THE APPLICATION OF “OPTSPACE” ALGORITHM AND COMPARISON WITH “LMAFIT” ALGORITHM IN THREE-DIMENSIONAL SEISMIC DATA RECONSTRUCTION VIA LOW-RANK MATRIX COMPLETION

A REPORT SUBMITTED TO THE DEPARTMENT OF ENERGY RESOURCES ENGINEERING OF STANFORD UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

By
Chen Xue
June 2016
I certify that I have read this report and that in my opinion it is fully adequate, in scope
and in quality, as partial fulfillment of the degree of Master of Science in Energy
Resources Engineering.

Prof. Tapan Mukerji
(Principal Advisor)
Abstract

This report is focused on three-dimensional seismic data reconstruction with randomly missing data on a regular grid. Ma (2013) developed a rank-reduction method that transforms three-dimensional seismic data reconstruction problem into low-rank matrix completion (MC) problem with the “texture-patch transformation”, and resolved the MC problem from the perspective of the nuclear-norm minimization. Aiming at achieving a higher-quality reconstruction with small computational complexity in time, this report followed the general framework and the low-rank matrix completion idea in Yang et al. (2011) and Ma (2013), generalized the three-dimensional texture-patch transform, and settled the low-rank matrix completion problem from two other perspectives, the rank-r matrix approximation problem and low-rank factorization problem. Furthermore, this report applied two corresponding matrix completion algorithms, a gradient descent algorithm on the Grassman manifold (OptSpace) and the low-rank matrix fitting algorithm (LMaFit), to resolving the MC problem, and finally compared the performance of the proposed methods by conducting numerical experiments on simulated seismic data and the field data set.
Acknowledgments

I would first and foremost like to express my deepest gratitude to my advisor, Tapan Mukerji, for his insightful suggestions and patient guidance for the past two years.

This research was funded by the Stanford Center for Reservoir Forecasting. I am also deeply grateful to the Department of Energy Resources Engineering for providing me the tuition of these two years’ master study.

I would like to sincerely thank Professor Jianwei Ma for his generous providing the North Sea Marine data set. Also, I would like to thank Dongzhao Li, who helped me a lot with the knowledge and the technical problems in the field of seismic modeling. This research would be impossible without their support.

I would like to thank all my colleagues who helped me out with valuable suggestions in my research group Stanford Center for Reservoir Forecasting. Especially I would like to thank Ogy Grujic, Peipei Li, and Yongduk Shin for their great help and support during my study life at Stanford.

I wish to express many thanks to my friends and family for their encouragement and comfort. Many thanks to my roommates, Yuran Zhang and Weiwei Shen. It’s you who helped create a sweet and warm “home” at Stanford so that I enjoyed my sufficient and happy study life with you together during my last two quarters at Stanford.
Contents

Abstract ......................................................................................................................... v
Acknowledgments .......................................................................................................... vii
Contents ....................................................................................................................... ix
List of Tables .................................................................................................................. xi
List of Figures ................................................................................................................ xiii
1. Introduction .............................................................................................................. 1
   1.1. Background ........................................................................................................... 1
   1.2. Literature Review ............................................................................................... 1
   1.3. Scope of this Work .............................................................................................. 3
   1.4. Dissertation Outline .......................................................................................... 3
2. Methodology and Algorithms ................................................................................... 5
   2.1. Motivation and Pipeline of the Method .............................................................. 5
   2.2. Generalized Three-dimensional Texture-Patch Transform ................................ 7
   2.3. Matrix Completion Theory ............................................................................... 9
        2.3.1. Exact low-rank matrix completion ......................................................... 10
        2.3.2. Approximately low-rank matrix completion ........................................ 11
        2.3.3. Matrix Completion Algorithms ............................................................. 12
   2.4. OptSpace Algorithm ......................................................................................... 12
        2.4.1. Trimming ................................................................................................. 13
        2.4.2. Rank r estimation .................................................................................... 14
        2.4.3. Rank- projection .................................................................................... 14
   2.5. LMaFit Algorithm ............................................................................................. 15
        2.5.1. Nonlinear Gauss-Seidel (GS) Scheme .................................................... 16
        2.5.2. Nonlinear successive over-relaxation (SOR) Scheme............................ 16
3. Simulated Pre-stack Seismic Data Example ............................................................ 19
   3.1. Construction of Pre-stack Seismic Volume ....................................................... 19
   3.2. Parameter Settings ............................................................................................ 23
   3.3. Results Comparison .......................................................................................... 24
        3.3.1. 50% missing traces case ...................................................................... 24
        3.3.2. 70% missing traces case ...................................................................... 31
4. Case Study: North Sea Marine Data Set .................................................................. 33
   4.1. Parameter Settings ............................................................................................ 33
   4.2. Results Comparison .......................................................................................... 33
        4.2.1. 50% missing traces case ...................................................................... 33
        4.2.2. 70% missing traces case ...................................................................... 38
5. Conclusion ................................................................................................................ 39
References ..................................................................................................................... 41
A. Reconstruction Results of Simulated Pre-stack Seismic Data in 70% Missing Traces Case ................................................................. 43
B. Reconstruction Results of North Sea Marine Data in 70% Missing Traces Case. 51
List of Tables

Table 1: Simulated seismic volume: Comparison in S/N and computational time among OptSpace algorithm with rank 6, OptSpace algorithm with rank 20, and LMaFit algorithm in the two cases of 50% and 70% missing traces. ................................................................. 31
Table 2: North Sea Marine seismic volume: Comparison in S/N and computational time among OptSpace algorithm with rank 6, OptSpace algorithm with rank 20, and LMaFit algorithm in the two cases of 50% and 70% missing traces. ................................................................. 38
List of Figures

Figure 1: The flowchart of the three-dimensional (3D) seismic data reconstruction via low-rank matrix completion ........................................................................................................ 6
Figure 2: Illustration of the three-dimensional “texture-patch transform” that transforms the volume into the matrix. The marker “a” and “b” means first scanning entries along the “a” route, and then go to the next step along the “b” route, and then again follow the “a” route and so forth. Remark: The hidden sub-block in the left-behind corner of the volume is in white. ...................................................................................................................... 8
Figure 3: A layered velocity model with the velocity of 2000 m/s in the upper layer, 2800 m/s in the middle layer, and 3200 m/s in the lower layer. ................................................................. 20
Figure 4: The first common source gather simulated on the velocity model with the source position at (128/2)th grid cell and the receiver positions at the first 128 grid cells on the ground................................................................. 21
Figure 5: The illustration of the translation of source and receiver positions and the formation of pre-stack seismic volume with common source gather. The red frame represents the first simulated region corresponding to the first set of source and receiver positions; the red “x” represents the corresponding source position. Blue represents the 65th set, and green represents the last 128th set ........................................................................................................ 22
Figure 6: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 50% missing traces (2nd column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 6 (3rd column) (S/N = 85.27) ........................................................................................................ 25
Figure 7: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 50% missing traces (2nd column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 20 (3rd column) (S/N = 84.78) ........................................................................................................ 26
Figure 8: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 50% missing traces (2nd column), and the reconstructed seismic volume by the LMaFit algorithm (3rd column) (S/N = 204.27)...

