Reduced Complexity Regularization of Geophysical Inverse Problems

A thesis

submitted by

Gregory Ely

In partial fulfillment of the requirements

for the degree of

Master of Science

in

Electrical Engineering

TUFTS UNIVERSITY

August 2013

ADVISER:

Shuchin Aeron

TUFTS UNIVERSITY

Abstract

Shuchin Aeron

Department of Electrical Engineering

Master of Science

by Gregory Ely

This thesis explores the application of complexity penalized algorithms to solve a variety of geophysical inverse problems: Hydraulic Fracture Monitoring (HFM), hyper-spectral imaging, and reflection seismology. Through these examples, the thesis examines how the physics of several systems gives rise to sparsity or low-dimensionality when posed in the proper basis. This low complexity can be quantified into several types of convex norms such as the ℓ_1 & nuclear norm. This paper demonstrates how minimization operations that encourage this reduced complexity by penalizing these convex norms can improve inversion. First & second order as well as stochastic algorithms are used to solve these minimization problems and I give details as to how the structure of the problem dictates the best technique to apply.

Acknowledgements

Many thanks to Zemin Zhang, Jason Gejie Liu, Ning Hao for their contributing work developing algorithms that I applied throughout thesis. Additional thanks to my paper collaborators Shuchin Aeron, Eric Miller, and Misha Kilmer. I am indebted to you for making this process much smoother.

Shuchin Aeron - Thesis advisor & co-author on original papers from which chapters 3-5 are derived.

Eric Miller - Thesis committee member & co-author on the original paper from which chapter 4 is derived.

Misha Kilmer - Thesis committee member & co-author on the original paper from which chapter 5 is derived.

Zemin Zhang - Algorithm collaborator.

Jason Gejie Liu - Algorithm collaborator.

Ning Hao - Algorithm collaborator & co-author on original paper from which chapter 5 is derived.

Stephanie Galaitsis - Future wife and expert & ruthless proof reader.

Contents

List of Figures

In loving memory of my father Richard Ely

Chapter 1

Introduction

All natural systems, no matter how complex, can be characterized using a basic set of laws governed by physics. The rich and complex wave-field of a concert hall arises from initial and boundary conditions and the wave equation. In this way the evolution of physical systems can be compressed to initial conditions and their corresponding physical laws. Our own intuition about these physical systems allows us to make highly accurate estimates of these partially observed systems on a daily basis. A right-fielder can predict and then catch a pop-fly ball with his brief observation of a batter's strike and innate knowledge of projectile motion and wind effects. In this case our prior knowledge about the physics of systems allows us to reconstruct and track the signal in real time from partial or incomplete measurements.

In this thesis we explore how signals can be significantly compressed according their physical model and how this compressibility or sparsity can be used to greatly improve the reconstruction of various geophysical inverse problems. My work extends the notion of compressed sensing to a more general theory of complexity length description: if a signal can be described in a compact form, then it can be recovered from a limited set of measurements that are proportional to the length of its description. Furthermore, we explore a similar extension to signal separation: if two signals are incoherent with each other in two different compression schemes, then they can be separated.

Although it may be easy to explain what we expect a system to look like given its physics, it is often difficult to express this idea in a concrete mathematical form. The challenge is then how form the physical prior of sparsity or low-dimensionality into one that is not only mathematical but that can also be relaxed into a convex function that can be efficiently solved.

In this thesis I will demonstrate how the physics of several systems result in three different forms of sparsity: sparsity in a basis, matrix low-rank, and tensor low-rank. Each of these forms of sparsity can be relaxed into convex signal norms and through iterative minimization techniques and the machinery of convex optimization can be used to denoise, unmix, and reconstruct signals from a limited set of measurements.

1.1 Organization of Thesis

The thesis is organized into six chapters starting with the introduction. The second chapter explains the mathematics and algorithms used to solve the various inverse problems. It outlines the different types of sparsities exploited throughout the paper and how these forms of sparsity can be relaxed into convex norms. In addition, the chapter outlines how each of these norms can be solved efficiently through the use of proximal (prox) or shrinkage functions. I then construct a host of algorithms based on prox functions and examine when each of the algorithms are most applicable.

The 3rd, 4th, & 5th chapters describe the applications of sparsity penalized algorithms to three geophysical domains: Hydraulic Fracture Monitoring (HFM), hyper-spectral imaging, and reflection seismology. Each of these three chapters illustrates how the physics of given system generates different forms of sparsity and how it can be practically exploited for geophysical inverse problems. These chapters are each based on papers submitted to conferences across several disciplines and as a result the notion across the three chapters may not be consistent and should be considered as separable entities.

Chapter 2

Algorithms

This chapter presents an overview of the algorithms used throughout the thesis and their mathematical background. I first present the inverse problems explored throughout this thesis and how they can be solved through several optimization problems involving the minimization of a sparse inducing norm. I then demonstrate how these non-convex sparse norms, like the ℓ_0 norm, can be relaxed to convex functions and describe several modular algorithms that heavily rely on shrinkage operations to solve them. Finally, I present an overview of the shrinkage operators and their rapid closed formed solutions.

2.1 Notation

In this chapter we will use a capital letter in non-bold, i.e X, text to denote an ambiguous object which could be a scalar, vector, matrix or tensor. A scalar is represented as a non-bold lower case later, i.e. x. A vector is denoted as a bold lower case letter, i.e. $\mathbf{x}$. A matrix is given as a bold upper case letter, i.e. $\mathbf{X}$, and a tensor is represented in capital script, i.e. $\mathscr{X}$. A summary of notion is given in Table 2.1.

	characteristics	example
object	uppercase	X
scalar	lowercase	x
vector	lowercase, bold	$\mathbf{x}$
matrix	uppercase, bold	$\mathbf{X}$
tensor	uppercase, script	$\mathscr{X}$

TABLE 2.1: Summary of notation

Forward Model	Optimization
Reconstruction $B = AX$	$\min_{X} \mathbf{F}(X)$ subject to $\|AX - B\|_F = 0$
De-noising $B = AX + N$	$\min_{X} \|AX - B\|_2^2 + \lambda \mathbf{F}(X)$
Separation $B = X + Y$	$\min_{X,Y} \mathbf{F}(X) + \lambda \mathbf{G}(Y)$ subject to $X + Y = B$

TABLE 2.2: The forward models and their corresponding optimization scheme for inversion.

2.2   Structure of Inverse Problems

In all of the examples throughout this paper, each of the three problems (reconstruction, denoising, and separation) can be expressed through the three different minimization operations given later in this section. Each of these minimization operations can be solved through a host of iterative algorithms, including both 1st and 2nd order techniques explored in Section 2.4. Table 2.2 summarizes the forward models for each of the three problems and the corresponding optimization problem used for inversion. The rest of this section describes the three forward problems and their inversion in detail.

Reconstruction without noise:

In this problem we observe an object X through an under-determined and likely ill-conditioned linear observation operation A resulting in the measured data B. A visual representation of the noiseless problem is shown in Figure 2.1 for reconstruction from a limited set of radon projections.

$$B = AX \tag{2.1}$$

The task is then to reconstruct or invert for X based on the partial measurement B. In the noiseless case, this is typically achieved through the pseudo-inverse, $A^{\dagger}$, which minimizes the ℓ_2 or Frobenius norm of X, $\|.\|_2$ and $\|.\|_F$, while satisfying the observation

$$\min_X ||X||_2 \qquad\qquad \text{subject to} \quad ||AX - B||_2 = 0$$
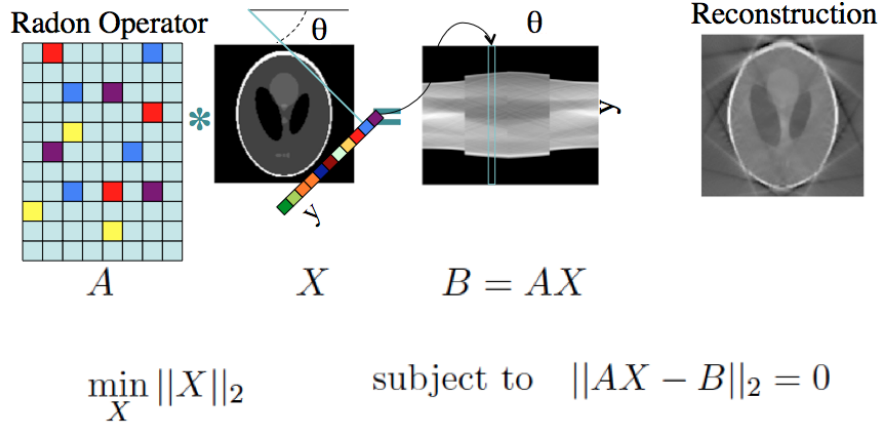
FIGURE 2.1: This figure shows the system setup for the noiseless reconstruction problem. The reconstruction shown in the right of the figure is achieved by minimizing the ℓ_2 norm of the reconstruction.

criteria of Equation 2.1.

$$\min_X ||X||_2 \qquad\qquad \text{subject to} \quad ||AX - B||_F = 0 \qquad (2.2)$$

An estimate of X, $\hat{X}$, is given by applying the pseudo-inverse to the observation.

$$\hat{X} = A^\dagger B \qquad\qquad A^\dagger = (A^T A)^{-1} A^T \qquad (2.3)$$

Because the pseduo inverse solves Equation 2.2, it effectively imposes a minimum energy prior on the estimate of X. This prior may be inaccurate and instead we will want to impose a prior that fully exploits the known structure of X. For example, if we expect X to be sparse we should penalize the ℓ_1 norm of X and solve Equation 2.4.

$$\min_X ||X||_1 \qquad\qquad \text{subject to} \quad ||AX - B||_F = 0 \qquad (2.4)$$

In the problems presented throughout this paper, we seek to solve the generalized version of Equations 2.2 & 2.4 given by Equation 2.5 where $\mathbf{F}(X)$ is a convex function on X that encourages sparsity or low-dimensionality.

$$\min_X \mathbf{F}(X) \qquad\qquad \text{subject to} \quad ||AX - B||_F = 0 \qquad (2.5)$$

Reconstruction with noise:

In this case we slightly alter the previous problem and now obtain an observation in the presence of Gaussian noise N as shown in Figure 2.2.
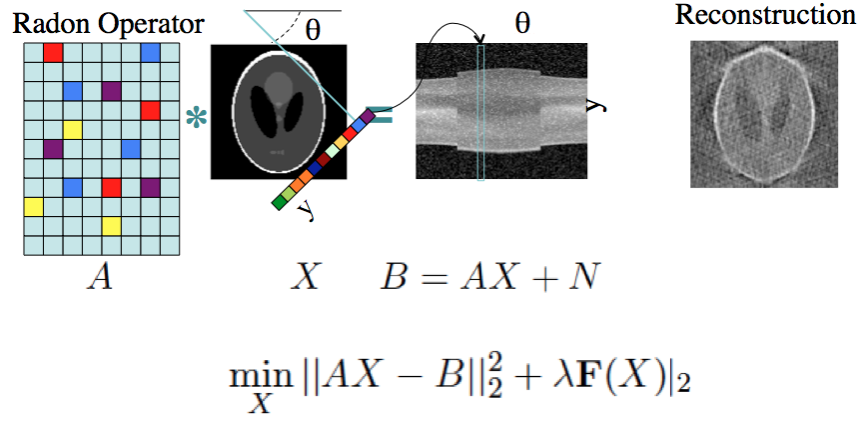
FIGURE 2.2: This figure shows the system setup for the noisy reconstruction problem. The reconstruction shown in the right of the figure is achieved by minimizing the ℓ_2 norm of the reconstruction.

$$B = AX + N \tag{2.6}$$

Given that the pseduo-inverse solves the noiseless case and the condition $AX = B$ no longer holds, application of $A^\dagger$ to B will give inaccurate results. Furthermore, if A is ill-conditioned the pseduo-inverse will magnify the noise and give a poor estimate of X [3]. Therefore, we seek to solve a relaxed version of Equation 2.2 that allows for the presence of noise. In order to do so we relax the problem to its unconstrained Lagrangian form and introduce a penalization constant λ which controls the relative importance of minimizing the ℓ_2 norm of X versus satisfying the observation criteria.

$$\min_X ||AX - B||_F^2 + \lambda||(X)||_2 \tag{2.7}$$

For high levels of noise λ should be set to a large number, allowing for large degree of mismatch between the observation criteria. In the case when the noise is very small, λ should be set to a very small value to put more weight on the observation criteria, resulting in a nearly identical solution to Equation 2.5. Equation 2.7 can then be solved using an altered version of the pseduo inverse, $A^{\dagger*}$ known as the Tikihonov regularized solution [4]. The regularized solution is achieved by adding a weighted identity matrix, $\lambda\mathbf{I}$, to the inverse used in Equation 2.3.

$$\hat{X} = A^{\dagger*}B \qquad\qquad A^{\dagger*} = (A^TA + \lambda\mathbf{I})^{-1}A^T \tag{2.8}$$

Like the constrained problem, throughout this paper we will wish to impose a different prior on X and solve the more general version of the unconstrained problem for a given

$$X + Y = B$$

Objects   Measurement

X   Y   B

$$\min_{X,Y} \mathbf{F}(X) + \lambda \mathbf{G}(Y) \quad \text{subject to} \quad X + Y = B$$
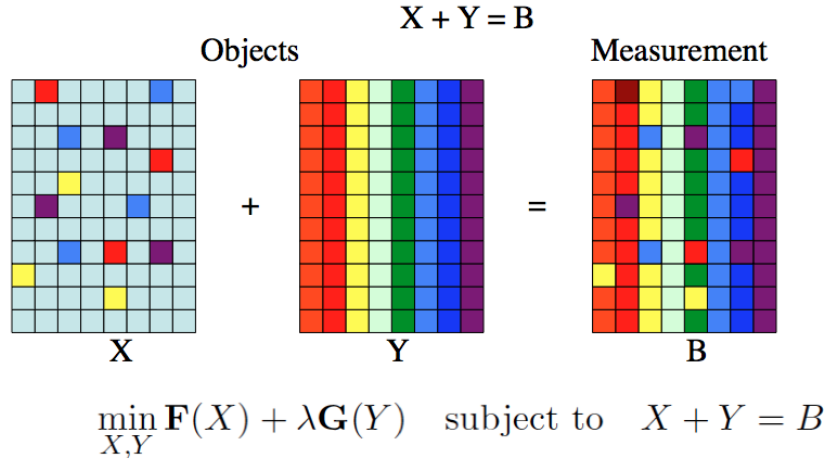
FIGURE 2.3: This figure shows the system setup for the separation problem in which a low-rank and sparse matrix are observed in a combined state.

convex penalty $\mathbf{F}(X)$.

$$\min_{X} ||AX - B||_F^2 + \lambda \mathbf{F}(X) \tag{2.9}$$

Separation:

In this problem we observe two signal X and Y added together to form an observation B and attempt to separate them from each other through a convex minimization operation. Figure 2.3 shows the separation problem for a low-rank X and sparse Y.

$$B = X + Y \tag{2.10}$$

The most well known form of this problem is known as robust Principle Component Analysis (PCA) in which a low-rank matrix $\mathbf{L}$ is combined with a sparse matrix $\mathbf{S}$ and we observe the $\mathbf{B}$ matrix [5]. For sufficiently low-rank $\mathbf{L}$ and sparse $\mathbf{S}$ of size $m \times n$, robust PCA can be provable solved through the convex optimization routine,

$$\min_{X,Y} ||\mathbf{L}||_* + \lambda ||\mathbf{S}||_1 \qquad \text{subject to} \quad \mathbf{L} + \mathbf{S} = \mathbf{B} \tag{2.11}$$

where λ is given by $\frac{1}{\sqrt{min(m,n)}}$ [6]. Several extensions of Robust PCA have been explored in the literature such as replacing the ℓ_1 penalty with a group sparse penalty [7], removal of Gaussian noise [8], and applications to high order tensors [9]. In this thesis we consider a more general form of Equation 2.11,

$$\min_{X,Y} \mathbf{F}(X) + \lambda \mathbf{G}(Y) \qquad\qquad \text{subject to} \quad X + Y = B \qquad (2.12)$$

Where $\mathbf{F}(X)$ & $\mathbf{G}(Y)$ are convex norms of X & Y that encourage sparsity or low-rank in some form.

2.3 Sparsity and Convex Relaxations
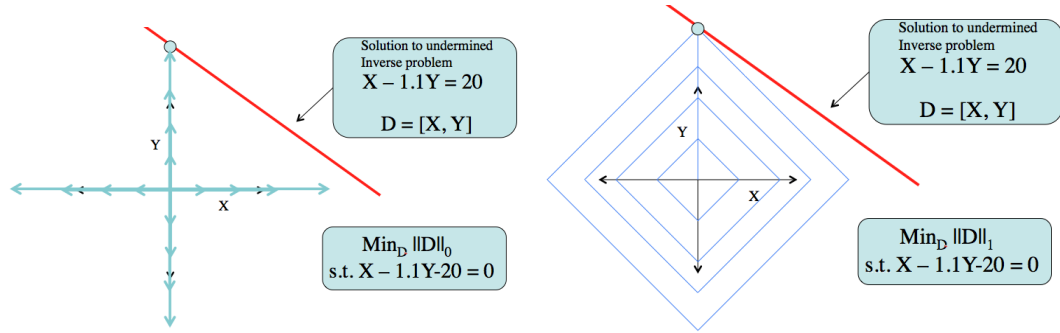
FIGURE 2.4: This figure illustrates how minimizing the support of a vector can be relaxed to a convex optimization problem that results in the same solution as the non-convex optimization problem.

All of the algorithms presented in this thesis rely on the measure of sparsity or low-dimensionality and its corresponding convex relaxation. Although the support of a vector, the rank of a matrix and tensor rank of high order tensor are all norms, they are non-convex and result in inherently combinatoric optimization problems. Figure 2.4 illustrates this issue for solving a simple ill-posed inverse problem in $\mathbb{R}^2$ where we wish to find a solution from all possible solutions, denoted by the red line, which is sparse. This problem can be solved in a combinatoric fashion by trying all of the possible sparse solutions from lowest to highest ℓ_0 norm, i.e. [x,0], [0,y], and then [x,y]. In this way the solution can be achieved by essentially walking the axes as shown in Figure 2.4 left. Although this process is cheap for two dimensional space, when the number of variables increases to several thousands or millions it becomes computational infeasible and the problem must be relaxed to a convex problem. For example the ℓ_0 norm, the number of non-zero entries in a vector or matrix X can be relaxed to the ℓ_1 norm, the sum of the absolute values of X, and results in provable equivalent solutions for minimization operations [1]. Minimizing the ℓ_1 norm of an object can be thought of as growing a diamond like hull with each vertex aligned with an axis, Figure 2.4 right. From the figure we can see that the hull will result in the same solution as the combinatoric problem unless one of the edges is perfectly aligned with the solution space.

