

Important Notice

This copy may be used only for the purposes of research and private study, and any use of the copy for a purpose other than research or private study may require the authorization of the copyright owner of the work in question. Responsibility regarding questions of copyright that may arise in the use of this copy is assumed by the recipient.

UNIVERSITY OF CALGARY

3-D data Interpolation and denoising by adaptive weighting rank-reduction method using
singular spectrum analysis algorithm

by

Farzaneh Bayati

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

November, 2021

© Farzaneh Bayati 2021

Abstract

A difficult challenge in seismic processing and imaging is to address insufficient and irregular sampling. Most processing algorithms require well-sampled data, which involves small sampling intervals with a regular distribution. This motivates us to find new techniques that are more efficient in interpolating seismic data.

The primary objective of this thesis is to study Singular Spectrum Analysis (SSA) as a tool for the reconstruction and denoising of seismic data. An overview of the methods of seismic interpolation and the potential use of SSA in time series is described. SSA as a rank-reduction method for 2-D and 3-D seismic data interpolation is studied. The rank-reduction step of SSA is improved by proposing an adaptive rank-reduction method. To improve the algorithm in denoising an adaptive weighting rank-reduction algorithm is proposed. SSA is compared with the Minimum Weighted Norm Interpolation (MWNI) algorithm.

Results obtained in this work demonstrate that SSA is a promising method for simultaneous denoising and reconstructing seismic data.

Preface

This thesis is an original work by the author. No part of this thesis has been previously published.

Acknowledgements

[Don't forget to thank someone.]

To my beloved family.

Table of Contents

Abstract	ii
Preface	iii
Acknowledgements	iv
Dedication	v
Table of Contents	vi
List of Figures and Illustrations	viii
List of Tables	xii
List of Symbols, Abbreviations and Nomenclature	xiii
Epigraph	xv
1 Introduction	1
1.1 Background	1
1.2 Seismic data interpolation methods	3
1.2.1 Spatial prediction filter	4
1.2.2 Wave equation based methods	4
1.2.3 Signal processing based methods	5
1.2.4 Machine learning based methods	5
1.2.5 Rank-reduction based methods	6
1.3 Importance of Dimensionality: from 2-D to 5-D and beyond	7
1.4 Organization of the thesis	9
2 Basics of Singular Spectrum Analysis	11
2.1 Background	11
2.2 The basic SSA algorithm	12
2.2.1 First stage: Decomposition	13
2.2.2 Second stage: Reconstruction	15
2.3 Potential applications for SSA	17
2.4 Summary	20

3	Background of SSA in seismic data processing	21
3.1	Basic SSA in seismic data processing	22
3.1.1	Projections Onto Convex Sets method	28
3.1.2	Reconstruction of missing traces	29
3.2	Multichannel Singular Spectrum Analysis	32
3.2.1	Methodology	32
3.2.2	Synthetic data example	37
3.2.3	Discussions	39
4	Adaptive rank reduction	41
4.1	Adaptive rank reduction for the Hankel matrix	42
4.1.1	Examples of adaptive rank-reduction for 2-D data	46
4.2	Adaptive rank-reduction for the block Hankel matrix	53
4.3	Discussions	60
5	Application of weighting operator on rank-reduction step	61
5.1	Motivation	61
5.2	Weighting rank-reduction (WRR)	62
5.3	Adaptive weighting rank-reduction	68
5.3.1	Results and discussions	72
5.3.2	Synthetic data	73
5.3.3	Optimally damped rank-reduction method	81
5.3.4	Real field data	86
5.4	Discussions	90
6	SSA comparison with Minimum Weighted Norm Interpolation	92
6.1	The binning approach	92
6.2	Theory	93
6.2.1	Synthetic data	96
7	Conclusions and Recommendations	101
7.1	Future Work	103
	Bibliography	105

List of Figures and Illustrations

1.1	Major types of seismic waves based on propagation characteristics.	2
1.2	Geometry of a seismic survey. The red star indicates source location and the blue dots represent offset. a) Irregularly sampled seismic survey. b) The same survey in a regular grid. c) the survey after interpolation. d) The highly regularized and interpolated survey.	3
1.3	Different ways in which seismic data are typically grouped for processing. The stars indicate shots' location and the triangles shows the location of receivers. a) Common receiver gather. b) Common shot gather. c) Common offset gather. d) Common mid point gather. e) Common depth point gather. f) Common azimuth gather.	9
2.1	Capability of SSA in finding trends with different resolutions: a) Rough trend. b) Accurate trend.	18
2.2	Capability of SSA on smoothing time series: a) Initial series (blue) smoothed series (red). b) Residuals.	19
2.3	Canada monthly live births. Initial series (blue), trend and cycle extracted by SSA series (red).	20
3.1	A 2-D data in a small window with a constant dip.	22
3.2	a) Singular values for noiseless data without decimated traces; b) singular values for noiseless data with 51% decimated traces.	31
3.3	Schematic of generating a block Hankel matrix from a frequency slice.	35
3.4	Interpolation of a 3-D noiseless data containing four linear events. (a) Input data with 51% gaped traces, (b) initial data, (c) interpolated data with MSSA rank=4, (d) residual errors.	37
3.5	Slices of data in inline and cross-line directions. (a) input data in inline direction, (b) input data in cross-line direction. (c) and (d) Initial data in inline and cross-line direction respectively. (e) and (f) Interpolated data by MSSA with rank=4 in inline and cross-line direction respectively. (g) and (h) The result of subtracting initial data from the interpolated data in inline and cross-line direction respectively.	38
3.6	Interpolation of a 3-D data containing four linear events. (a) Input data with SNR=1 and 51% missing traces, (b) initial data, (c) interpolated data with MSSA rank=4, (d) residual errors.	39

3.7	Slices of data in inline and cross-line directions. (a) input data in inline direction, (b) input data in cross-line direction. (c) and (d) Initial data in inline and cross-line direction respectively. (e) and (f) Interpolated data by MSSA with rank=4 in inline and cross-line direction respectively. (g) and (h) The result of subtracting initial data from the interpolated data in inline and cross-line direction respectively.	39
4.1	Analyzing the distribution of the singular values for data containing one linear flat event. a) Clean data. b) Normalized singular values of the Hankel matrix for the clean data. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for incomplete data.	43
4.2	Analyzing the distribution of the singular values for data containing one linear inclined event. a) Clean data. b) Normalized singular values of the Hankel matrix for the clean data in constant frequency 25 Hz. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for the incomplete data in constant frequency 25 Hz.	44
4.3	Analyzing singular values distribution of data containing four linear events. a) Clean data. b) Normalized singular values of the Hankel matrix for clean data in a frequency slice of 25 Hz. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for the incomplete data in constant frequency 25 Hz.	45
4.4	Comparing the interpolation results of SSA with TRR method and ARR method. a) Desired result, b) input data with $SNR = 100$ and 40% missing, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10, respectively. k) Interpolated data with SSA using ARR. d), f), h), j), and l) residual errors for c), e), g), i), and k), respectively.	48
4.5	Comparison of interpolation results of SSA using TRR with SSA using ARR method. a) Desired result, b) input data with $SNR = 2$ and 40% missing traces, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10, respectively. k) Interpolated data with SSA using ARR. d), f), h), j), and l) residual errors for c), e), g), i), and k), respectively.	50
4.6	Comparing the interpolation results of SSA using TRR method and ARR. a) Desired result, b) input data with $SNR = 1$ and 40% missing traces, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10 respectively. k) Interpolated data with SSA using adaptive rank reduction. d), f), h), j), and l), residual errors for c), e), g), i), and k), respectively.	52
4.7	a) Cube of 3-D data having three linear events. Slice of data in b) inline direction, c) cross-line direction.	53
4.8	Block Hankel matrix for frequency slice of a) 20 Hz, b) 50 Hz.	54
4.9	a) Singular values spectrum, b) Zoomed singular values spectrum.	55
4.10	Normalized singular values distribution of the block Hankel matrix of 3-D data with three linear events in the frequency slice a) 20 Hz, b) 50 Hz. The first 20 singular values of the same data for frequency slice of c) 20 Hz, d) 50 Hz.	56

4.11	Analysis of the frequency power for each singular value. The red line shows power per frequency, and the blue line indicates the estimated mean frequency. a) Power per frequency of data recovered by the first singular value. b), c), d), e) f), g), h), I), j), k), and l) power per frequency recovered by the 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th, 11th, and 12th singular value, respectively.	58
5.1	a) Cube of 3-D synthetic clean and complete data, b) and c) inline and cross-line sections of cube (a), respectively. d) Cube of 3-D synthetic data with $SNR = 2$ and 51% missing traces, e) and f) inline and cross-line sections of cube (d), respectively.	65
5.2	Block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with $SNR = 2$ and 51% missing traces, c) recovered data by TSVD rank-reduction after ten iterations, d) recovered data by WRR after ten iterations.	66
5.3	The first 25 singular values spectrum of the block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with $SNR = 2$ and 51% missing traces, c) The first 12 singular value of the data, d) singular values spectrum of the data after applying the weighting operator.	67
5.4	Block Hankel matrices for frequency of 50 Hz for a) clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered block Hankel matrix by TRR rank=4 after ten iterations, d) recovered data by WRR after ten0 iterations. e) Recovered data by applying the ARR method after ten iterations, d) recovered data after applying AWRR after ten iterations.	70
5.5	The first 25 singular values spectrum of the block Hankel matrices for frequency slice of 50 Hz for, a) clean and complete data, b) data with $SNR = 2$ and 50% killed traces, c) recovered data by TRR, d) recovered data by WRR, e) recovered data by ARR, f) recovered data by AWRR.	72
5.6	Comparison of different methods of rank-reduction. a) Clean and complete data, b) data with $SNR = 100$ and 60% missing traces. c), e), g), and i) Interpolated data by TRR, ARR, WRR, and AWRR methods, respectively. d), f), h), and j) Residual errors of TRR, ARR, WRR, and AWRR, respectively.	74
5.7	Comparison of different methods of rank-reduction. a) Clean and complete data, b) input data with $SNR = 2$ and 60% missing traces, c), e), g), and i) interpolated data by TRR, ARR, WRR, and AWRR, respectively. d), f), h), and j) residual errors of TRR , ARR, WRR, and AWRR, respectively.	76
5.8	$f - k$ spectra comparison of: a) clean and complete data, b) input data, c) TRR result, d) WRR result, e) ARR result, f) AWRR result.	77
5.9	The mean and standard error of the quality of reconstruction versus gap ratio for different methods. These results are obtained by running each method on 20 realizations of different data set for different gap ratios and rank=9 for the predefined rank methods.	78
5.10	The mean and standard error of the reconstruction's quality versus SNR for different methods. These results are obtained by running each algorithm on 20 realizations of the different data set for a signal-to-noise ratio with the gap ratio = 50%, and rank=9 for the predefined rank methods.	80

5.11	Diagram of the output Quality of interpolation using ORR with respect to the selected rank.	83
5.12	The quality of interpolation versus gap percent for ORR, and AWRR.	84
5.13	Comparison of the interpolation results for ORR method. a) Clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered data by ORR (output $QF = 11.12$ dB), d) residual errors for ORR, e) interpolated data by AWRR (output $QF = 9.44$ dB), f) residual errors for AWRR.	85
5.14	Zoomed images of figure 5.13 from the time 0.4 s to 1.2 s and trace number 400 to 460. a) Clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered data by ORR, d) residual errors for ORR, e) interpolated data by AWRR, f) residual errors for AWRR.	86
5.15	Geometry of the field data. (a) Initial distribution of the traces. (b) Distribution of traces after killing 41% of them.	88
5.16	Cube of 3-D field data. a) The input data. b) Interpolation result of applying TRR. c) Interpolation result of applying AWRR.	88
5.17	a) Input real data, b) interpolation result of applying TRR approach, c) interpolation result of applying AWRR method.	89
5.18	Zooming area of figure 5.17 from time 1.35 (s) to 1.75 (s) and trace number 160 to 190. a) input data, b) TRR result, c) AWRR result	90
6.1	2-D data with irregular offset. a) Input data with 30% randomly missing traces. b) $f - k$ spectrum of the input data. c) Desired data. d) $f - k$ spectrum of the desired data.	97
6.2	Geometry distribution of traces with irregular offset and different binning sizes. a) Bin size=18 (m), b) bin size=9 (m), c) bin size =6 (m), d) bin size=4.5 (m).	97
6.3	Results of MWNI interpolation on a 2-D data. a), d), g), and h) Regularized and interpolated data for different bin size of 18, 9, 6, and 4.5(m), respectively. b), e), h), and k) interpolated traces in their true coordinates. c), f), i), and l) Residual errors.	99
6.4	$f - k$ spectrum of the MWNI interpolation. a) Bin size=18(m), b) 9(m), c) 6(m), and d) 4.5(m).	100
6.5	Results of SSA interpolation a) Interpolated data in a rough grid of 18(m), b) residual errors, c) $f - k$ spectrum of the interpolated data.	100
7.1	Interpolating decimated traces by SSA or upsampling. a) Input data. b)Desired data. c) Interpolated by SSA. d) Residual errors. e) $f - k$ spectrum of input data. f) $f - k$ spectrum of desired data. g) $f - k$ spectrum of interpolates data.	104

List of Tables

5.1	Mean and standard deviation of the quality of interpolation for each method for different gap ratio inputs.	79
5.2	Mean and standard deviation of the quality of interpolation for each method for different SNR inputs.	81

List of Symbols, Abbreviations and Nomenclature

Symbol or abbreviation	Definition
ARMA	Auto-Regressive Moving Average
ARR	Adaptive Rank-Reduction
AWRR	Adaptive Weighting Rank-Reduction
FFT	Fast Fourier Transform
IHT	Iterative Hard Threshold
MSSA	Multichannel Singular Spectrum Analysis
MWNI	Minimum Weighted Norm Interpolation
NMO	Normal Move Out correction
POCS	Projection Onto Convex Sets
SNR	Signal to Noise Ratio
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
TRR	Traditional Rank-Reduction method that uses SVD
TSVD	Truncated SVD
WRR	Weighting Rank-Reduction
A	Amplitude of a waveform in the Fourier domain
\mathcal{B}	Hankelization operator to generate the block Hankel matrix of frequency
\mathbf{D}	Dreanlet transform
\mathcal{F}	Fourier operator
\mathcal{F}^{-1}	Inverse Fourier Operator
\mathcal{H}	Hankelization operator
\mathbf{H}	Matrix in a form of Hankel
\mathbf{H}_k	A reduced rank Hankel matrix
K	Number of lagged vectors
L	Length of the window of a vector
\mathbf{M}	A block Hankel matrix
N	Number of component of a vector
N_x	Number of channels in x direction
N_y	Number of channels in y direction
q	Dip of a waveform
QF	Quality factor of the interpolation result

\mathcal{R}	Rank-reduction operator
\mathbf{R}	Sampling operator
$s(x, t)$	2-D waveform in the time domain
$s(x, y, t)$	3-D waveform in the time domain
$\mathbf{S}_k(\omega, x)$	2-D waveform in the frequency domain
$\mathbf{S}_k(\omega, x, y)$	3-D waveform in the frequency domain
t	time dimension
$\mathcal{T}(\Sigma, k)$	The operator that gives the rank of the Hankel matrix
$T_{\mu n}$	The Hard threshold operator given element-wise multiplication
\mathbf{U}	Matrix of eigenvectors
\mathbf{V}	Matrix of eigenvectors
W	Wavelet
\mathbf{W}_s	Matrix of model weights
$\hat{\mathbf{W}}$	The weighting operator that reduces the noise effect of each singular value matrix
x	spatial dimension
α_n	An iteration based weighting factor
$\mathcal{B}(\Sigma, \tilde{k})$	The operator that gives the rank of the block Hankel matrix in the frequency domain
η	Random noise in data
λ	Singular value of a squared matrix
μ	Regularization factor
Σ	Matrix of eigenvalues
ω	Temporal frequency

Epigraph

Mask up, wash your hands, practice social distancing, and stay home.

Year 2020

Chapter 1

Introduction

1.1 Background

The seismic data exploration method provides data related to the physics of the earth. The information obtained from the seismic data maps the subsurface geologic structure and stratigraphic features. The fundamental procedure of seismic exploration consists of generating seismic waves and estimating the time expected for the waves to travel from the source to a series of receivers. These receivers are generally distributed on a straight line or lines (Yilmaz, 2001). With knowledge of travel times from the sources to the various receivers and the velocity at which waves travel, one can reconstruct the paths of the seismic waves. The energy produced by a source such as a seismic vibrator (on land) or an airgun (in water) discharges seismic waves that travel into the earth's layers and bounce back from different rock layers. The reflected and refracted seismic waves are recorded by a geophone (on land) or hydrophones (in water) and give the subsurface's first image (Yilmaz, 2001). The seismic method lets the ground underneath interpretation discover oil, gas, water, faults, folds etc. The seismic exploration method's power over other geophysical methods because of the high accuracy, high resolution, and excellent penetration. Several waveforms in a normal land seismogram can be classified to direct wave, refraction and reflection. Figure 1.1 shows a

schematic of the major seismic waves. Reflections waves are the main target in the seismic survey because they penetrate more into the subsurface than the other types of waves. Therefore, seismograms are often processed and modified to improve the reflections' signal, providing a more clear image of the subsurface.

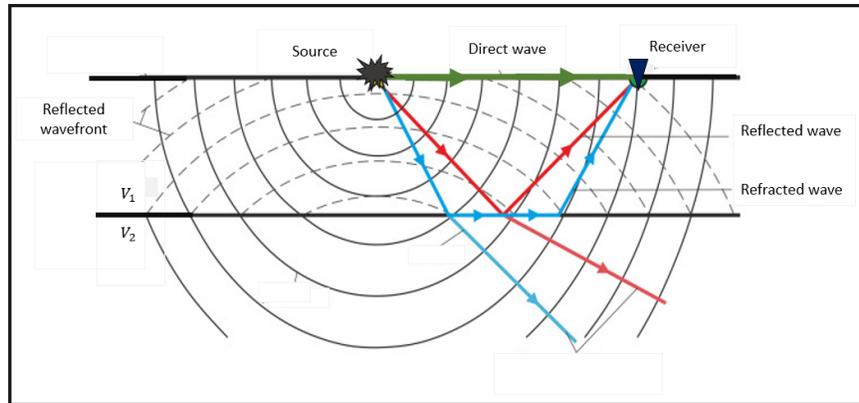


Figure 1.1: Major types of seismic waves based on propagation characteristics.

Seismic surveys are designed to keep a consistent grid of sources and receivers. In typical real-life scenarios, because of logistic obstacles or economic constraints regularly sampled seismic surveys are uncommon or rare. These limitations make large shot and receiver sampling intervals producing poorly irregularly sampled seismic data along with spatial coordinates with gaps without recorded traces.

Missing traces often have negative effects on many processing algorithms, producing sampling artifacts or noise on the resulting data or images that lessen accuracy and resolution. There are many seismic processing techniques for noise attenuation or imaging such as reverse-time migration, full waveform inversion, and surface-related multiple eliminations, that demand high-quality regularly sampled data to work properly. These demands make seismic data interpolation an important pre-processing step in the seismic data processing.

Figure 1.2 illustrates a typical land 3-D seismic survey. The red star is the source location, and the blue dots are the offset values. Figure 1.2 (a) shows the geometry of an irregularly sampled data survey. Figure (b) indicates the survey in figure 1.2 (a) after sitting in a

regular grid. Figure 1.2 (c) is the survey after interpolation, and figure 1.2 (d) represents a highly regularized and interpolated survey which is the desired survey for most of the seismic processing tools.

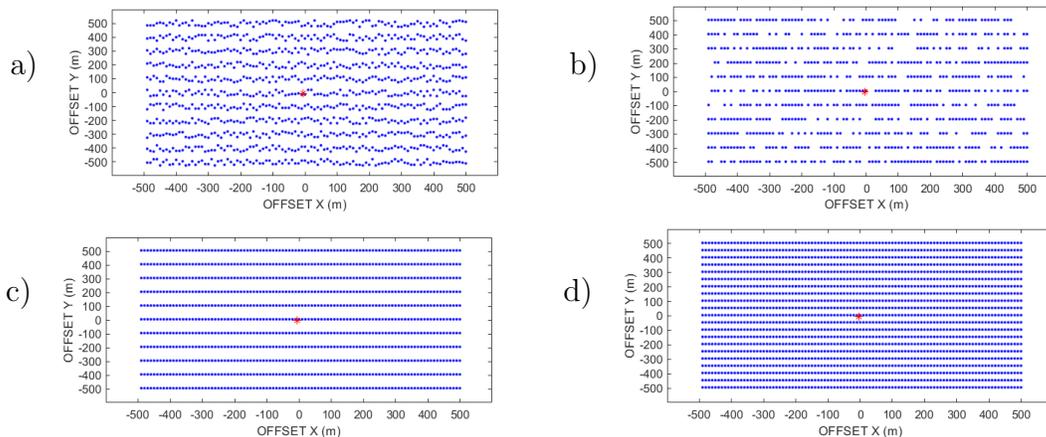


Figure 1.2: Geometry of a seismic survey. The red star indicates source location and the blue dots represent offset. a) Irregularly sampled seismic survey. b) The same survey in a regular grid. c) the survey after interpolation. d) The highly regularized and interpolated survey.

The main objective of this chapter is to present a background of different methods of seismic data interpolation. The next section is a review of a few common techniques for seismic data reconstruction, which leads to the motivation for this work.

1.2 Seismic data interpolation methods

Data reconstruction performs an important task in seismic data processing and imaging workflows. We often need to interpolate and extrapolate data at missing spatial coordinates. Some consequences of poorly sampled data without interpolation are degradation of amplitudes and sampling artifacts. Seismic reconstruction methods can be divided into five main classes:

1.2.1 Spatial prediction filter

Spitz (1991) proposed a method depending on the predictability of the linear events in the $f - x$ domain. The method considers linear events in a uniformly aligned traces section and can decrease the original spatial interval by integer factors without being given the actual dips. Porsani (1999) introduced a half-step prediction filter. The half prediction filter at frequency $f/2$ is computed by least-squares, predicting data elements at the odd locations from data at the even locations. Liu and Fomel (2011) interpolated aliased seismic data employing a method based on adaptive prediction-error filtering (PEF) and regularized non-stationary auto-regression. The proposed method includes an automatic selection of regularization parameters and fast iteration convergence. Liu and Chen (2018) developed a $f - x$ regularized non-stationary auto-regression interpolation method that can deal with the space-changing dips events. The method assumes that the coefficients of $f - x$ regularized non-stationary auto-regression are evenly changing along with the space axis. The method employs shaping regularization and its implementation in the frequency domain makes it efficient.

1.2.2 Wave equation based methods

Ronen (1987) introduced a trace-interpolation technique based on the wave equation with the assumption of a smooth spatial spectrum. The method includes an iterative inversion of the model linearly related to the Normal moveout (NMO) corrected seismic data. Fomel (2003) introduced a finite-difference filter built from the offset-continuation differential equation. The filter is employed on frequency slices in the log-stretch frequency domain and reconstructs gaps efficiently by iterative least-squares optimization. Trad (2003) proposed an extension of the hyperbolic Radon transform interpolating apex-shifted hyperbolic events. This transformation is performed by applying a sparse least-squares conjugate gradient algorithm in the time domain and a very efficient Stolt's $f - k$ domain operator.

1.2.3 Signal processing based methods

Most of the methods in the signal processing category are based on transformations across different domains. Sacchi et al. (1998) introduced an iterative non-parametric method to spectral estimation. For line spectra, Duijndam et al. (1999) proposed Fourier transform applying least squares estimation of Fourier components to reconstruct regularly sampled band-limited seismic signals along one spatial coordinate. Liu and Sacchi (2004) proposed an approach for spatially band-limited signals by solving an inverse problem with a $f-k$ domain regularization term. The method can be extended to multidimensional reconstruction in any spatial domain. Wang et al. (2011) introduced an l1-norm constrained trust region method for solving the compressive sensing problem based on the wavelet transform. The method succeeds in the limitations of random sampling.

