}^* f$. Then the solution $\hat{f}$ to (4.10) *satisfies*

$$||f - \hat{f}||_2 \leq c_1 \left(\omega||\tilde{D}^*_{T_0^c} f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c \cap T_0^c} f||_1\right) + c_2\epsilon \tag{4.14}$$

where c_1 and c_2 are some well behaved constants, and ϵ is an upper bound on the noise.

If we consider the case when D is a Parseval frame and the analysis operator $\tilde{D}$ is its canonical dual frame, then $B\tilde{B} = 1$. If we set $a = 3s$, $b = 12s$, and $\omega = 1$ then reduce to the case of non-weighting and achieve the condition $\delta_{2s} < .1398$ as given in [45]. However if set $\omega = .5$, $\rho = 1$, and $\alpha = .6$ then we require $\delta_{2s} < 0.1456$, a weaker condition. Moreover, whenever $\alpha > .5$, an estimate of the support which is at least 50% accurate, we have weaker conditions and also see a decrease in the error bound for the recovery. This means whenever one has a reasonably accurate estimate of the support, the weighted analysis formulation is preferred over it's non-weighted counterpart.

Proof of Theorem 4.3.1

An important contribution of this thesis is the proof of Theorem 4.3.1. The techniques for the proof are of the spirit given in Friedlander et al [23], Li et al [45], and Needell et al [9]. An important technique that has come up in various proofs in providing conditions for compressive sensing deals with partitioning indicies of whatever we consider our support set. We will follow the partitioning technique of Li et al [45], for which the following lemma is used.

Lemma 4.1 (The Shifting Inequality)**.** *Let q and r be positive integers satisfying $q \leq 3r$. Then for any non increasing sequence of real numbers $a_1 \geq \cdots \geq a_r \geq b_1 \geq \cdots \geq b_q \geq c_1 \geq \cdots \geq c_r \geq 0$, we have*

$$\sqrt{\sum_{i=1}^{q} b_i^2 + \sum_{i=1}^{r} c_i^2} \leq \frac{\sum_{i=1}^{r} a_i + \sum_{i=1}^{q} b_i}{\sqrt{q+r}} \tag{4.15}$$

Without loss of generality, we will assume the first s entries in $\tilde{D}^* f$ are largest in magnitude so that $T_0 = \{1, \ldots, s\}$. Making rearrangements if necessary we also assume the entries of $\tilde{D}^* f$ also satisfy

$$|\tilde{D}^* f(s+1)| \geq |\tilde{D}^* f(s+2)| \geq \ldots$$

We now partition T_0^c into smaller disjoint subsets. Let $T_1 = \{s+1, s+2, \ldots, s+1+a\}$ and $T_i = \{s+a+(i-2)b+1, \ldots, s+a+(i-1)b\}$ for $i = 2, 3, \ldots$ with the last subset of size less than or equal to b, where a and b are positive integers satisfying $0 < b - a \leq 3a$. Further, divide each T_i, $i \geq 2$ into two parts:

$$\begin{aligned} T_{i1} &= \{s+a+(i-2)b+1, \ldots, s+(i-1)b\} \\ T_{i2} &= \{s+(i-1)b+1, \ldots, s+(i-1)b+a\} \end{aligned}$$

We have $|T_{i1}| = b-a$ and $|T_{i2}| = a$ for $i \geq 2$ and notice $T_{i2} = T_i \backslash T_{i1}$. Denote $T_{01} := T_0 \cup T_1$. Let $h = f - \hat{f}$. We have

$$\begin{aligned} ||h||_2 &= ||D\tilde{D}^* h||_2 \\ &= || \left(D_{T_{01}} \tilde{D}^*_{T_{01}} + D_{T_{01}^c} \tilde{D}^*_{T_{01}^c} \right) h||_2 \\ &= ||D_{T_{01}} \tilde{D}^*_{T_{01}} h + D_{T_{01}^c} \tilde{D}^*_{T_{01}^c} h||_2 \end{aligned}$$

$$\leq ||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 + ||D_{T^c_{01}}\tilde{D}^*_{T^c_{01}}h||_2, \text{ by the triangle inequality.}$$
$$\leq ||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 + ||D_{T^c_{01}}||_2||\tilde{D}^*_{T^c_{01}}h||_2$$
$$\leq ||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 + \sqrt{B}||\tilde{D}^*_{T^c_{01}}h||_2, \text{ by (2.1)}$$

In other words,

$$||h||_2 \leq ||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 + \sqrt{B}||\tilde{D}^*_{T^c_{01}}h||_2 \tag{4.16}$$

What remains is to bound $||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2$ and $||\tilde{D}^*_{T^c_{01}}h||_2$.

Lemma 4.2 (Bound the tail $||\tilde{D}^*_{T^c_{01}}h||_2$)**.**

$$||\tilde{D}^*_{T^c_{01}}h||_2 \leq \sqrt{\frac{s}{b}}\left(\gamma\sqrt{\tilde{B}}||h||_2 + \eta\right) \tag{4.17}$$

where $\gamma = \omega + (1-\omega)\sqrt{1+\rho+2\alpha\rho}$ *and* $\eta = \frac{2}{\sqrt{s}}\left(\omega||\tilde{D}^*_{T^c_0}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T^c_0}f||_1\right)$

Proof. As in Li et al [45], if $0 < b-a \leq 3a$ then applying the shifting inequality to the vectors $\left[(\tilde{D}^*_{T_1}h)^*, (\tilde{D}^*_{T_{21}}h)^*), (\tilde{D}^*_{T_{22}}h)^*\right]^*$ and $\left[(\tilde{D}^*_{T_{(i-1)2}}h)^*, (\tilde{D}^*_{T_{i1}}h)^*), (\tilde{D}^*_{T_{i2}}h)^*\right]^*$ for $i = 3, 4, \ldots,$ we have

$$||\tilde{D}^*_{T_2}h||_2 \leq \frac{||\tilde{D}^*_{T_1}h||_1 + ||\tilde{D}^*_{T_{21}}h||_1}{\sqrt{b}}, \ldots$$
$$||\tilde{D}^*_{T_i}h||_2 \leq \frac{||\tilde{D}^*_{T_{(i-1)2}}h||_1 + ||\tilde{D}^*_{T_{i1}}h||_1}{\sqrt{b}}, \ldots$$

Using the above and the fact that $\sqrt{u^2+v^2} \leq u+v$ for $u, v \geq 0$, it then follows that

$$||\tilde{D}^*_{T^c_{01}}h||_2 \leq \sum_{i\geq 2}||\tilde{D}_{T_i}h||_2$$
$$\leq \frac{||\tilde{D}^*_{T^c_0}h||_1}{\sqrt{b}}$$

We then need to bound $||\tilde{D}^*_{T^c_0}h||_1$. We have $h = \hat{f} - f$. We now follow the strategy of Friedlander et al. By choice of weights and since f and $\hat{f}$ are feasible and $\hat{f}$ is the minimizer, we have

$$||\tilde{D}^*\hat{f}||_{1,\omega} \leq ||\tilde{D}^*f||_{1,\omega}$$
$$\Leftrightarrow \omega||\tilde{D}^*_{\tilde{T}}\hat{f}||_1 + ||\tilde{D}^*_{\tilde{T}^c}\hat{f}||_1 \leq \omega||\tilde{D}^*_{\tilde{T}}f||_1 + ||\tilde{D}^*_{\tilde{T}^c}f||_1$$

Further still,

$$\omega\left(||\tilde{D}^*_{\tilde{T}\cap T_0}\hat{f}||_1 + \tilde{D}^*_{\tilde{T}\cap T_0^c}\hat{f}||_1\right) + ||\tilde{D}^*_{\tilde{T}^c\cap T_0}\hat{f}||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}\hat{f}||_1$$
$$\leq \omega\left(||\tilde{D}^*_{\tilde{T}\cap T_0}f||_1 + ||\tilde{D}^*_{\tilde{T}}f||_1\right) + ||\tilde{D}^*_{\tilde{T}^c\cap T_0}f||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1$$

Using the reverse triangle inequality and $\hat{f} = f + h$,

$$\begin{aligned}\omega||\tilde{D}^*_{\tilde{T}\cap T_0^c}h||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1 \leq \omega||\tilde{D}^*_{\tilde{T}\cap T_0}h||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0}h||_1 \\ + 2\left(||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1 + \omega||\tilde{D}^*_{\tilde{T}\cap T_0^c}f||_1\right)\end{aligned} \tag{4.18}$$

On the left hand side of (4.18), we add and subtract $\omega||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1$. Similarly, we add and subtract $\omega||\tilde{D}^*_{\tilde{T}^c\cap T_0}h||_1$ and $2\omega||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1$ from the right hand side of (4.18). This, along with the fact that:

1. $||\tilde{D}^*_{T_0^c}h||_1 = ||\tilde{D}^*_{\tilde{T}\cap T_0^c}h||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1$
2. $||\tilde{D}^*_{T_0^c}f||_1 = ||\tilde{D}^*_{\tilde{T}\cap T_0^c}f||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1$
3. $||\tilde{D}^*_{T_0}h||_1 = ||\tilde{D}^*_{\tilde{T}\cap T_0}h||_1 + ||\tilde{D}^*_{\tilde{T}\cap T_0}h||_1$

we obtain:

$$\begin{aligned}\omega||\tilde{D}^*_{T_0^c}h||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1 \leq \omega||\tilde{D}^*_{T_0}h||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0}h||_1 \\ + 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1\right)\end{aligned}$$