Like the ℓ_0 and ℓ_1 relaxation, the rank of a matrix can also be approximated by a convex norm. A matrix can also be low-rank or sparse in the number of non-zero singular values. If $\mathbf{X}$ is a matrix, it can be decomposed into its singular value decomposition.

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{2.13}$$

Where $\mathbf{S}$ is a diagonal real matrix with the number of non-zero entries equal to the rank of the matrix $\mathbf{X}$. Like the relaxation of the ℓ_0 norm, we can define a relaxed convex norm on the matrix $\mathbf{X}$ as the ℓ_1 norm of diagonal $\mathbf{S}$ matrix. This norm, the nuclear norm denoted by $|.|_*$, in minimization operations, results in equivalent solutions to the minimization of the non-convex matrix rank.

Furthermore, in the case of tensor-rank, we adopt the standards of the tSVD to extend the notion of low-rank to higher dimensional data [2]. In the tSVD standard for an N dimensional object, the singular values take a form of $N-1$ dimensional object of positive scalars (see Chapter 5 for more details on the tSVD). Similar to the nuclear norm, we can apply a minimization operation to the sum of the singular values to recover low-rank tensors.
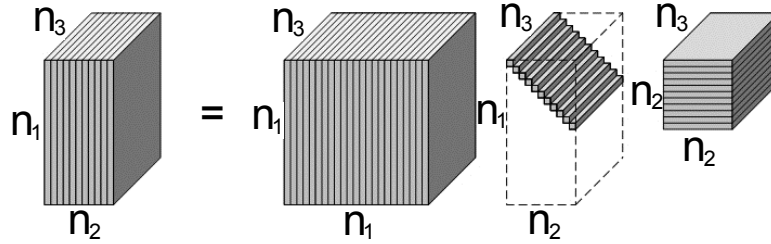
FIGURE 2.5: The t-SVD of an $n_1 \times n_2 \times n_3$ tensor. A tensor can be regarded as a matrix of fibers or tubes along the third dimension of a tensor $\mathscr{M}$. Then tSVD is analogous to a matrix SVD if we regard the diagonal tensor $\mathscr{S}$ as consisting of singular "tubes" or "vectors" on the diagonal analogous to singular values on the diagonal in the traditional SVD. For tensors of order p tSVD extends the notion of singular value to higher dimensions, in which each *tube* can be represented as $p-1$ dimensional tensor. For example, a 4D tensor of size $n_1 \times n_2 \times n_3 \times n_4$ has a tSVD decomposition in which each *tubal* singular value is a 3D tensor of size $n_1 \times 1 \times n_3 \times n_4$.

2.4 Iterative techniques for solving the optimization problems

All of the algorithms presented in the following section can be thought of consisting of two types of modular components: an operator driven step and a shrinkage operation. All of the algorithms presented consist of an iterative process in which an operator step or shrinkage step are applied until convergence as shown in Figure 2.6. In the case of
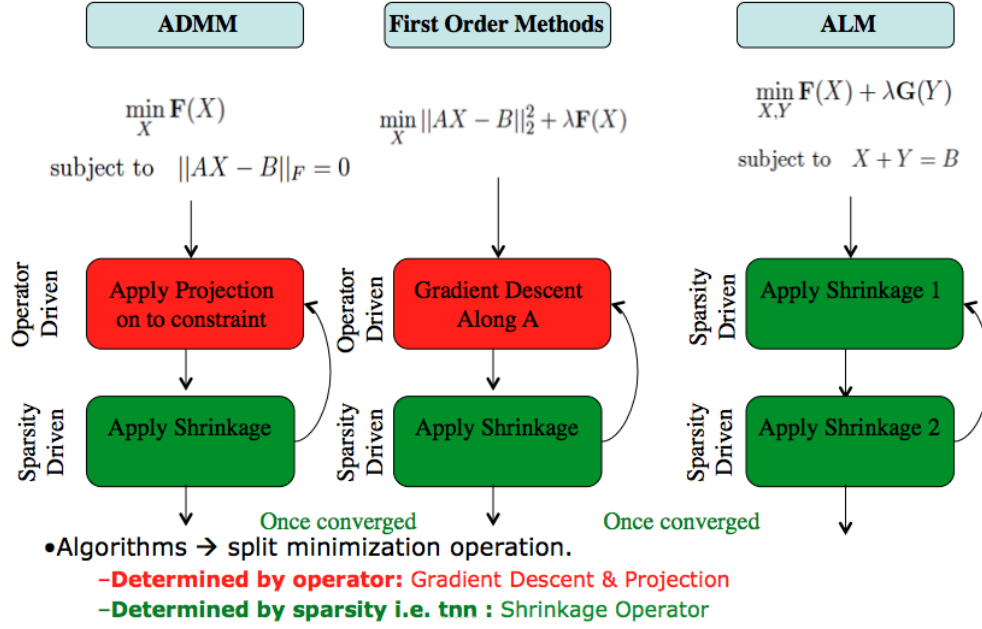
FIGURE 2.6: This figure summarizes the three types of algorithms used in thesis and their modular components.

reconstruction problems the measurement operator, $\mathbf{A}$, the operator driven step consists of either a projection onto the null-space of $\mathbf{A}$ for ADMM or a gradient descent for first order methods. The shrinkage operator is entirely determined by the type of convex norm being minimized and is independent of the given algorithm. These shrinkage operators or proximal functions, all have a closed form and typical fast solutions as given in Section 2.5. Because of significant use of the prox functions, it is very easy to apply a particular algorithm to any norm with little modification to the implemented code. For example, the algorithm to implement Iterative Shrinkage for ℓ_1 minimization differs only from the *nuclear* norm by the proximal function. To exploit this I collaborated with Shuchin Aeron and Zemin Zhang to implemented these algorithms in a highly modular and extensible set of codes stored on gitHub. Although the repository is currently closed, it can easily be made available to others interested in accessing the codes or contributing to the repository. It is my hope that this repository will eventually serve as a more extensible and open version of the currently available convex solvers, TFOCS and CVX. TFOCS was used significantly to implement the altered forms of the minimization operations described in Chapter 4.

2.4.1 Inversion & Reconstruction

For the task of inversion or reconstruction (solving Equations 2.5 & 2.9), I explored the application of three different types of solvers: Alternating Direction Multiplier Method

(ADMM), Fast Iterative Shrinkage (FISTA), and stochastic first order methods. These three methods involve two basic steps: a projection or gradient step and a shrinkage operation. A brief comparison of three algorithms is given in Table 2.3.

TABLE 2.3: Comparison of methods

	ADMM	**FISTA**	**Stochastic**
Pros:	fast convergence for all step-sizes	cheap cost per iteration (forward & back projection), good for large scale problems	very cheap cost per iteration for very large scale problems
Cons:	involves calculation of pseudo-inverse, infeasible for large problems	cannot solve the constrained problem, tuning of step-size required for convergence	convergence not guaranteed for constant step-size, difficult to determine rate of step-size decrease

2.4.1.1   ADMM

ADMM methods converge quickly in $O(\frac{1}{k^2})$ iterations and will converge for all step-sizes [10]. However, each iteration involves a projection onto the null space of A resulting in either high-computational cost per iteration or calculation of the pseudo-inverse. For large scale problems ADMM methods are computationally infeasible unless the measurement operator A is structured to allow fast projection on to the null space (see Chapter 5 for application of this method with a structured operator). In addition, ADMM methods offer a clear method of solving the constrained problem (equation 2.5) whereas first order methods require sub-gradient techniques or additional Lagrange multipliers. Algorithm 1 show the pseudo code for ADMM solving equations 2.5 and 2.9. The two optimization problems are solved by changing the choice of ε, for the constrained case $\varepsilon = \frac{1}{\rho}$ and for the unconstrained case $\varepsilon = \frac{\lambda}{\rho}$ where ρ is the step-size. The shrinkage operator $Sh_\varepsilon^{\mathbf{F}}[X]$ is one of the shrinkage operators described in Section 2.5 corresponding to the minimized norm, $\mathbf{F}(X)$.

Algorithm 1 ADMM: solves 2.5 & 2.9:
$\min_X \mathbf{F}(X)$ subject to $||AX - B||_2 = 0$ (constrained)
$\min_X ||AX - B||_2^2 + \lambda \mathbf{F}(X)$ (unconstrained)

$P = \mathbf{I} - A(A^T A)^{-1} A^T$ //Projects onto the null-space of the measurement tensor.
$Z = U = \mathbf{0}$ // Initialize internal variables.
while Not Converged **do**
　　$X = P(Z - U) + B$ //Apply Projection
　　$\hat{X} = X + Z$
　　// Apply Shrinkage operator.
　　// If constrained $\varepsilon = \frac{1}{\rho}$, If unconstrained $\varepsilon = \frac{\lambda}{\rho}$
　　$Z = Sh_\varepsilon^{\mathbf{F}}[(X + U)]$
　　$U = U + X - Z$
end while

2.4.1.2 First Order Methods: FISTA

In large scale problems where projection onto the null space is too costly, first order methods can be solved efficiently. In thesis we will apply two algorithms: Iterative Shrinkage (ISTA) Algorithm 2 and Fast Iterative Shrinkage (FISTA) Algorithm 3. ISTA converges in $O(\frac{1}{k})$ iterations and FISTA uses an interpolation procedure to reach convergence in $O(\frac{1}{k^2})$ [11]. Instead of calculating a costly pseudo inverse only the forward projection, A, and back-projection, A^T, needs to be calculated. If A is a sparse matrix, then this computation is extremely quick. However, because there is no projection onto the null space, it is difficult to solve the constrained problem (Equation 2.5) and these algorithms were used to solve only the unconstrained problem (Equation 2.9). Unlike ADMM methods, convergence of these methods are not guaranteed for all step sizes and in order to converge the inverse step size ρ must be larger than the Lipschitz constant, the largest eigenvalue of $A^T A$. If A is very large it can be infeasible to calculate the Lipschitz constant and implementing adaptable step sizes through the use of line search becomes necessary [12].

Algorithm 2 ISTA: solves $\min_X ||AX - B||_2^2 + \lambda \mathbf{F}(X)$ (Eq. 2.9)

  $X = \mathbf{0}$ //Initialize variables.
  while Not Converged **do**
    $Z = X - \frac{1}{\rho}A^T(AX - B)$ //Gradient calculation
    $X = Sh_{\frac{\lambda}{\rho}}^{\mathbf{F}}[Z]$ //Apply shrinkage operator
  end while

Algorithm 3 FISTA: solves $\min_X ||AX - B||_2^2 + \lambda \mathbf{F}(X)$ (Eq. 2.9)

  $X = \mathbf{0}$ //Initialize variables.
  $k = 1$
  while Not Converged **do**
    $X_{old} = X;\ k = k + 1;$
    $U = X + \frac{k-1}{k+2}(X - X_{old})$ //Interpolation
    $Z = U - \frac{1}{\rho}A^T(AU - B)$ //Gradient calculation
    $X = Sh_{\frac{\lambda}{\rho}}^{\mathbf{F}}[Z]$ //Apply shrinkage operator
  end while

2.4.1.3 Stochastic & Incremental Methods

Stochastic and Incremental methods are best used when the problem size is extremely large and the cost function can be expressed as separable summation operation. The algorithm presented in this section is essentially ISTA except that the gradient is only applied to a subset of the observed measurements and the proximal operator is only applied to several of the group indices. For a system of n measurements and m groups

at each iteration we choose a random subset of k measurements indices **i** and l group indices **j**. The gradient is then calculated only using the **i** indices and thus the gradient calculation only needs to access k rows of the A matrix, reducing the computational burden. Furthermore, in the case when the shrinkage function is highly separable, i.e the TNN shrinkage operator (Algorithm 4), the shrinkage is only applied to the l number of groups. This scheme is especially useful in the case when the shrinkage operator is very expensive to calculate such as the case of the TNN operator that requires calculation of numerous SVDs. By combining these two techniques the cost per iteration can be significantly faster than ISTA but result in a comparable number of iterations to reach convergence [13]. However, convergence to the minimum is not guaranteed for stochastic techniques and a decreasing step size is required to reach the true minimum.

Algorithm 4 Stochastic & Incremental Proximal: solves (Eq. 2.9)
$\min_X ||AX - B||_2^2 + \lambda \mathbf{F}(X)$

$X = \mathbf{0}$ //Initialize variables.
while Not Converged **do**
   $\mathbf{i} = randperm(k,n)$//generate measurement index
   $\mathbf{j} = randperm(l,m)$//generate group index
   $Z = X - \frac{1}{\rho}A(:,\mathbf{i})^T(A(:,\mathbf{i})X - B(:,\mathbf{i}))$ //Gradient calculation
   $X = Z$
   $X(\mathbf{j}) = Sh_{\frac{\lambda}{\rho}}[Z(\mathbf{j})]$ //Apply shrinkage operator only to **j** indices
end while

2.4.1.4   ALM: Separation

To solve the separation problem, I implemented a generalized version of the Augmented Lagrange Multiplier (ALM) method that can be applied to a set of arbitrary convex functions and solves the minimization problem given by Equation 2.12. This algorithm is the same as the inexact separation algorithm presented in [7] generalized to two convex functions rather than just the ℓ_1 and nuclear norm. The algorithm constructs a Lagrangian and applies two different shrinkage operator, $Sh_{\frac{\lambda}{\mu}}^{\mathbf{F}}[L]$ & $Sh_{\frac{\lambda}{\mu}}^{\mathbf{G}}[L]$ which are the proximal functions described in Section 2.5 of the corresponding convex functions of $\mathbf{G}()$ & $\mathbf{F}()$. For example, in the case where we wish to separate low-rank from sparse, the two prox operators would be given by Equations 2.17 and 2.15.

Algorithm 5 Augmented Lagrange Multiplier (ALM) for separation : solves Equation 2.12

$\min_{X,Y} \mathbf{F}(X) + \lambda \mathbf{G}(Y)$ subject to $X + Y = B$

$\hat{X} = \hat{Y} = \mathbf{0}$ //Initialize variables.
$Q = b; \mu = 1; \rho > 1$
while Not Converged **do**
    //Calculate Lagrangian and shrink to obtain X.
    $L = b - Y + \frac{1}{\mu}Q$
    $X = Sh^{\mathbf{F}}_{\frac{\lambda}{\mu}}[L]$
    //Calculate Lagrangian and shrink to obtain Y.
    $L = b - X + \frac{1}{\mu}Q$
    $Y = Sh^{\mathbf{G}}_{\frac{1}{\mu}}[L]$
    //update Lagrange multiplier.
    $Z = D - (X + Y)$
    $Q = Q + \mu Q$
    $\mu = \mu \rho$
end while

2.5   The Prox Operator

Beyond the relaxation of ℓ_0 norm or rank of a matrix, the algorithms presented in section 2.4 rely on a rapid closed formed solution to a sub-problem of the form,

$$\min_X ||X - Z||_2^2 + \varepsilon \mathbf{F}(X) \tag{2.14}$$

known as the Proximal (prox) function, where Z is a known object (vector, matrix or tensor), ε is the shrinkage factor and $F(X)$ is the convex function being minimized. The proximal function aids in optimization process by relaxing a non-smooth function $F(X)$ through the addition of a smoothing term $||X - Z||_2^2$. These types of functions arise in numerous types of minimization operations such as interior point methods [14], ADMM techniques [10], and first order methods such as iterative shrinkage [11]. In all of these methods the more complex optimization problem (Equations 2.5, 2.9, & 2.12) can be split into two or more simpler and easier to solve sub-problems. These class of algorithms originated from general methods of forward-backwards splitting and Breggman splitting [11]. Because these class of Proximal functions problems can be solved exactly and quickly, they frequently arise as sub-problems in optimization. For example, the optimization problems described in section 2.2 cannot be directly solved. Instead, each iteration of the minimization problem described in section 2.2 can be reduced to solving the above minimization operation. For the sake of clarity, the proximal functions are given for all of the norms used throughout this paper. For a vector $\mathbf{x} = [x_1, x_2, ...x_n]^T$ the shrinkage operator is given by elementwise operation.

proximal function: ℓ_1

$$Sh^1_\varepsilon[\mathbf{z}] = \min_{\mathbf{x}} ||\mathbf{x} - \mathbf{z}||^2_2 + \varepsilon||\mathbf{x}||_1 \qquad Sh^1_\varepsilon[\mathbf{z}] = \begin{cases} & x_i - \varepsilon, \text{ if } x_i > \varepsilon, \\ z_i = & x_i + \varepsilon, \text{ if } x_i < \varepsilon, \\ & 0, \text{otherwise}, \end{cases} \qquad (2.15)$$

The ℓ_{12} shrinkage operator can be conceptualized as applying a shrinkage operator to each of the n columns of $\mathbf{X}$ separately, for $[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3...\mathbf{x}_n] = \mathbf{X}$.

proximal function: ℓ_{12}

$$Sh^{12}_\varepsilon[\mathbf{Z}] = \min_{\mathbf{X}} ||\mathbf{X} - \mathbf{Z}||^2_2 + \varepsilon \sum_{i=1}^{n} ||\mathbf{x}_i||_2 \quad Sh^{12}_\varepsilon[\mathbf{Z}] = \begin{cases} \mathbf{z}_i = & \mathbf{x}_i(1 - \frac{\varepsilon}{||\mathbf{x}_i||_2}), \text{ if } ||\mathbf{x}_i||_2 > \varepsilon, \\ & \mathbf{x}_i = \mathbf{0}, \quad \text{otherwise}, \end{cases}$$
$$(2.16)$$

In the case of nuclear norm and Tensor Nuclear Norm (TNN) the operations involve the calculation of one or several SVDs.

proximal function: *nuclear*

$$Sh^*_\varepsilon[\mathbf{Z}] = \min_{\mathbf{X}} ||\mathbf{X} - \mathbf{Z}||^2_2 + \varepsilon||\mathbf{X}||_* \qquad \begin{matrix} Sh^*_\varepsilon[\mathbf{Z}] = \mathbf{U}Sh^1_\varepsilon[diag(diag(\mathbf{S}))]\mathbf{V}^T \\ \mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \end{matrix} \qquad (2.17)$$

For tensors (3 or more dimensions) we extend the notion of matrix rank to higher order spaces by adopting the t-SVD framework [2, 15, 16]. In the framework we view 3D tensors as a matrix of tubes (in the third dimension) and define a commutative operation (convolution) between the tubes. This commutative structure leads to viewing the tensor multiplication as a simple matrix-matrix multiplication where the multiplication operation is defined via the commutative operation. With this construction, one can now introduce the notion of a t-SVD which is similar to the traditional SVD, see Figure 2.5. A tensor $\mathscr{X}$ can be decomposed into three tensors having similar properties of ('orthogonal' & block diagonal see Appendix) the SVD,

$$\mathscr{X} = \mathscr{U} * \mathscr{S} * \mathscr{V}^T \qquad (2.18)$$

where $*$ denotes the tensor multiplication given by algorithm 11 in the appendix and $.^T$ denotes the tensor transpose given by Definition A.0.1 also in the appendix. In this context we identify a relaxed measures of rank, the sum of all singular tubal values, Tensor Nuclear Norm (TNN) [15] [17].

proximal function: TNN

$$Sh_{\varepsilon}^{tnn}[\mathscr{Z}] = \min_{\mathscr{X}} ||\mathscr{X} - \mathscr{Z}|_2^2|| + \varepsilon ||\mathscr{X}||_{tnn} \qquad (2.19)$$

The solution for the proximal function of the TNN can be thought of as a applying the nuclear norm shrinkage to each frontal slice of the tensor and is best understood through Algorithm 6 given for a general tensor of dimension p. Where $\mathscr{X}$ is a p dimensional tensor of size $n_1 \times n_2 \times ... \times n_p$. The shrinkage operation algorithm is nearly identical to the tSVD decomposition given by Algorithm 10 except that each slice decomposition $\mathbf{U}, \mathbf{S}, \mathbf{V}$ are shrunk and recombined in the main SVD loop rather than stored separately. Chapter 5 and [15] give more details on the use of the shrinkage algorithm and its origin.