1.2.4 Machine learning based methods

In recent years, machine learning has attracted much attention in geophysical studies. Deep learning (DL) has had a great impact on signal and image processing. However, Machine learning for geophysical problems is new and needs more development. Jin et al. (2018) introduced a novel denoising approach that utilizes deep residual networks (ResNets) learning framework. Wang et al. (2018) proposed a framework designed using ResNets block to interpolate anti-aliased seismic data. Generative Adversarial Networks (GANs) explained by Siahkoohi et al. (2018) can reconstruct heavily sub-sampled seismic data by giving up linearity and using an adaptive non-linear model. Convolutional Neural Networks (CNNs) is a novel strategy for the reconstruction of missing traces in pr-stack seismic images (Mandelli et al., 2018). A new framework for training Artificial Neural Networks (ANNs) was presented by Mikhailiuk and Faul (2018) to restore corrupted multidimensional seismic signals. Wang et al. (2019) utilized an eight-layer residual learning networks (ResNets) with a better back-propagation property for interpolation. Shi et al. (2020) proposed a regularization method using the random-initialized convolutional neural network (CNN) that parameterizes the

model. A method based on imaging techniques from single-image super-resolution (SISR) proposed by Greiner et al. (2019) that consist of a deep (CNN) and a periodic resampling layer for upscaling to the non-aliased wavefield.

1.2.5 Rank-reduction based methods

In rank-reduction-based methods, the linear events in a clean seismic data set are low rank data. However, the rank of data is increased by noise and missing traces (Trickett, 2008). The rank-reduction algorithm in the frequency domain is carried out in frequency slices by generating Hankel/Toeplitz matrix and applying a low-rank reduction method on the generated matrix. The singular spectrum analysis (SSA) method proposed by (Sacchi, 2009; Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011) works by rank-reduction of the Hankel matrix with an iterative algorithm in the frequency domain. Gao et al. (2013) extended the SSA method to higher seismic data dimensions and called a multichannel singular spectrum (MSSA). Interpolating regularly missing traces with a de-aliased MSSA method was proposed by Naghizadeh and Sacchi (2013). Kreimer et al. (2013) answered the complication of non-linear events in pre-stack data via low rank tensor completion. Kreimer and Sacchi (2012) used the higher order SVD for rank-reduction of the pre-stack seismic data tensor. Ely et al. (2015) used a statistical test to control the complexity of regularization by tuning a single regularization parameter. Interpreting the interpolation in the context of matrix completion (MC), Kumar et al. (2015) developed rank-reduction techniques to seismic data reconstruction. Carozzi and Sacchi (2021) proposed a method called (I-MSSA) that interpolates seismic data in its exact data coordinates which overcome the problem of vertical errors in rough binning.

1.3 Importance of Dimensionality: from 2-D to 5-D and beyond

Although seismic interpolation is not a new method of reconstructing data, using multi-dimensional interpolation simultaneously along all directions was not common until a few years ago (Trad and Cheadle, 2005). The potential use of several dimensions for interpolation brings many new questions like:

1-How to choose the best output geometry design (figure 1.3)?

Trad (2009) classified interpolation geometries in basically two types, each type with its variations. One type is acquisition-guided or surface-consistent geometry. In this group, by creating new shots and new receivers we make the output geometry similar to a real data acquisition geometry which consistent shot and receiver groups. This new geometry could follow the original design, preserving original traces and just adding more shots and receivers, making it better sampled and reducing gap sizes. Preserving original seismic traces provides opportunities for better quality control and prevents losing information not captured by the interpolation model. Alternatively, the output geometry could be completely new, which permits further improvements in sampling. Either way, these acquisition-guided geometries permit processing, migration and inversion with standard modules and algorithms. This is important because many migration tools like Reverse Time Migration require the geometries with shot/receiver consistent gathers. The second type of geometry design is midpoint-consistent data which is ideal for Kirchhoff and beam migrations. This design creates perfect common vectors that have advantages for creating multi azimuth angle gathers and multi-component processing. However, they also produce coarse offset/azimuth or inline crossline offset sampling and forbids techniques based on shots/receivers.

2- Which algorithms are more capable in terms of preserving/capturing signal in the presence of noise and poor sampling?

For 5-D interpolation, the algorithm should be capable of working with a large number of

input traces simultaneously. Minimum weighted norm interpolation (MWNI) was the first algorithm applied to 5-D interpolation (Liu and Sacchi, 2004; Trad, 2009). Although not a requisite, MWNI is typically applied on regular grids after binning in the 4 spatial dimensions with multidimensional Fast Fourier Transforms (FFTs). MSSA is another algorithm that is capable of using the information of the 4 spatial dimensions simultaneously. 5-D MSSA generates an order 4 Hankel matrix from the 4 spatial dimensions in each slice of frequency.

3- What is the behaviour of aliasing in multidimensional interpolation?

Due to the 5 dimensional nature of seismic data sets, it is usually the case that at least one or more of these dimensions are poorly sampled and also irregularly sampled. However, we can take advantage of the well-sampled dimensions to help constraint the interpolation model. Because non-aliased signal can be recorded along these directions, they can help to overcome the aliasing uncertainty along with the other dimensions. This is the main reason why simultaneous high-dimensional interpolation is more capable than lower-dimensional interpolations applied in sequence (Trad, 2009).

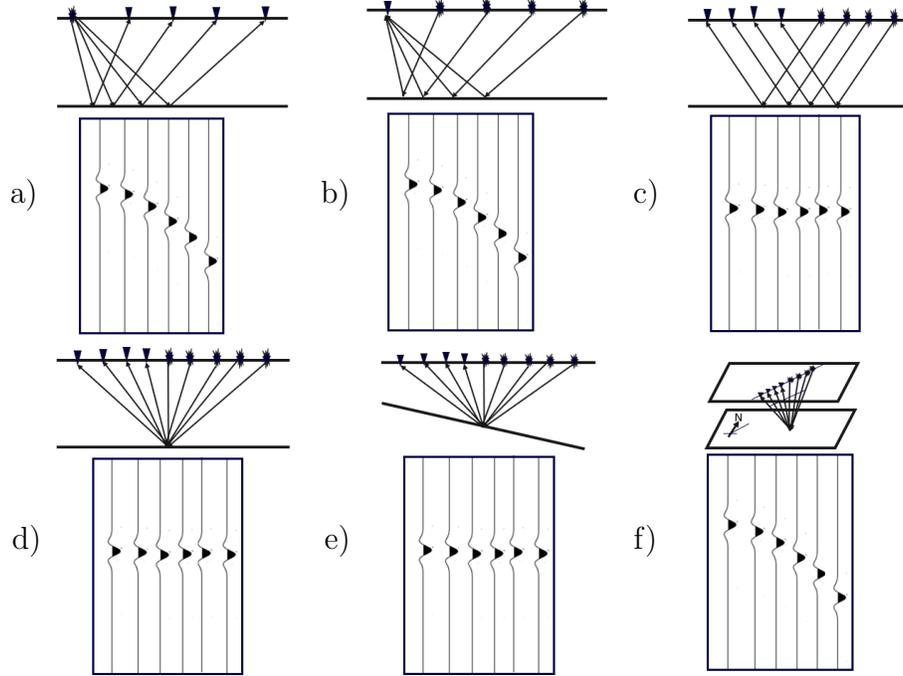


Figure 1.3: Different ways in which seismic data are typically grouped for processing. The stars indicate shots' location and the triangles shows the location of receivers. a) Common receiver gather. b) Common shot gather. c) Common offset gather. d) Common mid point gather. e) Common depth point gather. f) Common azimuth gather.

1.4 Organization of the thesis

The structure of this thesis is as follows:

Chapter 2 explores the basics of SSA. It presents the main steps of the algorithm and its implementation on a simple one-dimensional time series. This chapter analyses the potential applications of the algorithm, such as finding the trend or denoising a time series.

Chapter 3 explains the application of SSA in the seismic data processing. This chapter shows how the linear coherent events determine the rank of the Hankel matrix formed from the seismic traces. First, the chapter explains the application of the algorithm to 2-D data, and it shows the expansion of the algorithm to the 3-D data. It also explains an iterative algorithm and presents the capability of SSA for seismic data interpolation. I show some synthetic data 3-D interpolation examples by the multidimensional SSA algorithm.

Chapter 4 introduces a method of rank-reduction to apply to the rank-reduction step of the SSA algorithm. The method is called adaptive rank-reduction (ARR). Examples in this chapter demonstrate how ARR selects the rank without prior information about the number of linear events in each processing window.

Chapter 5 explains a weighting operator used in the rank-reduction step that helps to reduce residual noise from the data. The operator is a diagonal weighting matrix that reduces the effect of noise projection on the signal projection in each slice of frequency. The chapter introduces an adaptive weighting rank-reduction (AWRR) method that interpolates the data while removing the residual noise in the recovered data simultaneously. Several examples compare different methods of rank-reduction in the SSA algorithm.

Chapter 6 presents the MWNI method and compares it with SSA using several 2-D data.

Chapter 7 presents the conclusions and recommendations for further work.

Chapter 2

Basics of Singular Spectrum Analysis

2.1 Background

Singular spectrum analysis (SSA) is a common technique for the analysis and forecasting of time series. SSA can combine both model-free and parametric methods, which makes it a helpful method for a wide range of problems in different areas of study. It has been employed in various fields such as physics, economics, and business research. Depending on the field of study, the purpose of applying SSA varies (Golyandina and Korobeynikov, 2014).

One of the most important contributions to the development of SSA was made by Broomhead and King (1986), who used the method of delay vectors proposed by Takens (1981) to study dynamical systems using multivariate statistical analysis. Vautard et al. (1992) proposed SSA as a denoising algorithm. It identifies oscillatory components by selecting subsets of eigenvalues and eigenvectors and related principal components.

Read (1992) studied baroclinic processes by SSA. In the economic sciences, it has been applied to study and forecast time-series like daily market rate (Hassani et al., 2010). One of the advantages of SSA is that it can be extended to multiple dimensions. Read (1992) applied SSA on multiple time series simultaneously, called multichannel singular spectrum

analysis (MSSA). MSSA uses all the information of the time series in the trajectory matrix. Hsieh and Wu (2002); Ghil et al. (2002) applied MSSA to study the climate. MSSA is applied to signal reconstruction and forecasting of time series by Golyandina et al. (2001). Broomhead and King (1986) proposed SSA as a powerful noise attenuation method by recovering data from the largest singular values. Elsner and Tsonis (2013) demonstrated that the large singular values after the decomposition of the trajectory matrix could represent the signal while the small ones are related to the contaminated noise. Cody et al. (2018) studied SSA to extract leak detection components on an acoustic signal in a water pipeline with leakage. This chapter studies the main principles and ideas of SSA for one-dimensional data. The basic SSA algorithms can be divided into two stages of decomposition and reconstruction. In the decomposition stage, a time series is divided into a sequence of lagged vectors. Then these lagged vectors are embedded in a trajectory matrix in the form of a Hankel matrix, and the matrix is decomposed to its singular values and singular vectors using SVD. In the reconstruction stage, the trajectory matrix is recovered by the remaining few singular values and zeroing the small ones. The recovered low-rank matrix is a symmetric matrix where the recovered time series can be reconstructed by diagonal averaging.

2.2 The basic SSA algorithm

Low-rank approximations are often applied to data modelling and data compression. Since the rank of the model depends on the complexity of the generated model, low-rank approximations are useful as a method to simplify information, that is to extract the most important components of a matrix. Rank reductions are often applied to intermediate mathematical constructions using truncated SVD. A Hankel structure is one of such common mathematical constructs. The rank reduction theories are closely related to principal component analysis (PCA), matrix factorization, data compression, and feature selection.

Assume a time series $\mathbf{X} = \mathbf{X}_N = (x_1, \dots, x_N)$ containing N components and $N > 2$. Let find

L , ($1 < L < N$) as the window length and $K = N - L + 1$ as the number of lagged vectors. This algorithm consists of two stages of decomposition and reconstruction where there are four steps for each of these two stages.

1. Decomposition

- First, the elements of a time series become entries of a trajectory matrix with the window length of L and K lagged vectors.
- Second, using the singular value decomposition (SVD) of the trajectory matrix to decompose the matrix into singular matrices.

2. Reconstruction

- The third step is the rank reduction step by passing the large singular values and zeroing the small ones.
- The last step refers to recovering the rank-reduced series.

The next section describes the four main steps of these two stages in the basic SSA algorithm.

2.2.1 First stage: Decomposition

Embedding

We start with a sequence of lagged vectors with the length of L from the \mathbf{X} series mentioned in the previous section. As mentioned before, the number of the lagged vectors of the \mathbf{X} series is K . In other words; we will have $\mathbf{X}_i = (x_i, \dots, x_{i+L-1})$, ($1 \leq i \leq K$) with the length of L . In fact, the embedding step allows us to go from a one-dimensional space to a multidimensional series (Broomhead and King, 1986). The trajectory matrix generated from the lagged vectors is in the form of a Hankel matrix. A Hankel matrix is a $L \times K$ matrix in which each rising anti-diagonal from left to right is constant and symmetrical around the

main diagonals. The process of Hankelization or embedding can be summed as $\mathbf{X}_{i,j} = x_{i+j+1}$, where $1 \leq i \leq K$ and $1 \leq j \leq L$:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ x_2 & x_3 & \dots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{pmatrix}, \quad (2.1)$$

To simplify, let us call \mathcal{H} the Hankelization operator; therefore, we will have $\mathbf{X} = \mathcal{H}x$.

Singular Value Decomposition

In this step, we apply the singular value decomposition (SVD) of the trajectory matrix. SVD is popular for low-rank approximations (Eckart and Young, 1936) and is common also in linear algebra to decompose a matrix into its eigenvalues and eigenvectors.

To perform the SVD of the trajectory matrix of \mathbf{X} , let's set $\mathbf{S} = \mathbf{X}\mathbf{X}^H$ and let us call $\lambda_1, \dots, \lambda_L$ the nonzero eigenvalues of \mathbf{S} in descending order ($\lambda_1 > \lambda_2 > \dots > \lambda_L$). The eigenvectors ($\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L$) of the matrix \mathbf{S} are the orthonormal system corresponding to eigenvalues. We thus have $\mathbf{V}_i = \frac{\mathbf{X}^t \mathbf{U}_i}{\sqrt{\lambda_i}}$ ($i = 1, \dots, L$). The SVD of the trajectory matrix \mathbf{X} can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_L, \quad (2.2)$$

where $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^H$. In equation 2.2, \mathbf{X}_i have rank one and are called elementary matrices (Golyandina and Zhigljavsky, 2013).

$$\mathbf{X} = \sqrt{\lambda_1} \mathbf{U}_1 \mathbf{V}_1^H + \dots + \sqrt{\lambda_L} \mathbf{U}_L \mathbf{V}_L^H, \quad (2.3)$$

The number of non-zero eigenvalues ($\sqrt{\lambda_i}$) of \mathbf{X} represents its rank. In other words, the rank of matrix \mathbf{X} corresponds to the maximum number of linearly independent columns of \mathbf{X} . This is equal to the dimension of the vector space spanned by its rows.

In a more compact notation, the SVD of a matrix can be presented as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (2.4)$$

where the columns of \mathbf{U} represents singular vectors of $\mathbf{X}\mathbf{X}^H$, the singular vectors of $\mathbf{X}^H\mathbf{X}$ are the columns of \mathbf{V} , and $\mathbf{\Sigma}$ is a diagonal matrix of singular values of \mathbf{X} . The singular values are the diagonal entries of the matrix \mathbf{S} and are arranged in descending order. The singular values are always real numbers. If the matrix \mathbf{X} is a real matrix, then \mathbf{U} and \mathbf{V} are also real (Van Loan et al., 1996). To find the SVD of the matrix \mathbf{S} we will have:

$$\mathbf{S} = \mathbf{X}\mathbf{X}^H = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad (2.5)$$

If we rewrite equation 2.5 for $\mathbf{X}^H\mathbf{X}$ it leads to $\mathbf{X}^H\mathbf{X} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^H$ that explains how the singular values and singular vectors of \mathbf{X} are related to the eigenvectors of $\mathbf{X}^H\mathbf{X}$ and $\mathbf{X}\mathbf{X}^H$.

2.2.2 Second stage: Reconstruction

Rank-reduction

So far, we have decomposed the trajectory matrix to its singular values and singular vectors, but we still need to apply the rank-reduction process. As discussed earlier, the singular values matrix of a Hankel matrix is a diagonal matrix containing singular values across its diagonal in descending order. Therefore, in a full rank matrix, the greater singular values relate to the signal while the small singular values represent the noise or the signals with less energy or uncorrelated events. Recovering the Hankel matrix with a small subset of singular values provides us with a rank-reduced matrix. In equation 2.2 the rank k is the maximum number of the nonzero eigenvalues $k = \max\{i, \text{ for } \lambda_i > 0\}$. In other words, for a matrix with size $L \times M$, the rank $k = \min\{L, K\}$. The reduced rank matrix of \mathbf{X} can be written

as:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_k, \quad (2.6)$$

where k is the desired rank for the trajectory matrix. In matrix notation we will have:

$$\mathbf{X}_k = \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^H, \quad (2.7)$$

The reduced rank matrix \mathbf{X}_k is the recovered matrix with the desired rank k where performs the lowest $\|\mathbf{X} - \mathbf{X}_k\|_F$. Denoting the rank-reduction operator as \mathcal{R} , the rank reduction step will be: $\mathbf{X}_k = \mathcal{R}\mathbf{X}$.

Anti diagonal averaging

So far we generated a Hankel matrix of a series and applied SVD to reduce its rank. In this step, we recover the time series from the rank reduced matrix, which is in a Hankel form. Consequently, the result of this step is the desired recovered time series of length N . Let us \mathbf{X}_k be the reduced rank Hankel matrix with size $L \times K$ and entries $x_{i,j}$, $1 \leq i \leq L$, $1 \leq j \leq K$. The recovered series elements will be obtained along the secondary anti diagonal of the recovered Hankel matrix.

$$x_k(n) = \begin{cases} \frac{1}{n} \sum_{l=1}^n \mathbf{X}_k(l, n-l-1) & \text{for } 1 \leq n \leq L \\ \frac{1}{L} \sum_{l=1}^L \mathbf{X}_k(l, n-l-1) & \text{for } L+1 \leq n \leq K \\ \frac{1}{K+L-n} \sum_{l=n-K+L}^L \mathbf{X}_k(l, n-l-1) & \text{for } K+1 \leq n \leq N \end{cases} \quad (2.8)$$

With considering \mathcal{A} as the averaging diagonal operator, we will have $x_k = \mathcal{A}\mathbf{X}_k$.

2.3 Potential applications for SSA

This section discusses potential uses of the SSA algorithm for denoising, trend observation, smoothing, and period extraction of one-dimensional real data. To learn more on potential applications of SSA see Golyandina et al. (2001).

Trends

To show how SSA can be used for extracting trends, we use a dataset of Alberta's conventional light crude oil from January 2000 to June 2020. The conventional light crude oil refers to hydrocarbon produced from the Western Canadian Sedimentary Basin and is a common source for most Canadian oil production. In this example, by selecting a window length of $L = 245$ and two singular values, we can see a rough trend of the production. To observe more details in the trend of the production, we reconstruct the signal with more singular values. In this example selecting four singular values gives a more accurate trend.

Figure 2.1 shows two choices in the trend analysis. Figure 2.1 (a) represents the results of selecting two singular values that give a rough trend in the series of conventional light crude oil production rates. We can roughly see a decrease in the light crude oil production rate from 2000 to 2010 and then a jump in the production rate by the start of 2010. However, recovering data with four singular values represents the data's behaviour more carefully. Figure 2.1 (b) confirms a general decrease in the production rate by the end of 2010, while there is a growth by the start of 2011 continuing to 2014. The production rate starts to drop from 2014 until 2017 and then it starts to grow again in 2017.

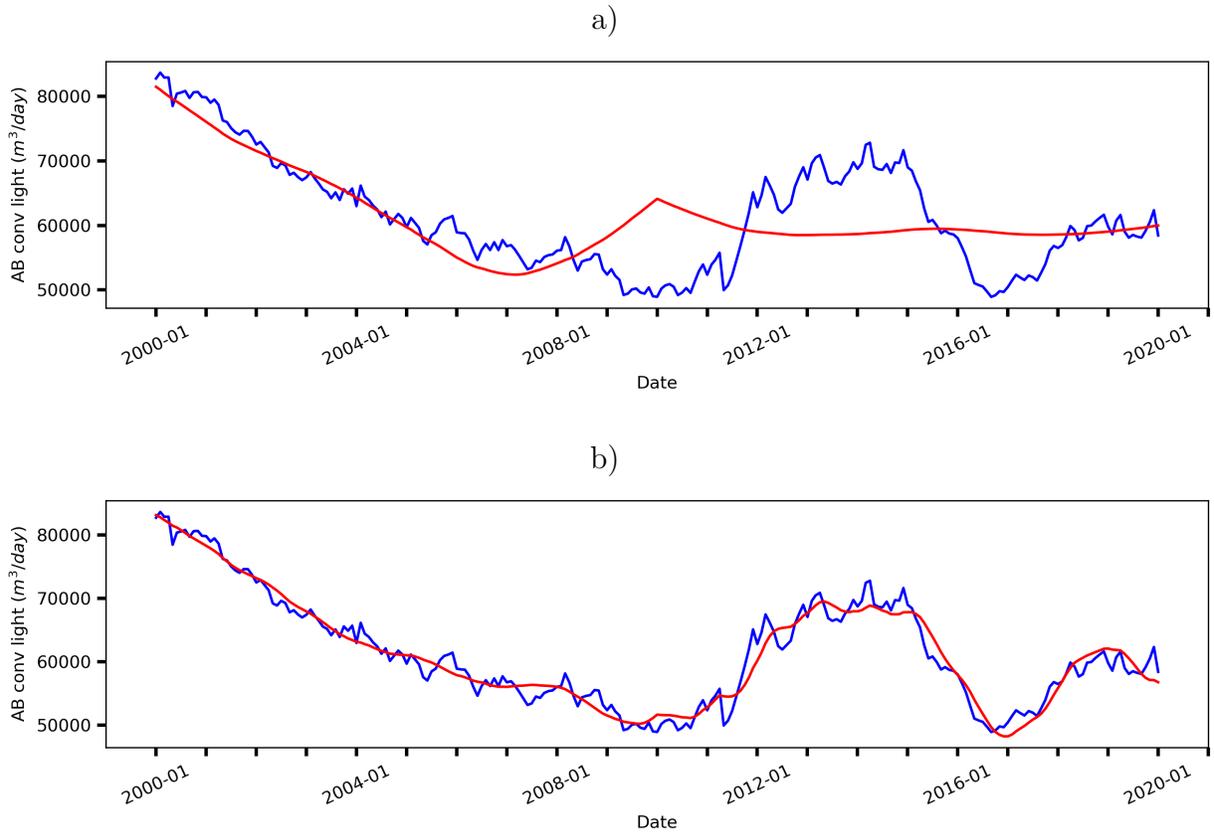


Figure 2.1: Capability of SSA in finding trends with different resolutions: a) Rough trend. b) Accurate trend.

Smoothing

To show the use of SSA on smoothing and extracting the periodicities of data, we will consider a data series containing tree rings information, taken from the International Tree Ring Data Bank (<http://www.ncdc.noaa.gov/paleo/treering.html>).

The *El Niño*/Southern Oscillation (ENSO) is the principal mode of weather variability for more than two years on Earth, fluctuating within irregularly warm (*El Niño*) and cold (*La Niña*) conditions in the tropical Pacific at periods of 2-8 years.

Figure 2.2 (a) displays the initial series and the result of the application of SSA on smoothing. The window size is set $L = 365$ and data is recovered with the first four singular values. The record of ENSO variability is for over 950 years, which the smoothed series indicates that ENSO amplitude presents a quasi-regular cycle of 50-90 years. Figure 2.2 (b) pictures the

residuals and we cannot observe any special trend of the initial data.