We also have $\omega||\tilde{D}^*_{T_0^c}h||_1 = ||\tilde{D}^*_{T_0^c}h||_1 - (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}h||_1 - (1-\omega)||\tilde{D}^*_{\tilde{T}\cap T_0^c}h||_1$ so that

$$||\tilde{D}^*_{T_0^c}h||_1 \leq (1-\omega)\left(||\tilde{D}^*_{\tilde{T}\cap T_0^c}h||_1 + ||\tilde{D}^*_{\tilde{T}^c\cap T_0}h||_1\right) + \omega||\tilde{D}^*_{T_0}h||_1$$

$$+ 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)$$

Recall, $\tilde{T}_\alpha = T_0 \cap \tilde{T}$ and noticing that $(\tilde{T}\cap T_0^c)\cup(\tilde{T}^c\cap T_0) = (T_0\cup\tilde{T})\backslash\tilde{T}_\alpha$ we arrive at the following:

$$\begin{aligned}||\tilde{D}^*_{T_0^c}h||_1 \leq{}& \omega||\tilde{D}^*_{T_0}h||_1 + (1-\omega)||\tilde{D}^*_{T_0\cup\tilde{T}\backslash T_\alpha}h||_1\\ &+ 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)\end{aligned}$$

Then we have

$$\begin{aligned}||\tilde{D}^*_{T_{01}^c}h||_2 &\leq \frac{1}{\sqrt{b}}\Bigg(\omega||\tilde{D}^*_{T_0}h||_1 + (1-\omega)||\tilde{D}^*_{T_0\cup\tilde{T}\backslash T_\alpha}h||_1\\ &\quad + 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)\Bigg)\\ &\leq \frac{1}{\sqrt{b}}\Bigg(\omega\sqrt{s}||\tilde{D}^*_{T_0}h||_2 + (1-\omega)\sqrt{s(1+\rho-2\alpha\rho)}||\tilde{D}^*_{T_0\cup\tilde{T}\backslash T_\alpha}h||_2\\ &\quad + 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)\Bigg)\\ &\leq \frac{1}{\sqrt{b}}\Bigg(\omega\sqrt{s}||\tilde{D}^*h||_2 + (1-\omega)\sqrt{s(1+\rho-2\alpha\rho)}||\tilde{D}^*h||_2\\ &\quad + 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)\Bigg)\\ &\leq \frac{1}{\sqrt{b}}\Bigg(\omega\sqrt{s}\sqrt{\tilde{B}}||h||_2 + (1-\omega)\sqrt{s(1+\rho-2\alpha\rho)}\sqrt{\tilde{B}}||h||_2\\ &\quad + 2\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)\Bigg)\\ &= \sqrt{\frac{s}{b}}\left(\gamma\sqrt{\tilde{B}}||h||_2 + \eta\right)\end{aligned}$$

where $\gamma = \omega + (1-\omega)\sqrt{1+\rho+2\alpha\rho}$ and $\eta = \frac{2}{\sqrt{s}}\left(\omega||\tilde{D}^*_{T_0^c}f||_1 + (1-\omega)||\tilde{D}^*_{\tilde{T}^c\cap T_0^c}f||_1\right)$

□

Lemma 4.3. $||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2$ *satisfies*

$$\begin{aligned}2\epsilon \geq \sqrt{1-\delta_{s+a}}||\Phi D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 \\ - \sqrt{(1+\delta_b)B}\left(\sqrt{\frac{s}{b}}\left(\gamma\sqrt{\tilde{B}}||h||_2+\eta\right)\right)\end{aligned}$$

Proof. Recall that $h = f - \hat{f}$ and since $\hat{f}$ is feasible we have

$$\begin{aligned}
2\epsilon &\geq ||\Phi h||_2 \\
&= ||\Phi D\tilde{D}^* h||_2 \\
&= ||\Phi D_{T_{01}}\tilde{D}^*_{T_{01}}h + \Phi D_{T^c_{01}}\tilde{D}^*_{T^c_{01}}h||_2 \\
&\geq ||\Phi D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - ||\Phi D_{T^c_{01}}\tilde{D}^*_{T^c_{01}}h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - ||\Phi(D_{T_2}\tilde{D}^*_{T_2} + D_{T_3}\tilde{D}^*_{T_3} + \dots)h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \left(||\Phi D_{T_2}\tilde{D}^*_{T_2}h||_2 + ||\Phi D_{T_3}\tilde{D}^*_{T_3}h||_2 + \dots\right) \\
&= \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \sum_{j\geq 2}||\Phi D_{T_j}\tilde{D}^*_{T_j}h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \sqrt{1+\delta_b}\sum_{j\geq 2}||D_{T_j}\tilde{D}^*_{T_j}h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \sqrt{1+\delta_b}\sum_{j\geq 2}||D_{T_j}||_2||\tilde{D}^*_{T_j}h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \sqrt{(1+\delta_b)B}\sum_{j\geq 2}||\tilde{D}^*_{T_j}h||_2 \\
&\geq \sqrt{1-\delta_{s+a}}||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 - \sqrt{(1+\delta_b)B}\left(\sqrt{\frac{s}{b}}\left(\gamma\sqrt{\tilde{B}}||h||_2+\eta\right)\right)
\end{aligned}$$

□

We are now in a position to prove the theorem. We have (4.16), i.e.,

$$||h||_2 \leq ||D_{T_{01}}\tilde{D}^*_{T_{01}}h||_2 + \sqrt{B}||\tilde{D}^*_{T^c_{01}}h||_2$$

and applying Lemma 4.2 and Lemma 4.3 we obtain

$$K_1||h||_2 \leq 2\epsilon + K_2\eta \tag{4.19}$$

where

$$K_1 = \sqrt{1-\delta_{s+a}} - \eta\left(\sqrt{\frac{sB\tilde{B}(1-\delta_{s+a})}{b}} + \sqrt{\frac{sB\tilde{B}(1+\delta_b)}{b}}\right)$$
$$K_2 = \sqrt{\frac{sB(1-\delta_{s+1})}{b}} + \sqrt{\frac{sB(1+\delta_b)}{b}}$$

We require K_1 to be positive, in which case

$$\left(1-\gamma\sqrt{\frac{sB\tilde{B}}{b}}\right)^2 \delta_{s+a} + \frac{sB\tilde{B}}{b}\delta_b < 1 - 2\gamma\sqrt{\frac{sB\tilde{B}}{b}} \tag{4.20}$$

and this proves our theorem.

4.3.3 Weighted Optimal Dual Analysis

Recall the optimal-dual ℓ_1 analysis formulation (3.14). We consider a weighted version of this problem:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d, D\tilde{D}^*=I}{\arg\min} ||\tilde{D}^*\tilde{f}||_{1,w} \text{ s.t. } ||y-\Phi\tilde{f}||_2 \leq \epsilon \tag{4.21}$$

As a similar derivation as in the non-weighted case will show, we reformulate this problem as follows:

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d, g\in\mathbb{R}^N}{\arg\min} ||\bar{D}^*\tilde{f} + Pg||_{1,w} \text{ s.t. } ||y-\Phi\tilde{f}||_2 \leq \epsilon \tag{4.22}$$

In the previous section, we gave the results from [44] that stated optimal dual-based ℓ_1-analysis and synthesis are equivalent. We extend their proof to the weighted case.

Theorem 4.3.2. *Weighted ℓ_1-synthesis and weighted optimal-dual-based ℓ_1-analysis are equivalent.*

Proof. First note that we can put the weighted-optimal-dual-based ℓ_1-analysis formulation (4.22) into matrix notation

$$\hat{f} = \underset{\tilde{f}\in\mathbb{R}^d, g\in\mathbb{R}^N}{\arg\min} ||\bar{D}^*\tilde{f} + Pg||_{1,w} \text{ s.t. } ||y - \Phi\tilde{f}||_2 \leq \epsilon \tag{4.23}$$

$$= \underset{\tilde{f}\in\mathbb{R}^d, g\in\mathbb{R}^N}{\arg\min} ||W\left(\bar{D}^*\tilde{f} + Pg\right)||_1 \text{ s.t. } ||y - \Phi\tilde{f}||_2 \leq \epsilon \tag{4.24}$$

where W is a diagonal matrix whose entries are ω or 1 on the diagonal, and 0 otherwise. Following the proof of Theorem 3.3.1, start with the weighted-optimal-dual-based ℓ_1-analysis formulation (4.24). Let $\tilde{x} = \bar{D}^*\tilde{f} + Pg$, then we have

$$\begin{aligned} D\tilde{x} &= D\bar{D}^*\tilde{f} + DPg \\ &= I\tilde{f} + 0 \\ &= \tilde{f} \end{aligned}$$

Since the columns of $[\bar{D}^*, P]$ span the whole N-dimensional space and both $\tilde{f}$ and g are free, then any $\tilde{x} \in \mathbb{R}^N$ can be put in this form. We then rewrite (4.24),

$$\hat{f} = D \underset{\tilde{x}\in\mathbb{R}^N}{\arg\min} ||W\tilde{x}||_1 \text{ subject to } ||\Phi D\tilde{x} - y||_2 \leq \epsilon \tag{4.25}$$

Conversely, starting with (4.25) we have that for any $x \in \mathbb{R}^N$, the following decomposition always holds

$$\begin{aligned} \tilde{x} = \tilde{x}_R + \tilde{x}_N &= D^*(DD^*)^{-1}D\tilde{x} + P\tilde{x} \\ &= \bar{D}^*D\tilde{x} + P\tilde{x} \end{aligned}$$

where $\tilde{x}_R$ and $\tilde{x}_N$ are the components of $\tilde{x}$ belonging to the row space and null space of D, respectively. By definition of weighted-ℓ_1-synthesis, we have $\tilde{f} = D\tilde{x} \in \mathbb{R}^d$. Again since g is free, let $g = \tilde{x} \in \mathbb{R}^N$, we can arrive at the weighted-optimal dual-based ℓ_1-analysis formulation. □

Recovery Conditions for Weighted ℓ_1-Synthesis via Weighted ℓ_1-Analysis

As a direct consequence of Theorem 4.3.1 and Theorem 4.3.2, we have a theorem for weighted ℓ_1-synthesis that holds even when using highly coherent dictionaries by simply replacing the general dual with an optimal dual. This is analogous to the non-weighted work of Li et al [44].