Figure 9: (i) The central slice of t-x cross-section seismogram of the original simulated seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 50% missing traces.

Figure 10: 50% missing traces case: All the left-hand figures are the same slices with Figure 9 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error between the reconstruction and original simulated seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 85.27); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 84.78); (v) and (vi) LMaFit (S/N = 204.27).

Figure 11: Comparison of an original trace in the simulated seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.

Figure 12: Part map view of the uniformly random sampling pattern with 50% missing traces in the case study of North Sea Marine data set.

Figure 13: (i) The central slice of t-x cross section of the original North Sea Marine seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 50% missing traces; (iii), (iv), and (v) are the corresponding slices of the reconstructed seismic volume by OptSpace with rank 6 (S/N = 25.92), OptSpace with rank 20 (S/N = 38.17), and LMaFit (46.11), respectively.

Figure 14: (i) Part plot of the central slice of t-x cross-section seismogram of the original North Sea Marine seismic volume; (ii) The corresponding part of the corrupted seismic volume with 50% missing traces.

Figure 15: 50% missing traces case: All the left-hand figures are the same part with Figure 14 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error of that part between the reconstruction and original North Sea Marine seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 25.92); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 38.17); (v) and (vi) LMaFit (S/N = 46.11).
Figure 16: Comparison of an original trace in the North Sea Marine seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.

Figure 17: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 70% missing traces (2nd column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 6 (3rd column) (S/N = 77.56).

Figure 18: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 70% missing traces (2nd column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 20 (3rd column) (S/N = 76.09).

Figure 19: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 70% missing traces (2nd column), and the reconstructed seismic volume by the LMaFit algorithm (3rd column) (S/N = 174.69).

Figure 20: (i) The central slice of t-x cross-section seismogram of the original simulated seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 70% missing traces.

Figure 21: 70% missing traces case: All the left-hand figures are the same slices with Figure 9 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error between the reconstruction and original simulated seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 77.56); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 76.09); (v) and (vi) LMaFit (S/N = 174.69).

Figure 22: Comparison of an original trace in the simulated seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.

Figure 23: Part map view of the uniformly random sampling pattern with 70% missing traces in the case study of North Sea Marine data set.

Figure 24: (i) The central slice of t-x cross section of the original North Sea Marine seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 70% missing traces; (iii), (iv), and (v) are the corresponding slice of the reconstructed seismic
volume by OptSpace with rank 6 (S/N = 18.2), OptSpace with rank 20 (S/N = 26.02), and LMaFit (S/N = 24.94), respectively.

Figure 25: (i) Part plot of the central slice of t-x cross-section seismogram of the original North Sea Marine seismic volume; (ii) The corresponding part of the corrupted seismic volume with 70% missing traces.

Figure 26: 70% missing traces case: All the left-hand figures are the same part with Figure 14 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error of that part between the reconstruction and original North Sea Marine seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 18.2); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 26.02); (v) and (vi) LMaFit (S/N = 24.94).

Figure 27: Comparison of an original trace in the North Sea Marine seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 70% missing traces case.
Chapter 1

1. Introduction

1.1. Background
In seismic survey, one of the common issues is that seismic data are sampled with random missing traces on a uniformed grid due to physical reasons such as empty or corrupted traces, and earth’s surface obstacles. Interpolation of missing traces or seismic data reconstruction is one of the essential steps prior to subsequent seismic processing steps, such as pre-stack migration, resolution improvement, and amplitude versus offset analysis.

1.2. Literature Review
Different approaches have been proposed to solve this seismic data reconstruction problem. Originally, Ronen (1987) proposed the wave-equation methods, which is to use the simulation of wave propagation to recover seismic data. With the flourishing development of computer science, many signal-processing methods have been proposed in succession. Most of these methods require transforming data into other spaces with certain special properties, such as the Radon transform (Kabir and Verschuur, 1995), Fourier transform (Abma and Kabir, 2006; Zwartjes and Sacchi, 2007; Naghizadeh and Innanen, 2011), curvelet transform (Herrmann and Hennenfent, 2008; Shahidi et al., 2012), and shearlet transform (Hauser and Ma, 2012). Another popular method is prediction filters (Spitz, 1991; Crawley et al., 1999; Porsani, 1999), where observed seismic data are used to build prediction filters and the filters are then applied to interpolate missing traces.

Recently, rank-reduction methods (Freire and Ulrych, 1988; Trickett, 2003; Trickett and Burroughs, 2009, Trickett et al., 2010, Oropeza and Sacchi, 2011) are extensively applied to the seismic data reconstruction. The underlying motivation of these methods is that the intact seismic data is of low rank, and the missing traces increase the rank of data.
Consequently, rank reduction is proposed as an approach to recover missing traces. Particularly, Trickett et al. (2010) performed the truncated singular-value decomposition on constant-frequency slices. Oropeza and Sacchi (2011) recovered three-dimensional seismic data by presenting multichannel singular spectrum analysis (MSSA), which organizes spatial data into a block Hankel matrix and then formulates the seismic data reconstruction problem as the rank reduction problem of Hankel matrix. The MSSA is generalized from singular spectrum analysis (SSA), which is the corresponding method applicable to two-dimensional seismic data reconstruction.

Similarly, Yang et al. (2012) reduced two-dimensional seismic data reconstruction problem into low-rank matrix completion (MC) problem by developing the so-called “texture-patch transformation” which converts seismic data to a low-rank matrix or an approximate low-rank matrix as the input of matrix completion problem. Yang et al. (2012) adopted two matrix-completion algorithms -- accelerated proximal gradient method (APG) and low-rank matrix fitting (LMaFit) to complete the low-rank matrix, and recovered the seismic data by “texture-patch” inverse transforming the completed low-rank matrix to the original space. Compared with Oropeza and Sacchi (2011)’s singular spectrum analysis (SSA), Yang et al. (2012)’s method costs much less time and gives reasonable result especially when the original seismic data is of complex structure.

Ma (2013) extended Yang et al. (2011)’s low-rank matrix completion methods to three-dimensional seismic data scenario. Ma (2013) designed the three-dimensional “texture-patch transform” that converts three-dimensional sub-blocks into column vectors that constitute a low-rank matrix or an approximately low-rank matrix. Ma (2013) modified the existing approximate message passing (AMP) algorithm and applied it and low-rank matrix fitting (LMaFit) algorithm to matrix completion. In comparison of one of the traditional Fourier transform-based methods, projection onto convex sets (POCS), LMaFit shows a super fast performance and AMP achieves a robust and higher-quality reconstruction.
1.3. Scope of this Work

This work is focused on three-dimensional seismic data reconstruction with randomly missing data on a regular grid. Aimed at achieving a higher-quality reconstruction with small computational complexity in time, this work followed the general framework and the low-rank matrix completion idea of Yang et al. (2011) and Ma (2013), generalized the three-dimensional texture-patch transform, and selected the LMaFit and another excellent matrix completion algorithm OptSpace to apply on the matrix completion problem.