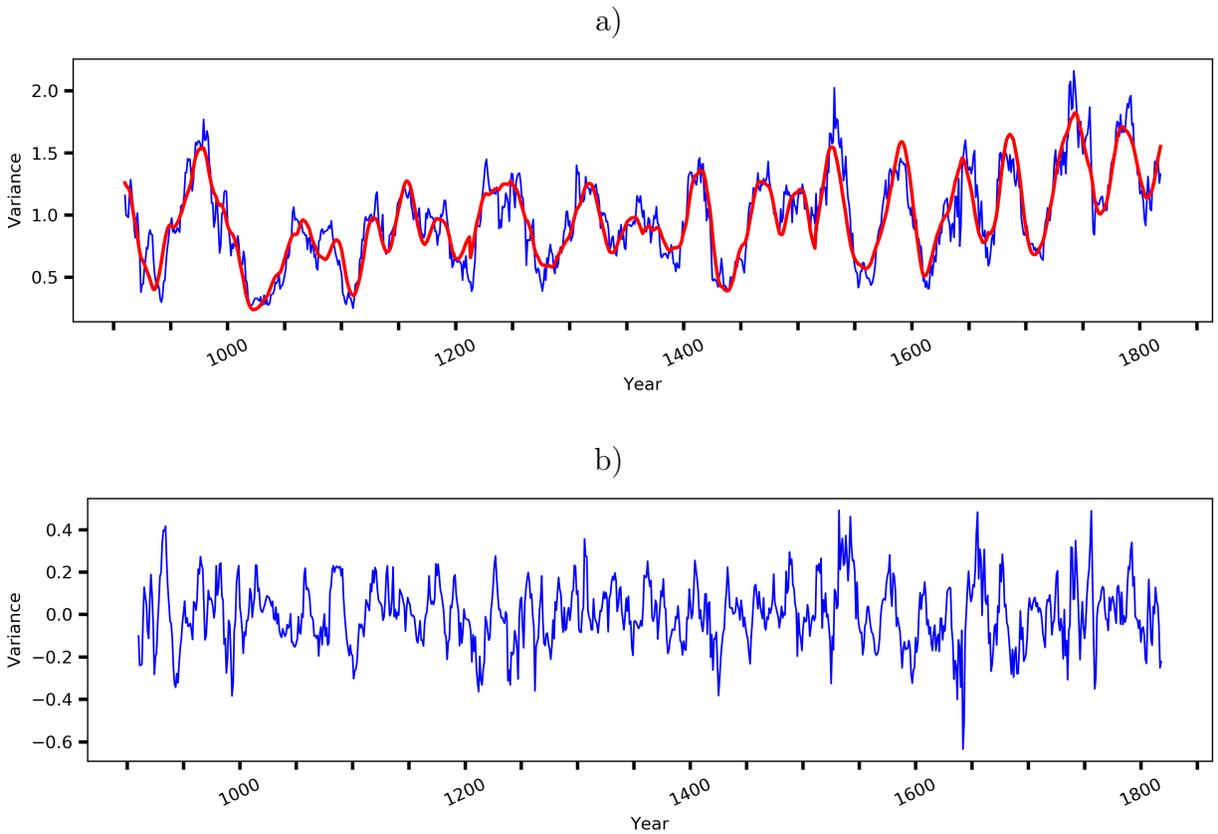


Figure 2.2: Capability of SSA on smoothing time series: a) Initial series (blue) smoothed series (red). b) Residuals.

Extraction of period

The following example demonstrates the capability of SSA on extracting periods. The test dataset contains the number of monthly births in Canada from 1991 to 2020 [Table 13-10-0415-01 Live births, by month]. The window length of input data is set $L = 100$ and recovered data is reconstructed with the first three singular values. The result shows the extraction of a smooth trend and a cycle with the length of one year repeating over the series. Figure 2.3 depicts a smooth trend for the number of births with a decrease in the year 2000. It can be seen a one-year period with the increase of birth in June in each year and a decrease of birth during December in each year.

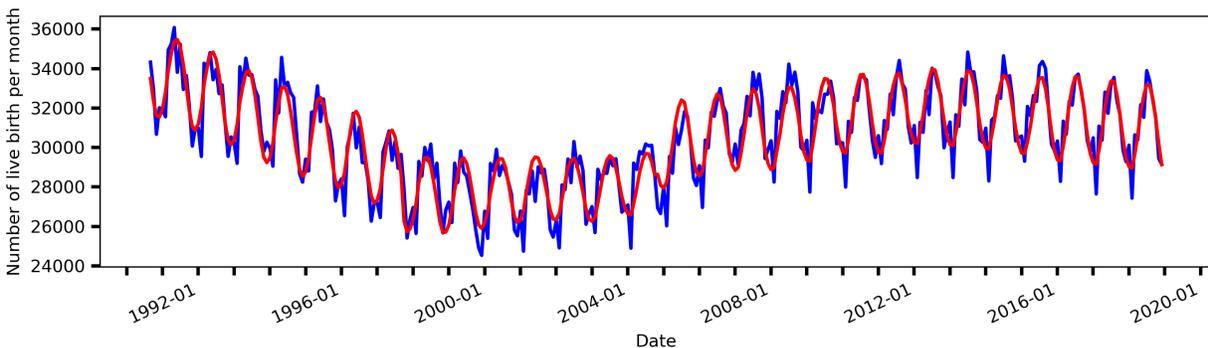


Figure 2.3: Canada monthly live births. Initial series (blue), trend and cycle extracted by SSA series (red).

2.4 Summary

This chapter introduced SSA for one-dimensional series and described the general structure of its algorithm. There are different problems that SSA can solve. We apply SSA to three real time series to evaluate its efficiency on some common problems such as denoising, trending or periodicities extraction. The results showed that SSA can be used as a tool for periodicities extraction, smoothing and trending algorithms in time series.

Chapter 3

Background of SSA in seismic data processing

Singular Spectrum Analysis (SSA) is a common time series analysis method. Multiple types of research have demonstrated SSA's ability to process the data contaminated with different kinds of noises. In the previous chapter, the basic SSA algorithm and its application on time series and its performance were discussed. However, as the geophysical field data processing technique, SSA is a novel method. It is worthy of analyzing geophysical data in terms of the Eigen features using SSA. In this chapter, I will study the application of SSA in geophysics. One of the advantages of SSA is that it can be easily expanded to multiple dimensions and also its predictability of time series in the frequency–space domain. SSA employs truncated singular value decomposition (TSVD) as the rank-reduction method to decompose the noisy signal into a signal subspace and a noise subspace. Using the basic SSA algorithm with the weighted projection onto the convex sets (POCS) method lets us fill the gaps in a data set with missing traces. SSA in the $f - x$ domain, also known as $f - x$ Cadzow filtering, was proposed by Trickett and Burroughs (2009).

SSA's application to seismic data processing has been originally introduced to signal denoising and seismic trace reconstruction in the $f - x - y$ domain (Oropeza, 2010; Oropeza and

Sacchi, 2011; Chen and Sacchi, 2015).

3.1 Basic SSA in seismic data processing

The former chapter presented four main steps of the basic SSA algorithm. The four steps were reviewed as embedding or Hankelization, singular value decomposition, rank-reduction and anti-diagonal averaging. However, in seismic data processing, two more steps of transforming data from the $t - x$ domain to the $f - x$ domain and reverse are required.

A 2-D data in a small window with a constant dip is depicted in figure 3.1.

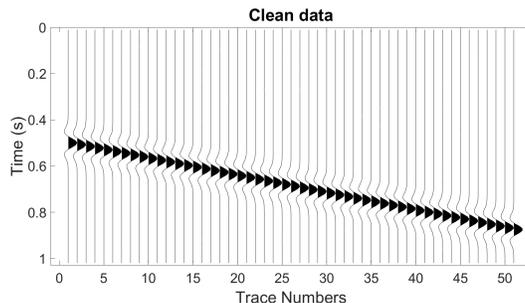


Figure 3.1: A 2-D data in a small window with a constant dip.

Figure 3.1 can be described as:

$$s(x, t) = W(t - qx), \tag{3.1}$$

where x is space, t is time, q is dip and $W(t)$ is a wavelet.

Fourier transform

To change the data domain in equation 3.1, from the $t - x$ domain to the $f - x$ domain, a Fourier transform is required. In the Fourier transform, an image is decomposed into its sine and cosine components by converting the input data in the time domain to the frequency domain. "A Fourier transformation of a time series consists of a complex valued function of frequency. The absolute value is the amplitude of the frequency, and the imaginary part is

the phase offset of the basic sinusoid in that frequency (Bochner et al., 1949).”

Data in the $f - x$ domain described as:

$$S(x, \omega) = A\omega e^{-i\omega qx}, \quad (3.2)$$

where ω is the temporal frequency and A is the amplitude. Let's analyze the seismic data with n number of channels with Δx interval in x space. Therefore for data with n number of channels, space x can be substituted by $x = n\Delta x$. Furthermore, we can write data for the n th channel as $S(x_n, \omega) = S_n$.

Equation 3.2 for the n th can be revised as following:

$$S_n = Ae^{-i\omega qn\Delta x}. \quad (3.3)$$

For the $n - 1$ th channel equation 3.2 will be as follow:

$$S_{n-1} = Ae^{-i\omega q(n-1)\Delta x} = Ae^{-i\omega qn\Delta x} e^{i\omega q\Delta x}. \quad (3.4)$$

From equation 3.4, one can note that S_n is related to its previous channel (s_{n-1}):

$$S_{n-1} = S_n e^{i\omega q\Delta x}. \quad (3.5)$$

In the beginning, we assumed that we have a plane wave; accordingly, the dip is uniform ($q = \text{constant}$), the algorithm is implemented in a constant frequency ($\omega = \text{constant}$); moreover, we assumed that the interval between each channel is ($\Delta x = \text{constant}$). Substituting the fixed parts with $X = e^{-i\omega\Delta x}$, equation 3.5 is described as:

$$S_n = XS_{n-1}. \quad (3.6)$$

Equation 3.5 describes 2-D data containing one linear event with constant dip. However, developing it for data with more than one event and different dips is quite straightforward. Let's consider a 2-D record displaying two events with different constant dips of (q_1 and q_2) and amplitudes (A_1 and A_2).

Data in the $f - x$ domain for the three channels next to each other will be explained as the following expression:

$$\begin{aligned}
S_n &= A_1 e^{-i\omega q_1 n \Delta x} + A_2 e^{-i\omega q_2 n \Delta x} \\
S_{n-1} &= A_1 e^{-i\omega q_1 (n-1) \Delta x} + A_2 e^{-i\omega q_2 (n-1) \Delta x} = A_1 e^{-i\omega q_1 n \Delta x} e^{i\omega q_1 \Delta x} + A_2 e^{-i\omega q_2 n \Delta x} e^{i\omega q_2 \Delta x} \\
S_{n-2} &= A_1 e^{-i\omega q_1 (n-2) \Delta x} + A_2 e^{-i\omega q_2 (n-2) \Delta x} = A_1 e^{-i\omega q_1 n \Delta x} e^{2i\omega q_1 \Delta x} + A_2 e^{-i\omega q_2 n \Delta x} e^{2i\omega q_2 \Delta x}.
\end{aligned} \tag{3.7}$$

Considering $\alpha_k = e^{i\omega q_k \Delta x}$ and substituting it to equation 3.7 leads to:

$$\begin{aligned}
S_n &= S_n^{(1)} + S_n^{(2)}, \\
S_{n-1} &= \alpha_1 S_n^{(1)} + \alpha_2 S_n^{(2)}, \\
S_{n-2} &= \alpha_1^2 S_n^{(1)} + \alpha_2^2 S_n^{(2)}.
\end{aligned} \tag{3.8}$$

Equation 3.8 can be written as:

$$\begin{pmatrix} S_{n-1} \\ S_{n-2} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1^2 & \alpha_2^2 \end{pmatrix} \begin{pmatrix} S_n^{(1)} \\ S_n^{(2)} \end{pmatrix} \tag{3.9}$$

To solve equation 3.9, one can use Carmer's rule method to solve a system of linear equations applying determinants.

The matrix $\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1^2 & \alpha_2^2 \end{pmatrix}$ will have an inverse $\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1^2 & \alpha_2^2 \end{pmatrix}^{-1}$ if and only if the determinant of $\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1^2 & \alpha_2^2 \end{pmatrix}$ is not equal to zero. It indicates that $\alpha_1 \neq \alpha_2$, where points to $e^{i\omega q_1 \Delta x} \neq e^{i\omega q_2 \Delta x}$ thus $q_1 \neq q_2$. Solving equation 3.9 leads to a linear relation between S_n, S_{n-1}, S_{n-2} :

$$S_n = Q_1 S_{n-1} + Q_2 S_{n-2}. \quad (3.10)$$

Embedding into a Hankel matrix structure

We continue our study by considering 2-D seismic data $S(\omega, x)$ in the frequency-space domain, assuming that the data are sampled regularly on the spatial dimension. At a given constant frequency ω_0 , the frequency slice can be indicated as $S(\omega_0, x) = [S_1, S_2, \dots, S_{N_x}]$, where N_x is the total number of traces along the spatial dimension. Embedding the monochromatic signal vector into a Hankel structured trajectory matrix leads to:

$$\mathbf{H} = \begin{pmatrix} S_1 & S_2 & \dots & S_{K_x} \\ S_2 & S_3 & \dots & S_{K_x+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{L_x} & S_{L_x+1} & \dots & S_{N_x} \end{pmatrix}, \quad (3.11)$$

where, as studied in chapter 2 the Hankel structured matrix generated from the lagged vectors is a $L_x \times K_x$ matrix and $K_x = N_x - L_x + 1$. Considering equation 3.6 for a 2-D data matrix with one linear event, the Hankel matrix will be:

$$\mathbf{H} = \begin{pmatrix} S_1 & X S_1 & \dots & X^{K_x-1} S_1 \\ S_2 & X S_2 & \dots & X^{K_x-1} S_2 \\ \vdots & \vdots & \ddots & \vdots \\ S_{L_x} & X S_{L_x} & \dots & X^{K_x-1} S_{L_x} \end{pmatrix}, \quad (3.12)$$

Considering equation 3.12, for a data set in the $f - x$ domain, it is apparent that the rank of the Hankel matrix is equal to one. However, the presence of random noise or missing traces will increase the rank of the Hankel matrix. For a data set presenting three linear events, equation 3.12 changes to equation 3.13, where the Hankel matrix is formed by the linear combination of the first three columns. Equation 3.13 can be extended to a record with k

linear events, leading to an increase in the Hankel matrix's rank to k . This indicates that the minimum Hankel matrix's rank can be established by identifying the number of linear events in a record.

$$\mathbf{H} = \begin{pmatrix} S_1 & S_2 & S_3 & Q_1S_3 + Q_2S_2 + Q_3S_1 & \dots & Q_1S_{k_x-1} + Q_2S_{k_x-2} + Q_3S_{k_x-3} \\ S_2 & S_3 & S_4 & Q_1S_4 + Q_2S_3 + Q_3S_2 & \dots & Q_1S_{k_x} + Q_2S_{k_x-1} + Q_3S_{k_x-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{L_x} & S_{L_x+1} & S_{L_x+2} & Q_1S_{L_x+2} + Q_2S_{L_x+1} + Q_3S_{L_x} & \dots & Q_1S_{N_x-1} + Q_2S_{N_x-2} + Q_3S_{N_x-3} \end{pmatrix}. \quad (3.13)$$

Determining the minimum rank by identifying the number of linear events could be one of the advantages of SSA; nonetheless, in real data or complicated records, it is hard to choose the exact number of linear events and, consequently, the Hankel matrix's rank. In the next chapter, I will discuss how to set the rank of the Hankel matrix automatically.

The next step is decomposing the Hankel matrix into its singular values and singular vector matrices. The rank-reduction method is performed by truncated SVD. We can summarize the embedding step as following:

$$\mathbf{H} = \mathcal{H}\mathbf{S}_{(\omega_0)}. \quad (3.14)$$

Singular value decomposition of the Hankel matrix

As discussed in chapter 2, the SSA algorithm's second step is decomposing the trajectory matrix into its singular value and singular vector matrices. The singular value decomposition of the Hankel matrix implemented by TSVD can be written as follow:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (3.15)$$

where, \mathbf{U} and \mathbf{V} are the singular vector matrices and $\mathbf{\Sigma}$ is the singular value matrix of the \mathbf{H} matrix.

The Hankel matrix's rank-reduction

In theory, for any given finite-rank Hankel matrices, SVD produces a basic description that explains the Hankel matrix's fundamental properties. The SVD decomposition of the Hankel matrix produces two orthogonal complex matrices, usually denoted \mathbf{U} and \mathbf{V} , and a diagonal matrix denoted $\mathbf{\Sigma}$ that contains the singular values in descending order. The number of non-zero singular values determines the Hankel matrix's rank. As mentioned before, the presence of uncorrelated noise and missing traces increase the Hankel matrix's rank. Consequently, if a record contains k linear events, a reduced Hankel matrix with information about only the signal events can be recovered by keeping only the first k singular values.

$$\mathbf{H}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H, \quad (3.16)$$

where \mathbf{U}_k and \mathbf{V}_k are the k first singular vectors of \mathbf{H} and $\mathbf{\Sigma}_k$ is a diagonal matrix with k largest singular values of \mathbf{H} . The rank-reduced Hankel matrix \mathbf{H}_k is in the form of Hankel; therefore, the signal in the constant frequency can be recovered by averaging along the anti-diagonals of its rank-reduced Hankel matrix. The rank-reduction step will be:

$$\mathbf{H}_k = \mathcal{R}\mathbf{H}. \quad (3.17)$$

Averaging along the anti-diagonals

The recovered Hankel matrix \mathbf{H}_k calculated earlier has a Hankel structure; therefore, averaging over the anti-diagonals leads to the recovered signal (Sacchi, 2009). Elements obtained from the diagonal averaging can be connected with the signal, whereas the residual series will be related mostly with noise or missing traces. The anti-diagonals averaging can be

written as:

$$\mathbf{S}_k = \begin{cases} \frac{1}{n} \sum_{l=1}^n \mathbf{H}_k(l, n-l-1) & \text{for } 1 \leq n \leq L_x \\ \frac{1}{L_x} \sum_{l=1}^{L_x} \mathbf{H}_k(l, n-l-1) & \text{for } L_x + 1 \leq n \leq K_x, \\ \frac{1}{K_x + L_x - n} \sum_{l=n-K_x+L_x}^{L_x} \mathbf{H}_k(l, n-l-1) & \text{for } K_x + 1 \leq n \leq N_x \end{cases} \quad (3.18)$$

where \mathbf{S}_k is the recovered signal in the constant frequency. Denoting \mathcal{A} as the operator that averages anti-diagonals; this step can be summed as:

$$\mathbf{S}_k = \mathcal{A}\mathbf{H}_k. \quad (3.19)$$

Inverse Fourier transform

SSA is implemented in the f-x domain for seismic data, so, with the inverse Fourier transform applying to each channel the rank-reduced data in the $t-x$ domain will be obtained. The \mathcal{F}^{-1} is the inverse Fourier operator:

$$\mathbf{S}_k(t, x) = \mathcal{F}^{-1}\mathbf{S}_k(\omega, x). \quad (3.20)$$

3.1.1 Projections Onto Convex Sets method

The six steps studied in the previous section are the essential SSA steps for random noise removal in seismic data. For data interpolation, we need to apply another algorithm that uses SSA to iteratively retrieve and incorporate information into the data. One of such algorithms is the Projection Onto Convex Sets (POCS), which can handle seismic data interpolation and random noise attenuation simultaneously. The POCS method is a helpful technique for domain transform-based interpolation methods. It was first proposed by Bregman (1965) for obtaining a common point of a finite collection of sets in a metric space. The first time Abma and Kabir (2006) introduced POCS for irregular seismic data interpolation. Gao et al. (2010) interpolated seismic data by using a Fourier transform-based POCS method with an

exponential threshold model. Oropenza and Sacchi (2011) used a weighted POCS method to interpolate seismic data using the SSA method.

Observed seismic data are always incomplete, irregular and contaminated with random noise, but the POCS-SSA algorithm can be used to alleviate these problems. One can connect the observed data with complete clean seismic data as follow:

$$\mathbf{S}_{obs} = \mathbf{R} \odot (\mathbf{S} + \eta), \quad (3.21)$$

where, \mathbf{S}_{obs} denotes the observed seismic data, \mathbf{S} depicts the full noiseless data, η represents the random noise and the residuals, and \mathbf{R} indicates a sampling matrix formed of zero and ones, given in equation 3.22.

$$\mathbf{R}(i, j) = \begin{cases} 1, & \text{for } \mathbf{S}(i, j) \neq 0, \\ 0, & \text{for } \mathbf{S}(i, j) = 0, \end{cases} \quad (3.22)$$

where, $\mathbf{R}(i, j) = 1$ is associated with observed traces, and $\mathbf{R}(i, j) = 0$ matches to a missing trace.

POCS iteratively replaces the result of the filtered data, which contains additional energy in the locations of the missing samples, into the original data set. In this iterative algorithm, the input data for the next iteration is the updated data from the last iteration. In each iteration, the amplitude of data is partially recovered therefore traces will be recovered after some iterations.

3.1.2 Reconstruction of missing traces

We discussed in the previous section that the random noise increases the rank of the Hankel matrix. The same happens for data with missing traces since the abrupt amplitude change on the zero traces requires additional data singular vectors to be represented. Therefore, it follows that reducing the rank of the data can be used to infill missing traces. This section

pays attention to the reconstruction of incomplete records. To reconstruct data with gaps, we utilize a weighted algorithm presented by Oropeza and Sacchi (2011).

First, let us assume a noiseless gapped record with nine channels ($N_x = 9$) in the Fourier domain in a constant frequency to begin the SSA algorithm for reconstructing the data with missing traces. Missing traces in the Fourier domain will be presented as zero value:

$$\mathbf{S} = \begin{pmatrix} S(\omega_0, x_1) & 0 & S(\omega_0, x_3) & S(\omega_0, x_4) & 0 & 0 & S(\omega_0, x_7) & 0 & S(\omega_0, x_9) \\ S(\omega_1, x_1) & 0 & S(\omega_1, x_3) & S(\omega_1, x_4) & 0 & 0 & S(\omega_1, x_7) & 0 & S(\omega_1, x_9) \\ \vdots & \vdots \\ S(\omega_p, x_1) & 0 & S(\omega_p, x_3) & S(\omega_p, x_4) & 0 & 0 & S(\omega_p, x_7) & 0 & S(\omega_p, x_9) \end{pmatrix}, \quad (3.23)$$

where p is the number of samples in the Fourier domain and zeros are the missing traces in the seismic record. Considering the signal in a constant frequency (ω_m), the data will be as follow:

$$\mathbf{S}(\omega_m, x_1) = [S(\omega_m, x_1), 0, S(\omega_m, x_3), S(\omega_m, x_4), 0, 0, S(\omega_m, x_7), 0, S(\omega_m, x_9)]. \quad (3.24)$$

In the next step, we require to generate a sampling operator that has ones for the locations corresponding to recorded components and zeroes for locations of missing values. The sampling operator of equation 3.22 now looks like:

$$\mathbf{R} = [1, 0, 1, 1, 0, 0, 1, 0, 1]. \quad (3.25)$$

In the embedding step for the constant frequency (ω_m), the Hankel matrix becomes:

$$\mathbf{S}_{(\omega_m)} = \begin{pmatrix} S(\omega_m, x_1) & 0 & S(\omega_m, x_3) & S(\omega_m, x_4) & 0 \\ 0 & S(\omega_m, x_3) & S(\omega_m, x_4) & 0 & 0 \\ S(\omega_m, x_3) & S(\omega_m, x_4) & 0 & 0 & S(\omega_m, x_7) \\ S(\omega_m, x_4) & 0 & 0 & S(\omega_m, x_7) & 0 \\ 0 & 0 & S(\omega_m, x_7) & 0 & S(\omega_m, x_9) \end{pmatrix}. \quad (3.26)$$

The missing traces increase the Hankel matrix's rank, which is similar to the presence of random noise in the record shown in figure 3.2. The rank reduction step leads to gradually reconstructing the missing traces and can be used together with an iterative algorithm to replace the missing traces with the reconstructed traces.