Theorem 4.3.3. *Consider a signal $f = Dx$, where $x \in \mathbb{R}^N$, and $D \in \mathbb{R}^{d\times N}$ is an arbitrary frame of $\mathbb{R}^d$ with frame bounds $0 < A \leq B < \infty$ and we have noisy linear measurements:*

$$y = \Phi Dx + z \tag{4.26}$$

Let $\tilde{D}_o$ be an optimal dual frame of D with frame bounds $0 < \tilde{A}_o \leq \tilde{B}_o < \infty$. Let $\tilde{T} \subset \{1, ..., N\}$ be an arbitrary set. Let T_0 index the support of the best s-term approximation of $\tilde{D}_o{}^* f$ and thus $|T_0| = s$. Define ρ and α such that $|\tilde{T}| = \rho s$ and $|\tilde{T} \cap T_0| = \alpha\rho s$. Suppose Φ has D-RIP with constant δ that satisfies

$$\left(1 - \gamma\sqrt{\frac{sB\tilde{B}_o}{b}}\right)^2 \delta_{s+a} + \frac{sB\tilde{B}_o}{b}\delta_b < 1 - 2\gamma\sqrt{\frac{sB\tilde{B}_o}{b}} \tag{4.27}$$

for some positive integers a and b satisfying $0 < b - a \leq 3a$ where $\gamma = \omega + (1-\omega)\sqrt{1 + \rho - 2\alpha\rho}$ and $0 \leq \omega \leq 1$ with our weights chosen as follows:

$$w_i = \begin{cases} 1 & \text{if } i \in \tilde{T}^c \\ \omega & \text{if } i \in \tilde{T} \end{cases} \tag{4.28}$$

and $\tilde{T}$ is a support estimate of $\tilde{D}_o{}^* f$. Then the solution $\hat{f}$ to (4.21) (or (4.1) after reconstruction via D) satisfies

$$||f - \hat{f}||_2 \leq c_1 \left(\omega||(\tilde{D}_o)^*_{T_0^c} f||_1 + (1-\omega)||(\tilde{D}_o)^*_{\tilde{T}^c \cap T_0^c} f||_1\right) + c_2\epsilon \tag{4.29}$$

where c_1 and c_2 are some well behaved constants and ϵ is an upper bound on the noise.

Chapter 5

Synthetic Experiments

5.1 Optimal Dual ℓ_1-Analysis

5.1.1 Weighted vs Non-Weighted

Recall the optimal-dual-ℓ_1-analysis can be formulated as follows:

$$\hat{f} = \underset{f\in\mathbb{R}^d, g\in\mathbb{R}^N}{\arg\min} ||\bar{D}^* f + Pg||_1 \text{ s.t. } ||y - \Phi f||_2 \leq \epsilon \tag{5.1}$$

where P is the orthogonal projection onto the nullspace of D. In this section we briefly outline an iterative algorithm developed by Li et al [45], based on the split Bregman iteration, to solve (5.1). Here is the algorithm they derived based on the split Bregman iteration[27]:

Algorithm 3 Split Bregman iteration for optimal dual analysis

Initialization: $f^0 = 0, d^0 = b^0 = Pg^0 = 0, c^0 = 0$, $\mu > 0$, $\lambda > 0$, $nOuter$,$nInner$,tol;
while $k < nOuter$ and $||\Phi f^k - y||_2 > tol$ **do**
  for $n = 1 : nInner$ **do**
    $f^{k+1} = (\mu\Phi^*\Phi + \lambda\bar{D}\bar{D}^*)^{-1}[\mu\Phi^*(y - c^k) + \lambda\bar{D}(d^k - Pg^k - b^k)]$
    $d^{k+1} = \text{shrink}(\bar{D}^* f^{k+1} + Pg^k + b^k, \frac{1}{\lambda})$
    $Pg^{k+1} = P(d^{k+1} - \bar{D}^* f^{k+1} - b^k)$
    $b^{k+1} = b^k + (\bar{D}^* f^{k+1} + Pg^{k+1} - d^{k+1})$
  end for
  $c^{k+1} = c^k + (\Phi f^{k+1} - y)$
  Increase k
end while

As a numerical contribution of this thesis we adapt this algorithm for the weighted version of (5.1), namely (4.24). Following the derivation of Li et al [45], this modification merely introduces weights into the soft threshholding operator *shrink* of **Algorithm 3**.

Algorithm 4 Weighted split Bregman iteration for optimal dual analysis

Initialization: $f^0 = 0, d^0 = b^0 = Pg^0 = 0, c^0 = 0$, $\mu > 0$, $\lambda > 0$, $nOuter$,$nInner$,tol;
while $k < nOuter$ and $||\Phi f^k - y||_2 > tol$ **do**
  for $n = 1 : nInner$ **do**
    $f^{k+1} = (\mu\Phi^*\Phi + \lambda\bar{D}\bar{D}^*)^{-1}[\mu\Phi^*(y - c^k) + \lambda\bar{D}(d^k - Pg^k - b^k)]$
    $d^{k+1} =$shrink$(\bar{D}^* f^{k+1} + Pg^k + b^k, w_\lambda)$ i.e., **Weighted shrink**
    $Pg^{k+1} = P(d^{k+1} - \bar{D}^* f^{k+1} - b^k)$
    $b^{k+1} = b^k + (\bar{D}^* f^{k+1} + Pg^{k+1} - d^{k+1})$
  end for
  $c^{k+1} = c^k + (\Phi f^{k+1} - y)$
  Increase k
end while

This form of weighting is not new, though . Indeed, it has been examined in papers such as Asif and Romberg [3], Bioucas-Dias et al [6], and in the context of approximate message passing for the purposes of compressed sensing by Navid Ghadermarzy in his thesis [24]. This form of weighting also comes up in the weighted SPGl1 algorithm [66].

Numerical Results

We now look at a numerical example. Consider the frame $D \in \mathbb{C}^{n\times 2n}$ made of the concatenation of the Fourier basis and the identity, ie $D = [F, I]/\sqrt{2}$ where $F \in \mathbb{C}^{n\times n}$ is the normalized discrete Fourier transform matrix and $I \in \mathbb{R}^{n\times n}$ is the identity. Letting $n = 128$ we will construct signals $f = Dx$ where x is 20-sparse with respect to D. We will collect 64 Gaussian measurements of the form

$$y = ADx$$

where A is a standard normal Gaussian. We now wish to compare weighted and non-weighted optimal dual ℓ_1-analysis using the Split Bregman algorithm. For this experiment we will suppose that we know apriori the signal is 20-sparse, and for example if we indicate that a weighting strategy is 50% accurate that means we've correctly identified 10 indices and incorrectly identified 10. That is we assign a weight to 20 indices, half of which are correct. The same logic applies to 25% and 75% accuracy levels.

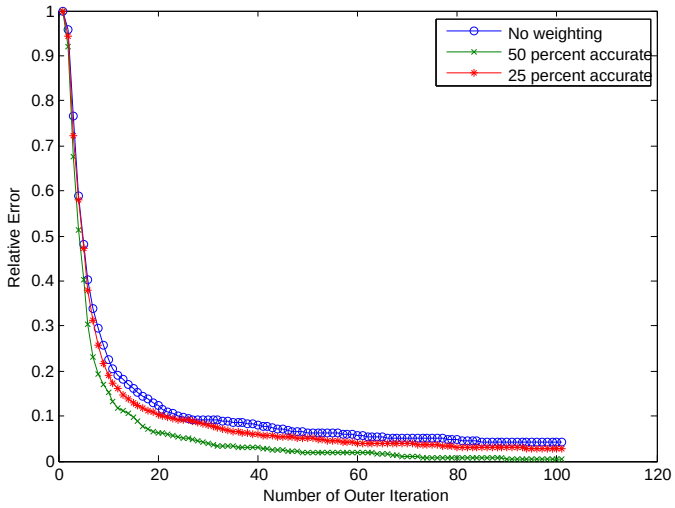


Figure 5.1: Average relative error of 10 reconstructed signals at a given outer iteration of the Split Bregman algorithm

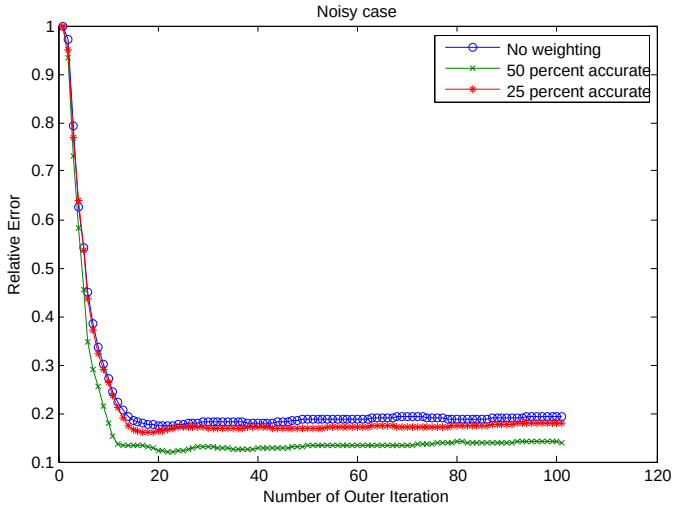


Figure 5.2: Measurements are now corrupted by a small amount of white noise η, i.e., $y = \Phi ADx + \eta$

5.1.2 Iterative Reweighting

In the case where we have no support estimate, we can still use weighting as a way of accelerating convergence of the Split Bregman algorithm by using the support of the current iterate as an estimate of the support of the actual signal, as in iterative reweighting for analysis and synthesis.