1.4. Dissertation Outline

This dissertation is organized as follows. Chapter 2 first presents the motivation and pipeline of the method, and then elaborates the theory and algorithms of this work, including the “generalized three-dimensional texture-patch transform”, matrix completion theory, and the two matrix completion algorithms OptSpace and LMaFit. Chapter 3 presents the numerical experiments of the proposed methods on simulated seismic data. Chapter 4 shows the performance of the proposed methods tested on the field data, North Sea marine data set. Finally, we will wrap up this work with a brief conclusion.
Chapter 2

2. Methodology and Algorithms

2.1. Motivation and Pipeline of the Method

The object of this work is a three-dimensional (3D) seismic volume with randomly missing traces on a regular grid. If simply scanning the volume slice by slice to form a matrix, there will exist columns whose elements are all zeros in the matrix, corresponding to the missing traces. In this scenario, matrix completion methods cannot be applicable, because matrix completion problem requires that all the rows and columns of the input matrix should have at least one non-zero element. Thus, we need to design a transform that converts the seismic volume into a matrix that satisfies this requirement of the matrix completion theory.

Three-dimensional texture-patch transform is such a transform that can convert the seismic volume to a matrix that meets the requirement of at least one non-zero element in each column. The transform first divides the three-dimensional seismic volume into a series of sub-blocks with the same appropriate size, and then scans the entries in each sub-block in the same ordering to form a column vector, and finally assigns these column vectors as the columns of a matrix. As long as there are no missing traces successively occurring in a region of the same length and width of sub-block, all the columns in the matrix will have at least one non-zero element after applying the three-dimensional texture-patch transform to the seismic volume.

Besides, three-dimensional texture-patch transform could convert the seismic volume into a potential (approximately) low-rank matrix. As a seismic survey is conducted in a continuous region, this nature guarantees that even though the subsurface geologic structure is very complex, intrinsically, only a limited set of basic textures constitutes the seismic record. The number of the basic textures is proportional to the complexity of the subsurface geologic structure. By setting the size of sub-block so appropriate as to be
able to involve local patterns and features in the seismic record, such as events, each sub-block will be a linear combination of the limited set of basic textures of the seismic record. Consequently, the columns of the new matrix will be highly linear-correlated so that the matrix is of low rank.

However, missing traces are shifted to the zero elements in the matrix after the texture-patch transform, and missing traces increase the rank of the seismic data. Thus, the three-dimensional seismic data reconstruction is converted to the low-rank matrix completion problem. After completing the missing entries of the underlying low-rank matrix by applying the matrix completion algorithm, we conducted the three-dimensional texture-patch inverse-transform to the completed matrix to get the completed three-dimensional seismic volume.

The flowchart in Figure 1 clearly illustrates every step in this method.

![Flowchart of 3D seismic data reconstruction](chart.png)

Figure 1: The flowchart of the three-dimensional (3D) seismic data reconstruction via low-rank matrix completion
2.2. Generalized Three-dimensional Texture-Patch Transform

In this report, we generalized the definition of three-dimensional texture-patch transform (Ma, 2013) as follows:

Given the seismic volume (i.e. tensor) \( X \in \mathbb{R}^{m \times n \times q} \) and the size of the sub-block \( a \times b \times c \) with the premise that \( m \) is divisible by \( a \), \( n \) is divisible by \( b \), and \( q \) is divisible by \( c \), we can divide the volume \( X \) into \( mnq/\text{abc} \) sub-blocks of the size \( a \times b \times c \). Denote each sub-block as \( B_i \in \mathbb{R}^{a \times b \times c} \), and then the set of sub-blocks is \( \{ B_i \in \mathbb{R}^{a \times b \times c}, i = 1, 2, \ldots, mnq/\text{abc} \} \). Each sub-block is then rearranged into a column vector with the same ordering. Denote each column vector as \( v_i \in \mathbb{R}^{abc \times 1} \), and then the new matrix \( M \in \mathbb{R}^{abc \times mnq/(abc)} \) is defined as:

\[
M = [v_1; v_2; \ldots; v_{mnq/(abc)}] 
\]

We denote this generalized three-dimensional texture-patch transform as \( \Psi : X \in \mathbb{R}^{m \times n \times q} \rightarrow M \in \mathbb{R}^{abc \times mnq/(abc)} \). The inverse mapping of the three-dimensional texture-patch transformation is defined as the texture-patch inverse-transform, denoted as \( \Psi^{-1} \), converting the matrix \( M \in \mathbb{R}^{abc \times mnq/(abc)} \) to the tensor \( X \in \mathbb{R}^{m \times n \times q} \).

The ordering of scanning the divided sub-blocks in the volume is not important, as long as the texture-patch transform and the inverse-transform share the same ordering. The ordering of rearranging the sub-block into column vector is neither important, as long as the rearrangement of each sub-block follows the consistent ordering. In this report, for both of these two orderings, we adopted the “column-first, row-second, and page-last scanning strategy”, which means we first scan or rearrange the entries by column on the first page, and then sequentially go to each of the remaining pages to scan entries in the same fashion. Figure 2 gives a simple illustration of the transform with this ordering in the case that the seismic volume is \( 4 \times 4 \times 4 \) and the sub-block size is assigned \( 2 \times 2 \times 2 \).
Figure 2: Illustration of the three-dimensional “texture-patch transform” that transforms the volume into the matrix. The marker “a” and “b” means first scanning entries along the “a” route, and then go to the next step along the “b” route, and then again follow the “a” route and so forth. Remark: The hidden sub-block in the left-behind corner of the volume is in white.

The size of the sub-block is determined by the size of the basic patterns or textures of the seismic volume. The larger the basic patterns’ size is, the larger the sub-block should be defined. However, before the seismic volume is recovered, we hardly foresee the basic patterns’ size. In this report, we generalized the sub-block’s shape to be a rectangular instead of a cube defined in Ma (2013). This generalization gives users more flexible space to try the sub-block’s size to achieve a better-quality performance when with the unknown size of the basic patterns.

Besides, the low-rank property of the new matrix could be perceived in the intuitive way that each sub-block is a linear combination of basic textures as mentioned in Chapter 2.1.

In order to tailor the matrix as the input of the matrix completion problem, we need guarantee that all the columns and rows should have at least one non-zero element.
Equivalently, each sub-block should have at least one non-zero entry. That is to say, missing traces could not be allowed to successively occur in the region that involves a sub-block.