Algorithm 6 TNN Shrinkage Solution to Equation 2.19

$\rho = n_3 n_4 ... n_p$
for $i = 3$ to p **do**
   $\mathscr{D} \leftarrow \text{fft}(\mathscr{X}, [\,], \text{i});$
end for
for $i = 1$ to ρ **do**
   $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathscr{D}(:,:,i));$
   $\mathbf{S} = Sh_{\varepsilon}[\mathbf{S}];$
   $\mathscr{X}(:,:,i) = \mathbf{U}\mathbf{S}\mathbf{V}^T$
end for
for $i = 3$ to p **do**
   $\mathscr{X} \leftarrow \text{ifft}(\mathscr{X}, [\,], \text{i})$
end for

These closed form solutions, summarized in Table 2.4 to the shrinkage or proximal operator can subsequently be used to solve the more complex minimization operations described in Section 2.2.

Sparse
$$Sh_\varepsilon^1[\mathbf{z}] = \min_{\mathbf{x}} ||\mathbf{x} - \mathbf{z}||_2^2 + \varepsilon||\mathbf{x}||_1 \qquad Sh_\varepsilon^1[\mathbf{z}] = \begin{cases} x_i - \varepsilon, & \text{if } x_i > \varepsilon, \\ z_i = & x_i + \varepsilon, & \text{if } x_i < \varepsilon, \\ 0, \text{otherwise}, \end{cases}$$

Block Sparse
$$Sh_\varepsilon^{12}[\mathbf{Z}] = \min_{\mathbf{X}} ||\mathbf{X} - \mathbf{Z}||_2^2 + \varepsilon \sum_{i=1}^n ||\mathbf{x}_i||_2 \quad Sh_\varepsilon^{12}[\mathbf{Z}] = \begin{cases} \mathbf{z}_i = & \mathbf{x}_i(1 - \frac{\varepsilon}{||\mathbf{x}_i||_2}), \text{ if } ||\mathbf{x}_i||_2 > \varepsilon, \\ & \mathbf{x}_i = \mathbf{0}, \quad \text{otherwise}, \end{cases}$$

Low Rank Matrix
$$Sh_\varepsilon^*[\mathbf{Z}] = \min_{\mathbf{X}} ||\mathbf{X} - \mathbf{Z}||_2^2 + \varepsilon||\mathbf{X}||_* \qquad \begin{aligned} Sh_\varepsilon^*[\mathbf{Z}] &= \mathbf{U}Sh_\varepsilon^1[diag(diag(\mathbf{S}))]\mathbf{V}^T \\ \mathbf{X} &= \mathbf{U}\mathbf{S}\mathbf{V}^T \end{aligned}$$

Low Rank Tensor
$$Sh_\varepsilon^{tnn}[\mathscr{X}] = \min_{\mathscr{X}} ||\mathscr{X} - \mathscr{Z}|_2^2|| + \varepsilon||\mathscr{X}||_{tnn} \rightarrow$$

```
p = n3n4...np
for i = 3 to p do
    D ← fft(X, [], i);
end for
for i = 1 to ρ do
    [U, S, V] = svd(D(:, :, i));
    S = Shε[S];
    X(:, :, i) = USV^T
end for
for i = 3 to p do
    X ← ifft(X, [], i)
end for
```

TABLE 2.4: A summary of the shrinkage operations for the norms used throughout this thesis and their solutions.

Chapter 3

Hydraulic Fracture Monitoring

In this chapter we propose a method for estimating the moment tensor and location of a micro-seismic based a group low-rank penalization. First, we propose a novel joint-complexity measure, namely the sum of nuclear norms which simultaneously imposes sparsity in the location of fractures over a large spatial volume, as well as captures the rank-1 nature of the induced wavefield distribution from a seismic source at the receivers. This feature is captured as the outer-product of the source signature with the amplitude pattern across the receivers, which in turn is a function of the seismic moment tensor and the array geometry, allowing us to drop any other assumption on the source signature. Second, we exploit the recently proposed first-order incremental projection algorithms for a fast and efficient implementation of the resulting optimization problem. We develop a hybrid stochastic & deterministic algorithm that results in significant computational savings and guaranteed convergence.

3.1 Introduction

Seismic hydraulic fracture monitoring (HFM) can both mitigate many of the environmental risks and improve reservoir effectiveness by providing real time estimates of locations and orientations of induced fractures. Determining the location of these microseimsic events remains challenging due to high levels of pumping noise, propagation of seismic waves through highly anisotropic shale, and the layered stratigraphy leading to complex wave propagation, [18–20]. Classical techniques for localization involves denoising of individual traces [21, 22] followed by estimating the arrival time of the events at each individual trace. The angle of arrival of the incident array, or polarization, is achieved via Hodogram analysis [23] or max-likelihood type estimation [24]. Once the angle and time arrival of the events has been estimated, the events are back-propagated
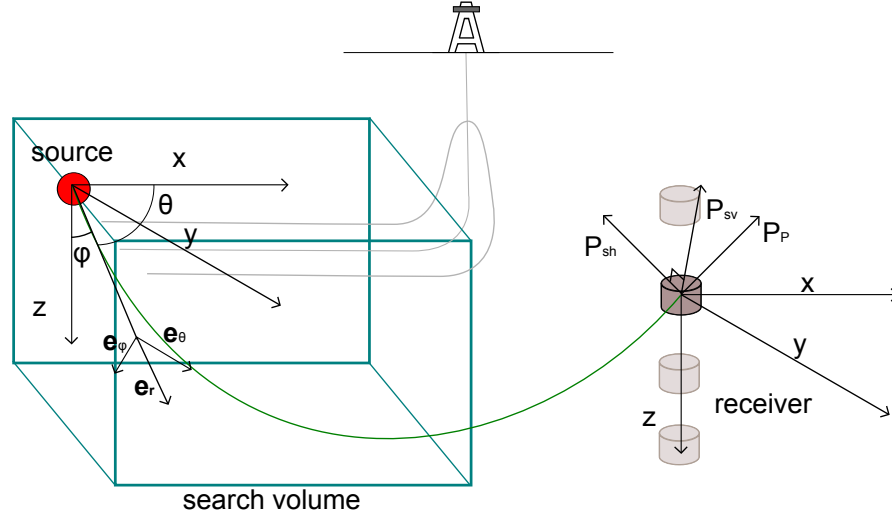
FIGURE 3.1: This figure shows the geometry and coordinate system used throughout this chapter.

using a forward model under known stratigraphy to determine the location, [24, 25]. In contrast to these approaches which tend to separate the de-noising of the signal from the physical model, recently the problem of moment tensor estimation and source localization was considered in [26] for general sources and in [27] & [28] for isotropic sources which exploit sparsity in the number of microseimsic events in the volume to be monitored. This approach is shown to be more robust and can handle processing of multiple time overlapping events.

Our approach, although similar to the technique proposed in [26] differs in that we do not use source waveform information from the Green's function and introduce a group low-rank penalization. Here we don't use the amplitude of the received waveform, but only the fact that the received signal across the seismometers is common across all seismometers with varying delays dictated by a known velocity model of the stratigraphy and the source receiver configuration. Since we are not using any amplitude information, we usually have more error in estimation and require more receivers for localization. Nevertheless, when the computation of Green's function is costly or accurate modeling of the stratigraphy is not available our method can be employed. Furthermore, due to amplitude independent processing our methods can be extended to handle the anisotropic cases using just the travel-time information for inversion, [29].

3.2 Physical model

In this paper we focus on propagation in isotropic media, although our approach can easily be extended to anisotropic and layered media. Figure 3.1 shows the physical setup

in which a seismic event with a symmetric moment tensor $\mathbf{M} \in \mathbb{R}^{3 \times 3}$ is recorded at a set of J tri-axial seismometers indexed as $j = 1, 2, ..., J$ with locations $\mathbf{r}_j$ and I denotes the location of the seismic event l. The seismometer record compressional wave denoted by p, and vertical and horizontal shear waves denoted by sv and sh respectively. Assuming ([19], [Chapter 4]) that the volume changes over time does not change the geometry of the source, Equation (3.1) describes the particle motion vector $\mathbf{u}_c(l, j, t)$ at the three axes of the seismometer j as a function of time t.

$$\mathbf{u}_c(l, j, t) = \frac{R_c(\theta, \phi)}{4 \pi d_{lj} \rho c^3} \mathbf{P}_c^{lj} \psi_c \left(t - \frac{d_{lj}}{v_c} \right) \tag{3.1}$$

where d_{lj} is the radial distance from the source to receiver; $c \in \{p, sh, sv\}$ is the given wave type, and ρ is the density, and R_c is the radiation pattern which is a function of the moment tensor, the take off direction parameters θ_j, ϕ_j with respect to the receiver j. $\mathbf{P}_c^{lj}$ is the unit polarization vector for the wave c at the receiver j. Up to a first order approximation [30] we assume that $\psi_c(t) \approx \psi(t)$ for all the wave types and henceforth will be referred to as the source signal. Note that for non-anisotropic formations the compressional waves $\mathbf{P}_p^{lj}$ aligns with the direction of ray propagation. The polarization vectors for the sh and sv correspond to the other mutually perpendicular directions. The radiation pattern depends on the moment tensor $\mathbf{M}$ and is related to the take off direction at the source with respect to the receiver j defined as the radial unit vector $\mathbf{e}_{r_j}$ relative to the source as determined by (θ_j, ϕ_j), see Figure 3.1. Likewise we denote the unit vectors $\mathbf{e}_{\theta_j}$ and $\mathbf{e}_{\phi_j}$ to be the radial coordinate system orthogonal to radial unit vector. The radiation pattern for a compressional source $R_p(\theta_j, \phi_j)$ is then given by,

$$R_p(\theta_j, \phi_j) = \mathbf{e}_{r_j}^T \mathbf{M} \mathbf{e}_{r_j} \qquad \mathbf{M} = \begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{xy} & M_{yy} & M_{yz} \\ M_{xz} & M_{yz} & M_{zz} \end{bmatrix} \tag{3.2}$$

The radiation energy at a receiver can then be simplified and described as the inner product of the vectorized compressional unit vector product, $\mathbf{e}_{p_j}$, and the vectorized moment tensor $\mathbf{m}$; where $(\cdot)^T$ denotes the transpose operation.

$$R_p(\theta_j, \phi_j) = \mathbf{e}_{p_j}^T \mathbf{m}$$

$$\mathbf{m} = [M_{xx}, M_{xy}, M_{xz}, M_{yy}, M_{yz}, M_{zz}]^T \tag{3.3}$$

$$\mathbf{e}_{p_j}^T = [e_{r_{jx}}^2, 2e_{r_{jx}} e_{r_{jy}}, 2e_{r_{jx}} e_{r_{jz}}, e_{r_{jy}}^2, 2e_{r_{jy}} e_{r_{jz}}, e_{r_{jz}}^2]^T$$

The above expression can then be extended to construct a vector of radiation pattern $\mathbf{a}_p \in \mathbb{R}^J$ across the J receivers, with take off angles of (θ_j and ϕ_j) corresponding to compressional unit vectors $\mathbf{e}_{p_j}$, given by $\mathbf{a}_p = \mathbf{E}_p \mathbf{m}$ where $\mathbf{E}_p = [\mathbf{e}_{p_1}, \mathbf{e}_{p_2}, ..., \mathbf{e}_{p_J}]^T$.

Similarly we have $\mathbf{a}_{sh} = \mathbf{E}_{sh}\mathbf{m}$ and $\mathbf{a}_{sv} = \mathbf{E}_{sv}\mathbf{m}$. Therefore we can write the radiation pattern across J receivers for the three wave types as the product of an augmented matrix with the vectorized moment tensor.

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_p \\ \mathbf{a}_{sh} \\ \mathbf{a}_{sv} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{E_p} \\ \mathbf{E_{sh}} \\ \mathbf{E_{sv}} \end{bmatrix}}_{\mathbf{E}} \mathbf{m} \tag{3.4}$$

Thus the radiation pattern across the receivers $\mathbf{a}$ can then be described as the product of the $\mathbf{E}$ matrix, which depends on the location of the event and the configuration of the array, and the vectorized moment tensor, derives solely from the geometry of the fault. Under the above model for seismic source and wave propagation, given the noisy data at the tri-axial seismometers, the problem is to estimate the event location and the associated moment tensor. This separability will be exploited in our dictionary construction to better recover the location and characteristics of the source.
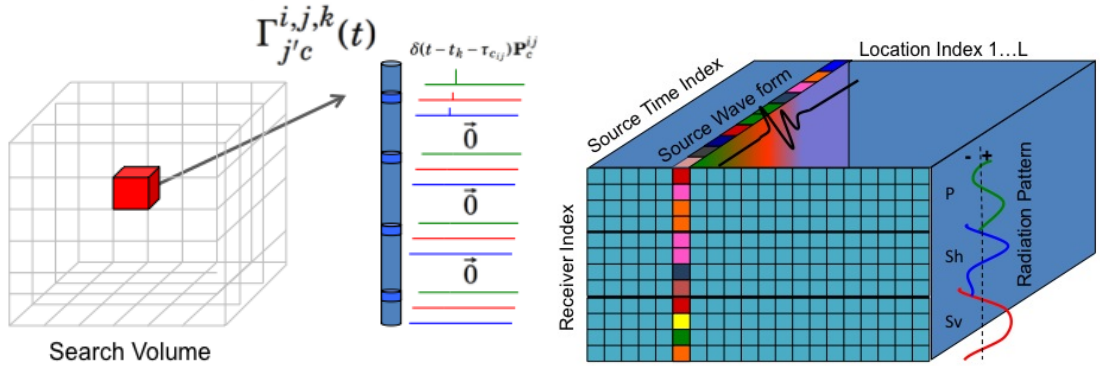
FIGURE 3.2: This figure shows an example propagator and the block sparsity we exploit in our dictionary construction. Note that the slice of the dictionary coefficients corresponding to the correct location of the event can be written as the outer product of the source signal and the amplitude pattern

3.3 Dictionary Construction

Our approach relies on the construction of a suitable representation of the data acquired at the receiver array under which seismic events can be *compactly represented*. We then exploit this compactness to robustly estimate the event location & moment tensor.

Under the assumption that the search volume I can be discretized into n_V locations indexed by $l = l_1, l_2, ..., l_i, .., l_{n_V}$ we construct an over complete dictionary of space time propagators $\mathbf{\Gamma}_c^{i,j,k}$. Where $\mathbf{\Gamma}_c^{i,j,k}$ describes the noiseless data at the *single* receiver, j, as excited by an *impulsive* hypothetical seismic event i at location l_i and time t_k of wave type c (p,sh or sv). Figure 3.2 shows a pictorial representation of a single propagator.

$$\mathbf{\Gamma}_{j'c}^{i,j,k}(t) = \begin{cases} \delta(t - t_k - \tau_{c_{ij}}) \, \mathbf{P}_c^{ij} & \text{if } j' = j \\ \vec{0} & \text{if } j' \neq j, \end{cases} \tag{3.5}$$

Note that $\tau_{c_{ij}} = \frac{d_{l_{ij}}}{v_c}$ is the time delay and $\mathbf{\Gamma}_{j'c}^{i,j,k} \in \mathbb{R}^{|\mathbb{T}_r| \times J \times 3}$. We then construct a dictionary Φ of propagators for all locations, source time indices, wave types, and receiver indices, where each column of the dictionary represents a vectorized propagator,

$$\Phi = [\mathbf{\Gamma}_c^{i,1,k}(:), \mathbf{\Gamma}_c^{i,2,k}(:), \ldots, \mathbf{\Gamma}_c^{n_v,J,k}(:)] \tag{3.6}$$

where (:) denotes the MATLAB colon operator which vectorizes the given matrix starting with the first dimension. Because the dictionary covers all possible locations, receiver responses, time support of the signal, and wave types, an observed seismic signal $\mathbf{Y}$ in the presence of Gaussian noise $\mathbf{N}$ can be written as the superposition of numerous propagators,

$$\mathbf{Y} = \Phi \mathbf{X}(:) + \mathbf{N} \tag{3.7}$$

where $\mathbf{X}$ is the coefficient tensor of size $3 \cdot J \times |\mathbb{T}_s| \times n_V$ and each of there dimensions correspond to 1st wave type receiver index, 2nd source time index, and 3rd location index as shown in figure 3.2.