Algorithm 5 Iterative reweighting split Bregman iteration for optimal dual analysis

Initialization: $f^0 = 0$,$d^0 = b^0 = Pg^0 = 0$,$c^0 = 0$, $\mu > 0$, $\lambda > 0$, $nOuter$,$nInner$,tol,$weights = 1$;
while $k < nOuter$ and $||\Phi f^k - y||_2 > tol$ **do**
  for $n = 1 : nInner$ **do**
    $f^{k+1} = (\mu\Phi^*\Phi + \lambda\bar{D}\bar{D}^*)^{-1}[\mu\Phi^*(y - c^k) + \lambda\bar{D}(d^k - Pg^k - b^k)]$
    d^{k+1} =shrink$(\bar{D}^* f^{k+1} + Pg^k + b^k, w_\lambda)$ i.e., **Weighted shrink**
    $b^{k+1} = b^k + (\bar{D}^* f^{k+1} + Pg^{k+1} - d^{k+1})$
    Update weights w_λ based on largest magnitude coefficients of d^{k+1}.
  end for
  $c^{k+1} = c^k + (\Phi f^{k+1} - y)$
  Increase k
end while

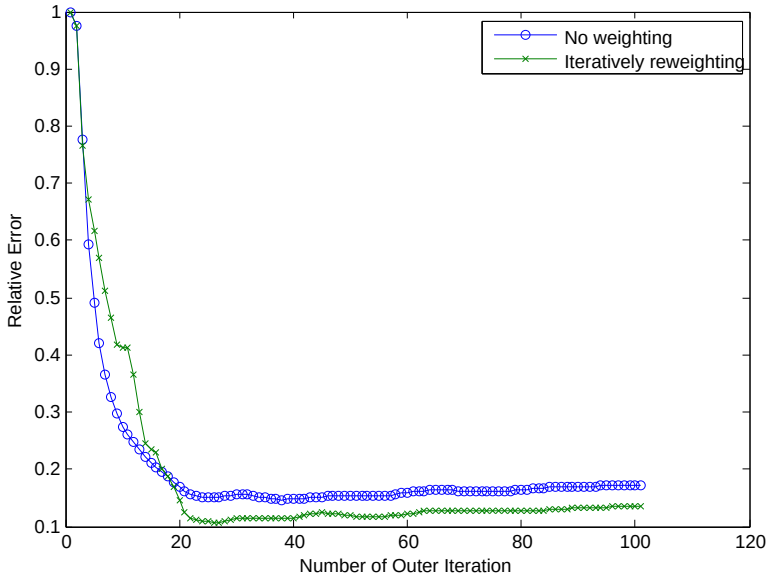


Figure 5.3: Iterative reweighting with no initial support estimate vs no weighting for the noisy case. We update our weights by calculating the coefficients of d^{k+1} which contribute 90 % of d^{k+1}'s cumulative energy.

5.1.3 The Optimal Dual Frame Bounds

As stated before, without actually calculating the optimal dual we cannot know apriori what the frame bounds of the optimal dual are without actually solving an optimization problem. This poses a serious problem theoretically since we have recovery conditions which depend on the upper frame bound of the optimal dual. Here we wish to provide some empirical evidence to show that for a particular frame, its optimal dual upper frame bound is small and in some sense largely independent of the signal.

Define the redundancy of a frame $D \in \mathbb{R}^{d\times N}$ as $\kappa = \frac{N}{d}$. Here we wish to provide evidence for which a particular frame has the property of having $\tilde{B}_o \approx \sqrt{\kappa}$. As in the previous section, we construct a set of signals and apply the split bregman algorithm, however here we calculate the optimal dual as a byproduct of the algorithm and compute its upper frame bound.

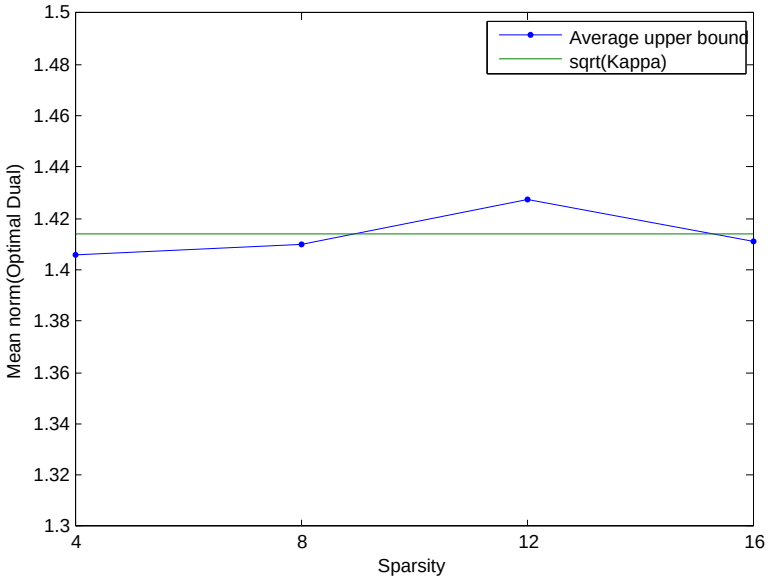


Figure 5.4: Average upper frame bound of the optimal dual reconstructed from signals of varying sparsity.

We will now take the example of a partial Fourier matrix, i.e., take the $N \times N$ discrete Fourier transform matrix, and keep only d rows. For a variety of d, we calculate the upper frame bound of the calculated optimal dual of the experiment and plot it in Figure 5.4 for various sparsity levels. We now have a reasonable expectation of having $\tilde{B}_o$ being small which is an important theoretical aspect of the recovery conditions for optimal dual ℓ_1 analysis, and thus synthesis.

Chapter 6

Seismic Wavefield Reconstruction

6.1 Introduction

Reflection seismology is a technique used by the oil industry for hydrocarbon reservoir exploration to explore and identify potential oil-rich areas in the earth. This technique involves propagating energy down into the earth and then collecting the reflected energy via some measurement device.

The energy propagated is generated via a given source. This source sends out a complicated, and largely unknown waveform, attenuates, reflects, transmits, changes modes, and scatters about while a set of receivers records what comes their way. On land, the source is typically given by either a Vibroseis truck or dynamite, and in marine environments airguns are used. The equipment used to record this energy on land is done via geophone receivers, whereas in marine environments hydrophone receivers are used. The recorded information gives an indirect measurement of the earth's structure. This recorded information is then processed to create an image of the subsurface.

The procedure of processing acquired data to develop an image of the subsurface can be described in a workflow chart. In Figure 6.1, we see a basic geophysical workflow. Though certainly non-exhaustive, these are some of the main movements from acquisition to interpretation of seismic data.

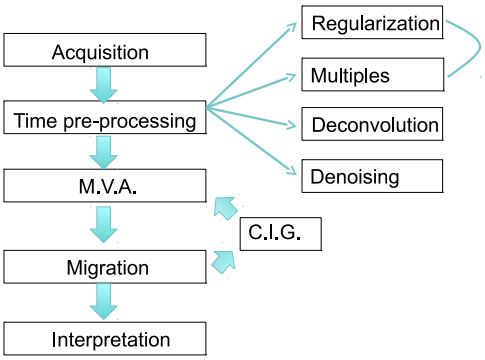


Figure 6.1: Basis seismic processing workflow. M.V.A stands for Migration Velocity Analysis and C.I.G. stands for Common Image Gathers

Let us briefly outline some of the key components in the geophysical workflow and note recent progress with respect to sparse approximation and compressive sampling, leaving regularization for the following section as it is the main objective of this chapter.

1. Acquisition: The collection of raw seismic data, where the objective is to probe the unknown geological subsurface. As stated above, this is done by sending seismic waves through the earth surface by a source. The seismic waves are then collected by the receivers. Leveraging principals of compressive sensing for seismic acquisition, Herrmann and Wason investigated randomization for ocean bottom sensors [68]. In another instance, Mansour et al [49] propose a practical randomized marine acquisition scheme for marine acquisition where the sequential sources fire airguns at only randomly time-dithered instances.

2. Multiples: When a seismic signal makes one reflection before being recorded by a receiver, we call this a primary reflection. On the other hand, when a seismic signal bounces more than once we call this a multiple reflection. Multiples can be characterized in a variety of ways, e.g., near-surface, interformational, and peg-leg [55]. Herrmann et al [34] exploit curvelet domain sparsity of seismic data and propose a data-adaptive method that corrects amplitude errors, which vary smoothly as a function of location, scale (frequency band), and angle.

3. Deconvolution: The premise of deconvolution is that seismic traces are the convolution of a reflectivity series with a seismic wavelet. Thus the

goal of deconvolution is to seperate a seismic trace into its constituent parts. For an investigation of deconvolution using advances in sparse inversion, see Herrmann and Lin [42]. Gholami, A. and Sacchi, M. use sparsity in a wavelet domain as a source prior for the purposes of blind seismic deconvolution [25].

4. Migration: Also referred to as event movement, the goal of migration is to give a true representation of the subsurface by converting from time to depth, e.g., moving primary reflections to their true spatial locations. The following papers all pose this problem as a sparsity-promoting formulation [31, 41, 62, 63].

What is not mentioned in the workflow is the fact that that a large majority of algorithms existing in seismic processing rely on the fact the data is given on a regular grid, we investigate this now.

6.2 Seismic Data Interpolation

In the remainder of this chapter, we will focus entirely on the regularization portion of the workflow. By regularization, we mean the process of taking seismic data on a irregular grid (e.g., Figure 6.2) and projecting it to a regularly sampled grid (e.g., Figure 6.3). The fact that acquired data is on an irregular grid is a byproduct of acquisition constraints, which are various in nature. Some physical constraints might be nearby oil wells, weather, or other seismic vessels. Even if ocean bottom sensors are used, in which placement of the node is more accurate, the ocean bottom could be jagged or sloped, causing further grief.