2.3. Matrix Completion Theory

Through the texture-patch transform in Chapter 2.2, the seismic volume with missing traces is transformed into the low-rank matrix or the approximately low-rank matrix \( M \in \mathbb{R}^{abc \times mnq/(abc)} \) with some missing entries. In order to recover this (approximately) low-rank matrix, we need to guarantee that this matrix satisfies the following three preliminary requirements of the matrix completion (Candes and Recht, 2008).

First, the number of the observed entries could not be too small so that they can provide sufficient information to infer the missing entries. Since the degree of freedom of a rank \( r \) matrix of dimension \( m \times n \) is \( (m + n - r) \), without the same number of observations is impossible to recover them.

Second, as mentioned before, it must be guaranteed that every row and column should have at least one observation (i.e. non-zero entry). If all the entries in one column are missing, there is no information at all about this column (i.e. column-representing sub-block). Because every sub-block is divided relatively independent from the seismic volume, we cannot infer any elements in this column with no individual information about this sub-block at all, and then cannot recover the matrix.

Third, the entries should be sufficiently uniformly distributed. Otherwise, the row or column with far more observations has far more information so that that row or column could be recovered more accurately, and vice versa.

Based on the above three general requirements, there are two cases of matrix completion that can be implemented. One is the case of exact low-rank matrix completion, where we assume that the matrix has exact low rank and the observations are known exactly without any noise. The other case is approximately low-rank matrix completion problem,
where the observations are with some noise or the original matrix is approximately low rank.

2.3.1. **Exact low-rank matrix completion**

In the first case, recovering the exact low-rank matrix is equivalent to solving the optimization problem:

\[
\begin{align*}
\min \quad & \text{rank}(X) \\
\text{s.t.} \quad & X_{ij} = M_{ij}, \quad \text{for } (i, j) \in \Omega
\end{align*}
\]

where \( \Omega \) is the set of locations corresponding to the observed entries, and \( X \) is the decision variable, which is the matrix as the recovery of matrix \( M \).

By introducing the projection operator \( \mathcal{P}_\Omega(\cdot) \) defined as follows,

\[
\mathcal{P}_\Omega(M_{ij}) = \begin{cases} M_{ij}, & \text{for } (i, j) \in \Omega \\
0, & \text{otherwise} \end{cases}
\]

the optimization problem (2.2) could be rewritten as (2.4):

\[
\begin{align*}
\min \quad & \text{rank}(X) \\
\text{s.t.} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M)
\end{align*}
\]

However, this optimization problem is NP-hard. One would consider the convex relaxation of \( \text{rank}(X) \) as the substitute of \( \text{rank}(X) \).

The nuclear norm \( \|X\|_* \) defined as (2.6) is the convex relaxation of the rank \( \text{rank}(X) \).

\[
\begin{align*}
\text{rank}(X) = & \sum_{k=1}^{\min(m,n)} \mathcal{I}(\sigma_k(X)) \\
\|X\|_* = & \sum_{k=1}^{\min(m,n)} \sigma_k(X)
\end{align*}
\]
where \( \mathcal{I}(\cdot) \) is the indicator function, and \( \{\sigma_k(X)\}_{1 \leq k \leq \min(m,n)} \) are the singular values of \( X \).

Therefore, the optimization problem (2.4) is substituted as the convex optimization problem (2.7):

\[
\begin{align*}
\min \quad & \|X\|_s \\
\text{s.t.} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \\
\end{align*}
\tag{2.7}
\]

\subsection*{2.3.2. Approximately low-rank matrix completion}

In the second case of approximately low-rank matrix completion problem, the condition \( \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \) must be relaxed. This results in the optimization problem (2.8):

\[
\begin{align*}
\min \quad & \|X\|_s \\
\text{s.t.} \quad & \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F \leq \Theta \\
\end{align*}
\tag{2.8}
\]

where \( \Theta \) is a certain positive number, and \( \|\cdot\|_F \) is the Frobenius norm of a matrix, defined as

\[
\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}
\tag{2.9}
\]

(2.8) can be converted to the problem (2.10) by the method of Lagrange multipliers:

\[
\begin{align*}
\min \quad & \mu \|X\|_s + \frac{1}{2} \|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(M)\|_F^2 \\
\end{align*}
\tag{2.10}
\]

Candes and Recht (2008) presented a rigorous statement of the matrix completion theorem that with a high probability, the matrix \( M \in \mathbb{R}^{m \times n} \) could be recovered or approximately recovered by a low-rank matrix \( X \) through solving (2.7) or (2.10) when satisfying some conditions including the bound of the observation number \( |\Omega| \geq \mathcal{O}(r \max(m,n)^{6/5} \log(max(m,n))) \).
2.3.3. Matrix Completion Algorithms

In practice, many algorithms have been developed to solve the convex relaxation problems (2.7) or (2.10). Cai et al. (2010) proposed the singular value thresholding (SVT) algorithm, which is an iterative algorithm for solving the (2.7) in the exact low-rank matrix completion problem. The efficient algorithms for the second case of approximately low-rank matrix completion problem (2.10) include the accelerated proximal gradient (APG) algorithm (Toh and Yun, 2009), fixed point continuation with approximate SVD (FPCA) algorithm (Ma et al., 2009), and Soft-Impute algorithm (Mazumder and Hastie, 2010).

However, both these two nuclear-norm minimization problem (2.7) and (2.10) require computing a series of singular value decompositions, which is of increasing complexity in time when the matrix sizes and ranks increase. Thus, this report introduces two novel algorithms, OptSpace algorithm (Keshavan and Oh, 2009; Keshavan et al., 2009) and LMaFit algorithm (Wen et al., 2012) to settle the matrix completion problem from other perspectives.

2.4. OptSpace Algorithm

According to Lee et al. (2010) and Jain and Meka (2010), the non-convex optimization problem (2.4) can be converted to the rank-r matrix approximation problem:

$$\min \| P_{\Omega}(X) - P_{\Omega}(M) \|_F$$

s.t. \( \text{rank}(X) \leq r \) ..................................... (2.11)

Both the two cases of the exact and approximately low-rank matrix completion can be settled through solving (2.11).

Keshavan and Oh (2009) proved that the matrix \( M \in \mathbb{R}^{m \times n} \) could be recovered or approximately recovered by a low-rank matrix \( X \) through solving (2.11) with an improvement of the bound requirement of the observation number from \( |\Omega| \geq O(r \max(m, n)^{6/5} \log(\max(m, n))) \) in Candes and Recht (2008) to
That is to say, (approximately) low-rank matrix can be completed with fewer observations though solving (2.11).