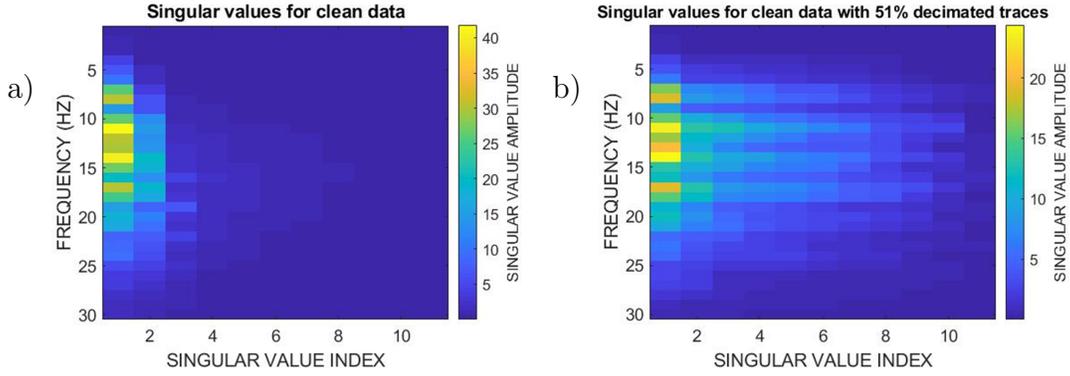


Figure 3.2: a) Singular values for noiseless data without decimated traces; b) singular values for noiseless data with 51% decimated traces.

Equation 3.27 contains the six steps of SSA combined with the iterative algorithm. It is a practical approach for random noise attenuation and seismic data amplitude reconstruction.

$$\tilde{\mathbf{S}}_{n+1}(\omega_m) = \alpha_n \mathbf{S}_{obs} + (\mathbf{I} - \alpha_n \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{H} \mathcal{F}(\mathbf{S}_n(\omega_m)) + (\mathbf{I} - \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{H} \mathcal{F}(\mathbf{S}_n(\omega_m)), n = 1, 2, \dots, N \quad (3.27)$$

where ω_m is the frequency that SSA is applied on the signals, \mathbf{I} is the matrix of ones, $\alpha_n \in (0, 1]$ is the weight factor and linearly changes with the number of iterations (Oropeza and Sacchi, 2011). \mathbf{S}_{obs} denotes the observed data, \mathcal{F} indicates the Fourier transform, and \mathcal{F}^{-1} shows the inverse Fourier transform. \mathcal{H} points to the Hankelization operator, \mathcal{R} reveals

rank reduction operator, and \mathcal{A} gives the averaging anti-diagonal operator. The algorithm converges when the maximum number of the iteration is reached or the energy of the change in the recovered traces is less than a threshold and the missing traces will be interpolated (Oropeza and Sacchi, 2011).

3.2 Multichannel Singular Spectrum Analysis

One of the advantages of SSA is its simplicity to extend to more dimensions (Oropeza and Sacchi, 2011). In the previous section we explained how to use SSA for seismic interpolation; in this section we will discuss multichannel singular spectrum analysis (MSSA) that permits to apply SSA in multidimensions. For 2D data, we just needed to apply SSA to each frequency slice; for 3D data we will apply MSSA to two spatial dimensions in each frequency slice. MSSA is similar to SSA but two more steps are required to include the second spatial dimension.

3.2.1 Methodology

Before studying the application of MSSA for 3-D interpolation, let us see how the algorithm of MSSA works.

Consider $\mathbf{S}(t, x, y)$ a block of 3-D data in the $t - x - y$ domain on N_t by N_x by N_y samples. ($t = 1, \dots, N_t$), ($x = 1, \dots, N_x$), ($y = 1, \dots, N_y$). In the frequency domain, the data are represented as $\mathbf{S}(\omega, x, y)$ and ($\omega = 1, \dots, N_\omega$). Each frequency slice of the data at a given

frequency ω_m , can be represented by the following matrix:

$$\mathbf{S}(\omega_m) = \begin{pmatrix} S(1,1) & S(1,2) & \dots & S(1,N_x) \\ S(2,1) & S(2,2) & \dots & S(2,N_x) \\ \vdots & \vdots & \ddots & \vdots \\ S(N_y,1) & S(N_y,2) & \dots & S(N_y,N_x) \end{pmatrix}. \quad (3.28)$$

To avoid notational confusion, let's ignore the argument ω_m . Then, construct a Hankel matrix from each inline of \mathbf{S} ; for the inline i th the Hankel matrix in the m th frequency slice would be:

$$\mathbf{H}_i = \begin{pmatrix} S(i,1) & S(i,2) & \dots & S(i,l) \\ S(i,2) & S(i,3) & \dots & S(i,l+1) \\ \vdots & \vdots & \ddots & \vdots \\ S(i,N_x-l+1) & S(i,N_x-l+2) & \dots & S(i,N_x) \end{pmatrix}. \quad (3.29)$$

At this step, we have N_y Hankel matrices of each inline. Then MSSA constructs a block Hankel matrix \mathbf{M} from Hankel matrices \mathbf{H}_i :

$$\mathbf{M} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_n \\ \mathbf{H}_2 & \mathbf{H}_3 & \dots & \mathbf{H}_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_y-n+1} & \mathbf{H}_{N_y-n+2} & \dots & \mathbf{H}_{N_y} \end{pmatrix}. \quad (3.30)$$

The size of \mathbf{M} is $I \times J$, where $I = (N_x - m + 1)(N_y - n + 1)$ and $J = mn$. The integers m and n are chosen to make the Hankel matrices of \mathbf{H} and the block Hankel matrix of \mathbf{M} square matrices or close to square. As we see, including the two spatial dimensions of the 3-D cube for each frequency can make the block Hankel matrix very large.

To study how the block Hankel matrix columns are related to each other, let us consider a 3-D record including a coherent event with a constant dip, that is a plane wave. The

waveform in the time domain with the spatial dimensions of x and y will be expressed as:

$$S(t, x, y) = w(t - p_x x - p_y y), \quad (3.31)$$

where p_x and p_y are the constant dips in the x and y dimension, respectively. The waveform in the $f - x - y$ domain can be described as:

$$S(\omega, x, y) = A(\omega)e^{-i\omega(p_x x + p_y y)}, \quad (3.32)$$

where ω is the temporal frequency, and A is the amplitude. Let's consider data with n number of channels in the x -direction and m number of channel in y -direction then we will have $x = n\Delta x$, and $y = m\Delta y$. Then equation 3.32 can be written as $S(\omega, x, y) = A(\omega)e^{-i\omega p_x n\Delta x}e^{-i\omega m\Delta y}$. To avoid notational clutter, let's assume that the waveform in the channel n th and m th is in the form of S_{nm} . The waveform for the adjacent channel of the channel m is represented as:

$$S_{n(m-1)} = A(\omega)e^{-i\omega p_x n\Delta x}e^{-i\omega p_y (m-1)\Delta y} = A(\omega)e^{-i\omega p_x n\Delta x}e^{-i\omega p_y m\Delta y}e^{i\omega p_y \Delta y}, \quad (3.33)$$

Simplifying equation 3.33 we obtain $S_{n(m-1)} = A(\omega)S_{nm}e^{i\omega p_y \Delta y}$ where $e^{i\omega p_y \Delta y}$ is a constant term. Denoting $R = A(\omega)e^{i\omega p_y \Delta y}$, we can write the relationship between the two neighbouring channels as:

$$S_{n(m-1)} = RS_{nm}. \quad (3.34)$$

Similarly to the previous section, equation 3.34 presents a linear relationship between each Hankel matrix and it's adjacent Hankel matrix. ($\mathbf{H}_i = X\mathbf{H}_{i-1}$). The block Hankel matrix

in equation 3.30 can be substituted by:

$$\mathbf{M} = \begin{pmatrix} \mathbf{H}_1 & X\mathbf{H}_1 & \dots & X^{n-1}\mathbf{H}_1 \\ \mathbf{H}_2 & X\mathbf{H}_2 & \dots & X^{n-1}\mathbf{H}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_y-n+1} & X\mathbf{H}_{N_y-n+1} & \dots & X^{n-1}\mathbf{H}_{N_y-n+1} \end{pmatrix}. \quad (3.35)$$

Equation 3.35 shows that the block Hankel matrix's rank for the given plane wave is equal to one. Moreover, similar to SSA, it can be extended to k plane waves for a 3-D data cube. The other steps of MSSA are the same as SSA. For the block Hankel matrix decomposition, one can apply TSVD and decompose the block Hankel matrix to its singular values and singular vector matrices. After the rank-reduction step, the averaging along the anti-diagonals is utilized to recover the data. The last step is to reverse the Fourier transformation to convert the data back to the $t - x - y$ domain. Figure 3.3 is a schematic of the MSSA algorithm in generating block Hankel matrix step for a 3-D data set.

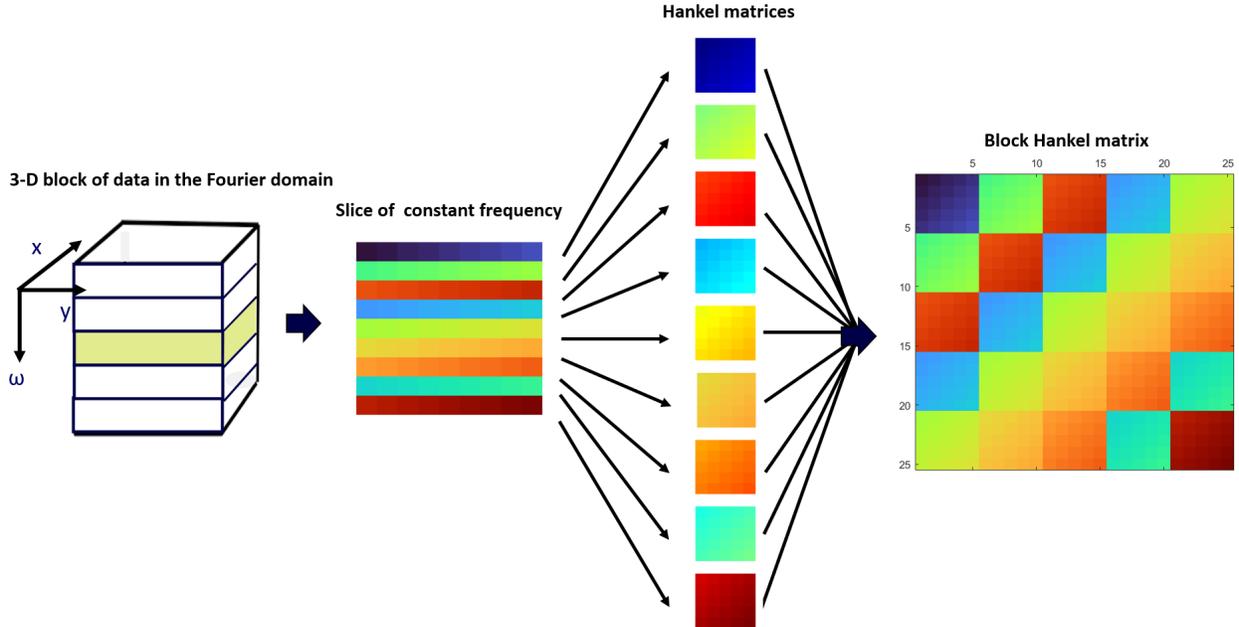


Figure 3.3: Schematic of generating a block Hankel matrix from a frequency slice.

By applying MSSA, up to this point data have been denoised but not interpolated. To reconstruct the missing traces, let's go back to equation 3.28 and consider a 3-D data cube with five traces in each spatial dimension. Let's assume the frequency slice of the data in $\omega = 0$ as:

$$\mathbf{S}(\omega_0) = \begin{pmatrix} S(1,1) & 0 & S(1,3) & S(1,4) & S(1,5) \\ S(2,1) & S(2,2) & 0 & S(2,4) & 0 \\ 0 & S(3,2) & S(3,3) & 0 & S(3,5) \\ S(4,1) & S(4,2) & 0 & S(4,4) & S(4,5) \\ S(5,1) & 0 & S(5,3) & 0 & S(5,5) \end{pmatrix}. \quad (3.36)$$

The sampling matrix for the observed data in the frequency slice of the data $\omega = 0$ is as follow:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \quad (3.37)$$

The same POCS algorithm discussed in the previous chapter can be applied to reconstruct missing traces or gaps in a 3-D data cube for each frequency as follows:

$$\tilde{\mathbf{S}}_{n+1}(\omega_m) = \alpha_n \mathbf{S}_{obs} + (\mathbf{I} - \alpha_n \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{B} \mathcal{F} \left(\tilde{\mathbf{S}}_n(\omega_m) \right) + (\mathbf{I} - \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{B} \mathcal{F} \left(\tilde{\mathbf{S}}_n(\omega_m) \right), \quad n = 1, 2, \dots, N \quad (3.38)$$

where ω_m is the temporal frequency, \mathbf{I} is a matrix of ones, $\alpha_n \in (0, 1]$ is a weight factor which decreases linearly with iterations, \mathcal{F} indicates the Fourier transform, and \mathcal{F}^{-1} shows the inverse Fourier transform. \mathcal{B} points to the Hankelization operator to generate the block Hankel matrix, \mathcal{R} reveals rank reduction operator, and \mathcal{A} gives the averaging anti-diagonal operator. For clean data without noise we consider $\alpha = 1$, so equation 3.38 changes to the

following:

$$\tilde{\mathbf{S}}_{n+1}(\omega_m) = \mathbf{S}_{obs} + (\mathbf{I} - \mathbf{R})\mathcal{F}^{-1}\mathcal{ARBF}\left(\tilde{\mathbf{S}}_n(\omega_m)\right). \quad (3.39)$$

3.2.2 Synthetic data example

In this section, we test the MSSA method for a 3-D post-stack synthetic data. The first test is a cube of 3-D noiseless data containing four linear events with 51% missing traces. Figure 3.4 (a) shows the cube of input data. Figure 3.4 (b) indicates the same 3-D cube without missing data (ground-truth). The interpolated data applying the MSSA algorithm with ten iterations is shown in figure 3.4 (c). The selected rank for the interpolation algorithm was set rank=4. Figure 3.4 (d) is the cube of residual errors derived from subtracting initial data from interpolated data. Figure 3.5 refers to the slices of the data in figure 3.4 in inline direction and one in cross-line direction. Figure 3.5 (a) and (b) shows a slice of input data in inline and cross-line directions respectively. Figure 3.5 (c) and (d) indicate slices of the initial data in inline and cross-line directions respectively. Figure 3.5 (e) and (f) are the result of interpolation with MSSA in inline and cross-line directions respectively. Figure 3.5 (g) and (h) are the slices of residual errors in inline and cross-line respectively.

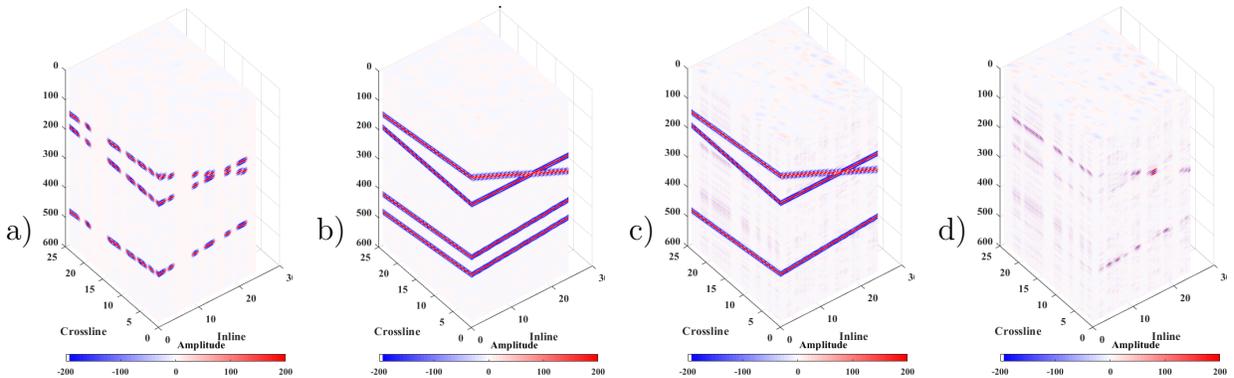


Figure 3.4: Interpolation of a 3-D noiseless data containing four linear events. (a) Input data with 51% gaped traces, (b) initial data, (c) interpolated data with MSSA rank=4, (d) residual errors.

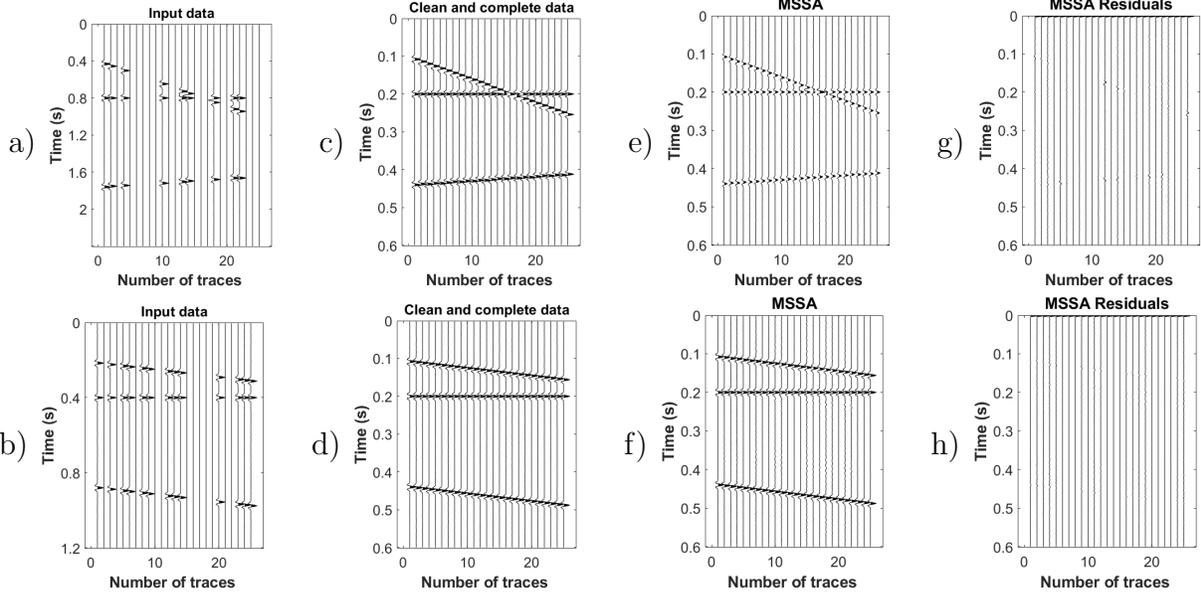


Figure 3.5: Slices of data in inline and cross-line directions. (a) input data in inline direction, (b) input data in cross-line direction. (c) and (d) Initial data in inline and cross-line direction respectively. (e) and (f) Interpolated data by MSSA with rank=4 in inline and cross-line direction respectively. (g) and (h) The result of subtracting initial data from the interpolated data in inline and cross-line direction respectively.

To test the capability of the algorithm in simultaneous denoising and interpolating the second example is 3-D data with additive random noise SNR=1 and 50% missing traces. The algorithm converged after 15 iterations. Figure 3.6 (a) is the cube of input data with SNR=1 and 50% missing traces. Figure 3.6 (b) is the 3-D cube of initial data. Figure 3.6 (c) shows the result of applying the MSSA algorithm with 15 iterations. Figure 3.6 (d) is the cube of residual errors derived from subtracting the initial data from the interpolated data. Figure 3.7 indicates the slices of the data in figure 3.6 in inline directions and one in cross-line direction. Figure 3.7 (a) and (b) shows a slice of input data in inline and cross-line direction respectively. Figure 3.7 (c) and (d) indicate slices of the initial data in inline and cross-line directions respectively. Figure 3.7 (e) and (f) are the result of interpolation with MSSA in inline and cross-line directions respectively. Figure 3.7 (g) and (h) are the slices of residual errors in inline and cross-line respectively.

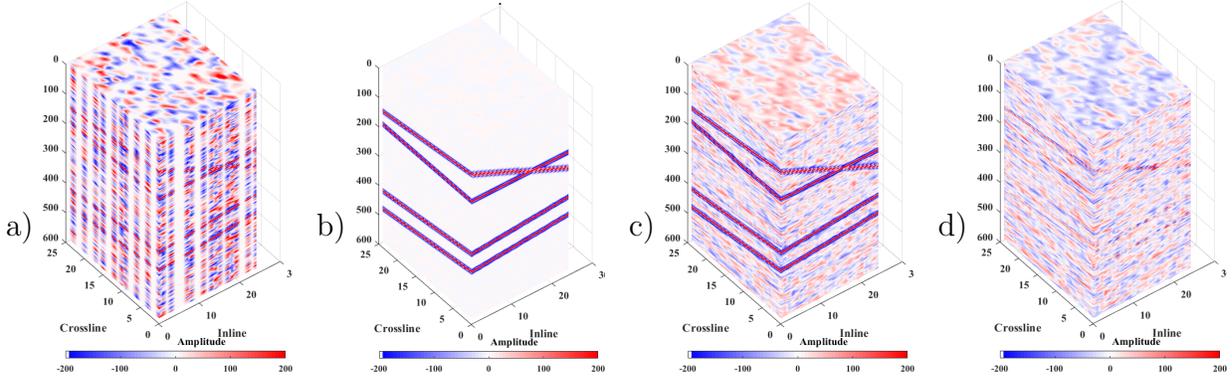


Figure 3.6: Interpolation of a 3-D data containing four linear events. (a) Input data with SNR=1 and 51% missing traces, (b) initial data, (c) interpolated data with MSSA rank=4, (d) residual errors.

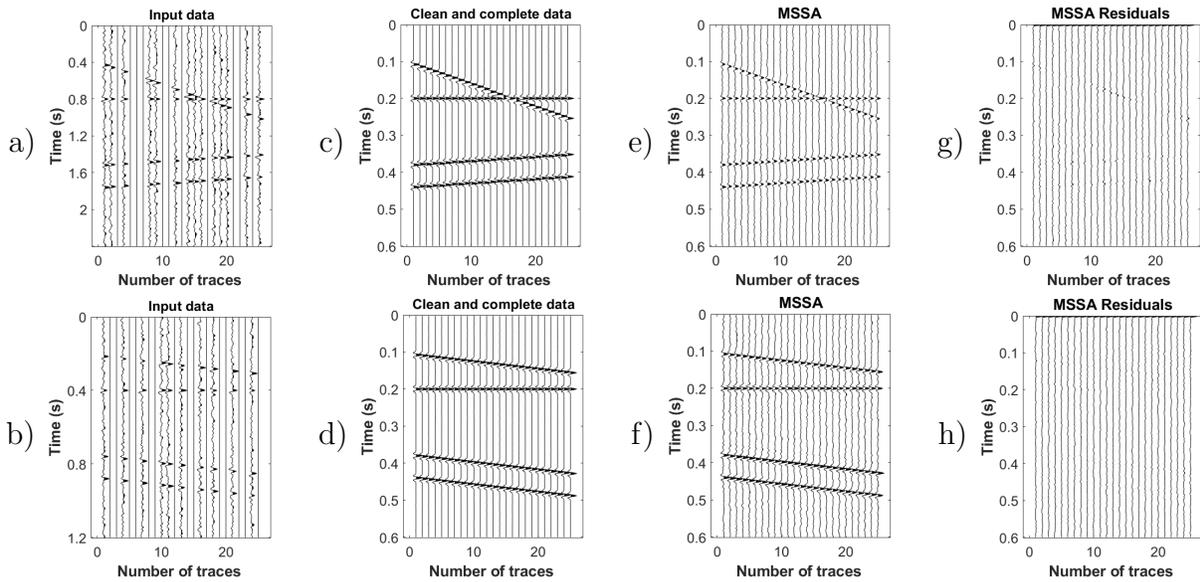


Figure 3.7: Slices of data in inline and cross-line directions. (a) input data in inline direction, (b) input data in cross-line direction. (c) and (d) Initial data in inline and cross-line direction respectively. (e) and (f) Interpolated data by MSSA with rank=4 in inline and cross-line direction respectively. (g) and (h) The result of subtracting initial data from the interpolated data in inline and cross-line direction respectively.