Therefore, a single seismic event having some radiation pattern R and arbitrary source signal will be block sparse along the lateral slice of dictionary corresponding to location L. Furthermore, the observed source signal will be common across all of receiver indices of the dictionary with its amplitude modulated by the radiation pattern. Therefore, the dictionary elements will not only be block sparse, but the active slice can be written as a rank 1 outer-product of the radiation pattern at the source wave signal $\boldsymbol{\psi} \, \mathbf{a}^T$. This notion can be extended to a signal of multiple events where $\mathbf{X}$ will have now have a few non-zero rank 1 slices. This is the key observation which we exploit in this chapter in the algorithm presented below.

3.4 Algorithm for location and moment tensor estimation

Under the above formulation and the assumption that for a given recorded signal only a few seismic events, we exploit the block-sparse/low-rank, structure of $\mathbf{X}$ for a high resolution localization. These priors can be expressed mathematically by regularizing the inversion of $\mathbf{X}$ to encourage simultaneous sparsity. The method corresponds to the following mathematical optimization problem also known as group sparse penalization in the literature [31, 32] and was taken in [28] for HFM.

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\arg \min} \ ||\mathbf{Y}(:) - \mathbf{\Phi}\mathbf{X}(:)||_2^2 + \lambda \sum_{i=1}^{n_V} ||\mathbf{X}(:,:,i)||_2 \qquad (3.8)$$

where $||\mathbf{X}(:,:,i)||_2^2$ denotes the ℓ_2 norm of the i-th slice, λ is a sparse tuning factor that controls the group sparseness of $\mathbf{X}$, i.e. the number of non-zero slices, versus the residual error. The parameter λ is chosen depending on the noise level and the anticipated number of events. The location estimate is then given by selecting the slices with the largest Frobenius norm above some threshold. In order to exploit the block low-rank structure of the dictionary coefficients the inversion can instead penalize the group nuclear norm,

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{argmin} ||\mathbf{Y}(:) - \mathbf{\Phi}\mathbf{X}(:)||_2^2 + \lambda \sum_{i=1}^{n_V} ||\mathbf{X}(:,:,i)||_* \qquad (3.9)$$

where $||\mathbf{X}(:,:,i)||_*$ represents the nuclear norm, i.e. the sum of the singular values of the i-th slice.

3.4.1 Numerical Algorithms

To solve either of the optimization problems given in equations 3.8 & 3.9 we implemented three different forms of first order algorithms, Iterative Shrinkage (ISTA), Fast Iterative Shrinkage (FISTA) [11] and stochastic gradient descent with incremental proximal methods [13]. ISTA being the simplest to implement is given by two operations: a gradient descent step, and a shrinkage operation like so,

$$\mathbf{X}^{k+1} = \mathbf{prox}_{\frac{\lambda}{\alpha}}(\mathbf{X}^{(k)} - \frac{1}{\alpha}\Phi^T(\Phi\mathbf{X}^k - \mathbf{Y})) \qquad (3.10)$$

where α is the step size and $\mathbf{prox}_\tau(z)$ is the shrinkage operator for one of the two norms. For the group sparse minimization the shrinkage operation is given by,

$$\mathbf{prox}_\tau(z) = \min_x \frac{1}{2}||x - z||_2^2 + \tau \sum_{i=1}^{n_V} ||z(:,:,i)||_2 \qquad (3.11)$$

and for the group low-rank the prox-operator is equivalent to a shrinkage on the singular values of each of the lateral slices of $\mathbf{X}$.

$$\mathbf{prox}_\tau(z) = \min_x \frac{1}{2}||x - z||_2^2 + \tau \sum_{i=1}^{n_V} ||z(:,:,i)||_* \qquad (3.12)$$

Iterative shrinkage can be increased in speed with minimal overhead by adding an interpolation term resulting in the FISTA algorithm.

$$\mathbf{Z} = \mathbf{X}^k + \frac{k-1}{k+2}(\mathbf{X}^k - \mathbf{X}^{k-1})$$

$$(3.13)$$

$$\mathbf{X}^{k+1} = \mathbf{prox}_{\frac{\lambda}{\alpha}}(\mathbf{Z} - \frac{1}{\alpha}\Phi^T(\Phi\mathbf{Z} - \mathbf{Y}))$$

The resulting FISTA algorithm achieves convergence in $O(1/k^2)$ iterations vs $O(1/k)$ for ISTA. In the case of the group low-rank penalization the proximal iteration can be to calculate given the large number of SVDs that need to be computed.

3.4.2 Incremental Proximal Method

For large scale problems it becomes computationally infeasible to calculate the full proximal iteration. As the problem scales the gradient also becomes more expensive to calculate at each iteration. Stochastic gradient descent with incremental proximal iterations can alleviate the computation burden by descending along a random subset of the full gradient and only applying the proximal shrinkage to a few random slices at each iteration. Adopting the MATLAB notion for matrices we can write the stochastic iteration along the set of random directions g and random subset of slices s,

$$\mathbf{X}^{k+1} = \mathbf{prox}_{\frac{\lambda}{\alpha}}^s(\mathbf{X}^{(k)} - \frac{1}{\alpha}\Phi(g,:)^T(\Phi(g,:)\mathbf{X}^k - \mathbf{Y}(g))) \qquad (3.14)$$

where $\mathbf{prox}_\tau^s(z)$ is the shrinkage operation given in Equation (3.10), except that it is only applied to subset of slices s instead of all slices. Because the cost of calculating the SVD of each slice is extremely burdensome, the stochastic gradient descent can drastically increase the speed of obtaining an approximate solution. Given that the cost function, Equation 3.9, can be written as the sum of several nuclear norms, the calculation of the shrinkage operation can be significantly reduced by only applying the shrinkage to

a few slices per iteration, greatly reducing the computational cost by several orders of magnitude. In this application of stochastic gradient descent, our forward operator Φ is sparse resulting in negligible difference in computational cost if the full or partial gradient is calculated. Therefore we can apply the full gradient at each iteration and the minimization operation to a subset of l indices $\mathbf{j}$ of m total groups, Algorithm 7.

Algorithm 7 Incremental Proximal: solves (Eq. 3.9)

$X = \mathbf{0}$ //Initialize variables.
while Not Converged **do**
  $\mathbf{j} = randperm(l, m)$//generate group index
  $\mathbf{Z} = \mathbf{X} - \frac{1}{\rho}\Phi^T(\Phi\mathbf{X} - \mathbf{Y})$ //Gradient calculation
  $\mathbf{X} = \mathbf{Z}$
  //Apply shrinkage operator only to $\mathbf{j}$ indices
  for $j \in \mathbf{j}$ **do**
    $\mathbf{X}(j) = \mathbf{prox}_{\frac{\lambda}{\rho}}[\mathbf{Z}(j)]$
  end for
end while

To estimate the moment tensor we use the estimated event location source-receiver array configuration to construct the matrix $\mathbf{E}$. Then using the estimate of the radiation pattern $\hat{\mathbf{a}}$ from the left singular vector of the active slice we construct the inverse problem $\hat{\mathbf{a}} = \mathbf{E}\mathbf{m}$ and apply Tikhonov regularization to mitigate the ill-conditioning of the $\mathbf{E}$ operator. The moment tensor vector $\mathbf{m}$ is estimated via,

$$\hat{\mathbf{m}} = ((\mathbf{E}^T\mathbf{E} + \lambda_m\mathbf{I})^{-1}\mathbf{E}^T)\hat{\mathbf{a}} \tag{3.15}$$

where λ_m is again tuned using some estimates on the uncertainty in estimation of $\mathbf{a}$ and according to the amount of ill-conditioning of $\mathbf{E}$.

3.5 Experiments

In order to test the effectiveness of the proposed algorithm we simulated an array of 10 seismometers equally spaced within a deviated well consisting of a 500 meter vertical and 500 meter horizontal section dipping at 20 degrees relative to horizontal and aligned with the Y axes, as shown in Figure 3.3 left. For the sake of simplicity the earth is considered to be isotropic with compressional velocity of 1500 and shear velocity of 1100 meters per second. A search volume of $500 \times 500 \times 500$ meters was placed perpendicular to well centered at (500, 300, 500) meters with varying resolution depending on the specific experiment conducted.
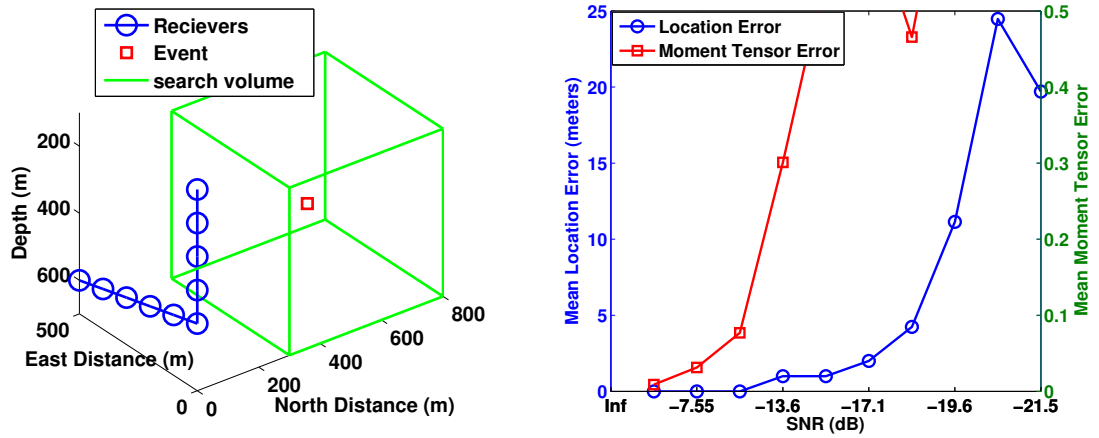
FIGURE 3.3: **Left:** This figure shows the setup for the deviated well and the search volume used in the experiment section. **Right:** This figure show location and moment tensor error as a function of SNR.

3.5.1 Performance in Noise

In order to determine the effectiveness of the algorithm in the presence of noise a single event, the same as the one in the previous section, was generated within the search volume with an increased grid resolution of 5 meters in the presence of various noise levels varying from 0 to -21 dB. The minimization operation given by Equation 3.9 was then applied to resulting simulations with a λ of .9 and the location index with the largest nuclear norm was taken to be the location of the event. Equation 3.15 with a λ of .01 was then used to invert for the moment tensor. This process was repeated 15 times for each noise level and the mean location error and RMS error in the estimate of the moment tensor vector are shown in Figure 3.3 right.

3.5.2 Algorithmic Speed

In order to test the Algorithmic Speed of the three algorithms, the search volume was configured with a coarse spatial resolution of 25 meters and an explosive event with a shear component event was generated in the center of the search volume in the presence of Gaussian noise with a resulting SNR of -18 dB. The three first order algorithms, ISTA, FISTA, and Incremental Proximal, were then applied to the group low-rank minimization problem, Equation (3.9), with a λ of .9 and step size of $.5 * 10^3$. Given that the search volume consisted of 9261 locations each iteration of both FISTA and ISTA would involve the computation of 9261 SVDs of matrices of size N_t x $3N_r$. In the case of incremental proximal method the number of SVD's taken per iteration could be set to 1 to 9261 per iterations. Furthermore, because the forward operator for this problem is sparse and thus fast to compute, the entire full gradient was calculated at each iteration.
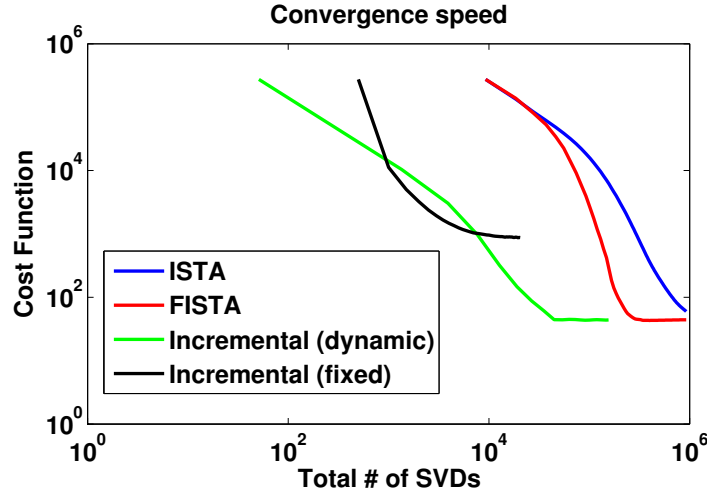
FIGURE 3.4: This figure shows the convergence of the objective function, Equation 3.9, as a function of number of SVDs computed.

Two variations of incremental proximal methods were used: one in which the number of SVD's taken per iteration was set to a constant 100 out of 9261 total, and one where the number of SVD's taken per iteration was increased from 5 at each iteration until reaching the maximum number of SVD and effectively becoming the ISTA. Both the dynamic and fixed version were implemented because only the dynamic version guarantees convergence to the minima for a fixed step size [13]. Figure 3.4 shows the convergence results for the various algorithms showing the cost function, Equation (3.9), as a function of total number of SVDs computed. As expected FISTA outperforms ISTA and the incremental fixed method results in early convergence. The incremental method with an increasing number of SVDs converges to the global minima in drastically fewer SVDs than either FISTA or ISTA.

3.5.3 Multiple Events

In order to test the algorithms ability to distinguish multiple events, three seismic events with varying moment tensors were generated in moderate noise within the search volume with a spatial resolution of 1.25 meters all with the same Y location such that the three events occupied a plane perpendicular to the X and Z axes. Both the group ℓ_2 sparse and group nuclear minimization operations were applied to the simulation with a λ of .9. Figure 3.5 shows resulting nuclear and Frobenius norms along the X-Z plane after the minimization operation have been applied. In the case of the nuclear norm minimization three distinct events are visible falling precisely on the location of true events. However, for the group sparse penalization the location of the two near incident events are impossible to separate and the outlying event's location is imprecise.
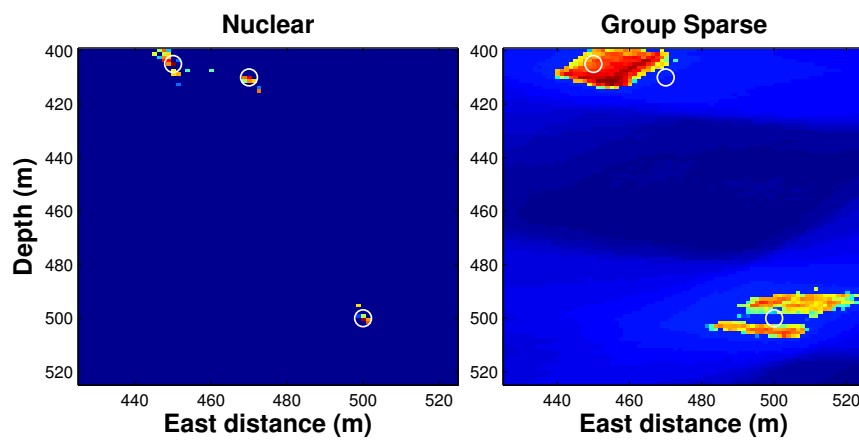
FIGURE 3.5: Performance in source localization for the group ℓ_2 sparse vs group nuclear sparse minimization algorithms. Image intensities are shown on a log scale.

Chapter 4

Hyperspectral Imaging

This chapter presents several strategies for spectral de-noising of hyperspectral images and hypercube reconstruction from a limited number of tomographic measurements. In particular we show that the non-noisy spectral data, when stacked across the spectral dimension, exhibits low-rank. On the other hand, under the same representation, the spectral noise exhibits a banded structure. Motivated by this we show that the de-noised spectral data and the unknown spectral noise and the respective bands can be simultaneously estimated through the use of a low-rank and simultaneous sparse minimization operation without prior knowledge of the noisy bands. This result is novel for for hyperspectral imaging applications. In addition, we show that imaging for the Computed Tomography Imaging Systems (CTIS) can be improved under limited angle tomography by using low-rank penalization. For both of these cases we exploit the recent results in the theory of low-rank matrix completion using nuclear norm minimization.

4.1 Introduction

This chapter addresses two specific image reconstruction challenges encountered in the field of hyperspectral imaging: de-noising in the presence of spectral noise and hypercube reconstruction from a limited set of Radon projections similar to angle limited Computed Tomography Imaging Systems (CTIS).

The first of these two problems is motivated by the desire to remove noise at specific frequency bands from hyperspectral image cubes. This problem frequently arises when using satellites or aircraft to capture hyperspectral images of the earth in which the light reflecting from the surface of the earth must travel through several kilometers of atmosphere to the sensor. The atmosphere even without the presence of clouds has

extremely high absorption bands, particularly at 1400 nm and 1900 nm due to water in the atmosphere [33]. This effect leads to numerous bands being discarded for many data classification and analysis algorithms [34] [35].

In order to mitigate the effects of both this spectral and electronic noise several de-noising techniques such as multidimensional Weiner filtering [36] and methods exploiting the use of high order singular value decompostion [37], curvelets [38], and wavelets [34] [39] have been used to de-noise these effects. However, both the intensity dependence the noise and the concentration across a few spectral bands makes the removal of optical noise challenging [40]. Many of these techniques are based on the premise of noise being AWGNG and performance can be poor [41]. Typically a preprocessing (whitening) step is needed to mitigate the effects of the Poisson noise and improve performance [41]. Recently, efforts to de-noise spectral bands have focussed on the use of sparse or joint penalizations in an appropriate basis such as wavelets [42] and dictionary learning techniques [43].

In this chapter we will explore a novel spectral de-noising technique based on a low-rank and simultaneously sparse matrix decomposition. The low-rank sparse matrix decomposition or Robust Principle Component Analysis (RPCA) has been well studied and theoretical limits well characterized in recent years [5] [44]. Furthermore, RPCA has been successfully employed in image and video processing to separate background from the foreground [45] and remove 'salt and pepper' noise from imagery [5]. However, little research has been done to explore variations of RPCA such as a low-rank group sparse decomposition and its potential applications. In particular, Tang proposed a feasible solution to solving the group RPCA problem through the method of Augmented Lagrange Multipliers [7] and Ji demonstrated the use of group RPCA to de-noise video data [46]. This chapter provides another potential application and extension of RPCA to CTIS systems.