Keshavan and Oh (2009) also proposed the OptSpace algorithm, which is the efficient algorithm to solve the matrix completion problem (2.11). OptSpace algorithm is based on sparse singular value decomposition followed by local manifold optimization (Keshavan et al., 2009). The basic idea is to minimize the function $F : \mathbb{R}^{m \times r} \times \mathbb{R}^{n \times r} \to \mathbb{R}$:

$$
\min_{X,Y} F(X,Y) = \min_{S \in \mathbb{R}^{r \times r}} \| \mathcal{P}_\Omega(M) - \mathcal{P}_\Omega(XSY^T) \|_F^2 \quad ............ (2.12)
$$

where $X$ and $Y$ are orthogonal matrices.

Since $F$ is non-convex function, (2.12) is NP-hard. Thus, Keshavan and Oh (2009) suggested that first estimating the rank $r$, and then performing the “rank-$r$ projection” of the trimmed matrix $\mathcal{P}_\Omega(M)$ defined in (2.13) to get the initial guess of $X, S,$ and $Y$ in (2.12), and finally solving (2.12) by gradient descent with the initial guess. The matrix multiplication $XSY^T$ consisting of the optimal $X, S,$ and $Y$ is the reconstruction of $M$ through OptSpace algorithm.

2.4.1. Trimming

By defining the number of observations in a row $i$ or a column $j$ as the degree of this row or column, respectively denoted as $d_{row}(i)$ and $d_{column}(j)$, the average degree of all the rows in the matrix is $|\Omega|/m$ and that of all the columns in the matrix is $|\Omega|/n$. If the number of observations in a row is more than twice of the average degree of row (i.e. $2|\Omega|/m$), this row is defined as over-represented. Similarly, if the number of observations in a column is more than $2|\Omega|/n$, this column is defined as over-represented. Trimming is a process that sets all the elements in the over-represented rows and columns to 0, defined in the mathematical expression (2.13):

$$
\mathcal{P}_\Omega(M)_{ij} = \begin{cases} 
0, & \text{if } d_{row} > 2|\Omega|/m \text{ or } d_{column} > 2|\Omega|/n \\
\mathcal{P}_\Omega(M)_{ij}, & \text{otherwise}
\end{cases} \quad ............ (2.13)
$$
The reason why trimming is needed before singular value decomposition is to avoid the existence of singular vectors which are the highly concentrated on the over-represented rows and columns and provide no valuable information about the missing entries of $M$. Furthermore, the information that has been trimmed will be recalled to join in the gradient descent algorithm when solving (2.12) after initialization. Thus, we didn’t miss any information overall.

2.4.2. Rank $r$ estimation

As for estimating the rank $r$, Keshavan et al. (2009) proved that when the rank $r$ matrix $M$ with dimension $m \times n$ has bounded condition number $\mathcal{K}$, then there exists a constant $C(\mathcal{K})$ s.t. if $|\Omega|/\sqrt{mn} > C(\mathcal{K})$, then (2.14) gives correct estimates of rank $r$, with high probability.

$$
\hat{r} = \arg \min_{1 \leq k \leq \min(m,n)} \frac{\sigma_{k+1} + \sigma_1 \sqrt{k \sqrt{mn} / |\Omega|}}{\sigma_k}
$$

\hspace{1cm} \text{......... (2.14)}

where $\{\sigma_k(X)\}_{1 \leq k \leq \min(m,n)}$ are the singular values of the trimmed matrix $\mathcal{P}_r(M)$.

The “OptSpace” Matlab function realizes the rank estimation procedure and uses the rank estimator when the input rank $r$ is given as null. However, when the matrix $M$ cannot satisfy the above conditions, the function still gives the rank estimation but it might be far less accurate. When the performance with function-estimating rank is too bad, it might because the matrix doesn’t satisfy the requirements of the above rank estimation theorem. Thus, in this case, it is necessary to assign a different value from the automatically estimated rank to the input $r$ of the “OptSpace” function by users.

2.4.3. Rank-$r$ projection

The rank-$r$ projection of a matrix is the sparse singular value decomposition of the matrix with a rescaling factor. In our problem, the rank-$r$ projection of $\mathcal{P}_r(M)$ is defined as (2.15):
where $\{\sigma_k(X)\}_{1 \leq k \leq r}$ are the first $r$ singular values of the matrix $\mathcal{P}_\Omega(M)$ with descending order, $\{u_i\}_{1 \leq k \leq r}$ and $\{v_i\}_{1 \leq k \leq r}$ are respectively the corresponding left-singular vectors and right-singular vectors.

The rescaling factor $\frac{mn}{|\Omega|}$ is designed to compensate the smaller average size of the non-zero entries of $\mathcal{P}_\Omega(M)$ with respect to $M$. (Keshavan et al., 2009)

(2.15) could be rewritten as (2.16):

$$\text{Proj}_r(\mathcal{P}_\Omega(M)) = X_0 S_0 Y_0 \quad \text{............... (2.16)}$$

where $X_0 = \sqrt{m}[u_1, ..., u_r]$, $Y_0 = \sqrt{n}[v_1, ..., v_r]$, and $S_0 = (\sqrt{mn}/|\Omega|)\text{diag}(\sigma_1, ..., \sigma_r)$.

The above $X_0$, $S_0$, and $Y_0$ are taken as the initial guess of $X$, $S$, and $Y$ in (2.12).

2.5. LMaFit Algorithm

Wen et al. (2012) proposed a low-rank factorization model and created the low-rank matrix fitting (LMaFit) algorithm, which shows a dominant advantage in computational time over nuclear-norm minimization algorithms.

The motivation of this low-rank factorization model (2.17) is that any rank up-to-$r$ matrix $M$ of dimension $m \times n$ can be factorized as the matrix multiplication $M = XY$ where $X \in \mathbb{R}^{m \times r}$ and $Y \in \mathbb{R}^{r \times n}$.

$$\min \frac{1}{2} \|XY - Z\|_F^2 \quad \text{s.t.} \quad \mathcal{P}_\Omega(Z) = \mathcal{P}_\Omega(M) \quad \text{............... (2.17)}$$

Thus, we should estimate the initial rank of the matrix $M$ first. However, the initial rank need not be close to the real rank $r$ of the matrix $M$, because the algorithm allows a
strategy of dynamically adjusting the rank by gradually decreasing or increasing the initial rank to achieve a better solution to the (2.17). The decreasing rank strategy is preferable for the first case of exact low-rank matrix completion mentioned in Chapter 2.3.1, and the increasing rank strategy is suitable for the second case of approximately low-rank matrix completion mentioned in Chapter 2.3.2.