3.2.3 Discussions

This chapter examined the efficiency of SSA (for 2-D seismic data) and MSSA (for 3-D seismic data) on seismic data interpolation. It studied how the rank of Hankel matrices depends on the number of linear events in the processing window. The drawback of this

method of selecting rank for the rank-reduction step is that real data are too complicated for us to decide the exact number of linear events in each processing window. Therefore, we need to search for a data-dependent strategy that does not require our guessing. In the next chapter, we will study an adaptive approach to choose a target rank in the rank-reduction step of the SSA algorithm.

Chapter 4

Adaptive rank reduction

SSA is a robust low-rank reduction method for simultaneous denoising and reconstruction. However, one of its weaknesses is that it requires to satisfy the plane wave assumption. For instance, complicated structures with large curvatures are not suitable for low rank-reduction methods like SSA. One solution for the problem of curved events is using NMO correction before applying the low rank-reduction methods. Another solution is using local spatial windows to assume that the events are linear. Furthermore, it is challenging to find the proper window size as it is not easy to determine whether the local window structure is linear or not. Besides, it is hard to estimate the rank of each window separately. All the above reasons make the estimation of the Hankel matrix's rank complicated. On the other hand, selecting the wrong rank will negatively affect the results: rank overestimation will make residuals to increase; rank underestimation will distort the signal and create artifacts. This section discusses a method that automatically chooses the rank for each local window suggested by Wu and Bai (2018) for 2-D data and continues the expansion of the method for 3-D data.

4.1 Adaptive rank reduction for the Hankel matrix

In seismic data, the observed data can be indicated as $\mathbf{S}_{obs} = \mathbf{R}(\mathbf{S}_0 + \eta)$. Where \mathbf{S}_{obs} is observed seismic data, \mathbf{S}_0 indicates the full noiseless data, η represents the random noise and the residuals, and \mathbf{R} indicates the sampling matrix formed of zeros and ones. Using the SSA algorithm, we can write the Hankel matrix of observed data as:

$$\mathbf{M} = \mathbf{S} + \mathbf{N}, \quad (4.1)$$

where \mathbf{S} represents low rank Hankel matrix of the desired signal and \mathbf{N} denotes the noise component which includes the gaps. In SSA, decomposition of the Hankel matrix in the rank-reduction step using TSVD will be as follow :

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H. \quad (4.2)$$

If one knows the desired rank of the Hankel matrix, the estimated signal desired signal is recovered by:

$$\hat{\mathbf{M}} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H, \quad (4.3)$$

where $\hat{\mathbf{M}}$ is the estimated signal, and k is the predefined rank of the Hankel matrix equals to the number of the linear events in each local window.

In this section, we are looking for an operator to choose the Hankel matrix's rank automatically without any information about the number of linear events of the processing window. The rank of the Hankel matrix is related to the number of distinctly large singular values. To prove this idea, figure 4.1 examines how the singular values are related to the number of events by investigating the distribution of the singular values. The first data set contains one flat linear event with 51 traces (figure 4.1 (a)). Figure 4.1 (b) presents the distribution of the singular values of the Hankel matrix of the clean data in a constant frequency of 25 Hz. The largest non-zero singular value is related to the flat event's energy. Figure 4.1 (c)

represents data with random noise and missing traces. Figure 4.1 (d) shows the singular value distribution of the Hankel matrix of incomplete data in 25 Hz. We can see that the presence of missing traces and random noise has increased the Hankel matrix's rank in the frequency of 25 Hz; however, the flat event's energy can be diagnosed with the first distinct large singular value.

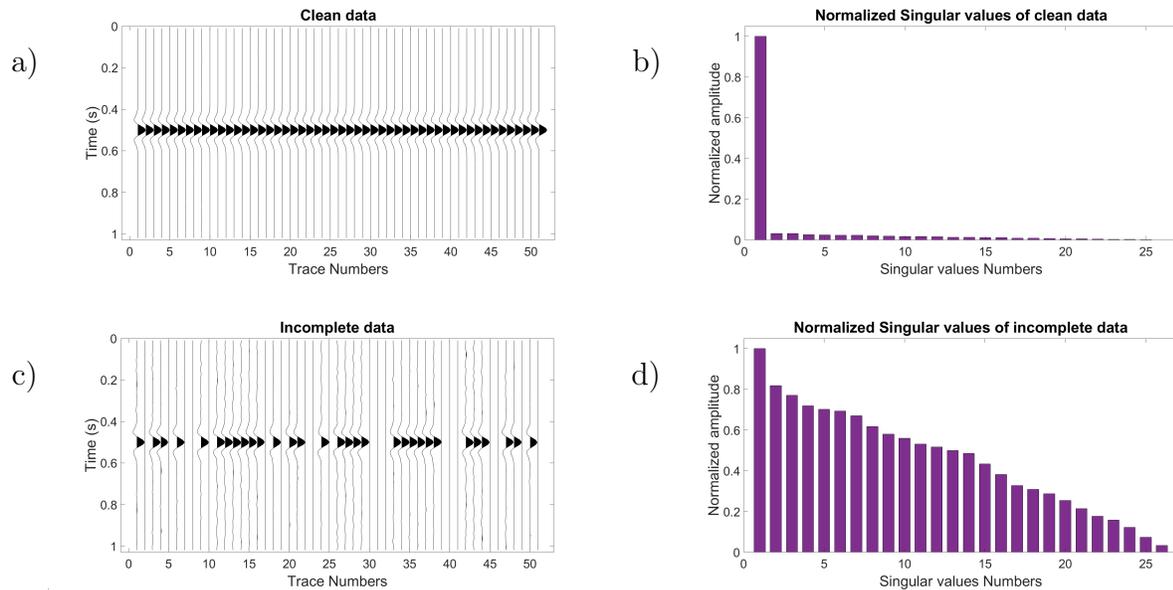


Figure 4.1: Analyzing the distribution of the singular values for data containing one linear flat event. a) Clean data. b) Normalized singular values of the Hankel matrix for the clean data. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for incomplete data.

Figure 4.2 demonstrates the same test with different data including an inclined linear event. Figure 4.2 (a) displays clean data. Figure 4.2 (b) shows the normalized singular value distribution of its Hankel matrix in 25 Hz; the non-zero singular values are related to the linear event's energy. Figure 4.2 (c) shows the incomplete data, and figure 4.2 (d) refers to the normalized singular values distribution of the incomplete data in 25 Hz that displays one clear singular value with others being.

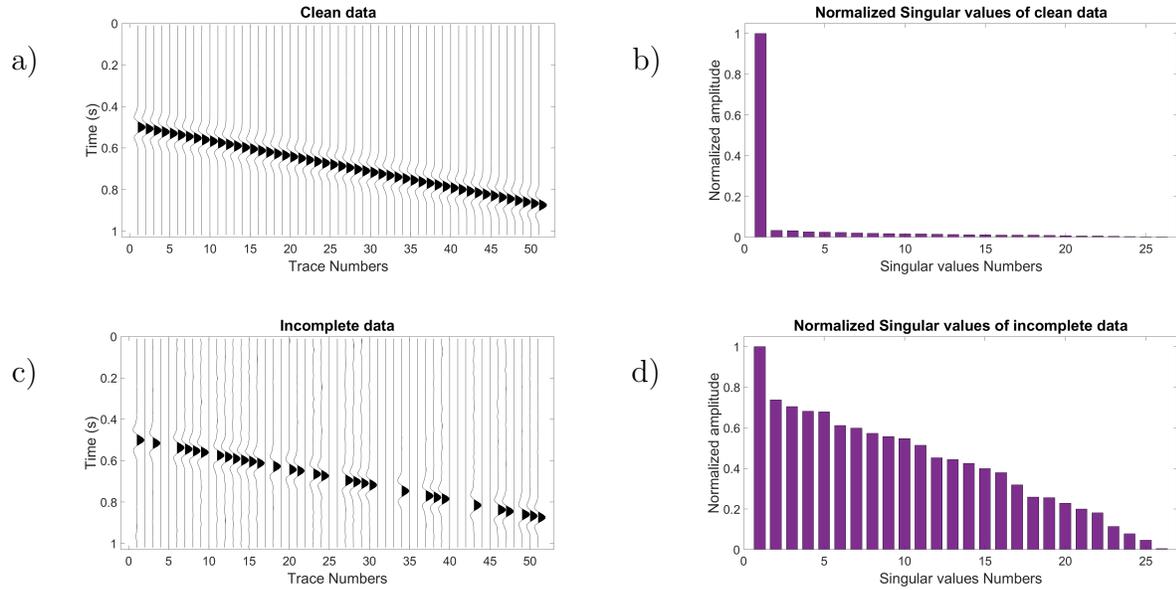


Figure 4.2: Analyzing the distribution of the singular values for data containing one linear inclined event. a) Clean data. b) Normalized singular values of the Hankel matrix for the clean data in constant frequency 25 Hz. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for the incomplete data in constant frequency 25 Hz.

We repeat the test with a 2-D data set containing four linear events to explore the effect of multi-event data. Figure 4.3 (a) represents clean data with four linear events. Figure 4.3 (b) shows the normalized singular value distribution of the Hankel matrix of clean data in a constant frequency of 25 Hz, where the first four large singular values represent the energy of events. Figure 4.3 (c) represents the data containing four linear events contaminated with random noise and missing traces. Figure 4.3 (d) shows the normalized distribution of singular values of the Hankel matrix for incomplete data in a constant frequency of 25 Hz. One can see that although the presence of missing traces and random noise has increased the Hankel matrix's rank, the linear events' energy can still be diagnosed with an abrupt drop of energy between the fourth singular value and the fifth singular value.

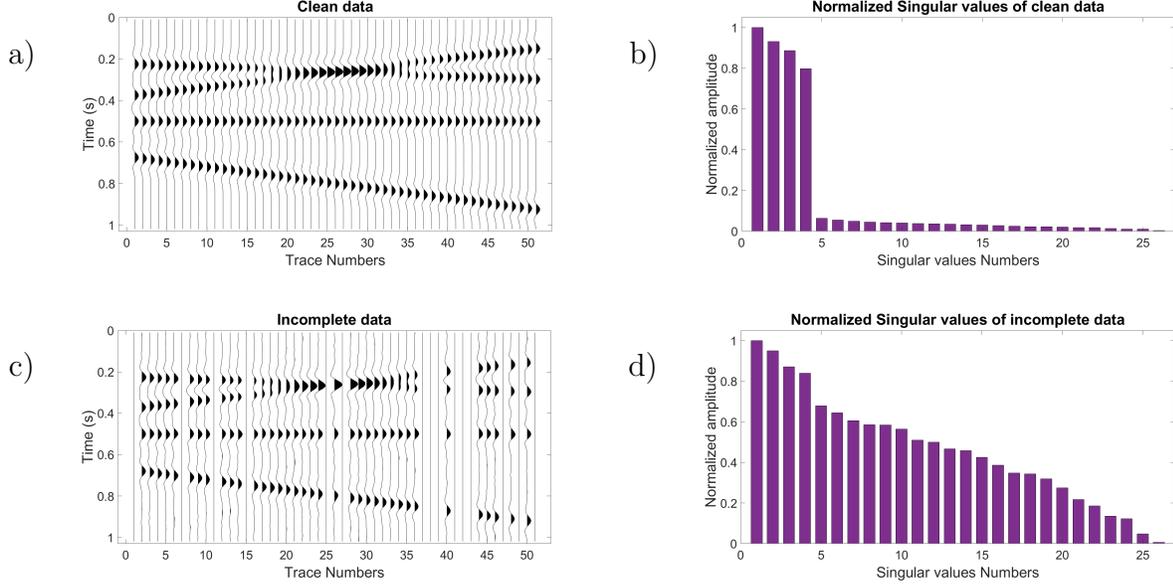


Figure 4.3: Analyzing singular values distribution of data containing four linear events. a) Clean data. b) Normalized singular values of the Hankel matrix for clean data in a frequency slice of 25 Hz. c) Incomplete data contaminated with random noise. d) Normalized singular values of the Hankel matrix for the incomplete data in constant frequency 25 Hz.

In the adaptive rank reduction (ARR) method, we want to find the cutoff number that indicates where the contribution from the signal becomes much less than the contribution of the missing traces or random noise.

To determine the rank of the Hankel matrix, a rule of thumb is that the rank related to the number of linear events equals to:

$$\frac{\sigma_{i+1}}{\sigma_1} \leq \frac{1}{\sqrt{N}} \quad \text{and} \quad \frac{\sigma_i}{\sigma_1} > \frac{1}{\sqrt{N}}, \quad i = 1, \dots, k \quad (4.4)$$

where N is the number of observations (linear events), σ_i is the i th singular value of the Hankel matrix in each frequency (Aoki, 2013).

Equation 4.4) proves that the desired rank is related to the number of linear observations. However from the tests in figures 4.1 to 4.3 , one can draw the conclusion that the number of events in each local window equals the number of large singular values. It satisfies the

rule that the desired rank of complete noiseless data is equal to the largest number of the non-zero singular values. We can conclude that the distribution of the singular values is divided into two clusters of the signal and residual energies. Wu and Bai (2018) proposed that the cutoff where the energy ratio between two adjacent singular values is maximum indicates the point where the signal is grouped from the residuals.

$$\mathcal{T}(\boldsymbol{\Sigma}, k) = \max_i \frac{\sigma_i^2}{\sigma_{i+1}^2}, \quad (4.5)$$

where σ_i is the i th singular value of the Hankel matrix in each frequency. $\mathcal{T}(\boldsymbol{\Sigma}, k)$ indicates the operator that finds the rank at the point where the two following singular values become more scattered. k can be introduced as the optimal rank of the Hankel matrix for each processing window in a constant frequency .

Using equation 4.5 by substituting $\boldsymbol{\Sigma}_k = \mathcal{T}(\boldsymbol{\Sigma}, k)$ we can improve the results of rank-reduction step in equation 4.2 as follow:

$$\hat{\mathbf{M}} = \mathbf{U}_k \mathcal{T}(\boldsymbol{\Sigma}, k) \mathbf{V}_k^H, \quad (4.6)$$

Utilizing equation 4.6 on the rank-reduction step of SSA, the recovered Hankel matrix is reconstructed. The next step is averaging anti-diagonals of the recovered Hankel matrix that recover the signal in the Fourier domain for each frequency slice. Applying the iterative algorithm will reconstruct the amplitude of missing traces.

4.1.1 Examples of adaptive rank-reduction for 2-D data

In this section, the performance of choosing constant rank and also applying the ARR method in denoising and interpolation is shown by applying SSA to 2-D synthetic data.

To judge the reconstruction results numerically, we use Quality Factor (QF) of interpolation defined by:

$$QF = 10 \log_{10} \left(\frac{\|d_0\|_2^2}{\|d_f - d_0\|_2^2} \right), \quad (4.7)$$

where d_0 is the clean and complete data, and d_f is the result after applying an interpolation algorithm. This allows us to test the efficiency of the interpolation method.

Figure 4.4 (a) Shows the synthetic data containing nine hyperbolic events, which is the desired output. Figure 4.4 (b) indicates input data with $SNR = 100$ and 40% of the muted traces. To compare the traditional rank-reduction (TRR) with the ARR method the local spatial window is set to 23 input traces with 11 traces overlap. The number of iterations for each method is 25, and the SSA algorithm is applied for the frequency range of 1 to 100 Hz for all tests. Figure 4.4 (c), (e), (g), and (i) show the results of applying SSA with constant rank $k = 4, 6, 8,$ and $10,$ respectively. The calculated quality of interpolation is $QF = 8.15, 5.4, 3.11, 2.12$ dB, respectively. Figures 4.4 (d), (f), (h), and (j) indicate the residual errors of applying SSA with the predefined rank $k = 4, 6, 8,$ and $10,$ respectively. Figure 4.4 (k) represents the result of SSA interpolation using ARR method, the interpolation quality is $QF = 10.83$ dB. Figure 4.4 (l) shows the residual error for the ARR method.

Comparing figures 4.4 (c), (e), (g), (i), and (k), the best reconstruction performance is when the rank is selected $k = 4$. The performance of the results decreases by increasing the rank. However, the best results come from the ARR method.

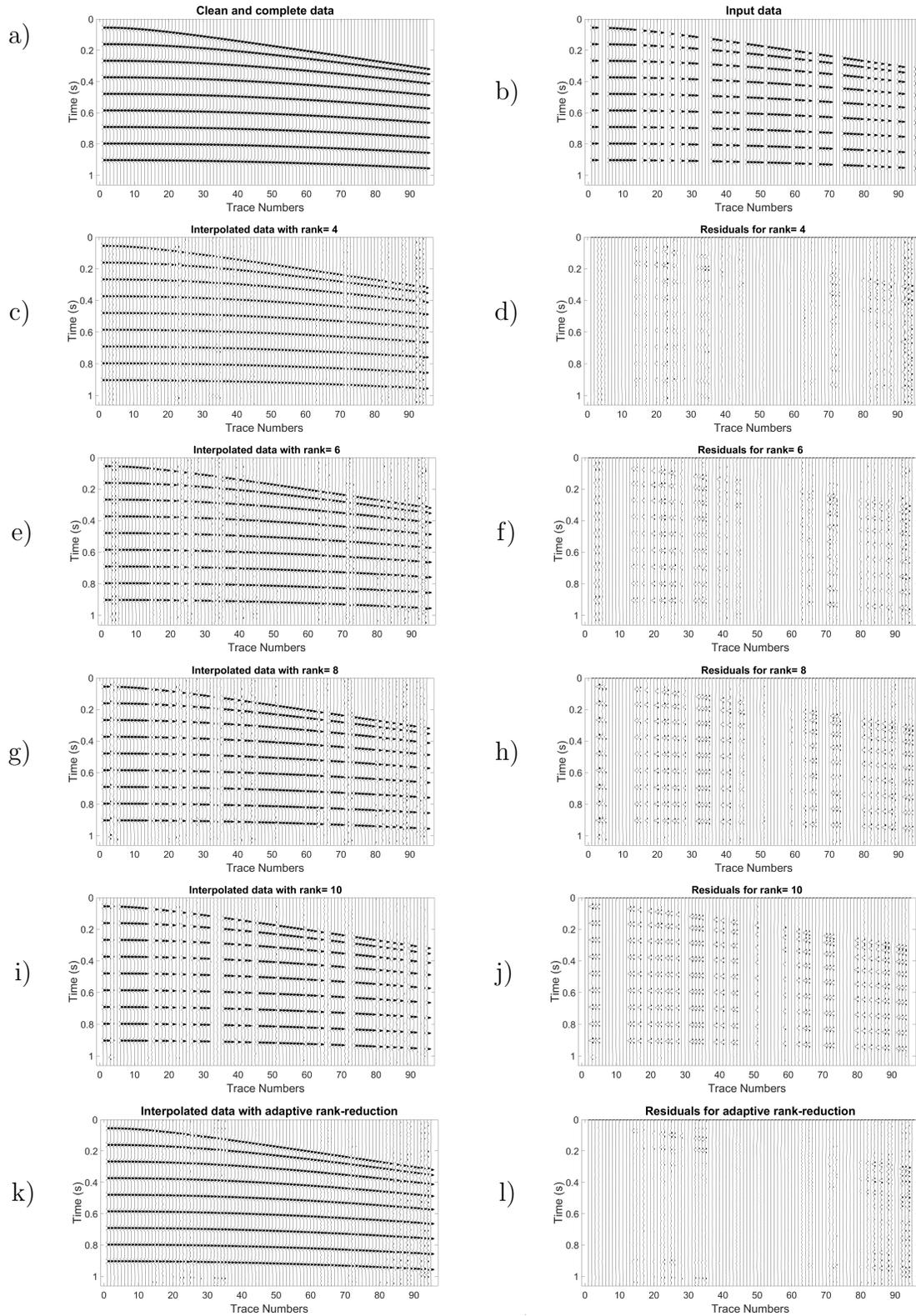


Figure 4.4: Comparing the interpolation results of SSA with TRR method and ARR method. a) Desired result, b) input data with $SNR = 100$ and 40% missing, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10, respectively. k) Interpolated data with SSA using ARR. d), f), h), j), and l) residual errors for c), e), g), i), and k), respectively.

In the presence of noise and missing traces, the behaviour of the distribution of singular values of the Hankel matrix would change. The reason is the projection of the noise component on the signal component. To understand the effect of noise on the rank, in this section we run tests with different levels of signal to noise ratio.

Figure 4.5 (a) Shows the clean and complete synthetic data. Figure 4.5 (b) represents input data with $SNR = 2$ and 40% of the traces were killed. Local spatial windows are set with 23 input traces each with an overlap of 11 traces. We run 25 iterations for each method in the range of frequencies 1 to 100 Hz. Figure 4.5 (c), (e), (g), and (i) shows the results of applying SSA with constant rank $k = 4, 6, 8,$ and 10 where the interpolation quality factor $QF = 4.81, 3.39, 2.17, 1.46$ dB, respectively. Figures 4.5 (d), (f), (h), and (j) indicate the residual errors of applying SSA with the predefined rank $k = 4, 6, 8,$ and $10,$ respectively. Figure 4.5 (k) represents the result of interpolation where SSA is applied using ARR it's output interpolation quality is $QF = 6.17$ dB. Figure 4.5 (l) shows the residual errors for ARR.

The results show that the ARR method works perfectly even when the signal quality is poor in the presence of noise.

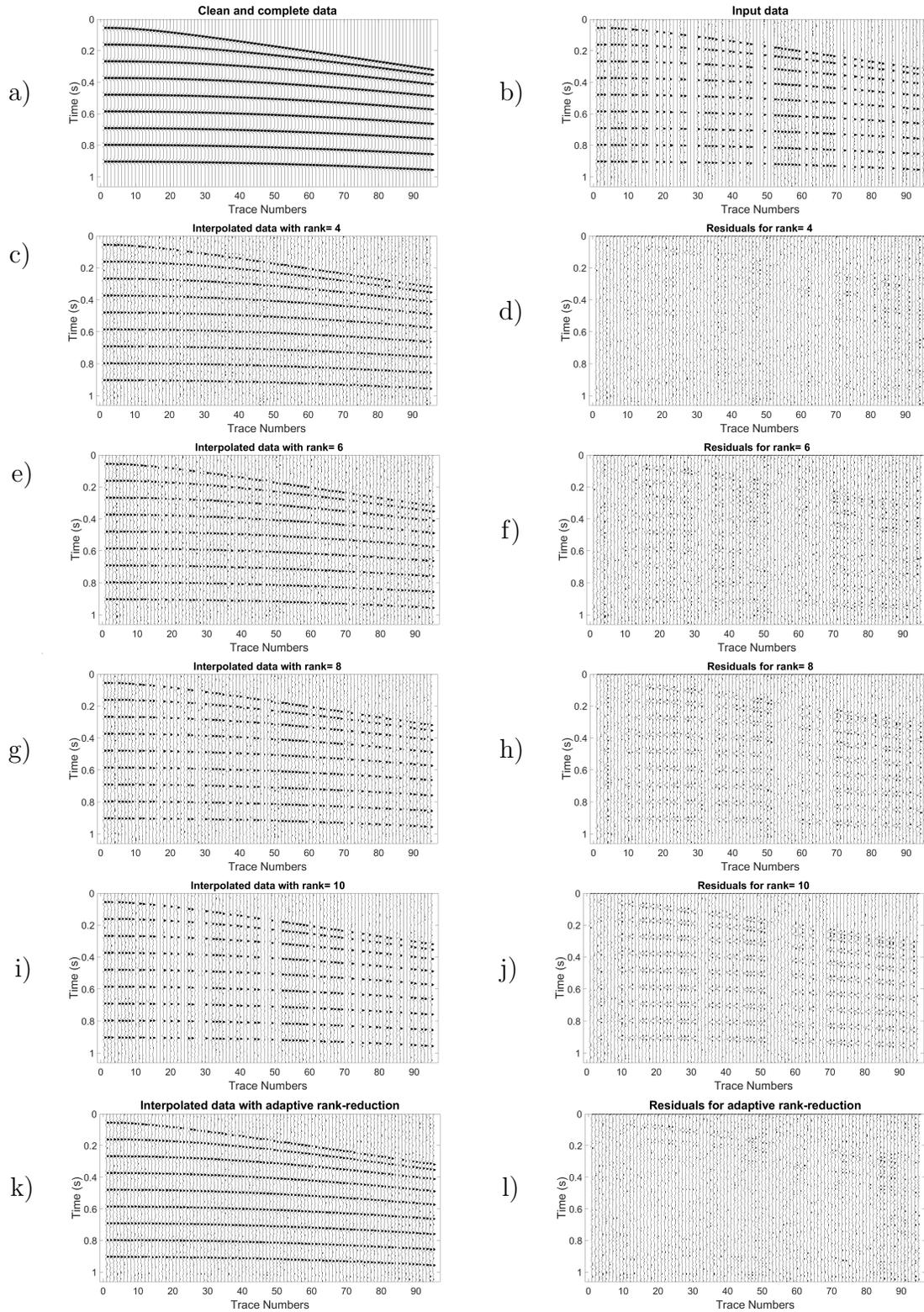


Figure 4.5: Comparison of interpolation results of SSA using TRR with SSA using ARR method. a) Desired result, b) input data with $SNR = 2$ and 40% missing traces, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10, respectively. k) Interpolated data with SSA using ARR. d), f), h), j), and l) residual errors for c), e), g), i), and k), respectively.