In the second part of this chapter we will focus on the problem of estimation of the hyperspectral data cube from limited number of tomographic projections. Here we show how the use of low-rank regularization can be used to improve an existing class of hyperspectral imagers. These hyperspectral imagers [47–49] sample the hyperspectral image cube by *simultaneously* (i.e. not sequentially) taking a number of Radon type projections of the 3D data cube onto a 2D focal plane array using diffractive optics. Traditionally filtered back-projection methods have been employed to recover the data cube form these tomographic projections. However, these techniques need a large number of projections to ensure accurate results, avoid the so called missing cone problem [50] and often fail in noisy environments. This need for a large number of projections increases the necessary focal plane size beyond what is often feasible. In this context we demonstrate how one

can exploit the low-rank regularization to improve the reconstruction under these classes of simultaneous and *compressive measurements*. Note that although some research has focussed on the use of both sparse and low-rank reconstructions of hyperspectral image cubes, these studies use *practically infeasible* sampling techniques such as randomly sampling a small set of pixels within the image cube [51].
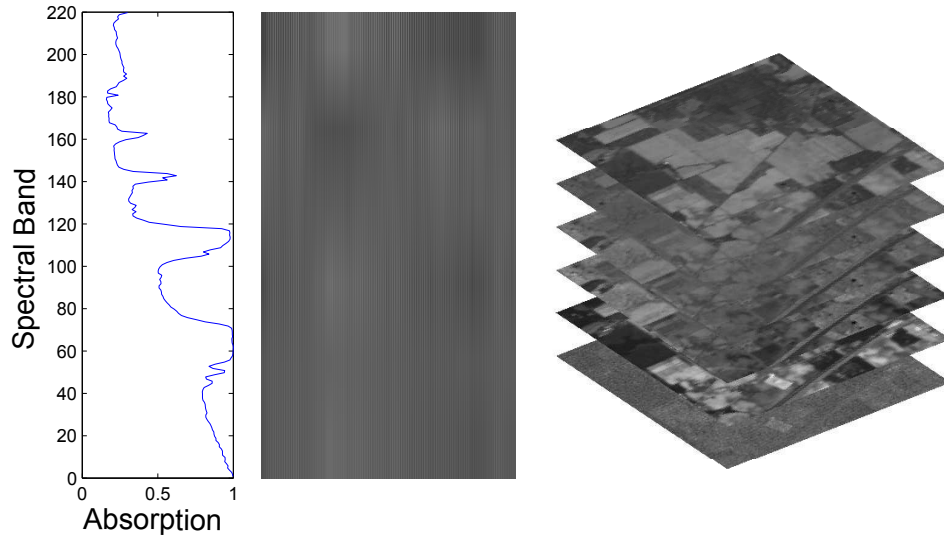
FIGURE 4.1: Left: Normalized total counts in the AVIRIS image as a function of band. We see two pronounced absorption bands. Right & Center: This figure shows a 3D and 2D representation of a hyperspectral image. One notices the horizontal bands of spectral noise in the two dimensional image that align with the absorption bands. Much of the structure in the matrix appears to be vertical but the horizontal bands are spectral noise at absorption bands.

4.2 Structural complexity of hyperspectral images

A hyperspectral image or data cube consists of many images of the same size collected over a number of spectral bands. Mathematically the hyperspectral image can be considered as a three-dimensional matrix $\mathscr{L} \in \mathbb{R}^{m \times n \times l}$ with spatial dimensions of m and n pixels and at l wavelengths. One can reshape the hyperspectral image as a two-dimensional array with a number of columns equal to the number of spectral bands and where each column is the vectored image at the given band, see Figure 4.1. With slight abuse of notation we denote the reshaped image by $\mathbf{L} \in \mathbb{R}^{mn \times l}$. We now present two observations regarding the structural complexity of the image data which will be exploited for recovery and de-noising.

4.2.1 Low-rank structure of the hyperspectral data cube

Although a hyperspectral image/data may have numerous bands, it has been shown that signal subspace is significantly smaller than the number of bands [52] [53]. In particular the eigenvalues of the reshaped hyperspectral cube $\mathbf{L}$ obey a power law decay. This means the vector of eigenvalues has a small weak-ℓ_p norm [54] which implies that image is compressible under the suitable transformation. This intuition can be physically explained by considering the Singular Value Decomposition (SVD) of the (reshaped) hyperspectral matrix $\mathbf{L}$.

$$\mathbf{L} = \mathbf{U\Sigma V}^* \tag{4.1}$$

We can think of the right singular vectors as giving the spectra of the common elements in the scene and the left singular values as the concentration map of these spectra. The singular values then give the relative amount each compound in the scene. Low-rank of the image can then be interpreted as presence of a few spectra with a correlated concentration profile across space.

4.2.2 Sparsity structure of hyperspectral noise

In hyperspectral imaging the atmosphere can lead to vastly different absorption rates across the spectrum of interest. In particular as shown in Figure 4.1, the two water absorption bands are attenuated, roughly at band 60 and 100. In a typical hyperspectral data processing the data from these two bands would be discarded. On the other hand we note that in the noisy reshaped image, the spectral noise exhibits a banded structure which is mathematically equivalent to saying that hyperspectral noise exhibits a simultaneous sparse structure under the given reshaping of the data cube.

Therefore, the noisy reshaped hyperspectral data cube can be represented as $\mathbf{Y} = \mathbf{L} + \mathbf{S}$ where $\mathbf{L}$ is the low-rank non-noisy image and $\mathbf{S}$ is the spectral noise which is simultaneously or group sparse across bands.

4.3 Robust & rapid hyperspectral imaging

Both the spectral de-noising and limited angle reconstruction problems can be viewed through the following framework in which we observe noisy measurement, $\mathbf{Y}$ of hyperspectral image cube $\mathbf{L}$ through a measurement system described by the linear operator

(matrix) $\mathcal{A}$, i.e.

$$\mathbf{Y}(:) = \mathcal{A}(\mathbf{L}(:) + \mathbf{S}(:)) \tag{4.2}$$

The problem is that given the observation $\mathbf{Y}$ and the *sensing* operator $\mathcal{A}$ (to be defined below for both problems of interest) we want to recover the de-noised image $\mathbf{L}$ while removing the noise $\mathbf{S}$.

4.3.1 Complexity penalized recovery algorithms

To de-noise and recover the hyperspectral data, one can exploit the low-rank and sparse structure of the data and noise and solve the following optimization,

$$\min_{\mathbf{L},\mathbf{S}} ||\mathcal{A}(\mathbf{L}(:) + \mathbf{S}(:)) - \mathbf{Y}||_2^2 + \lambda_L \text{rank}(\mathbf{L}) + \lambda_S ||\mathbf{S}||_{0,2} \tag{4.3}$$

where λ_L & λ_S control the relative strength of the sparsity and low-rank penalization and $||\mathbf{S}||_{p,q}$ is the p-norm of the vector formed by taking the q-norm along the rows of $\mathbf{S}$ or otherwise also known as $\ell_{p,q}$ norm. This optimization problem is known to be NP-hard. However, the rank and support penalties can be relaxed to the nuclear norm and $\ell_{1,2}$ norm, respectively, which makes the optimization tractable while still encouraging the desired structure for $\mathbf{L}$ and $\mathbf{S}$ [5]. Therefore we relax the above combinatorial optimization problem to the following convex optimization problem and consider three cases.

$$\min_{\mathbf{L}(:),\mathbf{S}(:)} ||\mathcal{A}(\mathbf{L} + \mathbf{S}) - y||_2^2 + \lambda_L ||\mathbf{L}||_* + \lambda_S ||\mathbf{S}||_{1,2} \tag{4.4}$$

Case I - Hyperspectral de-noising with raster scan data - In this case $\mathcal{A}$ is an identity operator and therefore the optimization problem becomes,

$$\min_{\mathbf{L},\mathbf{S}} ||\mathbf{L}(:) + \mathbf{S}(:) - \mathbf{Y}(:)||_2^2 + \lambda_L ||\mathbf{L}||_* + \lambda_S ||\mathbf{S}||_{1,2} \tag{4.5}$$

In Section 4.4 we will demonstrate the performance of this algorithm on real hyperspectral data and give experimental results that motivate why the sparse component is necessary for the de-noising.

Case II. - Image recovery from limited angle tomography: No spectral noise - As pointed out in the introduction the CTIS systems are limited by the size of focal plane array which limits the number of tomographic projections that can be obtained. In this case traditional reconstructions suffer from the missing cone problem [50]. These methods however do not exploit the low complexity of the underlying data cube. Assuming no spectral noise, given the limited number of Radon projections we propose

the following algorithm for estimation of the hyperspectral image which exploits the low-rank structure.

$$\min_{\mathbf{X}} ||\mathcal{A}(\mathbf{X}) - \mathbf{Y}||_2 + \lambda ||\mathbf{X}||_* \tag{4.6}$$

Case III. - Image recovery from limited angle tomography: Noisy case - Here we consider the most general case where the spectral data is corrupted by banded spectral noise and the data is acquired through a CTIS system with limited number of Radon projections. In this case simultaneous spectral cube recovery and spectral de-noising is affected by solving for the optimization problem given by equation 4.4. In the next section we will present detailed experimental results of the proposed algorithms on real data sets.

4.4 Experimental evaluation

In this section we will use a real hyperspectral image taken from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), far above an rural scene with a spatial dimension of 128 by 128 pixels. The imager uses 220 bands which cover the spectrum from IR to visible range. The two water absorption bands centered at 1400 and 1900 nm corrupt the image. **NOTE**: All the optimization problems below are implemented using TFOCS [55].

4.4.1 Case I. - Hyperspectral de-noising

- In some of the less noisy bands the structure of the image is still somewhat visible (figure 4.3). In order to improve the de-noising the AVIRIS data we first take and record the Frobenious norm of each frame to construct a $N_\lambda \times 1$ vector W. We then use this recorded vector to normalize each image at given wavelength such that the signal energy in each band is 1. Because we expect the noise in our experiment to be due to low photon counts in bands of high absorption, we can use the vector W to weight the minimization operation. In particular we want to encourage row sparsity along the bands with low counts. In order to do so we modify equation 4.5 to include the weighting factor W, that makes it more expensive for the intense bands to be decomposed into the sparse matrix.

$$\min_{\mathbf{L},\mathbf{S}} ||(\mathbf{L} + \mathbf{S}) - \mathbf{Y}||_2^2 + \lambda_L ||\mathbf{L}||_* + \lambda_S ||W\mathbf{S}||_{1,2} \tag{4.7}$$

This weighting factor allows our algorithm to be more robust to choices of λ_S and λ_L as it effectively decreases the coherence between the $\ell_{1,2}$ norm and the nuclear norm.

The minimization operation in Equation (4.7) is then applied to the hyperspectral image with λ_S of .06 and λ_L of 0.1. The proposed algorithm was successful in de-noising and was able to remove the spectral noise. Figure 4.3 show the results of algorithm applied to a few very noisy bands and figure 4.2 shows the results applied to all bands of the hyperspectral image. Like in the synthetic example, we can now see structure in the bands that were previously noisy.
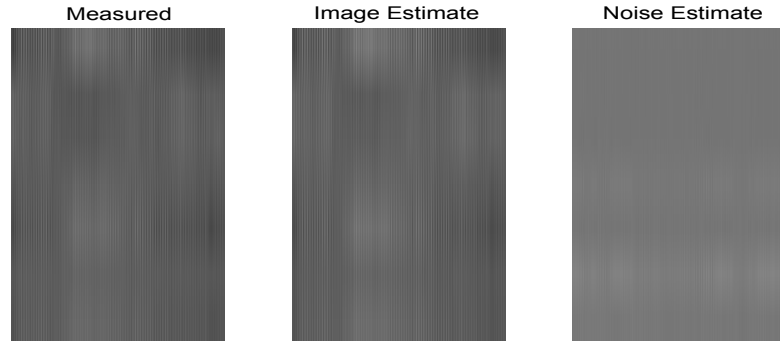
FIGURE 4.2: This figure shows 2D hyperspectral cube with noise and low-rank reconstruction.
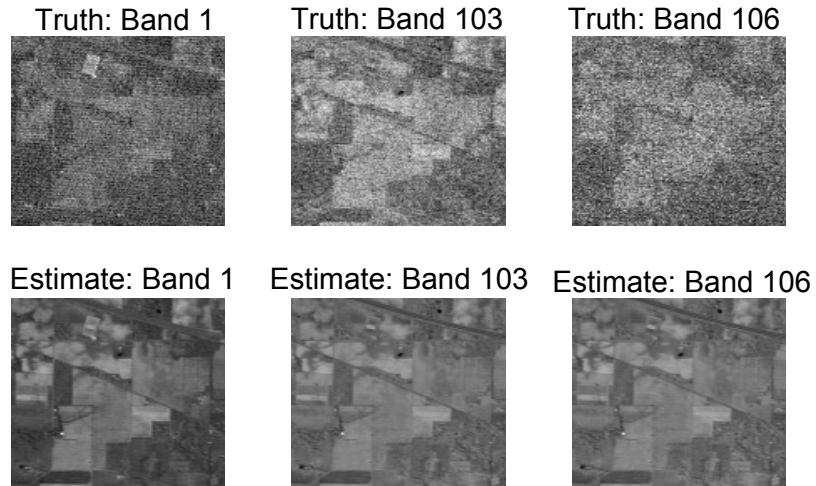
FIGURE 4.3: This figure shows images from AVIRIS data at various bands before de-noising and after de-noising.

4.4.2 Case II. - Hyperspectral imaging from limited Radon projections with no spectral noise

In the following example we attempt to reconstruct the 32 by 32 image of the hyperspectral flower [56] using a limited number of projections. The projections at various angles for a typical single-shot CTIS system [47, 57] are shown in Figure 4.4. Gaussian noise was then added to the measured projections, such that the resulting SNR of the projections was 4.5 dB. This projection operation can be represented through the underdetermined matrix $\mathcal{A}$. In this case we solve the optimization problem (4.6) for recovery. We compare the performance of this method to the standard Tikhonov regularization approach with ℓ_2 norm penalty instead of nuclear norm penalty. The choice of λ for both cases were determined using the Kolmogorov-Smirnoff (KS)-test method described in Section 4.4.4. As expected the low-rank minimization resulted in a better reconstruction of the hyperspectral image cube with normalized mean square error of .23 versus .35 for the Tikhonov reconstruction. The resulting reconstruction for the 12th band is shown in Figure 4.5.
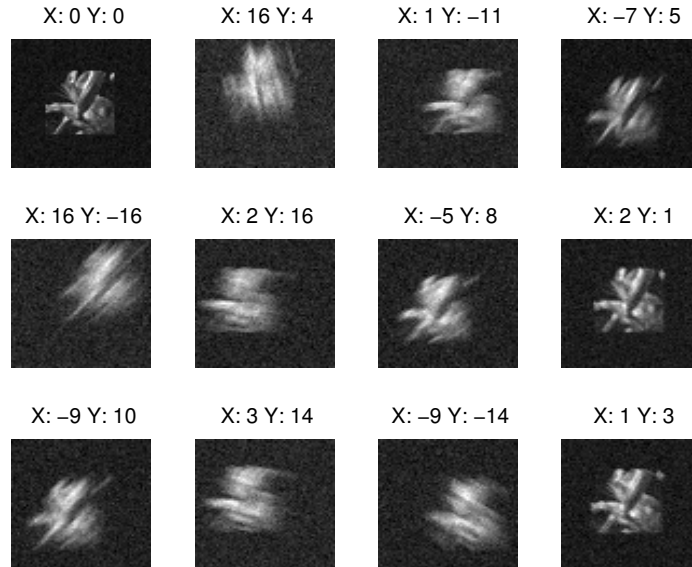
FIGURE 4.4: This figure shows the 12 noisy radon projections of the hyperspectral image cube. With 12 projections the system is underdetermined.

4.4.3 Case III.- Simultaneous tomographic reconstruction and de-noising

Here we attempted to remove spectral noise from a hyperspectal data cube and reconstruct the cube from a limited number of Radon projectioLns. In order to do so we use
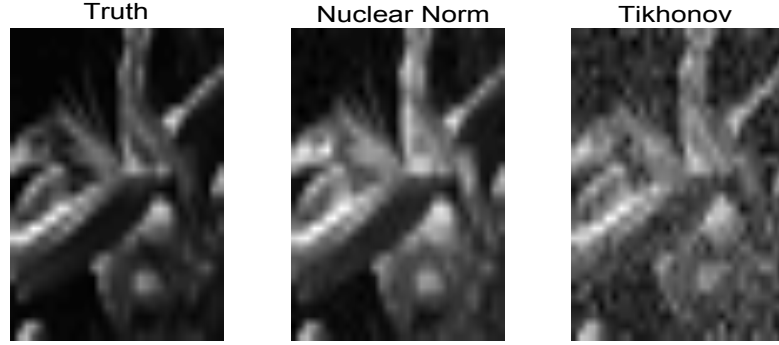
FIGURE 4.5: This figure shows an example of the true image, low-rank reconstruction, and least square reconstruction, from the hyperspectral flower at band 12.

a 64x64 section of the original AVIRIS image as used above and observe cube through the same Radon matrix as in the flower example. The simultaneous reconstruction and spectral noise was recovered by solving the optimization problem given in (4.4) the results are shown in Figure 4.6. A λ_S of .06 and λ_L of .1 were used. These reguliarization parameters were chosen using the KS-surface method described in the following section. We show good reconstruction outside of the noisy bands and significant reduction of noise within the spectrally corrupted bands. Although the noise was somewhat reduced in the corrupted bands, as is to be expected in this limited data case, the images still remained nosier than the case when the image hypercube was directly observed with the identity operator.
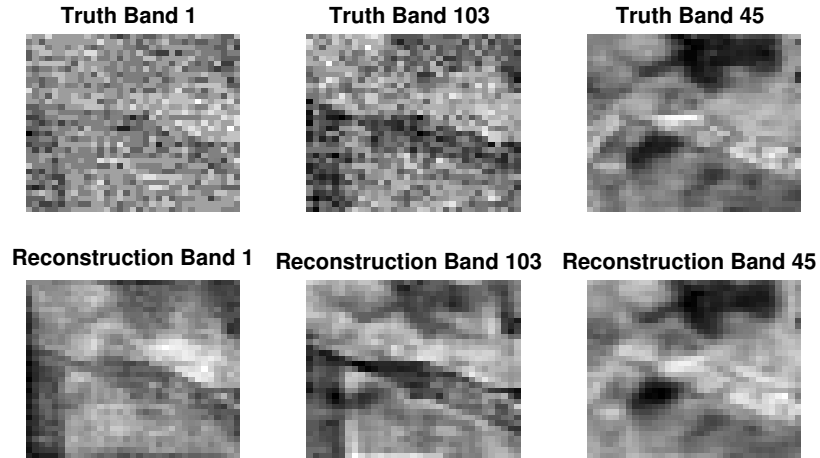
FIGURE 4.6: This figure shows the reconstructed and original hypercube at two noisy bands 1 & 103 and at the clean band 45. The reconstruction at the noiseless bands highly resemble the original image. Although somewhat de-noised, the the images at the corrupted bands remain somewhat blurry and the presence of noise is still visible.