2.5.1. **Nonlinear Gauss-Seidel (GS) Scheme**

Traditionally, the nonlinear Gauss-Seidel (GS) scheme can be applied to this problem (2.17) by updating one of $X, Y$ or $Z$ while fixing the other two variables each time. The procedure follows the scheme shown as (2.18):

$$
X_{(i+1)} = Z_{(i)} Y_{(i)}^T (Y_{(i)} Y_{(i)}^T)^g \\
Y_{(i)} = (X_{(i+1)}^T X_{(i+1)})^g (X_{(i+1)}^T Z_{(i)}) \\
Z_{(i+1)} = X_{(i+1)} Y_{(i+1)} + \mathcal{P}_\Omega (M - X_{(i+1)} Y_{(i+1)}) \quad \text{……………(2.18)}
$$

where $X_{(i)}, Y_{(i)},$ and $Z_{(i)}$ represent the values of $X, Y$ and $Z$ in the $i^{th}$ iteration step, and $A^g$ represents the generalized inverse of the matrix $A$.

2.5.2. **Nonlinear successive over-relaxation (SOR) Scheme**

Based on this GS method, Wen et al. (2012) introduced the nonlinear successive over-relaxation (SOR) scheme with a strategy to dynamically adjust the relaxation weight to solve (2.17). The nonlinear successive over-relaxation (SOR) scheme is to apply an extrapolation to the GS method. To be specific, the new iterate is the weighted average of the previous iterate and the corresponding new GS iterate.

First, (2.17) can be transformed to the Lagrange function (2.19):

$$
\mathcal{L}(X, Y, Z, \Lambda) = \frac{1}{2} \| X Y - Z \|_F^2 - \Lambda \cdot \mathcal{P}_\Omega (Z - M) \quad \text{………………(2.19)}
$$

where $\Lambda \in \mathbb{R}^{m \times n}$ is the Lagrange multiplier, and it satisfies that $\Lambda = \mathcal{P}_\Omega (\Lambda)$.
By differentiating the Lagrange function and introducing the method of successive over-relaxation (SOR) based on the nonlinear Gauss-Seidel scheme, we get the new trial points expressed as (2.20) in each iteration:

\[
X_{(i+1)} = Z_{(i)} Y_{(i)}^T (Y_{(i)} Y_{(i)}^T)^g \\
X_{(i+1)} (\omega) = \omega X_{(i+1)} + (1 - \omega) X_{(i)} (\omega') \\
Y_{(i+1)} = (X_{(i+1)} (\omega) X_{(i+1)} (\omega))^g (X_{(i+1)} (\omega) Z_{(i)} (\omega'),
Y_{(i+1)} (\omega) = \omega Y_{(i+1)} + (1 - \omega) Y_{(i)} (\omega') \\
Z_{(i+1)} (\omega) = X_{(i+1)} (\omega) Y_{(i+1)} (\omega) + P_{\Omega} (M - X_{(i+1)} (\omega) Y_{(i+1)} (\omega)) \ldots (2.20)
\]

where \(\omega\) and \(\omega'\) are respectively the weight value corresponding to \((i + 1)\)th and \(i\)th iterates. \(X_{(i+1)} (\omega), Y_{(i+1)} (\omega),\) and \(Z_{(i+1)} (\omega)\) are the new \((i + 1)\)th iterates. \(X_{(i+1)}, Y_{(i+1)},\) and \(Z_{(i+1)}\) are the \((i + 1)\)th GS iterates. \(X_{(i)} (\omega'), Y_{(i)} (\omega'),\) and \(Z_{(i)} (\omega')\) are the \(i\)th iterates.
Chapter 3

3. Simulated Pre-stack Seismic Data Example

3.1. Construction of Pre-stack Seismic Volume

In order to test our method and compare the performance of the two algorithms in a noise-free scenario, we simulated a pre-stack seismic volume with common source gather as the object of this numerical experiment.

We adopted the acoustic finite-difference seismic modeling facility (Youzwishen and Margrave, 1999) to construct the simulated pre-stack seismic volume. This facility is a very flexible facility that can model the acoustic P-waves propagation in heterogeneous, isotropic media, based on the variable-velocity scalar wave equation. This seismic modeling facility is available in the “finitedif” of the CREWES Matlab toolbox (Margrave, 2000), which can be downloaded from the URL: http://www.crewes.org/ResearchLinks/FreeSoftware/. In this toolbox, both 2nd and 4th order Laplacians are available and Clayton-Engquist absorbing boundaries have been implemented. Sources and receivers could be placed in arbitrary locations.

First, we defined a velocity model on a depth of 128 and a length of 255 grid with a spatial grid size of 4m, with the help of “afd_vmodel” Matlab function in Margrave (2000). This “afd_vmodel” function fills in a corners-defined polygonal region in a velocity matrix with a constant velocity. By using the “afd_vmodel” function three times, we superimposed three layers with different velocities together to form the velocity model shown in Figure 3.

Then, we simulated the shot records on this created velocity model, with the help of the “afd_shotrec” Matlab function in Margrave (2000). We adopted the 2nd order Laplacian to create the synthetic. The time step is defined as 0.0008s, the time-sample rate is defined as 0.004, and the maximum record time is 1.
Figure 3: A layered velocity model with the velocity of 2000 m/s in the upper layer, 2800 m/s in the middle layer, and 3200 m/s in the lower layer.

Because we aimed at getting a pre-stack seismic volume with common source gather on this velocity model, we defined the first set of source position and receiver positions respectively at the \((128/2)^{th}\) grid cell and the first 128 grid cells on the ground, and defined the next set of source and receiver positions by moving the preceding set of the positions one grid to the right, and so forth until there is receiver reaching the right boundary. In this way, we collected 128 slices (or pages) of such a common source gather in total. By arranging these slices sequentially from front to rear, we got a pre-stack seismic volume with common source gather, which is what we desired. Figure 4 is the first common source gather with the source position at \((128/2)^{th}\) grid cell and receiver positions at the first 128 grid cells on the ground, and Figure 5 is an illustration of the translation of source and receiver positions.
From Figure 4, we can see that from top down, the three distinct waves are respectively the primary reflection off of the contact of the first layer and the second layer, the first multiple reflection off of the contact of the first layer and the second layer, and the reflection off of the contact of the second layer and the third layer.
Figure 5: The illustration of the translation of source and receiver positions and the formation of pre-stack seismic volume with common source gather. The red frame represents the first simulated region corresponding to the first set of source and receiver positions; the red “x” represents the corresponding source position. Blue represents the 65th set, and green represents the last 128th set.
3.2. Parameter Settings

The created pre-stack seismic volume in Chapter 3.1 is of dimension $251 \times 128 \times 128$ in respectively row, column, and page dimension. The row dimension represents the time sampling ($t$), and both the column and page dimensions represent the spatial sampling ($x$ and $y$). Because the three dimensions of the seismic volume should be respectively divisible by the corresponding dimensions of the defined sub-block, we cut off the data corresponding to the 251st time samples and retained the seismic volume which involves the first 250 time samples as our experiment object, and set the size of sub-block as $10 \times 8 \times 8$.

In order to differentiate between the original zero elements and the zero missing traces in the sampled seismic volume, we first added the minimum positive number ("eps" in Matlab) to every elements in the original seismic volume, and then uniformly sampled a proportion of traces as missing traces by setting them as zero.