Figure 4.6 shows the performance of SSA using ARR in the presence of a strong level of noise $SNR = 1$. Figure 4.6 (a) demonstrates clean and complete synthetic data. Figure 4.6 (b) represents the input with $SNR = 1$ and 40% of the traces were killed. The local spatial window is set to 23 input traces with 11 traces overlap. The number of iterations for each method is set to 25, and the SSA algorithm is applied for the frequency range of 1 to 100 Hz. Figure 4.6 (c), (e), (g), and (i) represent the results of applying SSA with constant rank $k = 4, 6, 8, 10$ and the output quality factor $QF = 2.17, 1.64, 1.34, 0.99$ dB, respectively. Figures 4.6 (d), (f), (h), and (j) indicate the residual errors of applying SSA with the predefined rank $k = 4, 6, 8, 10$ respectively. Figure 4.6 (k) represents the result of interpolation employing SSA using ARR method the output quality factor $QF = 3.11$ dB. Figure 4.6 (l) represents the residual errors for ARR.

Comparing the interpolation results of SSA using predefined rank selection with ARR for different levels of noise in input data, we can conclude that the ARR method's performance in interpolation and denoising is more reliable than the TRR method.

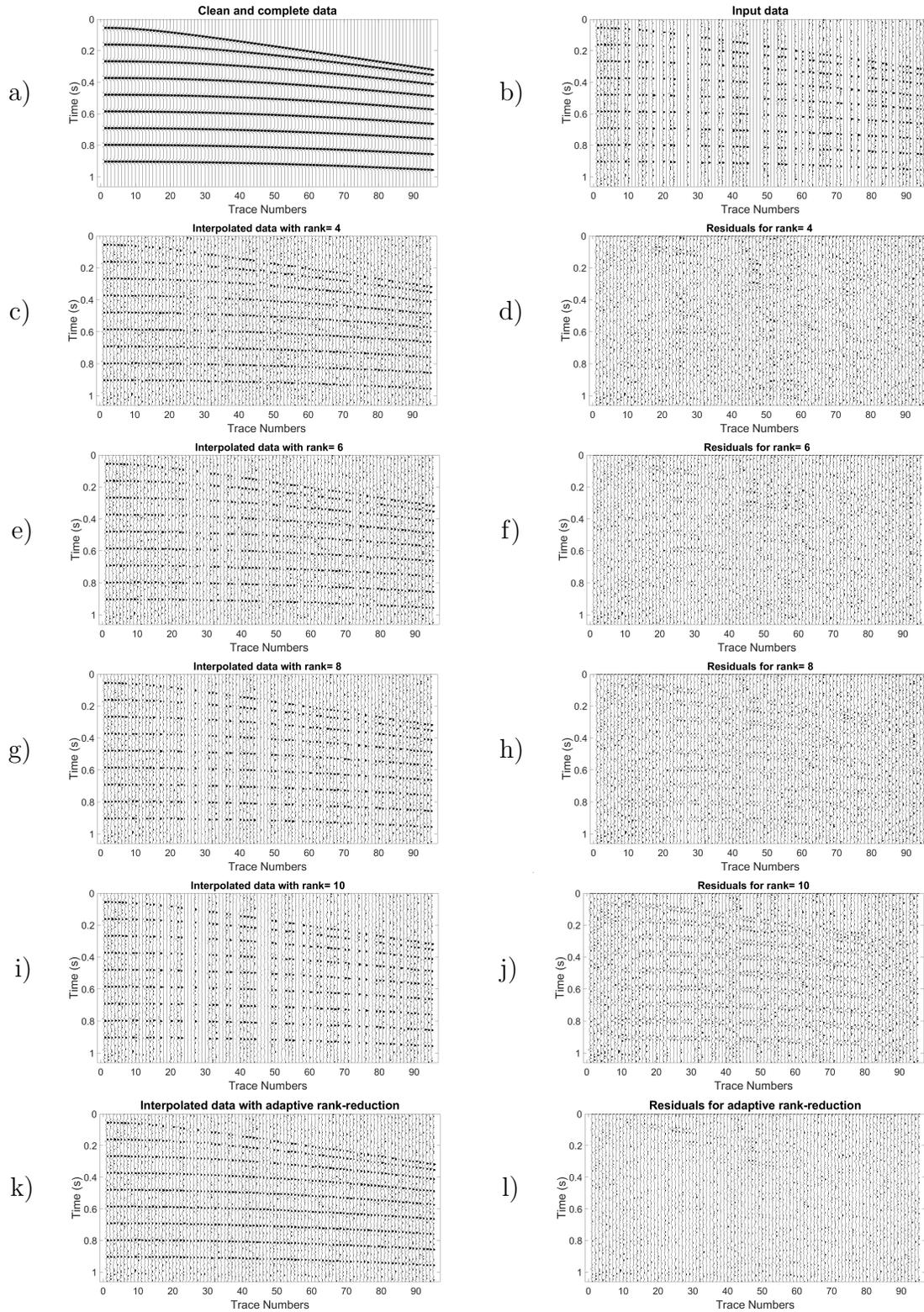


Figure 4.6: Comparing the interpolation results of SSA using TRR method and ARR. a) Desired result, b) input data with $SNR = 1$ and 40% missing traces, c), e), g), and i) interpolated data with constant rank=4, 6, 8, 10 respectively. k) Interpolated data with SSA using adaptive rank reduction. d), f), h), j), and l), residual errors for c), e), g), i), and k), respectively.

4.2 Adaptive rank-reduction for the block Hankel matrix

Although the ARR method can be extended to 3-D or 5-D, the overall rank-reduction procedure should be revised to some degree. For 3-D data, the rank-reduction selection is applied to the block Hankel matrix. For 5-D data is applied to a level-four block Hankel matrix.

In this section, we test the procedure of ARR on the block Hankel matrix on a cube of data containing three linear events with 31 inline and 31 cross-line. Figure 4.7 (a) shows the cube of 3-D data. Figure 4.7 (b) indicates an inline of the data, and figure 4.7 (c) represents a slice of data in cross-line direction.

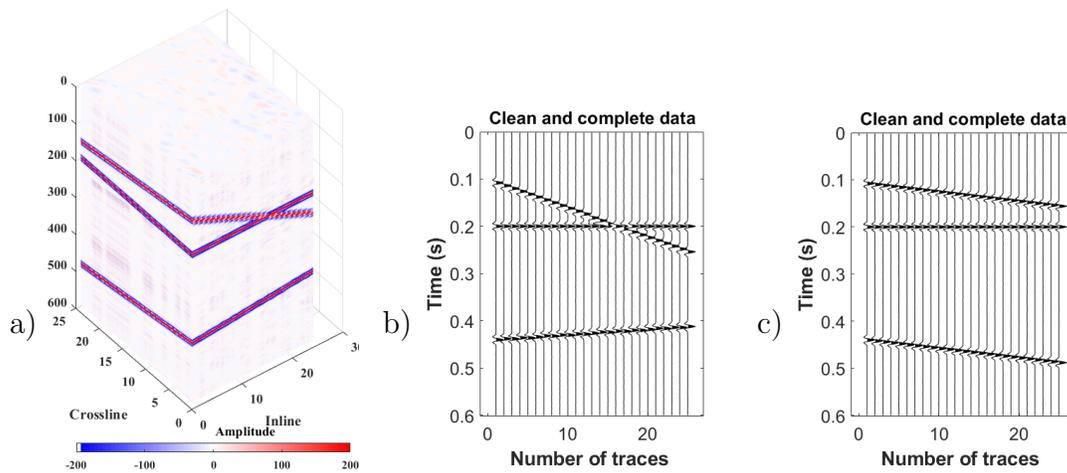


Figure 4.7: a) Cube of 3-D data having three linear events. Slice of data in b) inline direction, c) cross-line direction.

We investigate the singular value distribution of the data by converting it to the (f, x, y) domain. Next is generating the block Hankel matrix in each frequency slice. The size of the generated block Hankel matrix in each constant frequency would be 256×256 . Figure 4.8 (a) shows the block Hankel matrix for the frequency slice of 20 Hz. Figure 4.8 (b) displays the block Hankel matrix generated in the frequency slice of 60 Hz. According to figures 4.8 (a) and (b), it is clear that the block Hankel matrix for the frequency 20 Hz is

smoother than the one for the frequency 50 Hz. It signifies that the block Hankel matrices for the higher frequencies require a higher rank than those in the lower frequencies to recover.

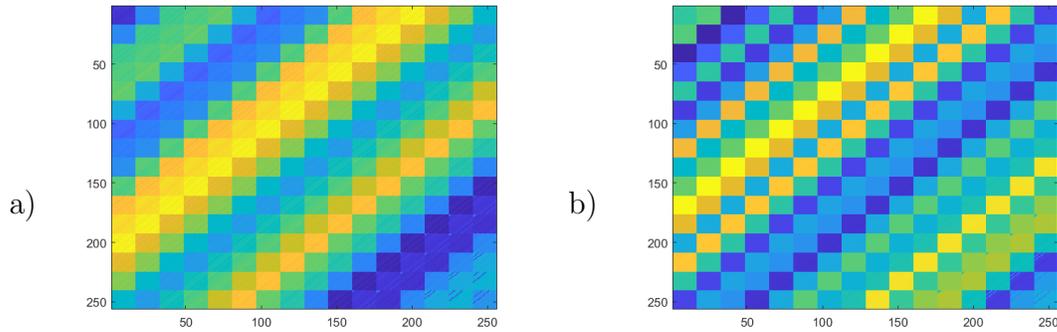


Figure 4.8: Block Hankel matrix for frequency slice of a) 20 Hz, b) 50 Hz.

The singular values spectrum of the data's block Hankel matrices for its frequency range is presented in figure 4.9. Figure 4.9 (b) represents a zoom into the first 15 singular values spectrum. This figure illustrates how the number of nonzero singular values of the block Hankel matrix at low frequencies is equal to the number of the linear events but increases with frequency.

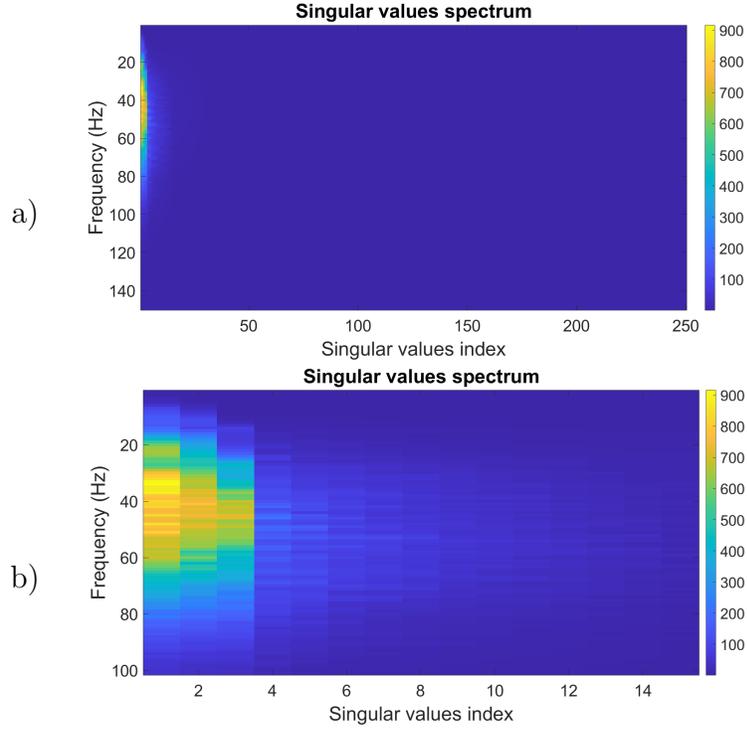


Figure 4.9: a) Singular values spectrum, b) Zoomed singular values spectrum.

Similarly, Figure 4.10 (a) represents the bar plot of the normalized singular values for the frequency slice of 20 Hz. Figure 4.10 (b) shows the zoomed image for the first 20 singular values. Figure 4.10 (c) shows the bar plot of the singular values for the frequency slice of 50 Hz. Figure 4.10 (d) is the zoomed image for the first 20 singular values. From the figures, one can see that for the frequency of 20 Hz the maximum difference of the energy between two adjacent singular values occurs in the third singular value, and it is the same as the number of the linear events. However, for the frequency slice of 50 Hz, there is a second group of nonzero singular values. It means that we need to keep more singular values to recover all the frequencies completely.

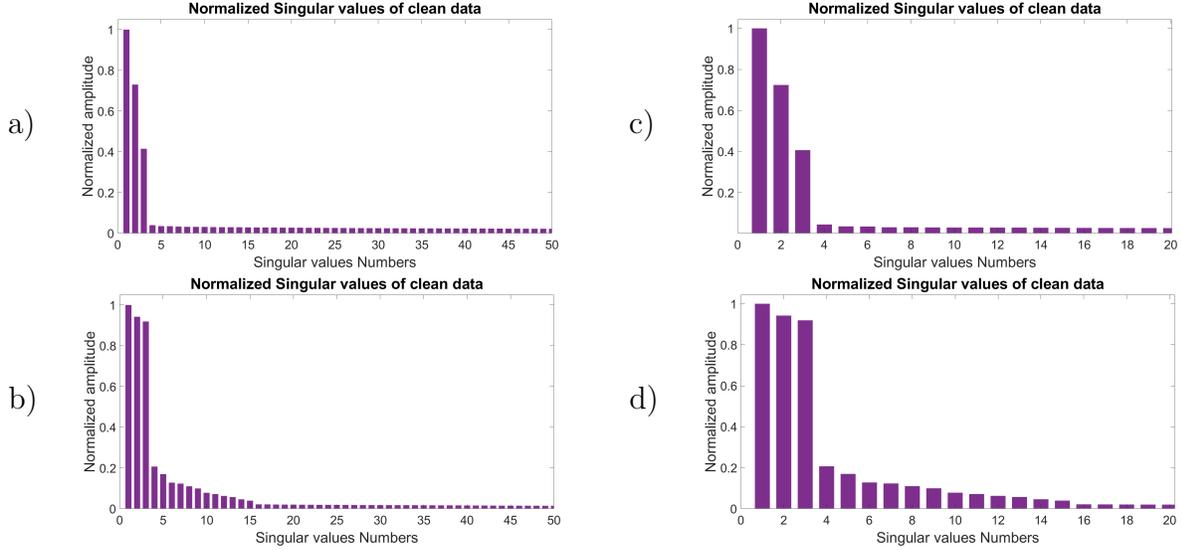


Figure 4.10: Normalized singular values distribution of the block Hankel matrix of 3-D data with three linear events in the frequency slice a) 20 Hz, b) 50 Hz. The first 20 singular values of the same data for frequency slice of c) 20 Hz, d) 50 Hz.

To analyze the effect of each singular value on the power of each frequency for the data presented in figure 4.7, data are decomposed into its singular matrices and recovered with each singular value from one to 15 separately. Then, the power per frequency of the recovered data with each singular value is calculated. In Figure 4.11 the red line represents the graph of frequency power per frequency and the blue line shows the estimate of the mean normalized frequency, of the power spectrum. Figure 4.11 (a) is the result of recovering data with the first singular value. Figure 4.11 (b) shows the power per frequency for the recovered data with the second singular value. Figure 4.11 (c) is the power per frequency for the recovered data including just the third singular value. The estimated mean frequency for these first three singular values is 77 Hz. Figures 4.11 (d), (e), and (f) represent the power per frequency for the recovered data including just the 4th, 5th, and 6th singular values, respectively. The estimated mean frequency for these singular values is 90 Hz. Figure 4.11g), h), and i) indicates the power per frequency for the recovered data including just the 7th, 8th and 9th singular values, respectively. The estimated mean frequency for these singular values is 88 Hz. Figure 4.11 (j), (k), and (l) shows the power per frequency for

the recovered data including just the 10th, 11th and 12th singular values, respectively. It is clear that the most power of the useful signal is recovered with the first three singular values. Nevertheless, the shift in the mean frequency from 77 Hz to 90 Hz states that there is a leakage for the higher frequencies in the data. We conclude from this test that despite the remaining residual errors in the presence of the additive noise, we need to choose the rank of the block Hankel matrix greater than the number of linear events in each processing window of frequency and space.

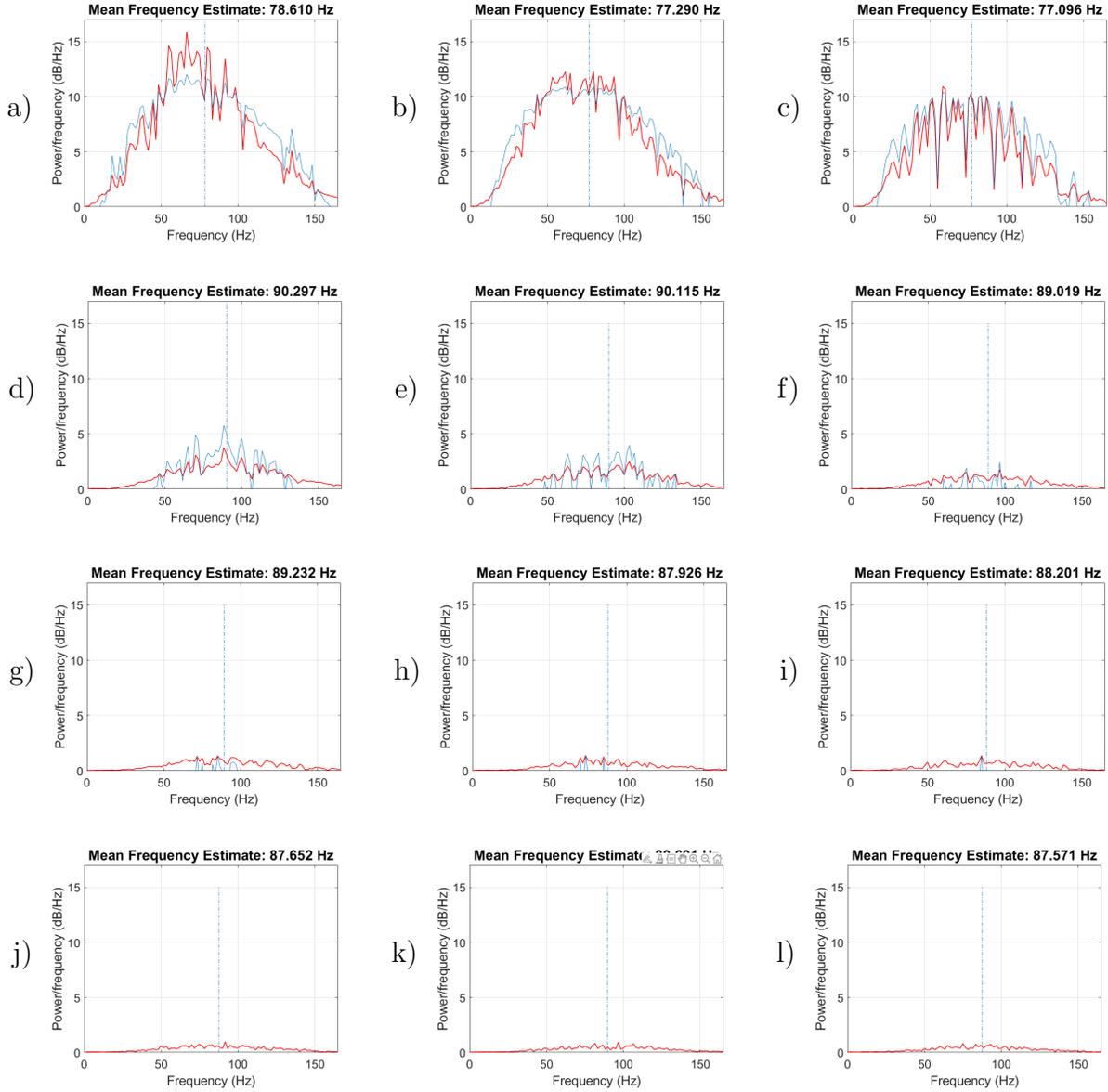


Figure 4.11: Analysis of the frequency power for each singular value. The red line shows power per frequency, and the blue line indicates the estimated mean frequency. a) Power per frequency of data recovered by the first singular value. b), c), d), e) f), g), h), I), j), k), and l) power per frequency recovered by the 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th, 11th, and 12th singular value, respectively.

To find the best rank of data that minimizes the residual errors, we tested different kinds of data containing various numbers of events with different slopes and amplitudes. The best result is when we choose the rank from the second cutoff of singular values instead of the

first cutoff .

In all tested data, it is possible to see the first cutoff with a distinct change of the energy of singular values. However, sometimes there is no abrupt change in the amplitude of the singular values for the second cutoff, especially in the presence of a high level of random noise. Finding the second cutoff could be a challenge when the quality of the signal is poor. With a careful look at where the first and the second cutoff in clean and complete data happen, we conclude that there is a linear relationship between the first and second cutoff. The best estimate for the second cutoff can be estimated as follows:

$$\mathcal{B}(\boldsymbol{\Sigma}, \tilde{k}) = 3 \times \mathcal{T}(\boldsymbol{\Sigma}, k), \quad (4.8)$$

where $\mathcal{T}(\boldsymbol{\Sigma}, k)$ is the operator that finds the point where the two following singular values become more scattered, $\mathcal{B}(\boldsymbol{\Sigma}, \tilde{k})$ indicates the second cutoff of the singular values spectrum of the block of Hankel matrix in each frequency slice. \tilde{k} can be introduced as the optimal rank of the block Hankel matrix that minimizes the Frobenius-norm difference between the approximated and the exact signal components.

Substituting equation 4.8 to the rank-reduction step will give us:

$$\tilde{\mathbf{M}} = \mathbf{U}_{\tilde{k}} \mathcal{B}(\boldsymbol{\Sigma}, \tilde{k}) \mathbf{V}_{\tilde{k}}^H, \quad (4.9)$$

Applying equation 4.9 on the rank-reduction step of MSSA, the rank reduced block Hankel matrix is reconstructed. The next step is averaging anti-diagonals of the recovered block Hankel matrix. It recovers the signal in the Fourier domain for each frequency slice. Applying the iterative algorithm will reconstruct the amplitude of missing traces. Selecting the rank using equation 4.9 makes the estimated signal contains all the signal components. However, the estimated signal contains all the signal components of the originally observed data including the noise component projection. The next chapter is a study on different

methods of rank reduction for adjusting the singular values so that the predicted signal can best approximate the correct signal.

4.3 Discussions

In this chapter, I proposed an adaptive rank-reduction method to deal with complex structures for interpolation and denoising of 3-D seismic data. The ARR method selects an optimal rank of the block Hankel matrix for each slice of the frequency and each spatial window. This strategy is more reliable than the traditional rank-reduction method because it lessens the risk of rank misinterpretation. The ARR method shows good signal reconstruction for the full frequency range, accounting for the need of higher rank on the high frequency part of the spectrum.