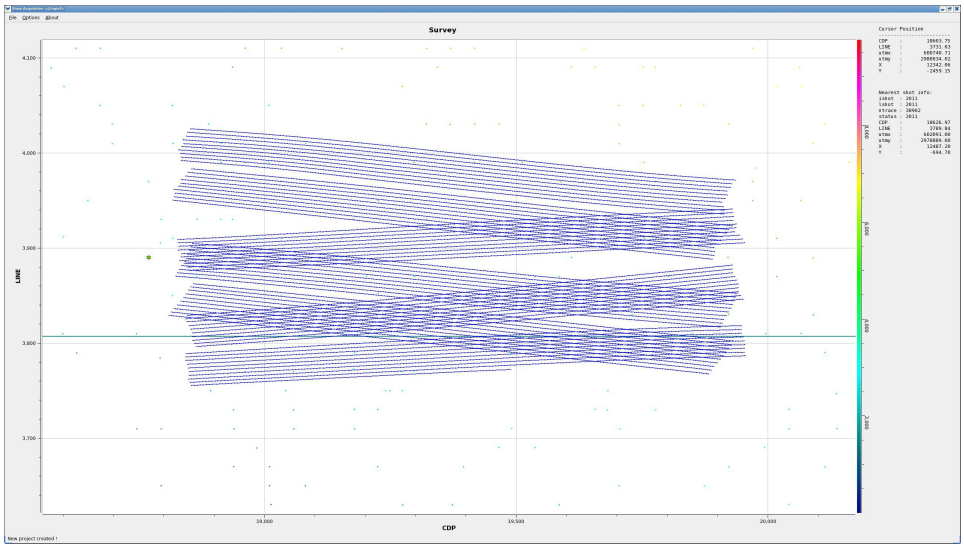


Figure 6.2: An example of a top down view of a seismic acquisition geometry provided by Total E & P USA. Each line represents a set of seismic receivers which are trailing a seismic vessel.

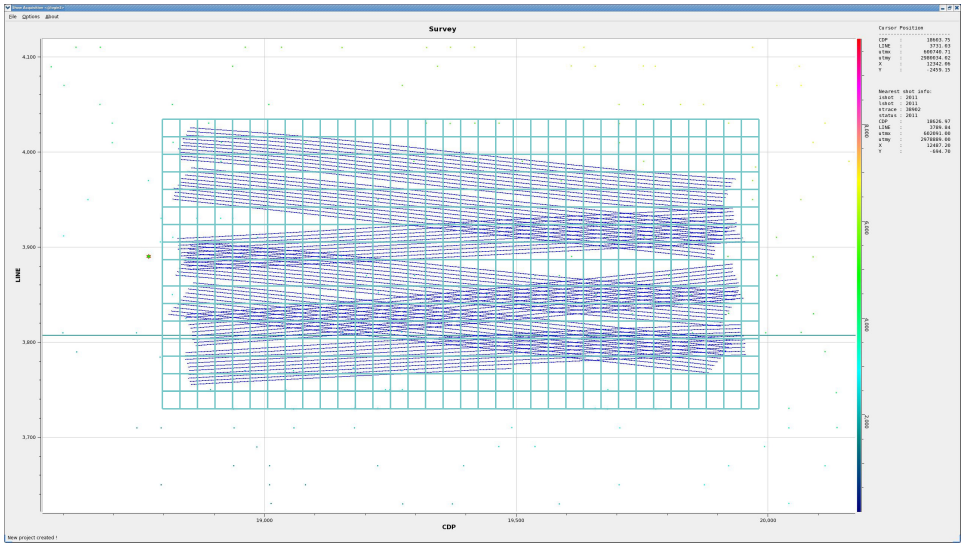


Figure 6.3: A regular grid overlaying the seismic acquisition. Each intersection of the grid corresponds to a physical location for which we require information.

There are a variety of strategies available to accomplish the process of projecting data to a regular grid. For example one might take nearest receiver approximations or weighted average of neighboring receivers to estimate what a receiver at a target location would look like. An issue with these methods is how to choose a threshold for what is considered a close neighbor. If the threshhold is taken too small, no interpolation will take place and the problem of missing receivers remains. Certainly those grid points located in the gaps shown in Figure 6.4 will be missed by a nearest neighbour algorithm. If a threshold is chosen too high, the accuracy of the recovery will be sacrificed.

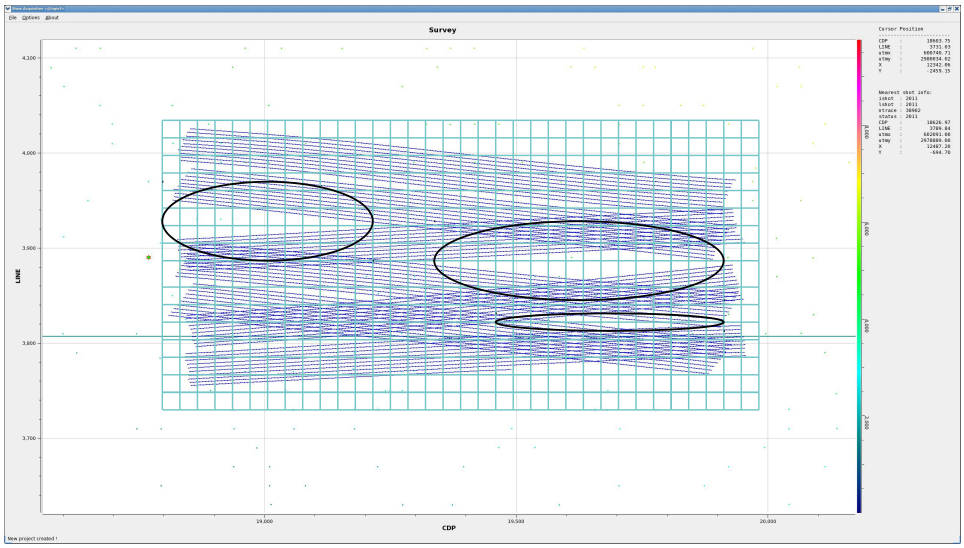


Figure 6.4: A regular grid overlaying the seismic acquisition. The circled areas represent regions for which there are no receivers collecting information.

Reconstruction methods can be divided into two main classes: wave equation based and signal processing based. The majority of methods in the latter category utilize transforms domains[50]. These transforms include Fourier[50],[43], Radon[60], and Curvelet[30]. Transform based methods can be easily adopted for irregularly sampled data, if the transform has a non-uniform counterpart.

6.3 Seismic Data Interpolation by Sparse Recovery

This section considers the methodology conducted in [48], however with the inclusion of the analysis formulation of the sparse recovery problem. Consider a seismic line with N_s sources, N_r receivers, and N_t time samples, thus there are $N = N_s N_r N_t$ datapoints under consideration. We assume that all sources see the same receivers. These N points can be represented in a 3-d cube as in Figure 6.6. In this particular experiment, the seismic line at full resolution has $N_s = 178$ sources, $N_r = 178$ receivers with a sample distance of 12.5 meters, and $N_t = 500$ time samples acquired with a sampling interval of 4 milliseconds.

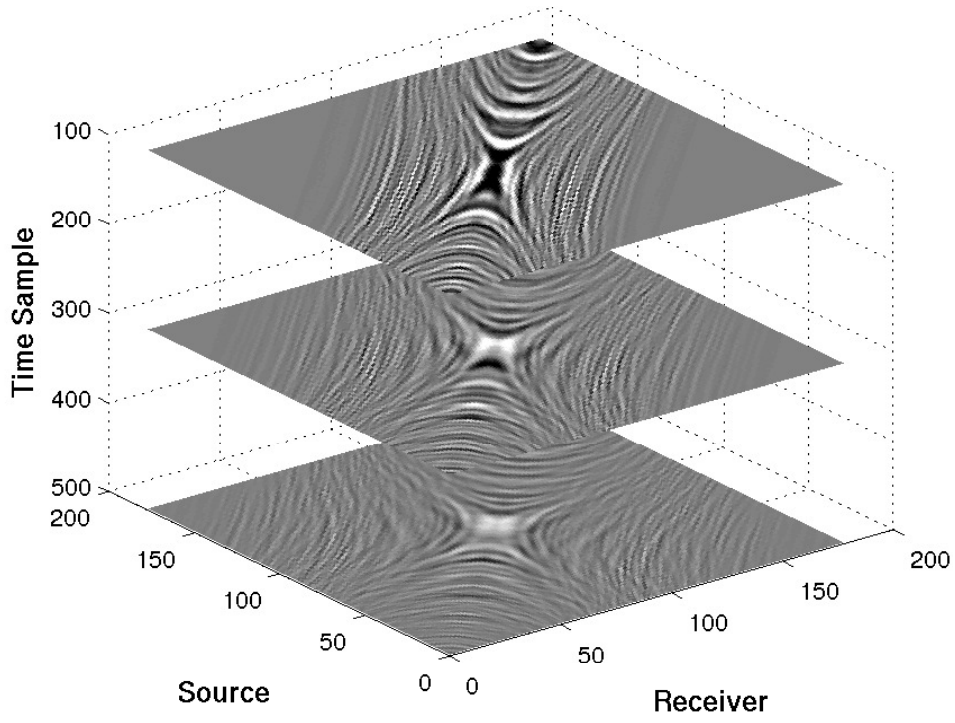


Figure 6.5: A seismic acquisition from the Gulf of Mexico expressed as a 3D volume, here we shows several time slices of the volume.

We consider the scenario for which each source is randomly missing 50% of its receivers. This is illustrated by removing a single shot gather corresponding to a slice in the 3D cube, as represented in Figure 6.7, with 50% of

its traces missing as shown in Figure 6.8. The problem is then interpolating those randomly missing traces to a complete periodic grid.