In the OutSpace algorithm (Keshavan and Oh, 2009) which can be downloaded from the URL: http://web.engr.illinois.edu/~swoh/software/optspace/code.html, we first allowed the program to guess the rank, which is 6. Then, we set the initial rank as 20 for the purpose of comparing the effect of different initial rank estimations on the performance of OutSpace in the quality of reconstruction and the computational cost.

As for the LMaFit algorithm (Wen et al., 2012) which can be downloaded form the URL: http://lmafitblogs.rice.edu/, we adopted the increasing rank strategy by setting the parameter “opts.est_rank” as 2 (“opts.est_rank” = 1 is for the decreasing rank strategy), because the created seismic volume is unlikely to be of exactly low rank.

In order to examine the applicability of our algorithms, we first tested the algorithms on the simulated seismic volume with 50% randomly missing traces, and then with 70% randomly missing traces.

All the experiments were performed on MacBook Pro with 2.8 GHz Intel Core i5 Processor and 16 GB of RAM.
3.3. Results Comparison

We adopted the signal-to-noise ratio (S/N) defined in (3.1) as the evaluation of the reconstruction quality.

\[ S/N = 20 \times \log\left( \frac{\|X\|_F^2}{\|X - \hat{X}\|_F^2} \right) \]  
\[ \text{.......................... (3.1)} \]

where \( X \) is the original seismic volume, and \( \hat{X} \) is the reconstructed seismic volume.

3.3.1. 50% missing traces case

Figure 6 – Figure 8 show slices of three cross sections of the original simulated seismic volume, the corrupted seismic volume with 50% missing traces, and the reconstructed seismic volume by three algorithms, OptSpace with initial rank 6, OptSpace with initial rank 20, and LMaFit, respectively. LMaFit performs good enough on the t-x cross section, while OptSpace with initial rank 6 and OptSpace with initial rank 20 have a relatively rougher reconstruction on that cross section. LMaFit still performs smooth and precise reconstruction on t-y and x-y cross sections, while OptSpace with initial rank 6 and 20 have a coarse grid-like reconstruction on those two cross sections.

LMaFit achieves the highest S/N among the three algorithms, followed by OptSpace with initial rank 6. This phenomenon indicates that the rank 20 is not suitable for this case, where the underlying rank of the original matrix might be far less than 20. The rank-20 completed matrix over-represents the exact simulated seismic volume. Thus, OptSpace with initial rank 20 gets a slightly smaller S/N value than that in OptSpace with initial rank 6.
Figure 6: Slices of three cross sections of the original simulated seismic volume (1\textsuperscript{st} column), the corrupted seismic volume with 50\% missing traces (2\textsuperscript{nd} column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 6 (3\textsuperscript{rd} column) (S/N = 85.27).
Figure 7: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 50% missing traces (2nd column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 20 (3rd column) (S/N = 84.78).
Figure 8: Slices of three cross sections of the original simulated seismic volume (1st column), the corrupted seismic volume with 50% missing traces (2nd column), and the reconstructed seismic volume by the LMaFit algorithm (3rd column) (S/N = 204.27).
Figure 9 – Figure 10 display that the central slice of the t-x cross-section seismogram of the original simulated seismic volume, the corrupted seismic volume with 50%, and the reconstruction and the error between the reconstruction and original data by the three algorithms. In this way, we can see more clearly that LMaFit indeed achieves the best quality of construction, because both the error figures of the two OptSpace algorithms show more noise than that of LMaFit.

Figure 9: (i) The central slice of t-x cross-section seismogram of the original simulated seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 50% missing traces.
Figure 10: 50% missing traces case: All the left-hand figures are the same slices with Figure 9 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error between the reconstruction and original simulated seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 85.27); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 84.78); (v) and (vi) LMaFit (S/N = 204.27).
Figure 11: Comparison of an original trace in the simulated seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.
In order to make a closer observation, we randomly extracted a trace to compare the performance of the three algorithms with regard to the exact trace. Shown as Figure 11, though OptSpace performs satisfactory results, LMaFit does a perfect match with the original trace.

3.3.2. 70% missing traces case

For the case of 70% missing traces, we conducted the same process as in the 50% missing traces case. Please refer to the supplemental material Figure 17 ~ Figure 22 in the Appendix A. The results show that the quality of reconstructions decreases for all the three algorithms because of less known information (i.e. more missing data), but LMaFit still achieves the best performance.

The S/N and the computational time of the three algorithms in the cases of simulated seismic volume with 50% and 70% missing traces are summarized in Table 1.

Table 1: Simulated seismic volume: Comparison in S/N and computational time among OptSpace algorithm with rank 6, OptSpace algorithm with rank 20, and LMaFit algorithm in the two cases of 50% and 70% missing traces.

<table>
<thead>
<tr>
<th>Missing Rate</th>
<th>50%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OptSpace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 6</td>
<td>S/N</td>
<td>85.27</td>
</tr>
<tr>
<td></td>
<td>Time (s)</td>
<td>316.59</td>
</tr>
<tr>
<td>OptSpace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 20</td>
<td>S/N</td>
<td>84.78</td>
</tr>
<tr>
<td></td>
<td>Time (s)</td>
<td>3649.20</td>
</tr>
<tr>
<td>LMaFit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S/N</td>
<td>204.27</td>
</tr>
<tr>
<td></td>
<td>Time (s)</td>
<td>15.04</td>
</tr>
</tbody>
</table>

From Table 1, we can see that in both two missing traces case, LMaFit occupies the dominant advantage over OptSpace in both aspects of reconstruction quality and computational time.
Chapter 4

4. Case Study: North Sea Marine Data Set

4.1. Parameter Settings

The field data is the North Sea Marine pre-stack seismic volume, which is of $512 \times 512 \times 16$. We set the size of sub-block as $8 \times 8 \times 8$.

For OptSpace algorithm in this case study, the program estimates the initial rank as 2, which results in a blurred reconstruction. That might because that the matrix transformed from the real data set with noise does not have the bounded condition number $\mathcal{K}$ so that the requirements of the rank estimation theorem cannot be satisfied. Thus, we set the initial rank in OptSpace as 6 and 20 respectively, followed the same parameter settings of LMaFit in Chapter 3, and then conducted the same two cases of 50% and 70% missing traces, and finally evaluated the quality of reconstruction by the signal-to-noise ratio (S/N) as well.

4.2. Results Comparison

4.2.1. 50% missing traces case

Figure 12 shows a part map view of the uniformly random sampling with 50% missing traces.

Figure 12: Part map view of the uniformly random sampling pattern with 50% missing traces in the case study of North Sea Marine data set.
Figure 13 shows the reconstructions of the central slice of t-x cross section of the North Sea Marine seismic volume by the three algorithms. All the three algorithms achieved excellent performance. Figure 14 ~ Figure 15 display a part of central slice of t-x cross section in details. Compared with OptSpace with initial rank 6, OptSpace with initial rank 20 and LMaFit complete the reconstructions much closer to the real data. Especially, LMaFit achieves the highest S/N.