Chapter 5

Application of weighting operator on rank-reduction step

5.1 Motivation

The previous chapter was a study on how to select the optimal rank in 3-D data to contain all the components of the signal's frequency. The strategy is to choose the second cutoff in the singular value spectrum of the block Hankel matrix. Using the truncated SVD is an optimal solution to measure the rank of a noisy matrix in a low-rank problem. However, in real data, because of the mix of the signal and noise, it is difficult to suppress the random noise from the desired signal by using only TSVD. This chapter is an effort to develop the rank-reduction step described in chapters 3 and 4 to adjust the singular values to estimate the best-desired signal.

When using TSVD in the low signal-to-noise ratio, the reconstructed data carries a meaningful amount of noise. That is because of the subset projection on the signal component projection. Nadakuditi (2014) introduced an algorithm that provides a weighted approximation and focuses on the reduction and thresholding form of the optimal weights, on the singular values. Their method mitigates the effect of rank overestimation. If the rank is

properly estimated, the algorithm will better estimate weak subspace components of the signal matrix.

Chen et al. (2016) introduced a damping operator that shrinks the singular values containing significant particles of residual noise. Their method in the reconstruction of highly noisy incomplete 5-D data set reached better results than the Cadzow rank-reduction method (Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011).

5.2 Weighting rank-reduction (WRR)

Lets assume \mathbf{M} an $n \times m$ block Hankel matrix of the observed data whose $\sigma_1, \dots, \sigma_n$ are its singular values as below:

$$\mathbf{M} = \sum_{i=1}^n \sigma_i u_i v_i^H, \quad (5.1)$$

where u_i and v_i are the left and right singular vectors of \mathbf{M} . let \mathbf{S} be an $n \times m$ matrix having rank k with $\theta_1, \dots, \theta_k$ singular values. Moreover, the singular value distribution of the noise only part can be estimated from the singular value distribution of the signal plus noise matrix (\mathbf{M}). The observed data can be written as:

$$\mathbf{M} = \mathbf{S} + \mathbf{N}, \quad (5.2)$$

where \mathbf{M} is the observed data, \mathbf{S} represents the desired signal, and \mathbf{N} denotes the noise component which includes the missing traces. Let's assume that the rank of the signal is predefined k . The signal can be modelled as:

$$\mathbf{S} = \sum_{i=1}^k \theta_i u_i v_i^H, \quad (5.3)$$

where θ_i are the true singular values of the signal only (\mathbf{S}) part of the \mathbf{M} . Considering the rank of the true signal is k , using the TSVD of \mathbf{M} the calculated signal will be as follow:

$$\tilde{\mathbf{M}} = \sum_{i=1}^k \sigma_i u_i v_i^H, \quad (5.4)$$

where $\tilde{\mathbf{M}}$ is the rank reduced matrix of \mathbf{M} that solves the problem of equation 5.5. However, the threshold k of the observed data cannot serve as the only signal part of data. It is yet infected with the noise characteristics because of the projection of the noise component on the projection of the signal's only component.

$$\left\| \tilde{\mathbf{M}} - \mathbf{S} \right\| \leq \epsilon. \quad (5.5)$$

In this section, we are looking for a weighting operator $\hat{\mathbf{W}}$, to adjust the singular values of $\tilde{\mathbf{M}}$ to calculate the best estimation of the desired signal. Nadakuditi (2014) proposed an algorithm for low-rank matrix denoising that can be summarized as follow:

- 1: $\mathbf{M}_{n \times m}$ is the signal plus noise Hankel matrix.
- 2: k is the best effective rank that can represent the signal,
- 3: for $i = 1 : k$

$$\text{compute } \hat{w}_i = \left(-\frac{2}{\sigma_i} \frac{\mathcal{D}(\sigma_i; \boldsymbol{\Sigma})}{\mathcal{D}'(\sigma_i; \boldsymbol{\Sigma})} \right), \quad (5.6)$$

- 4: end for loop,
- 5: compute results as $\hat{\mathbf{M}} = \sum_{i=1}^k \hat{w}_i \sigma_i u_i v_i^H$

In this algorithm $\mathcal{D}(\sigma; \boldsymbol{\Sigma})$ is computed as:

$$\mathcal{D}(\sigma; \boldsymbol{\Sigma}) = \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^H)^{-1}) \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \boldsymbol{\Sigma}^H \boldsymbol{\Sigma})^{-1}) = \left[\frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \boldsymbol{\Sigma}^2)^{-1}) \right]^2, \quad (5.7)$$

where \mathcal{D} represents D -transform, and $Tr(\cdot)$ denotes the trace operator of the input.

The \mathcal{D}' represents the derivative of \mathcal{D} with respect to σ :

$$\begin{aligned}\mathcal{D}'(\sigma; \Sigma) &= 2 \left[\frac{1}{k} Tr(\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-1}) \right] \\ &= \left[\frac{1}{k} Tr((\sigma^2 \mathbf{I} - \Sigma^2)^{-1} - 2\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-2}\sigma) \right] \\ &= \frac{2}{k^2} \left[Tr(\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-1}) \right] \left[Tr((\sigma^2 \mathbf{I} - \Sigma^2)^{-1} - 2\sigma^2(\sigma^2 \mathbf{I} - \Sigma^2)^{-2}) \right],\end{aligned}\tag{5.8}$$

The D -transform describes how the distribution of the singular values of the sum of the independent matrices is related to the distribution of the singular values of the individual matrices (Benaych-Georges, 2009).

Benaych-Georges and Nadakuditi (2012) indicate that the principal singular values and vectors of a large matrix can be set apart as the singular values of the signal matrix and the D -transform of the limiting noise-only singular value distribution. From Equation 5.6 the weighting operator $\hat{\mathbf{W}}$ can be written as:

$$\hat{\mathbf{W}} = \text{diag}(\hat{w}_1, \hat{w}_2, \dots, \hat{w}_k).\tag{5.9}$$

We can substitute Equation 5.9 on Equation 5.4 to have:

$$\tilde{\mathbf{M}} = \sum_{i=1}^k \hat{w}_i \sigma_i \mathbf{u}_i \mathbf{v}_i^H,\tag{5.10}$$

Applying operator $\hat{\mathbf{W}}$ or the weighting coefficients on the matrix of singular values will loosen the noise projection's effect. In the matrix notion we can write Equation 5.10 as below:

$$\tilde{\mathbf{M}} = \mathbf{U}_k \hat{\mathbf{W}} \Sigma_k \mathbf{V}_k^H,\tag{5.11}$$

The weighting operator satisfies the equation below (Nadakuditi, 2014):

$$\left\| \tilde{\mathbf{M}} - \mathbf{S} \right\| \leq \epsilon. \quad (5.12)$$

To understand the effect of the weighting operator on the singular values of the block Hankel matrices, the predefined rank-reduction method (TRR) and the WRR method are applied to a block Hankel matrix. Figures 5.1 (a) shows clean and complete 3-D synthetic data having four linear events, figure 5.1 (b) indicates the same 3-D data with $SNR = 2$ and 51% missing traces. Figures 5.1 (c) and 5.1 (d) show the slices of data in inline directions for the cube of Figures 5.1 (a) and 5.1 (b), respectively. Figures 5.1 (e) and 5.1 (f) are the slices of the cubes of Figures 5.1 (a) and 5.1 (b) in crossline direction, respectively.

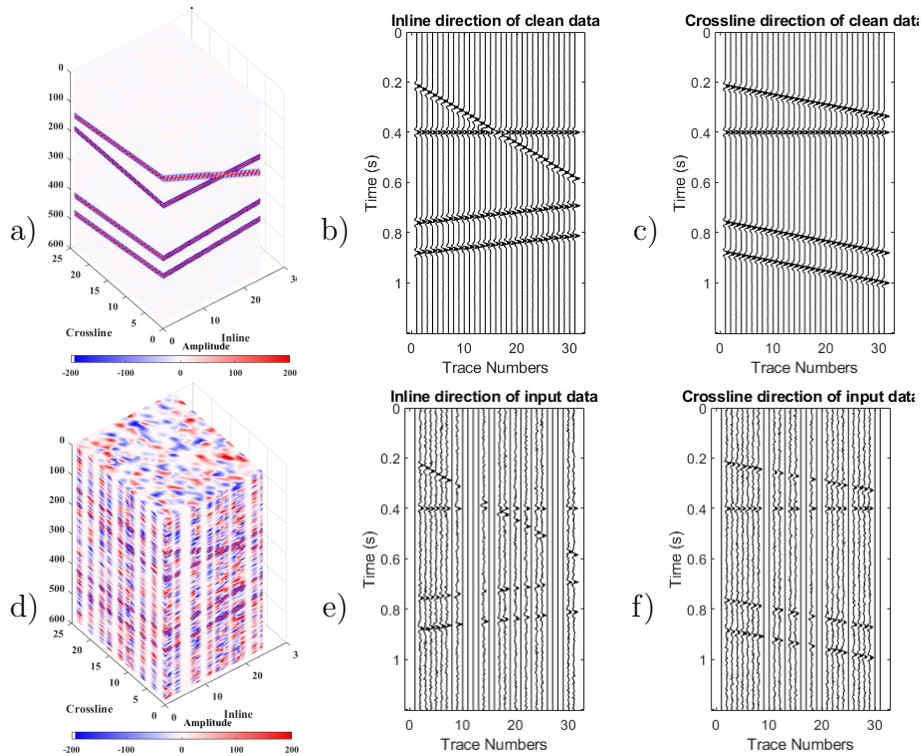


Figure 5.1: a) Cube of 3-D synthetic clean and complete data, b) and c) inline and cross-line sections of cube (a), respectively. d) Cube of 3-D synthetic data with $SNR = 2$ and 51% missing traces, e) and f) inline and cross-line sections of cube (d), respectively.

Figure 5.2 shows the Hankel matrices of data in figure 5.1 for the frequency slice of 20

Hz. Figure 5.2 (a) is the block Hankel matrix of clean and complete data. Figure 5.2 (b) corresponds to the block Hankel matrix of data with $SNR = 2$ and 51% missing traces. Figure 5.2 (c) shows the block Hankel matrix after applying the predefined rank=12 after ten iterations. Figure 5.2 (d) indicates the block Hankel matrix after applying the weighting operator to adjust the singular values of the data using TSVD with predefined rank=12 after ten iterations. We can see from the figure 5.2 that the block Hankel matrix of the WRR method is smoother than the result of applying the predefined rank=12 after ten iterations.

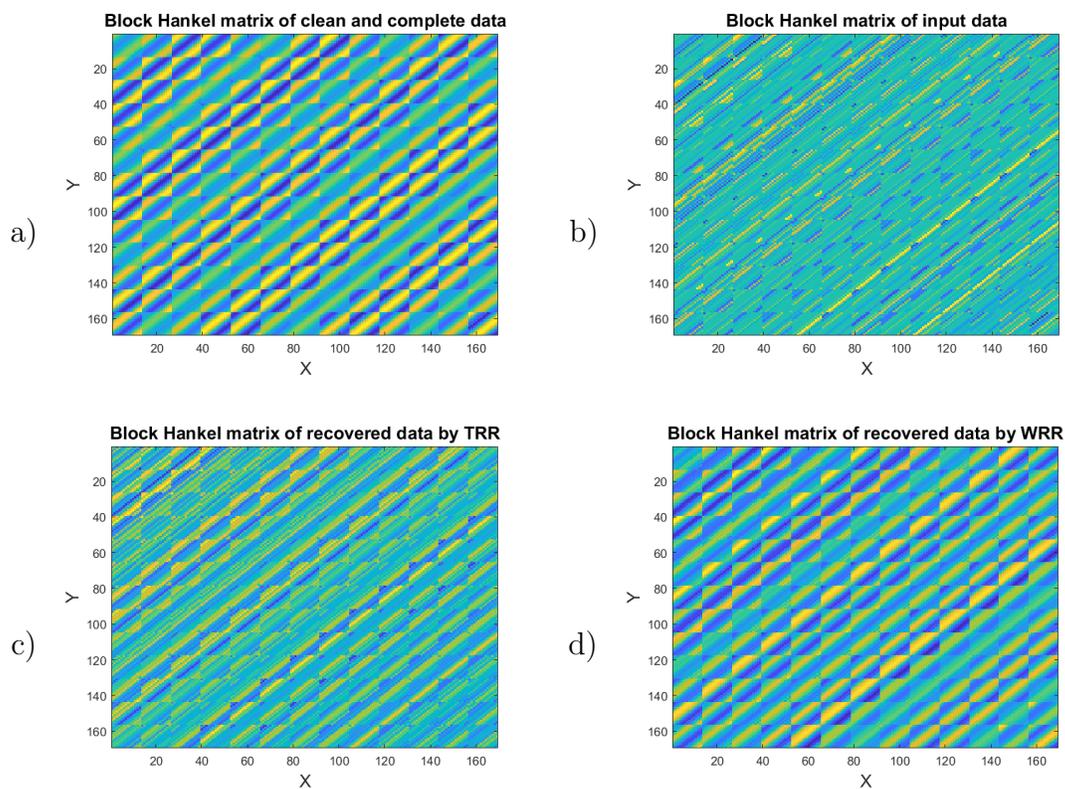


Figure 5.2: Block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with $SNR = 2$ and 51% missing traces, c) recovered data by TSVD rank-reduction after ten iterations, d) recovered data by WRR after ten iterations.

Figure 5.3 corresponds to the first 25 singular values spectrum of the block Hankel matrices of figure 5.1. Figure 5.3 (a) represents the singular values spectrum of the block Hankel matrix of clean and complete data for the frequency of 20 Hz, the first 3 singular values

indicate the useful signal that relates to the coherent events in the data. Figure 5.3 (b) relates to the singular values spectra of the noisy and incomplete data, figure 5.3 (c) shows the singular values spectra of the noisy and incomplete data after applying TSVD with rank =12. Figure 5.3 (d) indicates the singular values spectra of the noisy and incomplete data after applying TSVD with rank=12 and the weighting operator. We can see that the singular value spectrum of the data after applying the weighting operator is much closer to the one relates to the desired signal in figure 5.3 (a).

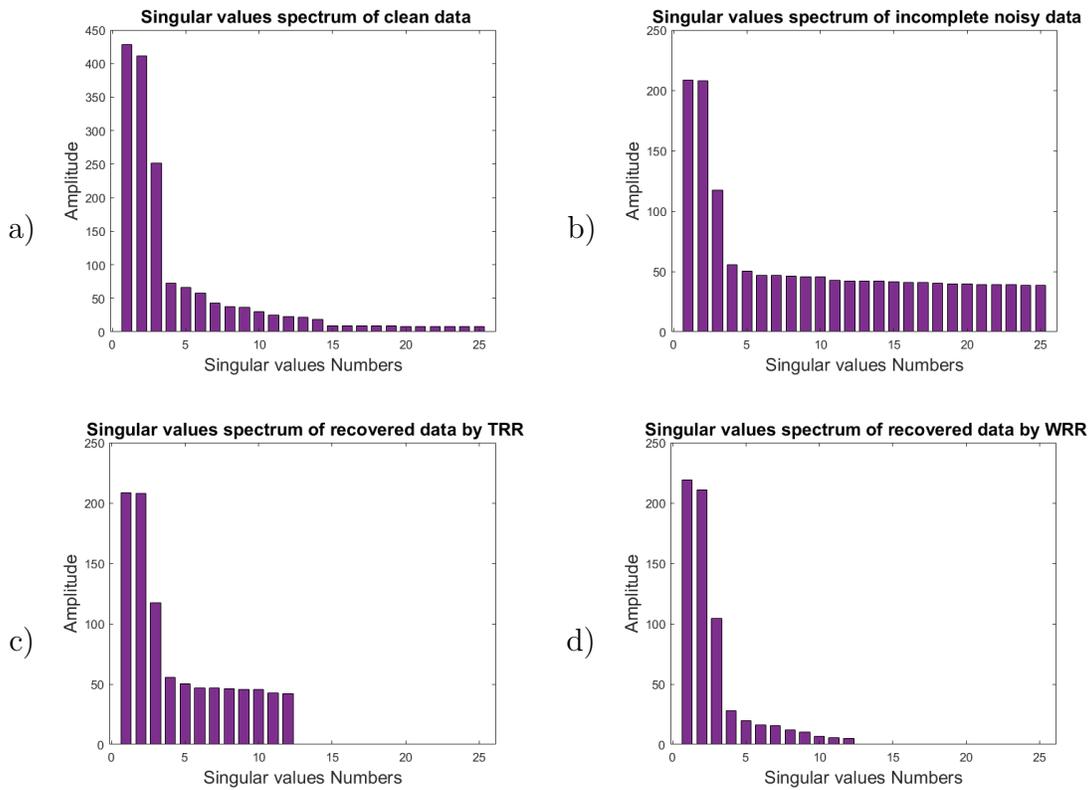


Figure 5.3: The first 25 singular values spectrum of the block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with $SNR = 2$ and 51% missing traces, c) The first 12 singular value of the data, d) singular values spectrum of the data after applying the weighting operator.

5.3 Adaptive weighting rank-reduction

In the previous chapter, we examined how to select the optimum rank of the Hankel matrix in 2-D data and block Hankel Matrix in 3-D data without any prior information about the number of linear events in each window. To recover the higher frequencies in 3-D data, we need to select the second cutoff in the singular values spectrum of data. One of the advantages of this rank-reduction method is that it is adaptive and data-driven, and there is no need to set any parameter for the rank-reduction step in each processing window. Moreover, choosing the second cutoff as a substitute for the first cutoff lets the high frequencies recover completely. However, selecting large ranks leads to remain more residual errors in the recovered data. But, applying the weighting operator reduces the effect of noise on the projected signal components and improves data results by recovering higher frequencies.

Let's get back to equation 5.9 and equation 4.9 as follow:

$$\hat{\mathbf{W}} = \text{diag}(\hat{w}_1, \hat{w}_2, \dots, \hat{w}_{\tilde{k}}). \quad (5.13)$$

$$\tilde{\mathbf{M}} = \mathbf{U}_{\tilde{k}} \mathcal{B}(\boldsymbol{\Sigma}, \tilde{k}) \mathbf{V}_{\tilde{k}}^H. \quad (5.14)$$

We can substitute the weighting algorithm obtained from equation 5.13 in equation 5.14 to enhance the results of the rank-reduction step as:

$$\hat{\mathbf{M}} = \mathbf{U}_{\tilde{k}} \hat{\mathbf{W}} \mathcal{B}(\boldsymbol{\Sigma}, \tilde{k}) \mathbf{V}_{\tilde{k}}^H, \quad (5.15)$$

where $\hat{\mathbf{M}}$ indicates the reduced rank block Hankel matrix. Missing traces can be interpolated completely by applying the iterative algorithm. This Adaptive Weighting Rank-Reduction (AWRR) method leads the way that sorts out the rank of the block Hankel matrix automatically while denoising data.

Figure 5.4 corresponds to the block Hankel matrices of the same data in the constant fre-

quency slice of 50 Hz. Figure 5.4 (a) shows the block Hankel matrix of the clean and complete data, figure 5.4 (b) is the block Hankel matrix of data with $SNR = 2$ where 51% of traces are killed. Figure 5.4 (c) displays the recovered block Hankel matrix after applying TRR and rank=4 after ten iterations. Figure 5.4 (d) indicates the block Hankel matrix after applying the WRR with predefined rank=4. Figure 5.4 (e) represents the recovered block Hankel matrix of the ARR method, and figure 5.4 (f) is the block Hankel matrix recovered by the weighting ARR (AWRR) method. Generally, the results of AWRR and WRR are smoother than the output of TRR and ARR, and the results of ARR and AWRR include more details than TRR and WRR.

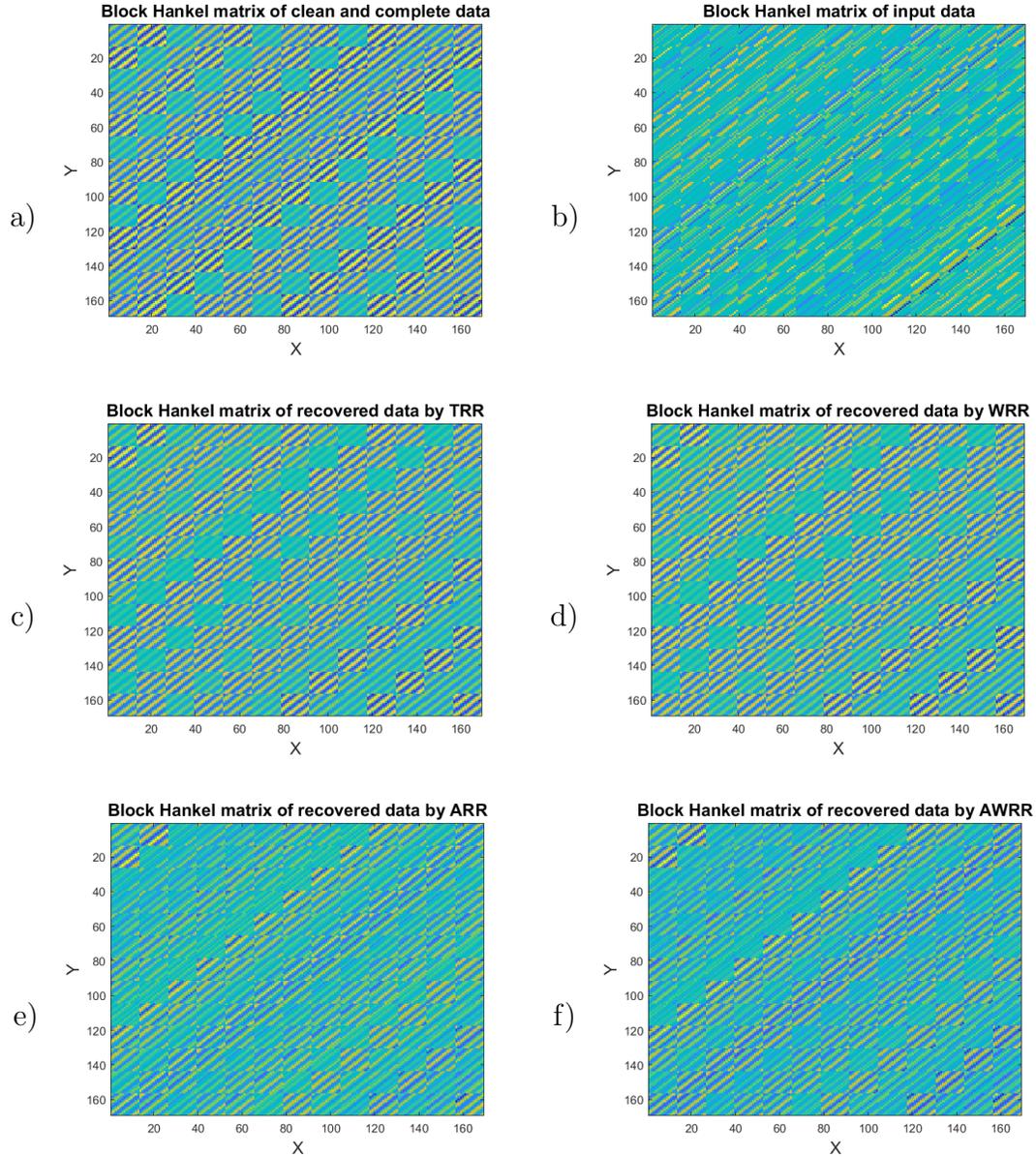


Figure 5.4: Block Hankel matrices for frequency of 50 Hz for a) clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered block Hankel matrix by TRR rank=4 after ten iterations, d) recovered data by WRR after ten iterations. e) Recovered data by applying the ARR method after ten iterations, d) recovered data after applying AWRR after ten iterations.

Figure 5.5 corresponds to the first 25 singular values spectrum of the block Hankel matrices of figure 5.4. Figure 5.5 (a) represents the singular values spectrum of the block Hankel matrix of clean and complete data for the frequency of 50 Hz. The abrupt drop of energy

of the third and the fourth singular value relates to the useful signal. However, there are still nonzero singular values that are related to the useful signal. Figure 5.5 (b) is the singular values spectra of the block Hankel matrix of noisy and incomplete data. Figure 5.5 (c) shows the singular values spectrum of the noisy and incomplete data after TRR with rank=4. Figure 5.5 (d) displays the singular values spectrum of the noisy and incomplete data after applying WRR with rank=4. Figure 5.5 (e) depicts the singular values spectrum after applying ARR. Figure 5.5 (d) indicates the singular values spectrum of the noisy and incomplete data after implementation of the AWRR method. From the figures, one can see that the result of AWRR is more comparable to the one relates to the desired signal in Figure 5.5 (a).