4.4.4 Selection of parameters λ_L and λ_S

In all of the above algorithms the issue of selection of complexity regularization parameters λs is of practical importance. In the following experiments that we carried out on real and synthetic data sets the choice of both λ_S and λ_L was determined using a one and two dimension variation of the Komolgrov-Smirnov test method proposed in [58]. The method essentially computes the KS test statistics of errors for a particular value of regularization parameter with respect to error residuals at extreme values of regularization parameter(s) and generates two curves. The operating point is then picked at the intersection of these two curves.

Selection of regularization parameter for limited angle tomography: no spectral noise - The KS plot was generated with logarithmic spaced choice of λ from 10^{-1} to $10^2$. The KS statistic values and the associated p-values are shown in the top of the Figure 4.7 and their intersection yield a slight suboptimal λ of 12.6. For reference the KS-test was preformed for Tikhonov regularization and the optimal, see Figure 4.7-bottom left plot. From this plot we can see that for all feasible values of λ Tikhonov regularization results in a poorer reconstruction than the nuclear norm reconstruction. In addition to performing the KS-test for selection of regularization parameter the L-curve method [59], commonly used for the selection of the λ was also generated, see Figure 4.7 bottom right. The L-curve method results in a shallow curve without the presence of sharp knee typical of L-curve plots. The lack of the knee makes it very difficult to select a λ and introduces an opportunity for user bias in the selection of the regularization parameter.
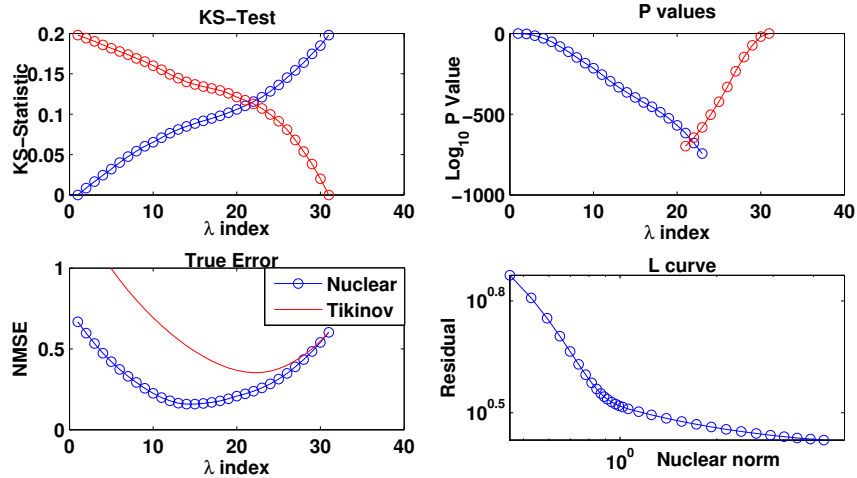
FIGURE 4.7: Top plots - KS test plot for recovery under limited Radon projections for the case considered. Bottom plots: (Left) - MMSE computed using the true image for various values of λ for Tikhonov and RPCA methods; (Right) - L-curve for the RPCA method.

Selection of regularization parameter for limited angle tomography with spectral noise - For this we extend the one dimensional KS test method in [58] to a two dimension variation by generating a KS test surface. In order to generate the KS-surface many KS-tests were run with a fixed λ_S and the λ_L was varied from 10^{-3} to 10^{-1}. This process was then repeated for a range of λ_S from 10^{-3} to 10^{-1}, effectively generating a KS-plot for the selection of λ_L for a given value of λ_S. We can then view these multiple KS-tests as two surfaces of KS statistics as shown in Figure 4.8 (left plot), where the intersection of the two surfaces represents best choice of λ_L as a function of λ_S. From this KS-surface we can then interpret the line defining the intersection of the two surfaces line of optimal regularization pairs. We can then take the pair corresponding to the smallest value and largest value of λ_L and generate a conventional one dimensional KS-test along the intersection of λ pairs as shown in Figure 4.8 right. In this was the one dimensional KS-test selects the best pair set among the intersection of pair sets.
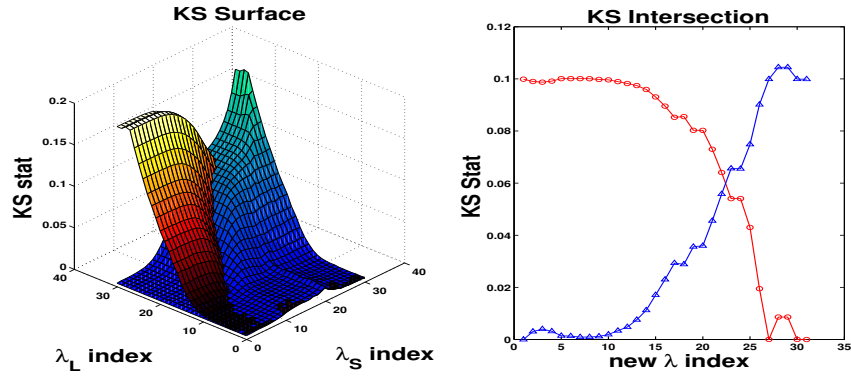
FIGURE 4.8: KS Surface for selecting regularization parameters for simultaneous data cube recovery and hyperspectral noise elimination.

Chapter 5

Reflection Seismology

In this chapter we present novel strategies for completion of 5D pre-stack seismic data, viewed as a 5D tensor or as a set of 4D tensors across temporal frequencies. In contrast to existing complexity penalized algorithms for seismic data completion, which employ matrix analogues of tensor decompositions such as HOSVD or use overlapped Schatten norms from different unfoldings (or matricization) of the tensors, our approach uses a recently proposed decomposition called tensor SVD or tSVD for short, proposed in [2]. We show that seismic data exhibits low complexity under tSVD, i.e. is compressible under tSVD representation, and we subsequently propose a new complexity penalized algorithm for pre-stack seismic data completion under missing traces. This complexity measure which we call the Tensor Nuclear Norm (TNN) is motivated by algebraic properties of the tSVD. We test the performance of the proposed algorithms on synthetic and real data and show that missing data can be reliably recovered under heavy down-sampling.

5.1 Introduction

This chapter will mainly discuss recovery of pre-stack seismic records from incomplete spatial measurements. Ideally, a seismic survey would have receiver and source geometries at all of the possible spatial indices, however this rarely occurs due to financial and physical constraints. Instead, only a fraction of the possible shot and receiver locations are recorded and the data is spatially under-sampled, confounding interpretation of the seismic data, see [60] and references therein.

The pre-stack seismic data can be viewed as a 5D data or a fifth order *tensor* consisting of one time or frequency dimension and four spatial dimensions describing the location of

the detector and the receiver in a two dimensional plane. This data can then be described in terms of the original (r_x, r_y, s_x, s_y) coordinate frames or in terms of midpoint receivers and offsets (x, y, h_x, h_y). Alternatively, one can also view this data as a collection of 4D tensors across the temporal frequencies as considered in [60] and [61]. In this chapter we present a novel method for seismic data completion from the limited survey information using complexity penalized recovery algorithms which measure complexity under the appropriate tensor decompositions. We assume that the seismic data has low information complexity in the sense that there is high redundancy or correlation in the traces from one grid point to the next. These tensor complexity measures capture this redundancy across the traces. From the perspective of optimal sampling and recovery, low complexity data can be reliably recovered from a measurement rate in proportion to the information rate of the data, which is a generic theme in compressive sensing theory and methods [62–64]. In the context of data viewed as a matrix (or a 2nd order tensor) one complexity measure that has shown promise in matrix completion problems is matrix rank [65] (computed via its Singular Value Decomposition (SVD)). The notion of rank extends to higher order tensors to capture the redundancy of the data but unlike the 2D case, for higher order tensors rank depends highly on the type of decomposition or factorization. Many methods exist for tensor decompositions together with their associated notions of rank, primary among them being the Higher Order SVD (HOSVD) and Canonical Decomposition (CANDECOMP), see [66] for a survey of these decompositions. Other notions are based on flattening of the tensor into matrices and then computing the ranks of the resulting matrices, e.g. the tensor p-rank of order p tensors which is the sum of ranks of the p matrices obtained by flattening the tensor along the p different directions [67].

Seismic data completion using an appropriate convex relaxation of these complexity measures have been proposed recently in [61, 68], [69] and [70, 71].In this chapter we exploit a novel tensor singular value decomposition (tSVD) proposed in [2] for third order tensors and its extensions to higher order tensors in [16]. With synthetic data we show that the pre-stack seismic data is compressible, i.e. has low informational complexity in the tSVD domain and hence can be reliably recovered under limited sampling using an appropriate complexity penalized algorithm in the tSVD domain. The following section describes mathematical preliminaries to state the problem and background needed for understanding the proposed algorithms.

5.2   Method

The true seismic data $\mathscr{M}$ is spatially under-sampled and this operation can be represented by a linear operator $\mathscr{A}$ resulting in the sparse observed data $\mathscr{Y}$ under additive noise.

$$\mathscr{Y} = \mathscr{A}(\mathscr{M}) + \mathscr{N} \tag{5.1}$$

The problem of seismic data completion becomes to reliably estimate $\hat{\mathscr{M}}$ from $\mathscr{Y}$ under the sampling operator $\mathscr{A}$. However, because the number of observed measurements is significantly less than the number of elements in $\mathscr{M}$, the problem is severely ill-posed and cannot be solved directly without placing some constraints on $\mathscr{M}$. As noted in the introduction, these constraints arise in the form of complexity measures on the underlying true data $\mathscr{M}$ and one can reliably recover the $\mathscr{M}$ using the complexity penalized algorithms of the type,

$$\min\ h(\mathscr{M})$$
$$s.t. \|\mathscr{Y} - \mathscr{A}(\mathscr{M})\| \leq \sigma_n \tag{5.2}$$

where $h(\mathscr{M}) \in \mathbb{R}^+ \cup \{0\}$ is a non-negative real valued mapping and measures the complexity of the true data $\mathscr{M}$ and σ_n^2 is the additive noise variance. In the following, we assume that the noise variance is negligible. Since our complexity measure derived from tSVD, in the next section we provide a brief overview of the decomposition.

5.2.1   Math Background: tSVD

Herein we provide a brief overview of the tSVD framework. For more details, please refer to [2] and [16]. For the sake of visual and analytical interpretation, we discuss the 3D case here briefly. The 3D tSVD extends in a recursive manner to higher order tSVD, as shown in [16]. Let $\mathscr{M}$ be an order p tensor $\in \mathbb{R}^{n_1 \times n_2 \times ... \times n_p}$, then $\mathscr{M}$ can be decomposed as the following form,

$$\mathscr{M} = \mathscr{U} * \mathscr{S} * \mathscr{V}^{\mathrm{T}} \tag{5.3}$$

where $\mathscr{U}$ is an *orthogonal* tensor of size $n_1 \times n_1 \times n_3 \times ... \times n_p$, $\mathscr{S}$ is a block-diagonal tensor of size $n_1 \times n_2 ... \times n_p$, $\mathscr{V}$ is an *orthogonal* tensor of size $n_2 \times n_2 \times n_3 \times ... \times n_p$ and $*$ denotes the t-product. The orthogonality of $\mathscr{U}$ and $\mathscr{V}$ is understood in terms of the t-product and a tensor transpose operation, namely $\mathscr{U} * \mathscr{U}^T = \mathscr{I}$ where $\mathscr{I}$ denotes the identity tensor such that for all tensors of appropriate dimension $\mathscr{M} * \mathscr{I} = \mathscr{M}$, see

[72] for details. This SVD like decomposition is best understood by looking at Figure 2.5 where we show the decomposition for 3D case. As explained in [2], due to the nature by which the tensor product $*$ is defined, this tensor factorization can be obtained by **Algorithm 8** using FFT, taken verbatim from [16], where in line with notation and convention there, we have chosen the convention in MATLAB which stores higher order tensors as a stack of matrices with the stack index running from 1 to $\rho = n_3 n_4 ... n_5$.

Algorithm 8 tSVD

Input: $\mathscr{M} \in \mathbb{R}^{n_1 \times n_2 ... \times n_p}$
$\rho = n_3 n_4 ... n_p$
for $i = 3$ to p **do**
  $\mathscr{D} \leftarrow \text{fft}(\mathscr{M}, [\,], i)$;
end for
for $i = 1$ to ρ **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathscr{D}(:,:,i))$
  $\mathscr{U}(:,:,i) = \mathbf{U}$; $\mathscr{S}(:,:,i) = \mathbf{S}$; $\mathscr{V}(:,:,i) = \mathbf{V}$;
end for
for $i = 3$ to p **do**
  $\mathscr{U} \leftarrow \text{ifft}(\mathscr{U}, [\,], i)$; $\mathscr{S} \leftarrow \text{ifft}(\mathscr{S}, [\,], i)$; $\mathscr{V} \leftarrow \text{ifft}(\mathscr{V}, [\,],)$;
end for

The main advantage of this type of decomposition stems from the optimality property of a "truncated-tSVD" as a best *t-rank k* approximation, where $k \leq l = \min\{n_1, n_2\}$, to a tensor in the Frobenius norm where the rank is defined with respect to tSVD as the number of non-zero singular "tubes" in SVD, [2]. This optimality results from the natural ordering of the singular tubes by their energy captured by the ℓ_2 norms of the tubal vectors. This key feature distinguishes tSVD from HOSVD[1] where there is no natural ordering of the entries in the core tensor. Other tensor analysis methods, such as tensor *p*-rank used in [67], operate by flattening the tensor and do not respect the *orientation* of the tensor and the structural complexity aspects inherent in the orientation are lost. In contrast, the tSVD based approach preserves this orientation and can be taken along the most natural orientation of the data.

In this context, tensor completion using tSVD exploits the compressibility of natural data in the tSVD domain in the sense that it can be well approximated by a low-rank tSVD, where rank is the notion presented in [72]. There the *multi-rank* of a tensor using tSVD was defined to be a *vector of the ranks* of the frontal slices $\mathscr{D}(:,:,i)$ after taking the Fourier transform along the third dimension of the 3D tensor, see **Algorithm 8**. This suggests a complexity penalized algorithm for recovery of $\mathscr{M}$ from linear measurements,

---

[1]HOSVD is the only reasonably computable decompositions for higher order tensors. CANDECOMP may or may not be unique generally and when it is unique finding the right decomposition is usually computationally intensive.

that minimizes the sum of the elements of the tubal-rank vector of $\mathscr{M}$.

$$\text{CP1}: \quad \min \sum_{i=1}^{\rho} \text{rank}(\mathscr{D}(:,:,i))$$

$$s.t. \mathscr{Y} = \mathscr{A}(\mathscr{M}) \tag{5.4}$$

However, as in the case of recovery of matrices under-sampling via linear operators by minimizing rank, the problem C1 is NP-hard. We therefore *relax* the complexity measure to a norm which we call the tensor nuclear norm (TNN) into the constrained (noiseless) and unconstrained (noisey) problem [73],

$$\text{OPT\_TNN(1)}: \quad \min \sum_{i=1}^{\rho} ||\mathscr{D}(:,:,i)||_{nuc}$$

$$s.t. \ \mathscr{Y} = \mathscr{A}(\mathscr{M}) \tag{5.5}$$

$$\text{OPT\_TNN(2)}: \quad \min \ |\mathscr{Y} - \mathscr{A}(\mathscr{M})|_2^2 + \lambda \sum_{i=1}^{\rho} ||\mathscr{D}(:,:,i)||_{nuc} \tag{5.6}$$

where $|| \cdot ||_{nuc}$ denotes the Schatten-1 norm on the singular values of the matrix in the argument [74], also known as the nuclear norm in the literature [65] and λ is the sparse tuning factor. It is easy to see that the proposed optimization problem is a convex optimization problem, see also [73] for details, and therefore can be solved using existing techniques. Before we present the algorithm, in the next section we first show that the pre-stack seismic data is compressible in the tSVD domain.
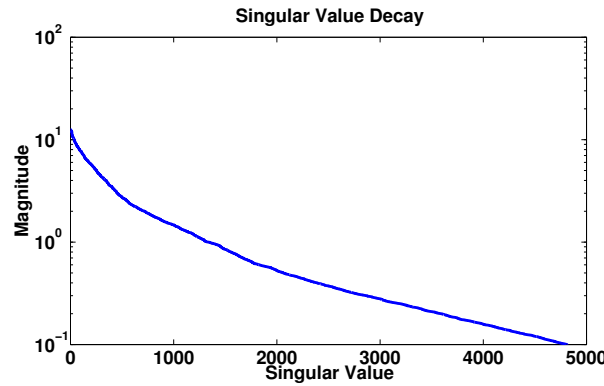

FIGURE 5.1: This figure shows the decay of singular values of the synthetic seismic data which empirically obey a power law decay.