Figure 13: (i) The central slice of t-x cross section of the original North Sea Marine seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 50% missing traces; (iii), (iv), and (v) are the corresponding slices of the reconstructed seismic volume by OptSpace with rank 6 (S/N = 25.92), OptSpace with rank 20 (S/N = 38.17), and LMaFit (46.11), respectively.
Figure 14: (i) Part plot of the central slice of t-x cross-section seismogram of the original North Sea Marine seismic volume; (ii) The corresponding part of the corrupted seismic volume with 50% missing traces.
Figure 15: 50% missing traces case: All the left-hand figures are the same part with Figure 14 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error of that part between the reconstruction and original North Sea Marine seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 25.92); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 38.17); (v) and (vi) LMaFit (S/N = 46.11).
Figure 16: Comparison of an original trace in the North Sea Marine seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.
Figure 16 shows the recovery of a randomly chosen trace by three algorithms. In comparison of OptSpace with rank 6 and 20, LMaFit achieves a better match with the original trace.

4.2.2. 70% missing traces case

In the case of 70% missing traces, we conducted the same process as in the 50% missing traces case. Please refer to the supplemental material Figure 23 ~ Figure 27 in the Appendix B. The results show that OptSpace with rank 20 achieved the best performance with the highest S/N, which indicates that the rank 20 is much closer to the real rank of the matrix than the rank 6 for this case. Figure 27 shows that OptSpace with rank 20 does a good match with the real trace, while LMaFit always overestimated or exaggerated the amplitude of the trace, resulting in a decrease in the quality of reconstruction.

The S/N and the computational time of the three algorithms in the cases of North Sea Marine seismic volume with 50% and 70% missing traces are summarized in Table 2.

Table 2: North Sea Marine seismic volume: Comparison in S/N and computational time among OptSpace algorithm with rank 6, OptSpace algorithm with rank 20, and LMaFit algorithm in the two cases of 50% and 70% missing traces.

<table>
<thead>
<tr>
<th>Missing Rate</th>
<th>50%</th>
<th>70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OptSpace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 6</td>
<td>S/N</td>
<td>25.92</td>
</tr>
<tr>
<td></td>
<td>Time (s)</td>
<td>316.83</td>
</tr>
<tr>
<td>OptSpace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r = 20</td>
<td>S/N</td>
<td>38.17</td>
</tr>
<tr>
<td></td>
<td>Time (s)</td>
<td>3795.50</td>
</tr>
<tr>
<td>LMaFit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/N</td>
<td>46.11</td>
<td>24.94</td>
</tr>
<tr>
<td>Time (s)</td>
<td>10.34</td>
<td>19.00</td>
</tr>
</tbody>
</table>

From Table 2, we can see that in the 50% missing traces case, LMaFit occupies the advantage over OptSpace in both aspects of reconstruction quality and computational time. However, in 70% missing traces case, even though LMaFit still has a speed advantage, LMaFit lost to OptSpace with rank 20 in the quality of reconstruction. Thus, the results in the 70% missing trace case confirm that the OptSpace with a good estimation of rank could complete a (approximately) low-rank matrix with fewer observations more accurately (Keshavan and Oh, 2009).
Chapter 5

5. Conclusion

In this report, we followed the framework and the low-rank matrix completion idea of Yang et al. (2001) and Ma (2013), generalized the three-dimensional texture-patch transform, and compared the performance of two matrix completion algorithms OptSpace with different parameter settings and LMaFit in simulated and field seismic data sets with different proportion of missing traces. LMaFit is very fast and generally achieves the reconstruction of good quality. However, OptSpace with a good estimation of rank catches up and even exceeds the performance of LMaFit when the matrix with fewer observations, for example, in the 70% missing traces case.
References


Appendix A

A. Reconstruction Results of Simulated Pre-stack Seismic Data in 70% Missing Traces Case
Figure 17: Slices of three cross sections of the original simulated seismic volume (1\textsuperscript{st} column), the corrupted seismic volume with 70\% missing traces (2\textsuperscript{nd} column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 6 (3\textsuperscript{rd} column) (S/N = 77.56).
Figure 18: Slices of three cross sections of the original simulated seismic volume (1\textsuperscript{st} column), the corrupted seismic volume with 70\% missing traces (2\textsuperscript{nd} column), and the reconstructed seismic volume by the OptSpace algorithm with the initial estimated rank as 20 (3\textsuperscript{rd} column) (S/N = 76.09).
Figure 19: Slices of three cross sections of the original simulated seismic volume (1\textsuperscript{st} column), the corrupted seismic volume with 70\% missing traces (2\textsuperscript{nd} column), and the reconstructed seismic volume by the LMaFit algorithm (3\textsuperscript{rd} column) (S/N = 174.69).
Figure 20: (i) The central slice of t-x cross-section seismogram of the original simulated seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 70% missing traces.
Figure 21: 70% missing traces case: All the left-hand figures are the same slices with Figure 9 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error between the reconstruction and original simulated seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 77.56); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 76.09); (v) and (vi) LMaFit (S/N = 174.69).
Figure 22: Comparison of an original trace in the simulated seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 50% missing traces case.
Appendix B

B. Reconstruction Results of North Sea Marine Data in 70% Missing Traces Case

Figure 23: Part map view of the uniformly random sampling pattern with 70% missing traces in the case study of North Sea Marine data set.
Figure 24: (i) The central slice of t-x cross section of the original North Sea Marine seismic volume; (ii) The corresponding slice of the corrupted seismic volume with 70% missing traces; (iii), (iv), and (v) are the corresponding slice of the reconstructed seismic volume by OptSpace with rank 6 (S/N = 18.2), OptSpace with rank 20 (S/N = 26.02), and LMaFit (S/N = 24.94), respectively.
Figure 25: (i) Part plot of the central slice of t-x cross-section seismogram of the original North Sea Marine seismic volume; (ii) The corresponding part of the corrupted seismic volume with 70% missing traces.
Figure 26: 70% missing traces case: All the left-hand figures are the same part with Figure 14 of the seismogram of the reconstructed seismic volume and all the right-hand figures are the error of that part between the reconstruction and original North Sea Marine seismic data. (i) and (ii) OptSpace with initial estimated rank = 6 (S/N = 18.2); (iii) and (iv) OptSpace with initial estimated rank = 20 (S/N = 26.02); (v) and (vi) LMaFit (S/N = 24.94).
Figure 27: Comparison of an original trace in the North Sea Marine seismic volume and the corresponding predicted trace by (i) OptSpace with rank 6, (ii) OptSpace with rank 20, and (iii) LMaFit in the 70% missing traces case.