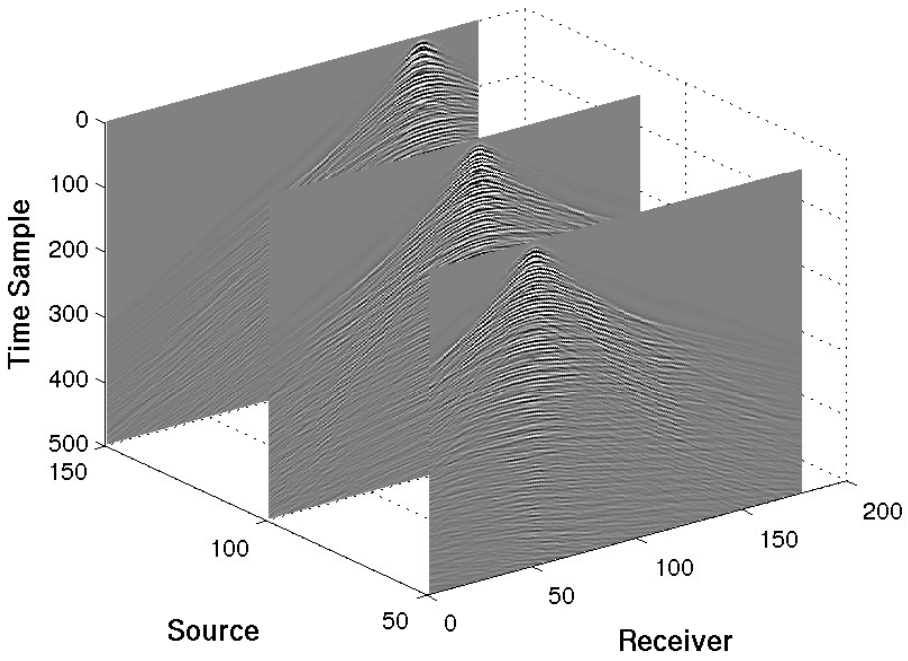


Figure 6.6: A seismic acquisition from the Gulf of Mexico expressed as a 3D volume, here we show several common shot gathers.

Expressing the shot gather as a single N-dimensional vector, our goal then is to recover an approximation $\hat{f}$ of the discretized wavefield f from the measurements

$$b = RMf \tag{6.1}$$

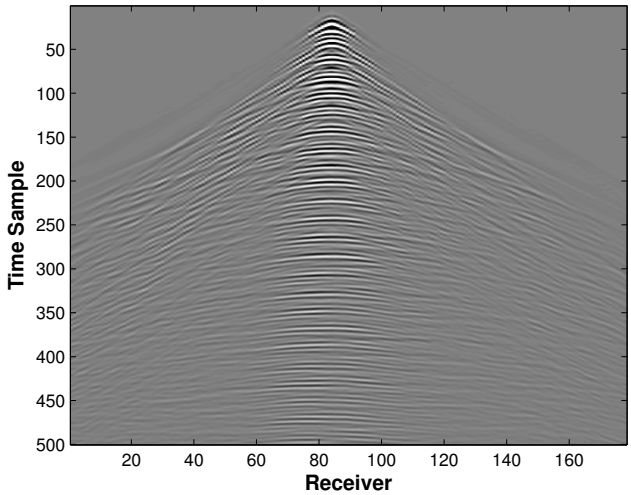


Figure 6.7: A common shot gather sliced from the 3D cube.

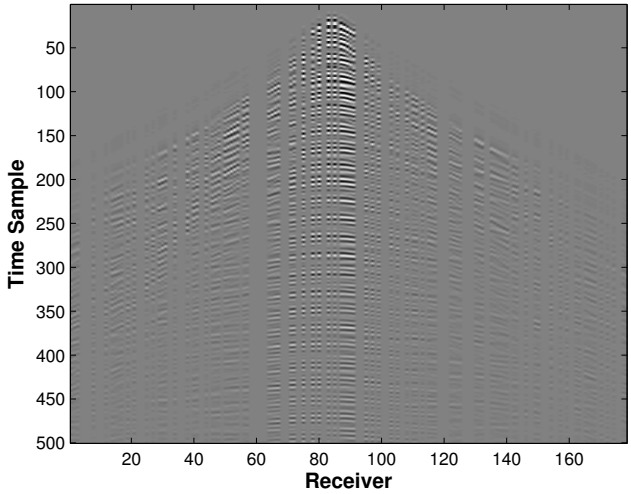


Figure 6.8: The shot gather sliced from the 3D cube in Figure 6.7 with randomly missing receivers.

where RM is a sampling operator composed of the product of a restriction matrix R with a measurement basis matrix M. The measurement matrix M represents the basis in which the measurements are taken and corresponds

to the Dirac (identity) basis in the missing trace interpolation problem.

The synthesis formulation attempts to find a sparse representation of the wavefield in a transform domain whose inverse transform coefficients fit the acquired data,

$$\hat{f}_{synthesis} = D \cdot \arg\min_{\tilde{x}} ||\tilde{x}||_1 \text{ subject to } ||RMD\tilde{x} - b||_2 \leq \epsilon \tag{6.2}$$

One issue with this approach is that the coefficients recovered may not be the forward transform of an actual wavefield, i.e., those coefficients may not represent an actual physical phenomenon. On the other hand, the traditional analysis formulation attempts to directly find a wavefield which fits the acquired data whose forwards transform coefficients are sparse.

$$\hat{f}_{analysis} = \arg\min_{\tilde{f}} ||D^\dagger \tilde{f}||_1 \text{ subject to } ||RM\tilde{f} - b||_2 \leq \epsilon \tag{6.3}$$

where $D^\dagger$ is the pseudoinverse (canonical dual) of D i.e., $D^\dagger = \bar{D}^*$. For the Curvelet transform, this is simple the transpose for which fast matrix vector products exist. Note here that it would likely be difficult to find a general dual frame for the Curvelet frame for which there still exists a fast implementation, which is the reason we only consider the canonical dual.

6.3.1 Incorporating Weights

As in [48], we incorporate weighting by performing the sparse recovery in the frequency domain. Rather than consider the wavefield as an $N = N_s N_r N_t$ dimensional vector, we convert take the Fourier transform of the data volume in the time domain and partition the volume into frequency slices. Each frequency slice is missing 50% of its traces randomly, and we wish to recover those slices. An example of this is shown in Figure 6.9 and Figure 6.10. After all the frequency slices have been recovered, one simply takes the inverse Fourier transform to see the recovered time domain solution.

We begin with no initial support estimate and then sweeping from low frequency to high frequencies, develop a support estimate and using the support of the current frequency slice as an estimate for the next frequency slice as illustrated in Figure 6.11. The continuity of wavefronts across adjacent partitions of a seismic line manifests itself as a high correlation in the

support of the Curvelet coefficients of the partitions, and thus the support of adjacent solutions will overlap.

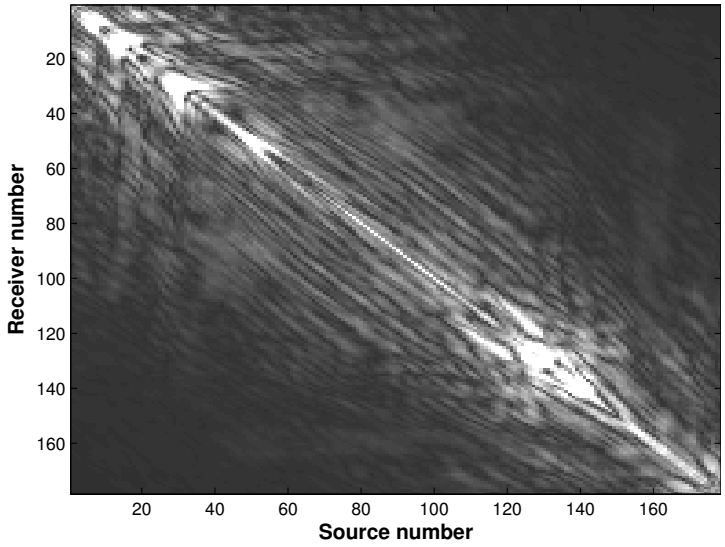


Figure 6.9: A frequency sliced from the 3D cube.

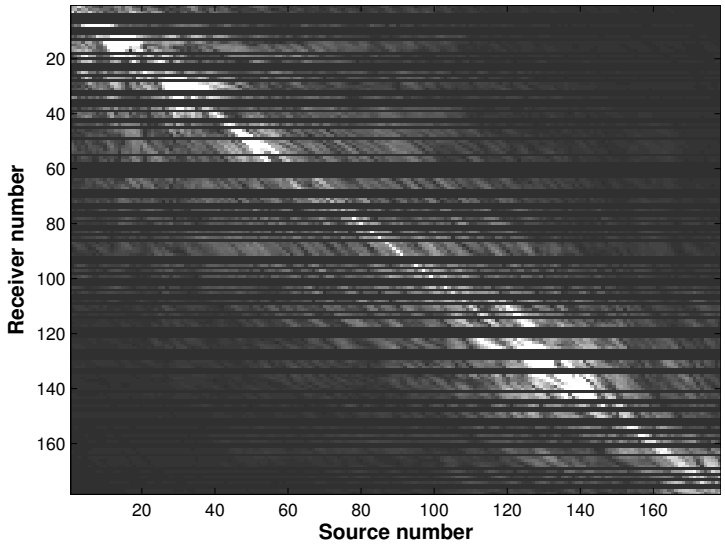


Figure 6.10: The frequency sliced from the 3D cube in Figure 6.9 with randomly missing receivers.

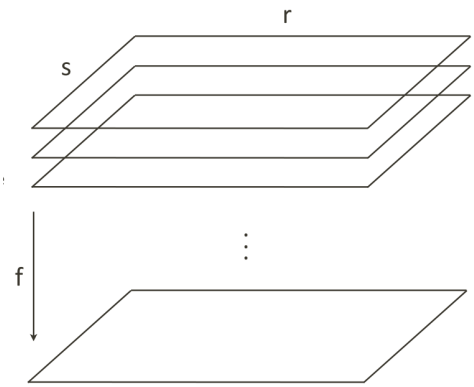


Figure 6.11: Illustration of weighting strategy, s: source , r: receiver , f: frequency . One begins by recovering low frequency slices, which are hopefully non-aliased, and then using those recovered slices to generate weights for recovering higher frequency slices

6.4 Gulf of Mexico Experiment Results

The experiment conducted here is exactly the same as described in [48] however we consider both analysis and synthesis approaches for solving the associated optimization problem. Note that if one acquires seismic data and then applies a nearest neighbor algorithm to approximate missing receivers, it would be a natural problem to consider if the nearest neighborhood algorithm misses some traces. Also, this scenario is certainly a reality if a jittersampling strategy is employed during acquisition. There are several benefits of applying such a strategy and for details the reader is referred to [57] and [30].