5.2.2 Compressibility of seismic data in the tSVD domain

In order to demonstrate the compressibility of the seismic data, we generated a synthetic 5D survey in which sources and receivers were placed on a 12×12 grid spaced 100 meters apart. Three synthetic Born scatters were placed below the surface and traces consisting of N_t samples were generated for all of the possible receivers source geometry to generate a 5D tensor with dimensions $12 \times 12 \times 12 \times 12 \times N_t$. Several r_y directional slices of the original tensor are shown in Figure 5.3 for different values r_x with source location being held constant at s_x 3 and s_y 7. The tSVD was applied to the synthetic data. Figure 5.1 shows the decay of the singular values of the matrices $\mathscr{D}(:,:,i)$ as computed using the tSVD and for the synthetic data case they obey a power law decay, which implies that $\sum_{i=1}^{\rho} ||\mathscr{D}(:,:,i)||_{nuc}$ is sufficient measure of complexity of the seismic data in the tSVD domain. As a result, we expect the data to be recoverable from a limited number of measurements by using the algorithm OPT_TNN.

5.2.3 An ADMM algorithm for solving OPT_TNN

Motivated by recently proposed ADMM algorithms for solving nuclear norm minimization problems [75], we employ **Algorithm 9** for solving OPT_TNN. Where Sh_ϵ is an

---

Algorithm 9 ADMM: TNN

$\mathscr{P} = null(\mathscr{A})$ // Projects onto the nullspace of the measurement tensor.
$\mathscr{X} = \mathscr{L} = \mathscr{B} = 0$ // Initialize variables.
$\rho = n_3 n_4 ... n_p$
while Not Converged **do**
  $\mathscr{X} = \mathscr{P} \cdot (\mathscr{L} - \mathscr{B}) + \mathscr{Y}$;
  $\mathscr{L}_{old} = \mathscr{L}$
  $\hat{\mathscr{X}} = \alpha \hat{\mathscr{X}} + (1 - \alpha)\mathscr{L}_{old}$
  $\mathscr{M} = \hat{\mathscr{X}} + \mathscr{B}$
  // Calulcate tSVD and shrink in the tSVD domain.
  for $i = 3$ to p **do**
    $\mathscr{D} \leftarrow \text{fft}(\mathscr{M}, [], i)$;
  end for
  for $i = 1$ to ρ **do**
    $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathscr{D}(:,:,i))$;
    $\mathbf{S} = Sh_\varepsilon[\mathbf{S}]$;
    $\mathscr{U}(:,:,i) = \mathbf{U}$; $\mathscr{S}(:,:,i) = \mathbf{S}$; $\mathscr{V}(:,:,i) = \mathbf{V}$;
  end for
  for $i = 3$ to p **do**
    $\mathscr{U} \leftarrow \text{ifft}(\mathscr{U}, [], i)$; $\mathscr{S} \leftarrow \text{ifft}(\mathscr{S}, [], i)$; $\mathscr{V} \leftarrow \text{ifft}(\mathscr{V}, [],)i$;
  end for
  $\mathscr{L} = \mathscr{U} * \mathscr{S} * \mathscr{V}^T$
  $\mathscr{B} = \mathscr{B} + \hat{\mathscr{X}} - \mathscr{L}$
end while

---

element-wise shrinkage function that applies a soft thresholding to $\mathscr{S}$ in the Fourier domain, see Equation 5.7. ε determines the step size and α controls the tolerance of the equality constraint. For all of the instances of the algorithm, a value of 1 & 1 was chosen for these two parameters and the algorithm ran for 100 iterations. It is observed that the result is robust to the choice of these parameters although for some extreme values the algorithm performance degrades.

$$Sh_\varepsilon[x] = \begin{cases} x - \varepsilon, \text{ if } x > \varepsilon, \\ x + \varepsilon, \text{ if } x < \varepsilon, \\ 0, \text{otherwise}, \end{cases} \tag{5.7}$$
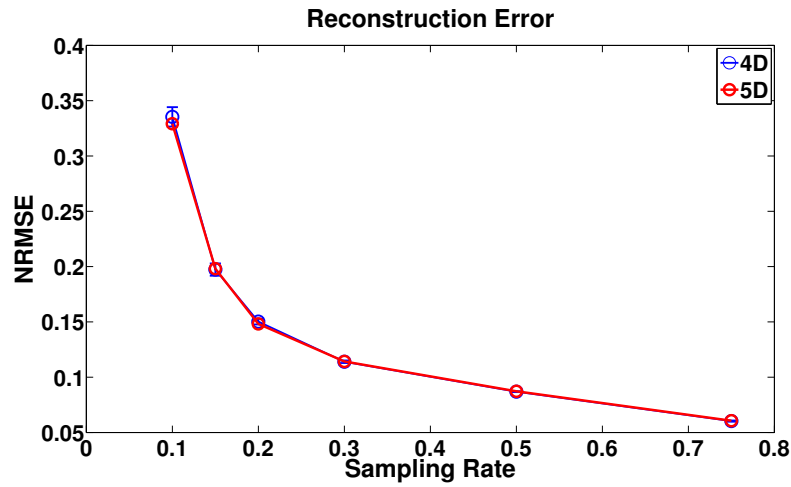
FIGURE 5.2: This figure shows the reconstruction error as function of sampling fraction for both the 4D frequency by frequency and full 5D reconstruction. For severely under-sampled data, below 20 percent, the 5D reconstruction provides marginally better results than the 4D reconstruction.

5.3 Performance evaluation: Synthetic data

To evaluate our algorithm we used the synthetic dataset described previously and removed 25, 50, 70, 80, 85, and 90 percent of the measured traces, resulting in highly under-sampled data as shown in Figure 5.3B for 90% of traces removed. **Algorithm 9** was then applied in two ways: (1) Recovering the data by completion of 4D tensors frequency by frequency (2) Recovering the entire 5D seismic volume at once. Since the downsampling operator is chosen at random, we averaged the performance in recovery over 10 randomly selected downsampling operators for each level of under-sampling. The error in recovery is measured via Normalized Root Mean Square Error (NRMSE). For the synthetic data, the performance as function of sampling rate is shown in Figure 5.2. Although the NRMSE is significant for the highly under-sampled data, the reflections are still highly visible in the reconstruction, Figure 5.3. Therefore, the NRMSE

appears to be a pessimistic norm of the quality of the reconstruction and one should focus on the errors at the time support of the arrivals instead of the entire time window. For highly under-sampled data, the reconstruction using the full 5D data rather than the frequency by frequency data has slightly less NRMSE, as shown in the Figure 5.2. Unlike the methods considered so far in [61, 68] which work on 4D data frequency by frequency, we observed a small benefit to processing the data as a whole.
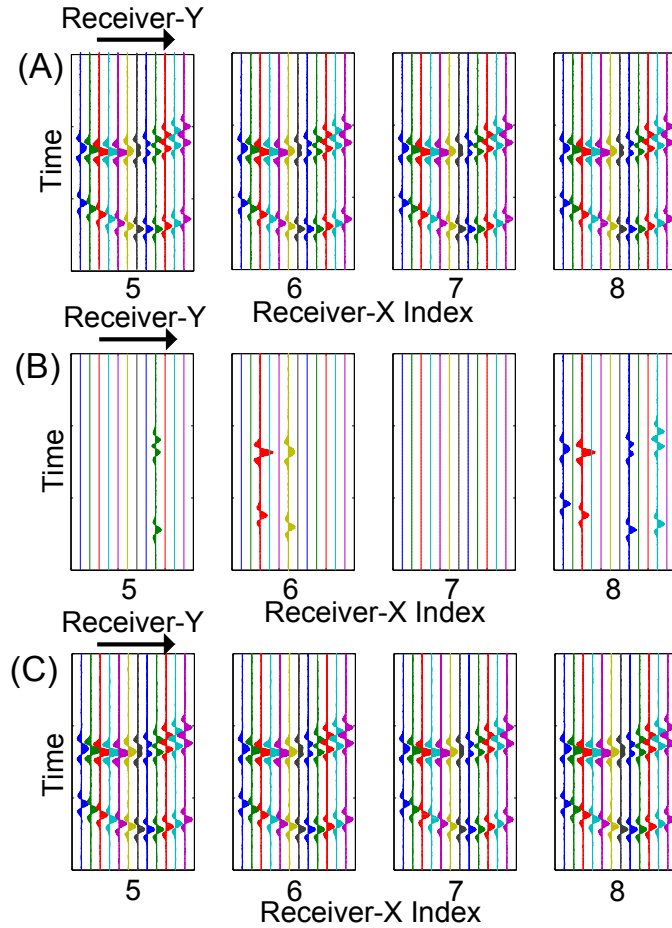
FIGURE 5.3: This figure shows the full synthetic data (A) for four different receiver source slices as well as the under-sampled measured data for the case when 90% of the traces were removed (B). In addition, the reconstruction for the 5D (C) slices are shown as well.

5.4 Performance on field data

In addition to synthetic data, the algorithm was tested on a survey of the Western Canadian Sedimentary Basin. The dataset consists of grid with 29 midpoint gather locations with 12×12 offset coordinates to generate a data set with spatial dimensions of $29 \times 29 \times 12 \times 12$. The midpoint grid spacing is 26 meters and 52 meters for x and y with an offset spacing from 0 to 1400 meters. Given the geometry roughly 121,000 source

receiver geometries are possible, however only 16060 traces were recorded resulting in sampling rate of approximately 8%. The algorithm was applied to the field data for both the constrained & unconstrained problems. The reconstructed results for the a fixed y-midpoint gather of 11 and y-offset of 6 are shown in figure 5.4 (constrained) & figure 5.5 (unconstrained). Because the field data contains noise the unconstrained reconstruction provides better results 5.5. These traces are the same as in [68] which used HOSVD was used to reconstruct the same dataset.
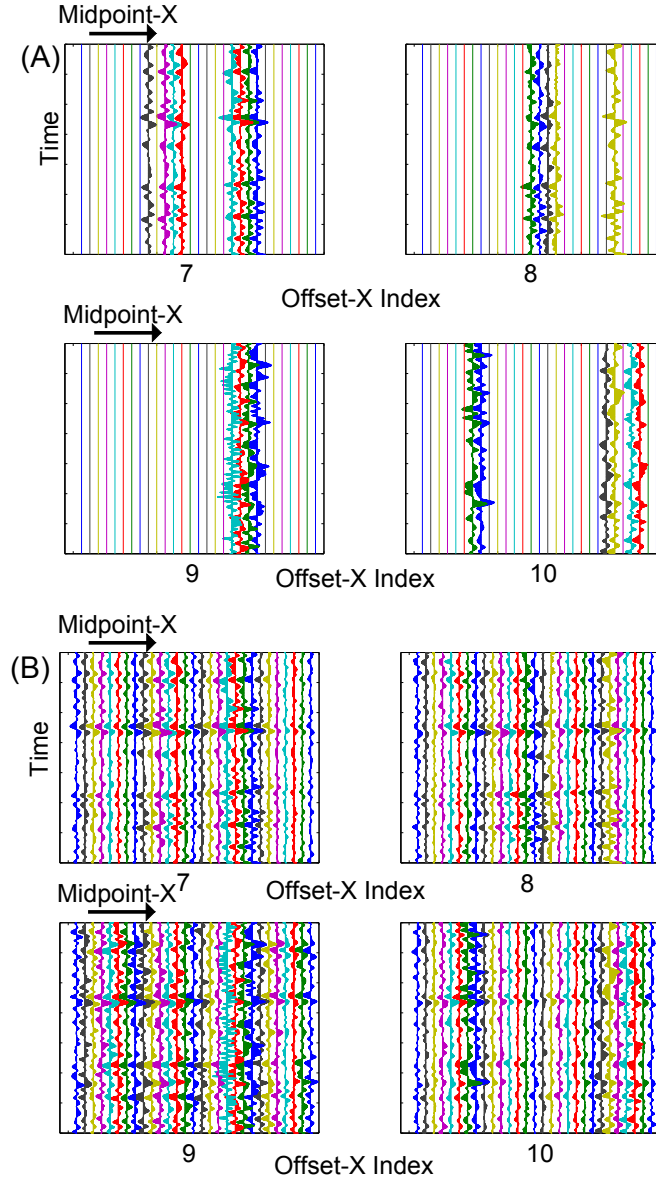
FIGURE 5.4: This figure (A) shows the sparsely sampled field data from the Western Canadian Sedimentary Basin and the reconstructed traces (B) using a frequency by frequency procedure.
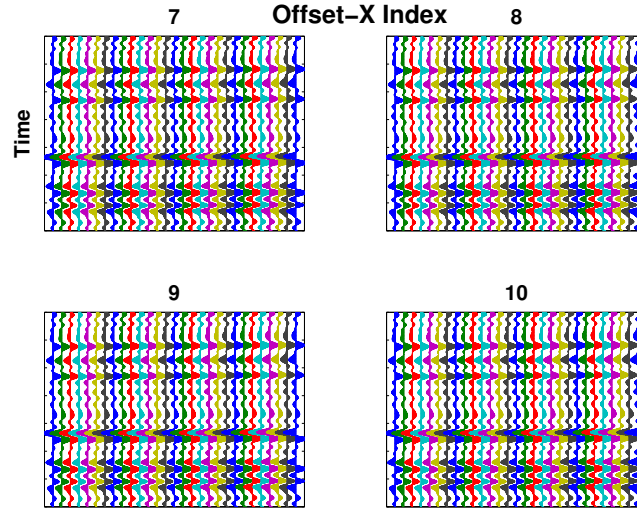
FIGURE 5.5: This figure the reconstructed traces using a frequency by frequency procedure using the unconstrained optimization.

5.5 Conclusion

We have presented a novel method for the reconstruction of incomplete seismic through the use of tSVD rank minimization algorithm. The algorithm was tested both in a frequency by frequency manner as well as applied to the full 5D tensor. For highly under-sampled data, completing in the full 5D space results in slightly better reconstructions. Due to the construction of the tSVD and the shrinkage operator in the tSVD domain, both the 5D and 4D have equivalent computational load and therefore the full 5D reconstruction should be used.

5.6 acknowledgment

We would like to thank CGG, Canada and Prof. Mauricio Sacchi at University of Alberta, Canada for providing us with the field data.

Chapter 6

Conclusion

In this thesis I presented several novel complexity penalized algorithms for improving the recovery of various geophysical inverse problems. Furthermore, I demonstrated how the physics of a given system generates sparsity. In particular, Chapter 3 shows how the spatially compact nature of microseismic events creates a block-low rank structure. In Chapter 4, the limited number of end-members in a scene results in a low-rank image cube. In Chapter 5 the seismic traces can be well represented as shift operations on several unknown common signals, causing the data to be compressible in the tSVD domain. In addition to demonstrating how the physics of these systems result in low complexity, I've developed a collection of highly modular algorithms that can be used to solve a number of denoising, reconstruction, or separation problems. Although I have presented some of the applications of these algorithms, there are numerous opportunities and extensions to alternate domains.

In particular, the tSVD shrinkage operator is well suited to the application of the incremental proximal methods discussed in Chapter 3. Because the tSVD shrinkage operator involves the calculation of numerous SVDs (Equation 6) the speed per iteration could be drastically reduced by only shrinking on a few tensor slices. In addition, I have done some preliminary work with the separation of low-rank tensor objects with sparse objects using ALM but I am still looking for systems that are separable in this framework. One possible application is spurious electronic noise in seismic surveys: this data typically needs to be removed by hand could be automated. For hyper-spectral imaging, the low-rank of the image arises from a positive mixture of a few all positive endmembers. Nuclear norm regularization exploits this limited subspace but does not fully capture the generative model of the image. There may be an opportunity to apply non-negative matrix factorization techniques to improve the reconstruction and separation of the sources of noise in imaging.

Appendix A

Appendix: tSVD Background

The section gives a brief background of the tSVD for an arbitrary dimensional tensor and presents several definitions such as the tSVD transpose. This chapter is largely derived from [15] submitted to NIPS 2013.

A **Slice** of an N-dimensional tensor is a 2-D section defined by fixing all but two indices. Denoted by using the Matlab notation $\mathscr{A}(k, :, :)$ to describe the k_{th} frontal slices.

Definition A.0.1. *Tensor Transpose*. Let $\mathscr{A}$ be a p dimensional tensor of size $n_1 \times n_2 \times n_3... \times n_p$, then $\mathscr{A}^{\mathrm{T}}$ is the $n_2 \times n_1 \times n_3... \times n_p$ tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through p.

Definition A.0.2. *Orthogonal Tensor*. A tensor $\mathscr{Q} \in \mathbb{R}^{n_1 \times n_1 \times n_3... \times n_p}$ is orthogonal if

$$\mathscr{Q}^{\mathrm{T}} * \mathscr{Q} = \mathscr{Q} * \mathscr{Q}^{\mathrm{T}} = \mathscr{I} \tag{A.1}$$

Definition A.0.3. *f-diagonal Tensor*. A tensor is called f-diagonal if each frontal slice of the tensor is a diagonal matrix.

Definition A.0.4. *t-product*. The t-product $\mathscr{C}$ of $\mathscr{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3... \times n_p}$ and $\mathscr{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3... \times n_p}$ is a tensor of size $n_1 \times n_4 \times n_3... \times n_p$ where the $(i, j)_{th}$ tube denoted by $\mathscr{C}(i, j, :)$ for $i = 1, 2, ..., n_1$ and $j = 1, 2, ..., n_4... \times n_p$ of the tensor $\mathscr{C}$ is given by $\sum_{k=1}^{n_2} \mathscr{A}(i, k, :) * \mathscr{B}(k, j, :)$. The product can be efficiently computed in the Fourier given by Algorithm 10.

A.0.1   Tensor Singular Value Decomposition (t-SVD)

The new t-product allows us to define a tensor Singular Value Decomposition (t-SVD).

Theorem A.1. *For $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathcal{M}$ is given by*

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}} \tag{A.2}$$

where $\mathcal{U}$ and $\mathcal{V}$ are orthogonal tensors of size $n_1 \times n_1 \times n_3 ... \times n_p$ and $n_2 \times n_2 \times n_3 ... \times n_p$ respectively. $\mathcal{S}$ is a rectangular f-diagonal tensor of size $n_1 \times n_2 \times n_3 ... \times n_p$, and $*$ denotes the t-product.

We can obtain this decomposition by computing matrix SVDs in the Fourier domain, see Algorithm 10. Figure 2.5 illustrates the decomposition for the 3-D case.