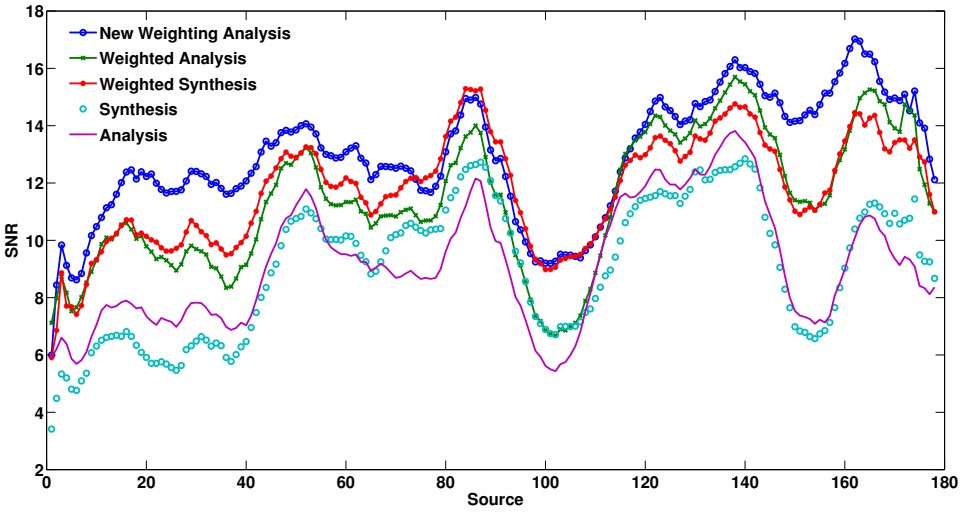


Figure 6.12: Seismic wavefield reconstruction of the Gulf of Mexico Data using analysis and synthesis with various weighting techniques. New weighting analysis refers to (4.8), and the others are the traditional weighting techniques.

It is clear that the non-weighted analysis and synthesis formulations perform poorer than their weighted counterparts. An interesting observation is that non-weighted analysis only outperforms non-weighted synthesis in some shot ranges and vice versa, a phenomenon that does not occur in the weighted methods in the same magnitude.

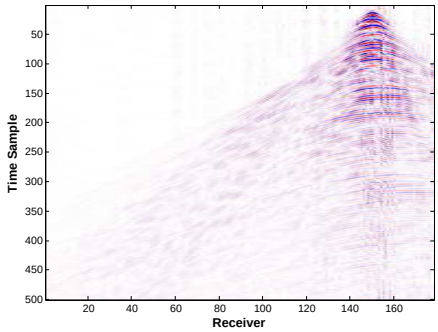


Figure 6.13: Recovered Synthesis

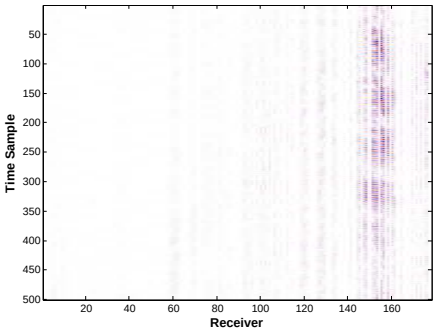


Figure 6.14: Difference plot

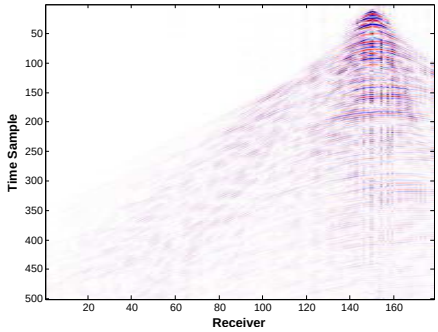


Figure 6.15: Recovered Analysis

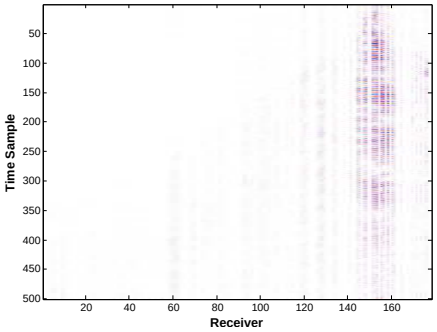


Figure 6.16: Difference plot

Note that upon inspecting Figure 6.12, the novel weighted method for analysis from Section 4.3.1 of Chapter 4 has a larger SNR than all other methods investigated among the majority of shots. It is only in the shot range of 80-100 where weighted synthesis performs slightly better. However in lowest SNR point of the shot range 140-160, the novel weighted method for analysis has at least a 2dB increase to it's closest competitor. Figures 6.13-6.22 show the recovery of the various algorithms for the 150th shot in the acquisition.

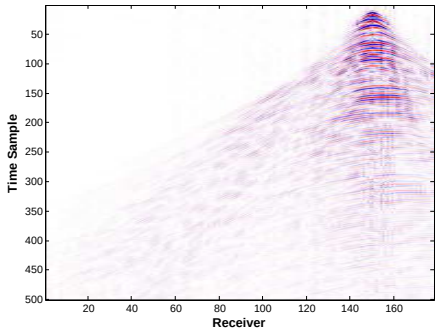


Figure 6.17: Recovered Weighted Synthesis

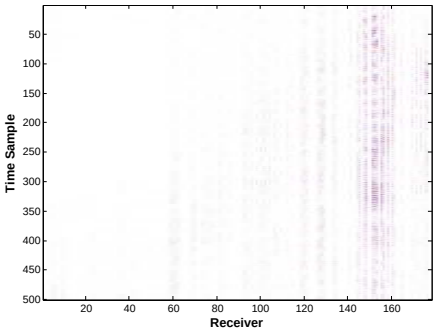


Figure 6.18: Difference plot

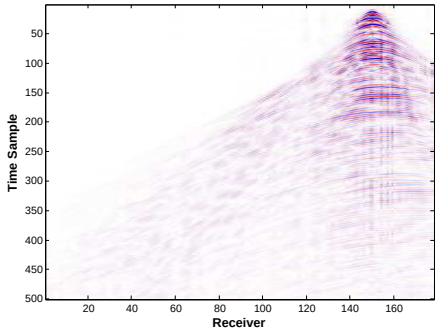


Figure 6.19: Recovered Weighted Analysis

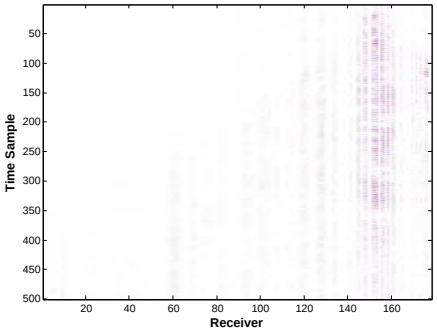


Figure 6.20: Difference plot

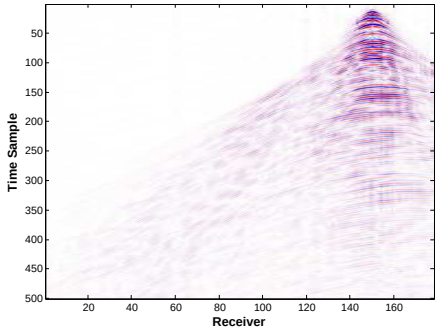


Figure 6.21: Recovered New Weighted Analysis

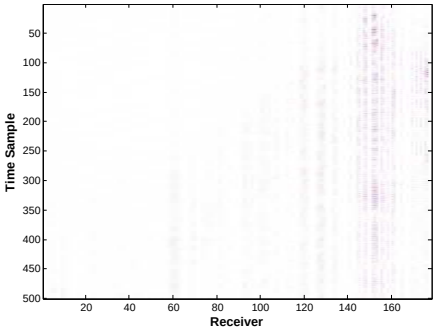


Figure 6.22: Difference plot

6.4.1 Computational Considerations

It should be noted that [48] use SPGL1 [66] to solve the optimization technique. Unfortunately, SPGL1 cannot solve the analysis formulation of the basis pursuit problem. A strength of NESTA[5] is that it can solve both the synthesis and analysis formulations. Thus for a fair comparison we have opted for NESTA.

One of the main goals of using compressive sensing in a seismic setting is to drive down the expense of computation. Calculating pseudoinverses is an expensive computation and the new weighting method considered here has to recalculate how the pseudoinverse acts on a vector at every iteration. This overhead might be too expensive for only a 1db increase in the SNR for the majority of the sources.

6.4.2 Conclusions and Future Work

We applied the method of [48], i.e., weighted one-norm minimization in a framework for the recovery of missing-traces caused by physical obstacles or deliberate subsampling in otherwise regularly-sampled seismic wavefields, to the analysis and synthesis formulations of the seismic interpolation problem along with their weighted extensions discussed in this thesis. Weighted one-norm minimization exploits correlations between the locations of significant coefficients of different partitions. Though we only considered source-receiver gathers in this experiment, it was shown in [48] that converting the data to time-midpoint and midpoint offset increases the SNR of the recovery of synthesis and weighted synthesis. We expect the same results for analysis, weighted analysis, and the novel weighted analysis method discussed in this thesis which we leave for future work.

Bibliography

[1] A. Aravkin, M. P. Friedlander, F. Herrmann, and T. van Leeuwen. Robust inversion, dimensionality reduction, and randomized sampling. *Mathematical Programming*, 134(1):101–125, 2012.

[2] Aleksandr Y. Aravkin, Michael P. Friedlander, and Tristan van Leeuwen. Robust inversion via semistochastic dimensionality reduction. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 5245 –5248, march 2012.

[3] Muhammad Salman Asif and Justin K. Romberg. Fast and accurate algorithms for re-weighted l1-norm minimization. *CoRR*, abs/1208.0651, 2012.

[4] Christopher A. Baker. A note on sparsification by frames. *CoRR*, abs/1308.5249, 2013.

[5] S. Becker, J. Bobin, and E. Candes. NESTA: A Fast and Accurate First-order Method for Sparse Recovery. *ArXiv e-prints*, April 2009.

[6] Jose M. Bioucas-Dias and Mario A. T. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007.

[7] T. Tony Cai and Anru Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *CoRR*, abs/1306.1154, 2013.

[8] E.J. Candes and M.B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21 –30, march 2008.

[9] Emmanuel J. Candès, Yonina C. Eldar, Deanna Needell, and Paige Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31(1):59–73, 2011.

[10] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[11] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted l_1 minimization. *J. Fourier Anal. Appl.*, 14(5-6):877–905, 2008.

[12] Peter G. Casazza, Gitta Kutyniok, and Friedrich Philipp. Introduction to finite frame theory. In *Finite frames*, Appl. Numer. Harmon. Anal., pages 1–53. Birkhäuser/Springer, New York, 2013.

[13] Ole Christensen. *Frames and bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2008. An introductory course.

[14] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[15] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[16] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[17] P.L. Dragotti and M. Vetterli. Wavelet footprints: theory, algorithms, and applications. *Signal Processing, IEEE Transactions on*, 51(5):1306 – 1323, may 2003.

[18] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. Technical report, 2005.

[19] Y.C. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Compressed Sensing: Theory and Applications. Cambridge University Press, 2012.

[20] Massimo Fornasier and Holger Rauhut. Compressive sensing, in. *Handbook of Mathematical Methods in Imaging*, 2010.

[21] Simon Foucart. A note on guaranteed sparse recovery via - minimization. *Applied and Computational Harmonic Analysis*, 29(1):97 – 103, 2010.

[22] Simon Foucart and Holger Rauhut. A mathematical introduction to compressive sensing. *Appl. Numer. Harmon. Anal. Birkhäuser, Boston*, 2013.

[23] Michael P. Friedlander, Hassan Mansour, Rayan Saab, and Özgür Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inform. Theory*, 58(2):1122–1134, 2012.

[24] Navid Ghadermarzy. Using prior support information in compressed sensing. Master's thesis, University of British Columbia, 2013.

[25] A. Gholami and M. Sacchi. Fast 3d blind seismic deconvolution via constrained total variation and gcv. *SIAM Journal on Imaging Sciences*, 6(4):2350–2369, 2013.

[26] R. Giryes and M. Elad. Can we allow linear dependencies in the dictionary in the sparse synthesis framework? *ArXiv e-prints*, March 2013.

[27] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM J. Img. Sci.*, 2(2):323–343, April 2009.

[28] Y. Gousseau and J. Morel. Are natural images of bounded variation? *SIAM Journal on Mathematical Analysis*, 33(3):634–648, 2001.

[29] Karlheinz Gröchenig. *Foundations of time-frequency analysis*, chapter 4.3, pages 61–63. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2001.

[30] Gilles Hennenfent and Felix J. Herrmann. Simply denoise: wavefield reconstruction via jittered undersampling. *Geophysics*, 73(3):V19–V28, 2008.

[31] Felix J. Herrmann, Cody R. Brown, Yogi A. Erlangga, and Peyman P. Moghaddam. Curvelet-based migration preconditioning and scaling. *Geophysics*, 74:A41, 09 2009.

[32] Felix J. Herrmann and Xiang Li. Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical Prospecting*, 60(4):696–712, 2012.

[33] F.J. Herrmann, M.P. Friedlander, and O. Yilmaz. Fighting the curse of dimensionality: Compressive sensing in exploration seismology. *Signal Processing Magazine, IEEE*, 29(3):88 –100, may 2012.

[34] F.J. Herrmann, D. Wang, and D. Verschuur. Adaptive curvelet-domain primary-multiple separation. *GEOPHYSICS*, 73(3):A17–A21, 2008.

[35] Laurent Jacques. A short note on compressed sensing with partially known signal support. *CoRR*, abs/0908.0660, 2009.

[36] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and Babak Hassibi. Weighted l1 minimization for sparse recovery with prior information. In *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 1*, ISIT'09, pages 483–487, Piscataway, NJ, USA, 2009. IEEE Press.

[37] Jelena Kovacevic and Amina Chebira. An introduction to frames. *Signal Processing*, 2(1):1–94, 2008.

[38] Gitta Kutyniok. Compressed sensing: Theory and applications. *CoRR*, abs/1203.3815, 2012.

[39] M. Lammers, A. Powell, and O. Yilmaz. Alternative dual frames for digital-to-analog conversion in sigmadelta quantization. *Advances in Computational Mathematics*, 32(1):73–102, 2010.

[40] Shidong Li. On general frame decompositions. *Numerical functional analysis and optimization*, 16(9-10):1181–1191, 1995.

[41] Xiang Li and Felix J. Herrmann. Sparsity-promoting migration accelerated by message passing. In *SEG Technical Program Expanded Abstracts*. SEG, SEG, 04 2012.

[42] Tim T.Y. Lin and Felix J. Herrmann. Robust source signature deconvolution and the estimation of primaries by sparse inversion. In *SEG Technical Program Expanded Abstracts*, volume 30, pages 4354–4359. Dept. of Earth and Ocean Sciences, University of British Columbia, Dept. of Earth and Ocean Sciences, University of British Columbia, 04 2011.

[43] B. Liu and M. Sacchi. Minimum weighted norm interpolation of seismic records. *GEOPHYSICS*, 69(6):1560–1568, 2004.

[44] Yulong Liu, S. Li, Tiebin Mi, Hong Lei, and Weidong Yu. Performance analysis l1-synthesis with coherent frames. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2042–2046, 2012.

[45] Yulong Liu, Tiebin Mi, and Shidong Li. Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis. *IEEE Trans. Inform. Theory*, 58(7):4201–4214, 2012.

[46] Wei Lu and Namrata Vaswani. Exact reconstruction conditions for regularized modified basis pursuit. *IEEE Transactions on Signal Processing*, 60(5):2634–2640, 2012.

[47] Stephane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Access Online via Elsevier, 2008.

[48] H. Mansour, F.J. Herrmann, and Ö. Yilmaz. Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization. *GEOPHYSICS*, 78(5):V193–V206, 2013.

[49] Hassan Mansour, Haneet Wason, Tim T.Y. Lin, and Felix J. Herrmann. Randomized marine acquisition with compressive sampling matrices. *Geophysical Prospecting*, 60(4):648–662, 2012.

[50] M. Naghizadeh and K. Innanen. Seismic data interpolation using a fast generalized fourier transform. *GEOPHYSICS*, 76(1):V1–V10, 2011.

[51] Sangnam Nam, Mike E. Davies, Michael Elad, and Rémi Gribonval. The Cosparse Analysis Model and Algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30–56, 2013. Preprint available on arXiv since 24 Jun 2011.

[52] Deanna Needell. Noisy signal recovery via iterative reweighted l1-minimization. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pages 113–117. IEEE, 2009.

[53] H. Nyquist. Certain topics in telegraph transmission theory. *American Institute of Electrical Engineers, Transactions of the*, 47(2):617 –644, april 1928.

[54] Von Borries R., Miosso C. J., and Cristhian Potes. Compressed sensing using prior information. *The Second International Workshop on Computational Advances in MultiSensor Adaptive Processing*, 2007.

[55] Enders A. Robinson and Sven Treitel. *Digital Imaging and Deconvolution: The ABCs of Seismic Exploration and Processing*. Society of Exploration Geophysicists, Tulsa, USA, November 2008.

[56] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[57] Reza Shahidi, Gang Tang, Jianwei Ma, and Felix J. Herrmann. Application of randomized sampling schemes to curvelet-based sparsity-promoting seismic data recovery. *Geophysical Prospecting*, 61(5):973–997, 2013.

[58] Claude E. Shannon. Communication in the presence of noise. *Proc. I.R.E.*, 37:10–21, 1949.

[59] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6):670–684, 2002.

[60] D. Trad, T. Ulrych, and M. Sacchi. Accurate interpolation with highresolution timevariant radon transforms. *GEOPHYSICS*, 67(2):644–656, 2002.

[61] Daniel Trad, Tadeusz Ulrych, and Mauricio Sacchi. Latest views of the sparse radon transform. *Geophysics*, pages 386–399, 2003.

[62] Ning Tu, Aleksandr Y. Aravkin, Tristan van Leeuwen, and Felix J. Herrmann. Fast least-squares migration with multiples and source estimation. In *EAGE*, 06 2013.

[63] Ning Tu and Felix J. Herrmann. Least-squares migration of full wavefield with source encoding. In *EAGE technical program*. EAGE, EAGE, 01 2012.

[64] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569 –587, april 2000.

[65] Fleet Van. *Discrete Wavelet Transformations An Elementary Approach with Applications*. Wiley-Interscience. Hoboken, NJ: John Wiley & Sons. xxiv, 2008.

[66] 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.

[67] Namrata Vaswani and Wei Lu. Modified-cs: modifying compressive sensing for problems with partially known support. *Trans. Sig. Proc.*, 58(9):4595–4607, September 2010.

[68] Haneet Wason and Felix J. Herrmann. Time-jittered ocean bottom seismic acquisition. In *SEG*, 9 2013.

[69] G Zimmermann. Normalized tight frames in finite dimensions. In *Recent Progress in Multivariate Approximation, Birkhauser (2001*, 2001.

[70] P. M. Zwartjes and M. D. Sacchi. Fourier reconstruction of nonuniformly sampled, aliased seismic data. *Geophysics*, 72(1):V21–V32, 2007.