Algorithm 10 t-SVD

Input: $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 ... \times n_p}$
$\rho = n_3 n_4 ... n_p$
for $i = 3$ to p **do**
    $\mathcal{D} \leftarrow \mathrm{fft}(\mathcal{M}, [\,], i);$
end for
for $i = 1$ to ρ **do**
    $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathcal{D}(:,:,i))$
    $\hat{\mathcal{U}}(:,:,i) = \mathbf{U}; \; \hat{\mathcal{S}}(:,:,i) = \mathbf{S}; \; \hat{\mathcal{V}}(:,:,i) = \mathbf{V};$
end for
for $i = 3$ to p **do**
    $\mathcal{U} \leftarrow \mathrm{ifft}(\hat{\mathcal{U}}, [\,], i); \; \mathcal{S} \leftarrow \mathrm{ifft}(\hat{\mathcal{S}}, [\,], i); \; \mathcal{V} \leftarrow \mathrm{ifft}(\hat{\mathcal{V}}, [\,],)i;$
end for

Algorithm 11 t-Product

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 ... \times n_p}$, $\mathcal{B} \in \mathbb{R}^{n_2 \times \psi \times n_3 ... \times n_p}$
$\rho = n_3 n_4 ... n_p$
for $i = 3$ to p **do**
    $\mathcal{A} \leftarrow \mathrm{fft}(\mathcal{A}, [\,], i);$
    $\mathcal{B} \leftarrow \mathrm{fft}(\mathcal{B}, [\,], i);$
end for
for $i = 1$ to ρ **do**
    $\mathcal{C}(:,:,i) = \mathcal{A}(:,:,i)\mathcal{B}(:,:,i);$
end for
for $i = 3$ to p **do**
    $\mathcal{C} \leftarrow \mathrm{ifft}(\mathcal{C}, [\,], i);$
end for

Bibliography

[1] JA Tropp. Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. *SIGNAL PROCESSING*, 86(3):589–602, MAR 2006. ISSN 0165-1684. doi: {10.1016/j.sigpro.2005.05.031}.

[2] Misha E. Kilmer and Carla D. Martin. Factorization strategies for third-order tensors. *Linear Algebra and Its Applications*, 435(3):641–658, August 2011. ISSN 0024-3795. doi: 10.1016/j.laa.2010.09.020. WOS:000291142900013.

[3] Per Christian Hansen, James G Nagy, and Dianne P O'leary. *Deblurring images: matrices, spectra, and filtering*, volume 3. Siam, 2006.

[4] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary. Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21(1):185–194, 1999.

[5] Emmanuel J. Candes, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the Acm*, 58(3), May 2011. ISSN 0004-5411. doi: 10.1145/1970392.1970395. WOS:000291246000003.

[6] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.

[7] Gongguo Tang and A. Nehorai. Robust principal component analysis based on low-rank and block-sparse matrix decomposition. In *2011 45th Annual Conference on Information Sciences and Systems (CISS)*, pages 1 –5, March 2011. doi: 10.1109/CISS.2011.5766144.

[8] Chenlu Qiu and Namrata Vaswani. Reprocs: A missing link between recursive robust pca and recursive sparse recovery in large but correlated noise. *arXiv preprint arXiv:1106.3286*, 2011.

[9] Yudong Chen, Constantine Caramanis, and Shie Mannor. Robust high dimensional sparse regression and matching pursuit. *arXiv preprint arXiv:1301.2725*, 2013.

[10] T Goldstein, BRENDAN ODonoghue, and Simon Setzer. Fast alternating direction optimization methods. *CAM report*, pages 12–35, 2012.

[11] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[12] KATYA SCHEINBERG, DONALD GOLDFARB, and XI BAI. Fast first-order methods for composite convex optimization with backtracking.

[13] Dimitri P Bertsekas. Incremental gradient, subgradient, and proximal methods for convex optimization: a survey. *Optimization for Machine Learning*, page 85, 2011.

[14] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[15] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel factorization strategies for higher order tensors: Implications for compression and recovery of multi-linear data. arXiv e-print 1307.0805, July 2013. URL `http://arxiv.org/abs/1307.0805`.

[16] Carla D. Martin, Richard Shafer, and Betsy LaRue. An order-p tensor factorization with applications in imaging. *SIAM Journal on Scientific Computing*, 35(1): A474–A490, January 2013. ISSN 1064-8275, 1095-7197. doi: 10.1137/110841229. URL `http://epubs.siam.org/doi/abs/10.1137/110841229`.

[17] O. Semerci, N. Hao, M. Kilmer, and E. Miller. Tensor-based formulation for spectral computed tomography. In *SIAM Conference on Imaging Science*, Philadelphia, PA, 2012.

[18] Leo Eisner and Peter M. Duncan. Uncertainties in passive seismic monitoring. *The Leading Edge 28*, 28:648–655, 2009.

[19] Keiiti Aki and Paul G. Richards. *Quantitative Seismology, 2nd Edition*. University Science Books, 2002.

[20] Peter M. Shearer. *Introduction to Seismology*. Cambridge University Press, 2009.

[21] Qiuhua Liu, S. Bose, H.-P. Valero, R.G. Shenoy, and A. Ounadjela. Detecting small amplitude signal and transit times in high noise: Application to hydraulic fracture monitoring. In *IEEE Geoscience and Remote Sensing Symposium*, 2009. doi: 10.1109/IGARSS.2009.5417430.

[22] Ismael Vera Rodriguez, David Bonar, and Mauricio Sacchi. Microseismic data denoising using a 3c group sparsity constrained time-frequency transform. *Geophysics*, 77(2):V21–V29, 2012. doi: 10.1190/geo2011-0260.1. URL http://geophysics.geoscienceworld.org/content/77/2/V21.abstract.

[23] Lejia Han. *Microseismic Monitoring and Hypocenter Location*. PhD thesis, Department of Geoscience, Calgary, Alberta, Canada, 2010.

[24] Bassem Khadhraoui, David Leslie, Julian Drew, and Rob Jones. Real-time detection and localization of microseismic events. *SEG Technical Program Expanded Abstracts*, 29(1):2146–2150, 2010. doi: 10.1190/1.3513270. URL http://link.aip.org/link/?SGA/29/2146/1.

[25] Donald N Burch. Live hydraulic fracture monitoring and diversion. *Oilfield Review*, 21, Autumn 2009.

[26] I. V. Rodriguez, M. Sacchi, and Y. J. Gu. Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory. *Geophysical Journal International*, 2012.

[27] G. Ely and S. Aeron. Robust hydraulic fracture monitoring (hfm) of multiple time overlapping events using a generalized discrete radon transform. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 622 –625, july 2012. doi: 10.1109/IGARSS.2012.6351517.

[28] Gregory Ely and Shuchin Aeron. Complexity penalized hydraulic fracture localization and moment tensor estimation under limited model information. In *Proceedings of Meetings on Acoustics (POMA)*, volume 19, page 045051, Montreal, Canada, 2013. Acoustical Society of America. doi: 10.1121/1.4799336. URL http://link.aip.org/link/?PMA/19/045051/1.

[29] R. G. Pratt and C. H. Chapman. Traveltime tomography in anisotropic mediaii. application. *Geophysical Journal International*, 109(1):20–37, 1992. ISSN 1365-246X. doi: 10.1111/j.1365-246X.1992.tb00076.x. URL http://dx.doi.org/10.1111/j.1365-246X.1992.tb00076.x.

[30] R Madariaga. Seismic source theory. In G Schubert, editor, *Treatise on Geophysics*, volume 4, pages 59–82. Elsevier, 2007. ISBN 978-0-44-452748-6. URL http://dx.doi.org/10.1016/B978-044452748-6.00061-4.

[31] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part II: Convex relaxation. *Signal Processing, special issue on Sparse approximations in signal and image processing*, 86:572–588, April 2006.

[32] A. Majumdar and R.K. Ward. Fast group sparse classification. *Electrical and Computer Engineering, Canadian Journal of*, 34(4):136 –144, fall 2009. ISSN 0840-8688. doi: 10.1109/CJECE.2009.5599420.

[33] AVIRIS-spectra. aviris.jpl.nasa.gov/html/aviris.spectrum.html. URL http://aviris.jpl.nasa.gov/html/aviris.spectrum.html.

[34] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi. *A wavelet-based PCA reduction for hyperspectral imagery*. Ieee, New York, 2002. ISBN 0-7803-7536-X. WOS:000179116800839.

[35] C. A. Shah, P. Watanachaturaporn, P. K. Varshney, and M. K. Arora. *Some recent results on hyperspectral image classification*. Ieee, New York, 2004. ISBN 0-7803-8350-8. WOS:000222142800050.

[36] Damien Letexier and Salah Bourennane. Multidimensional wiener filtering using fourth order statistics of hyperspectral images. In *2008 Ieee International Conference on Acoustics, Speech and Signal Processing, Vols 1-12*, pages 917–920. Ieee, New York, 2008. ISBN 978-1-4244-1483-3. WOS:000257456700230.

[37] Guangyi Chen and Shen-En Qian. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *Ieee Transactions on Geoscience and Remote Sensing*, 49(3):973–980, March 2011. ISSN 0196-2892. doi: 10.1109/TGRS.2010.2075937. WOS:000287658000007.

[38] Sun Lei and Luo Jian-shu. Junk band recovery for hyperspectral image based on curvelet transform. *Journal of Central South University of Technology*, 18 (3):816–822, June 2011. ISSN 1005-9784. doi: 10.1007/s11771-011-0767-6. WOS:000291185400032.

[39] P. Scheunders and J. Driesen. Least-squares interband denoising of color and multispectral images. In *Icip: 2004 International Conference on Image Processing, Vols 1- 5*, pages 985–988. Ieee, New York, 2004. ISBN 0-7803-8554-3. WOS:000228043501069.

[40] Nicola Acito, Marco Diani, and Giovanni Corsini. Signal-dependent noise modeling and model parameter estimation in hyperspectral images. *Ieee Transactions on Geoscience and Remote Sensing*, 49(8):2957–2971, August 2011. ISSN 0196-2892. doi: 10.1109/TGRS.2011.2110657.

[41] Xuefeng Liu, Salah Bourennane, and Caroline Fossati. Nonwhite noise reduction in hyperspectral images. *Ieee Geoscience and Remote Sensing Letters*, 9 (3):368–372, May 2012. ISSN 1545-598X. doi: 10.1109/LGRS.2011.2169041. WOS:000301236900010.

[42] Adam C. Zelinski and Vivek K. Goyal. *Denoising Hyperspectral Imagery and Recovering Junk Bands using Wavelets and Sparse Approximation*. Ieee, New York, 2006. ISBN 978-0-7803-9509-1. WOS:000260989400100.

[43] Zhengming Xing, Mingyuan Zhou, Alexey Castrodad, Guillermo Sapiro, and Lawrence Carin. Dictionary learning for noisy and incomplete hyperspectral images. *Siam Journal on Imaging Sciences*, 5(1):33–56, 2012. ISSN 1936-4954. doi: 10.1137/110837486. WOS:000302220800002.

[44] Daniel Hsu, Sham M. Kakade, and Tong Zhang. Robust matrix decomposition with sparse corruptions. *Ieee Transactions on Information Theory*, 57(11): 7221–7234, November 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2158250. WOS:000297046100001.

[45] G. Pope, M. Baumann, C. Studer, and G. Durisi. Real-time principal component pursuit. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1433 –1437, November 2011. doi: 10.1109/ACSSC.2011.6190254.

[46] Hui Ji, Sibin Huang, Zuowei Shen, and Yuhong Xu. Robust video restoration by joint sparse and low rank matrix approximation. *Siam Journal on Imaging Sciences*, 4(4):1122–1142, 2011. ISSN 1936-4954. doi: 10.1137/100817206.

[47] M. R. Descour, C. E. Volin, D. S. Sabatke, E. L. Dereniak, K. J. Thome, A. B. Schumacher, D. W. Wilson, and P. D. Maker. Demonstration of a high speed non-scanning imaging spectrometer. In E. L. Dereniak and R. E. Sampson, editors, *Infrared Detectors and Focal Plane Arrays V*, volume 3379, pages 250– 258. Spie-Int Soc Optical Engineering, Bellingham, 1998. ISBN 0-8194-2828-0. WOS:000075913800024.

[48] Ashwin A. Wagadarikar, Nikos P. Pitsianis, Xiaobai Sun, and David J. Brady. Spectral image estimation for coded aperture snapshot spectral imagers. *Proceedings of SPIE*, 7076(1):707602–707602–15, August 2008. ISSN 0277786X. doi: doi:10.1117/12.795545. URL http://spiedigitallibrary.org/proceedings/resource/2/psisdg/7076/1/707602_1?isAuthorized=no.

[49] J. M. Mooney, V. E. Vickers, M. An, and A. K. Brodzik. High-throughput hyperspectral infrared camera. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 14(11):2951–2961, November 1997. ISSN 0740-3232. doi: 10.1364/JOSAA.14.002951. WOS:A1997YD31900009.

[50] Ping Feng, Sze Fong Yau, and Y. Bresler. A multicoset sampling approach to the missing cone problem in computer-aided tomography. In *Circuits and Systems*,

1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium on, volume 2, pages 734 –737 vol.2, may 1996. doi: 10.1109/ISCAS.1996.541830.

[51] A.E. Waters, A.C. Sankaranarayanan, and R.G. Baraniuk. SpaRCS: recovering low-rank and sparse matrices from compressive measurements. *Neural Information Processing Systems (NIPS)*, page 10891097, 2011.

[52] J.C. Harsanyi and C.-I. Chang. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):779 –785, July 1994. ISSN 0196-2892. doi: 10.1109/36.298007.

[53] D. Landgrebe. Hyperspectral image data analysis. *Signal Processing Magazine, IEEE*, 19(1):17 –28, January 2002. ISSN 1053-5888. doi: 10.1109/79.974718.

[54] David L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1(1):100 – 115, 1993. ISSN 1063-5203. doi: 10.1006/acha.1993.1008.

[55] S. Becker, E. J. Candes, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3, 2011.

[56] D. H. Foster, S. M. C. Nascimento, and K. Amano. Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience*, 21(3): 331–336, June 2004. ISSN 0952-5238. doi: 10.1017/S0952523804043330.

[57] William R. Johnson, Daniel W. Wilson, and Greg Bearman. Spatial-spectral modulating snapshot hyperspectral imager. *Applied Optics*, 45(9):1898–1908, March 2006. doi: 10.1364/AO.45.001898. URL http://ao.osa.org/abstract.cfm?URI=ao-45-9-1898.

[58] Shuchin Aeron, Sandip Bose, Henri-Pierre Valero, and Venkatesh Saligrama. Broadband dispersion extraction using simultaneous sparse penalization. *IEEE Transactions on Signal Processing*, 59(10):4821–4837, 2011.

[59] Per Christian Hansen and Dianne Prost O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14(6):1487–1503, November 1993. ISSN 1064-8275. doi: 10.1137/0914086. URL http://dx.doi.org/10.1137/0914086.

[60] Daniel Trad. Five-dimensional interpolation: Recovering from acquisition constraints. *GEOPHYSICS*, 74(6):V123–V132, November 2009. ISSN 0016-8033, 1942-2156. doi: 10.1190/1.3245216. URL http://apps.webofknowledge.com.ezproxy.

library.tufts.edu/CitedFullRecord.do?product=WOS&colName=WOS&SID=
3Cl9gaLmgm2Alkhkf@G&search_mode=CitedFullRecord&isickref=WOS:
000273037800025.

[61] Nadia Kreimer and Mauricio D. Sacchi. *SEG Technical Program Expanded Abstracts 2011*, chapter A tensor higher order singular value decomposition (HOSVD) for pre-stack simultaneous noise reduction and interpolation, pages 3069–3074. SEG, 2011. doi: 10.1190/1.3627833. URL http://library.seg.org/doi/abs/10.1190/1.3627833.

[62] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52 (4):1289–1306, April 2006.

[63] E. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12): 5406–5425, December 2006.

[64] Shuchin Aeron, Venkatesh Saligrama, and Manqi Zhao. Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56 (10):5111–5130, Oct. 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2059891.

[65] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Commun. ACM*, 55(6):111–119, 2012.

[66] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[67] Silvia Gandy, Benjamin Recht, and Isao Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011. URL http://stacks.iop.org/0266-5611/27/i=2/a=025010.

[68] Nadia Kreimer and Mauricio D. Sacchi. *SEG Technical Program Expanded Abstracts*, chapter Tensor completion via nuclear norm minimization for 5D seismic data reconstruction, pages 1–5. SEG, 2012. doi: 10.1190/segam2012-0529.1. URL http://library.seg.org/doi/abs/10.1190/segam2012-0529.1.

[69] Jianjun Gao, Mauricio D. Sacchi, and Xiaohong Chen. A fast rank reduction method for the reconstruction of 5D seismic volumes. In *SEG Technical Program Expanded Abstracts*, pages 3622–3627, 2011. doi: 10.1190/1.3627953. URL http://library.seg.org/doi/abs/10.1190/1.3627953.

[70] Rajiv Kumar, Aleksandr Y. Aravkin, Hassan Mansour, Ben Recht, and Felix J. Herrmann. Seismic data interpolation and denoising using svd-free low-rank matrix factorization. https://www.slim.eos.ubc.ca/Publications/Public/Conferences/

EAGE/2013/kumar2013EAGEsind/kumar2013EAGEsind.pdf, 2013. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/kumar2013EAGEsind/kumar2013EAGEsind.pdf.

[71] Curt Da Silva and Felix J. Herrmann. Hierarchical tucker tensor optimization - applications to 4d seismic data interpolation. https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/dasilva2013EAGEhtucktensor/ dasilva2013EAGEhtucktensor.pdf, 2013. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/dasilva2013EAGEhtucktensor/dasilva2013EAGEhtucktensor.pdf.

[72] Misha E. Kilmer, Karen Braman, Ning Hao, and Randy C. Hoover. Third order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 2012. Accepted for publication, Oct. 2012.

[73] Oguz Semerci, Ning Hao, Misha E. Kilmer, and Eric L. Miller. Tensor based formulation and nuclear norm regularization for multienergy computed tomography. submitted to IEEE Transactions on Image Processing, October 2012.

[74] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170(0):33 – 45, 1992. ISSN 0024-3795. doi: 10.1016/0024-3795(92)90407-2. URL http://www.sciencedirect.com/science/article/pii/0024379592904072.

[75] J F Cai, E J Candès, and Z Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.