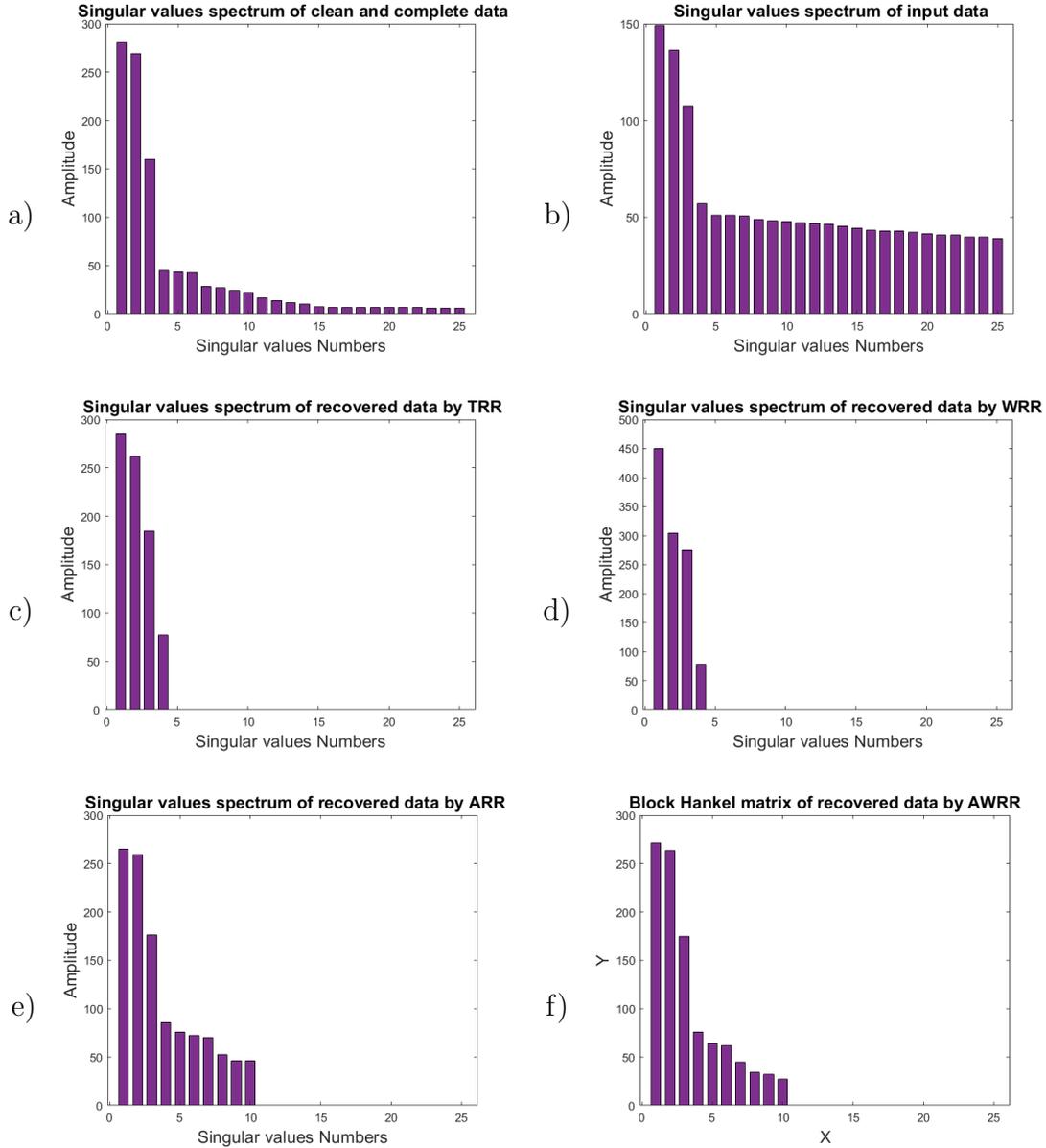


Figure 5.5: The first 25 singular values spectrum of the block Hankel matrices for frequency slice of 50 Hz for, a) clean and complete data, b) data with $SNR = 2$ and 50% killed traces, c) recovered data by TRR, d) recovered data by WRR, e) recovered data by ARR, f) recovered data by AWRR.

5.3.1 Results and discussions

Several evaluations are presented in this section to estimate the proficiency of the different methods of rank-reduction for the MSSA algorithm.

First, the methods are tested on a synthetic shot gather containing nine hyperbolic events with different curvatures. Then the efficiency of the methods on a shot gather of a 3-D field data set is tested.

5.3.2 Synthetic data

Figure 5.6 shows the result of applying the methods of rank-reduction to a synthetic shot gather holding nine hyperbolic events with different curvatures. The first test is a cube of 100 inline and 11 crosslines with $SNR = 100$ and 60% decimated traces. The local window is set to 23 traces in the inline direction and 11 in the crossline direction with half of the input traces overlapping. The number of iterations is fixed for all of the methods. The MSSA algorithm is applied for the frequency range of 1 to 100 Hz. The predefined rank for TRR and WRR methods is set to nine where equals the number of observations in each window of processing.

Figure 5.6 (a) shows a 3-D shot gather prior to noise contamination and killing traces arranged into a 2-D matrix. 5.6 (b) is the the same data with $SNR = 100$ and 60% missing traces. The results of applying TRR, ARR, WRR, and AWRR are shown in figures 5.6 (c), (e), (g), (i), respectively. Figures 5.6 (d), (f), (h), and (j) are the noise estimators resulting from subtracting the clean and complete data from the recovered data of figures 5.6 (c), (e), (g), and (i) respectively. We see in figure 5.6 that all the methods succeed in recovering the missing traces. The quality of the input data is $QF = -1.71 \text{ dB}$. The output quality of interpolation of the TRR, ARR, WRR, and AWRR methods is $QF = 8.82, 9.31, 10.21, 11.45, \text{ dB}$, respectively. The results of AWRR and WRR are comparable, but it is important to notice that for real data it is hard to decide the correct rank in each window. The benefit of the AWRR rather than WRR is that it is adaptive, and we do not need to know the rank.

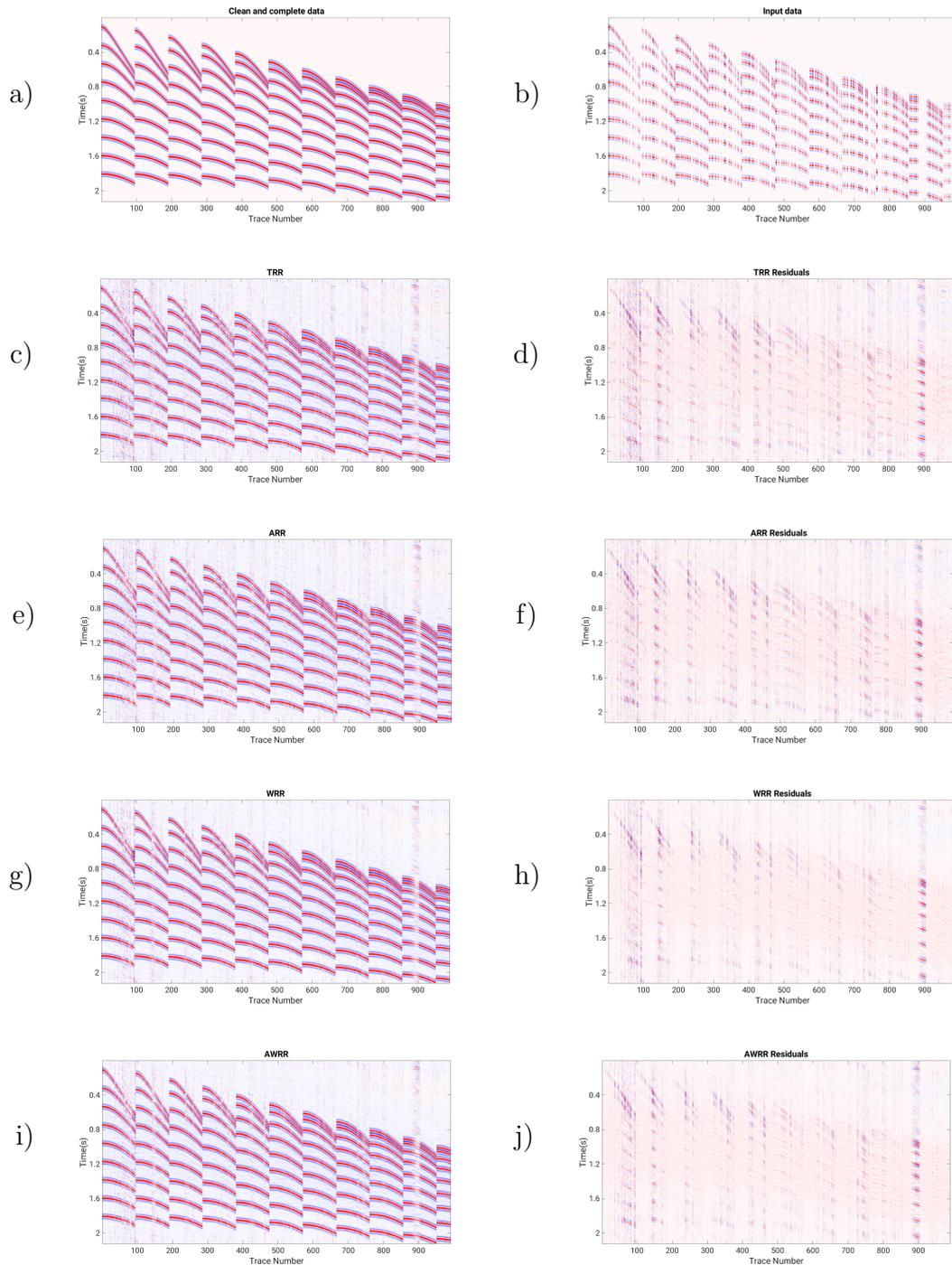


Figure 5.6: Comparison of different methods of rank-reduction. a) Clean and complete data, b) data with $SNR = 100$ and 60% missing traces. c), e), g), and i) Interpolated data by TRR, ARR, WRR, and AWRR methods, respectively. d), f), h), and j) Residual errors of TRR, ARR, WRR, and AWRR, respectively.

The next test represents the same data in the previous example but in presence of random

noise. Missing traces and the random noise both increase the rank of the Hankel matrix which means an increase in the non-zero singular values. Figure 5.7 shows the result of applying different rank-reduction methods to the same data in figure 5.6 but contaminated with random noise. Alike to the early test, we choose the local window with 23 traces in the inline direction and 11 in the crossline direction for each method and half a window overlapping in each direction. We set the number of iterations constant for all of the methods. The frequency range of the MSSA algorithm is 1 to 100 Hz. Furthermore, the rank of TRR and WRR is set to 9. Figure 5.7 (a) is the desired data arranged into a 2-D matrix. Figure 5.7 (b) shows input data with $SNR = 2$ and 60% missing traces. The results of applying TRR, ARR, WRR, and AWRR methods are shown in figures 5.7 (c), (e), (g), and (i), respectively. Figures 5.7 (d), (f), (h), and (j) are residual errors of figures 5.7 (c), (e), (g), and (i), respectively. All four methods have recovered the amplitude of the signals. However, TRR remains severe residual errors, the result from ARR causes significant residual noise and AWRR is much cleaner than TRR, ARR and WRR. The input data quality is calculated $QF = -1.44 \text{ dB}$. The output quality factor of the TRR, ARR, WRR, and AWRR methods are $QF = 5.23, \quad 6.58, \quad 7.56, \quad 8.12, \text{ dB}$, respectively.

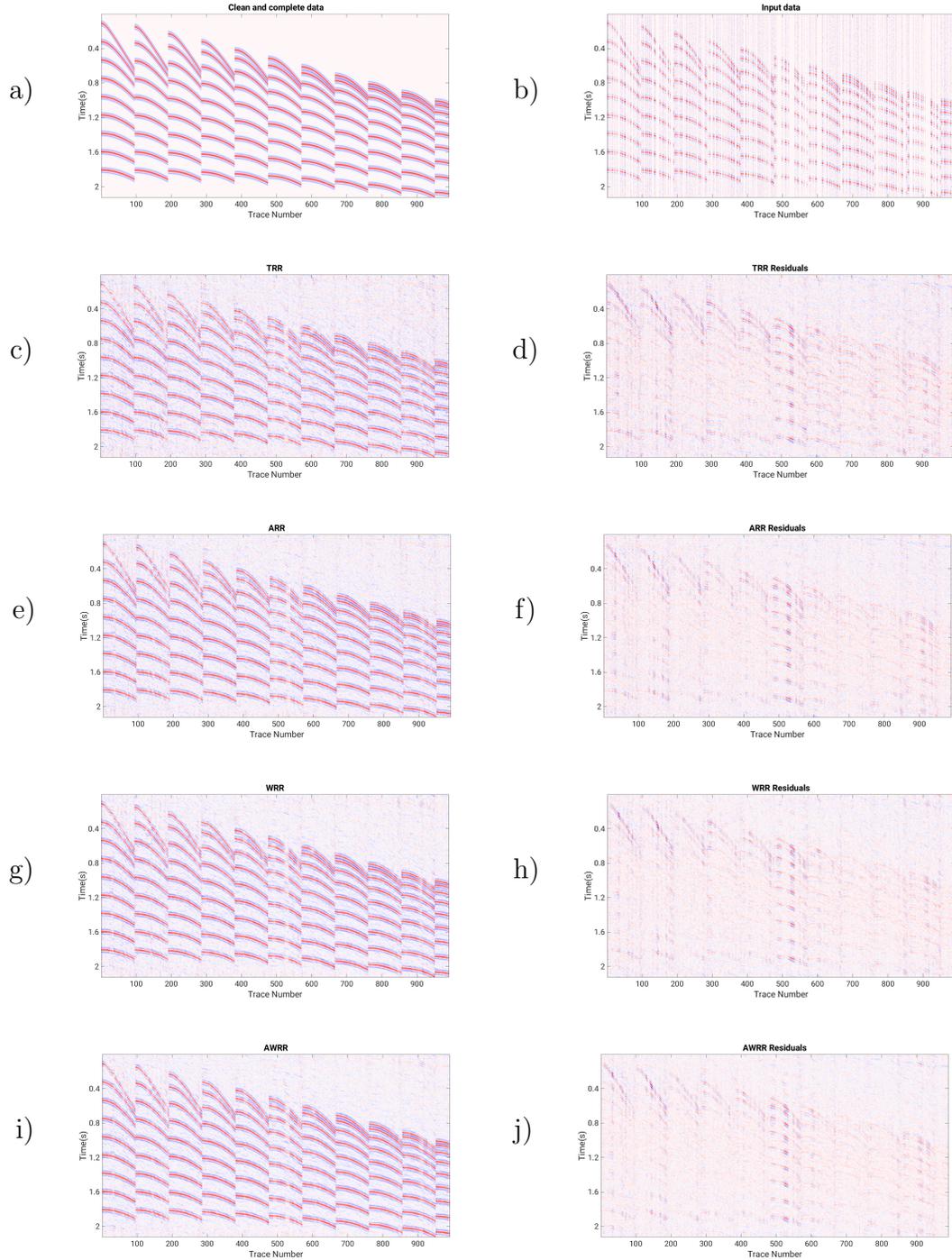


Figure 5.7: Comparison of different methods of rank-reduction. a) Clean and complete data, b) input data with $SNR = 2$ and 60% missing traces, c), e), g), and i) interpolated data by TRR, ARR, WRR, and AWRR, respectively. d), f), h), and j) residual errors of TRR, ARR, WRR, and AWRR, respectively.

Figure 5.8 compares the $f - k$ spectra of the discussed rank-reduction methods. Figure

5.8 (a) shows the $f - k$ spectra of clean data. Figure 5.8 (b) represents the $f - k$ spectra of data with $SNR = 2$ and 50% missing traces. Figures 5.8 (c), (d), (e), and (f) are the $f - k$ spectra of the TRR, WRR, ARR, and AWRR methods, respectively.

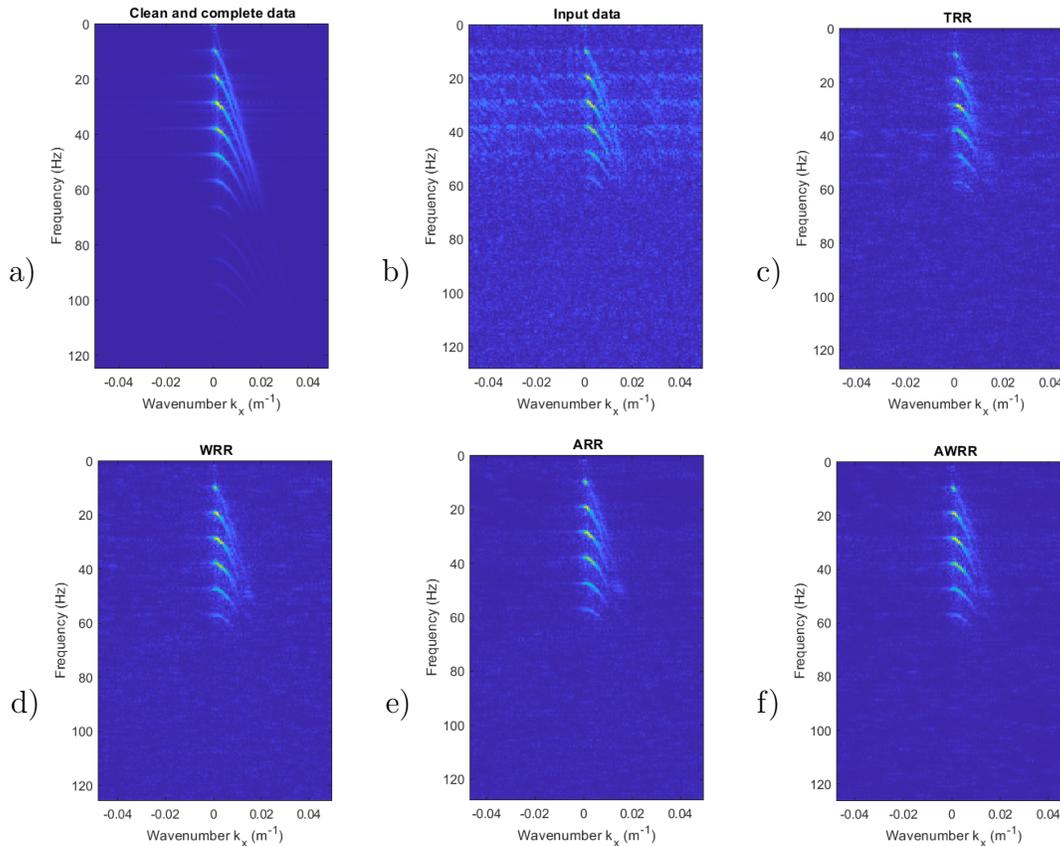


Figure 5.8: $f - k$ spectra comparison of: a) clean and complete data, b) input data, c) TRR result, d) WRR result, e) ARR result, f) AWRR result.

For the stability analyses of each algorithm, we run them under different levels of gaps percentage with 20 several realizations of sampling ratio for a given $SNR = 2$. The other parameters are set constant for each run. Figure 5.9 shows the graph of the interpolation quality factor for each method versus the gap ratio for $SNR = 2$. The length of the error bars is the standard deviation of the interpolation quality factor, and each coloured lines connect the mean value of the interpolation quality factor for each method at each gap ratio. As we can see from figure 5.9, the input quality factor decreases with increasing gap ratio

also the output quality factor decreases with increasing gap ratios. Moreover, the output quality factor curve of the AWRR method is always above the other curves.

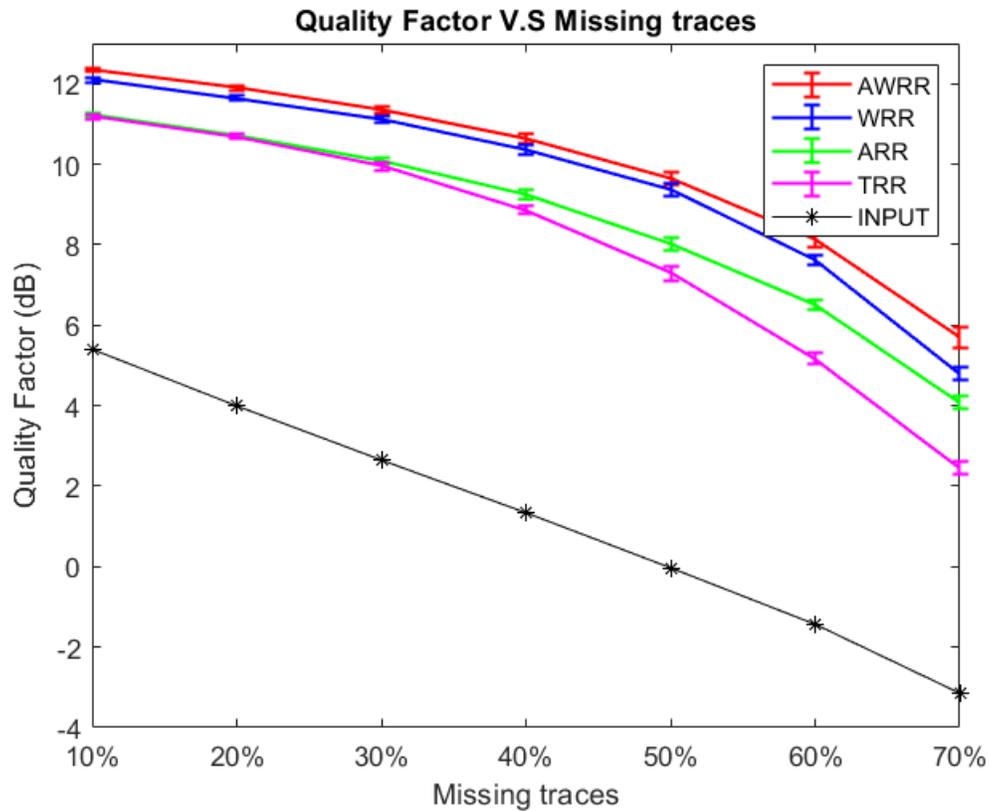


Figure 5.9: The mean and standard error of the quality of reconstruction versus gap ratio for different methods. These results are obtained by running each method on 20 realizations of different data set for different gap ratios and rank=9 for the predefined rank methods.

Table 5.1 indicates the values of the mean and standard deviation of the interpolation quality of each method for each gap percent. We can see that the average value of the AWRR quality factor is higher than the other methods. Comparing the discussed methods, AWRR has lower standard deviation values that indicate the values tend to be close to the mean value, especially when the data is more sparse.

Table 5.1: Mean and standard deviation of the quality of interpolation for each method for different gap ratio inputs.

Gap Ratio	Output Quality Factor (dB)							
	AWRR		WRR		ARR		TRR	
	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD
10	12.34	0.04	12.10	0.04	11.22	0.06	11.17	0.04
20	11.90	0.05	11.64	0.05	10.71	0.09	10.69	0.06
30	11.36	0.05	11.12	0.05	10.08	0.08	8.86	0.08
40	10.45	0.07	10.35	0.07	9.24	0.10	8.90	0.11
50	9.64	0.16	9.38	0.20	8.01	0.17	7.28	0.20
60	8.13	0.13	7.61	0.13	6.51	0.14	5.16	0.20
70	5.70	0.16	4.78	0.21	4.10	0.21	2.46	0.25

In the next experiment, the stability analysis is repeated to test the sensitivities of each method to the additive noise. The test is performed by changing the level of signal to noise ratio with 20 different realizations of SNR for a gap ratio = 50%. We set the other parameters constant for each run. In figure 5.10 we see that the proposed method (AWRR) outperforms the other methods even in the poor quality of the signal.

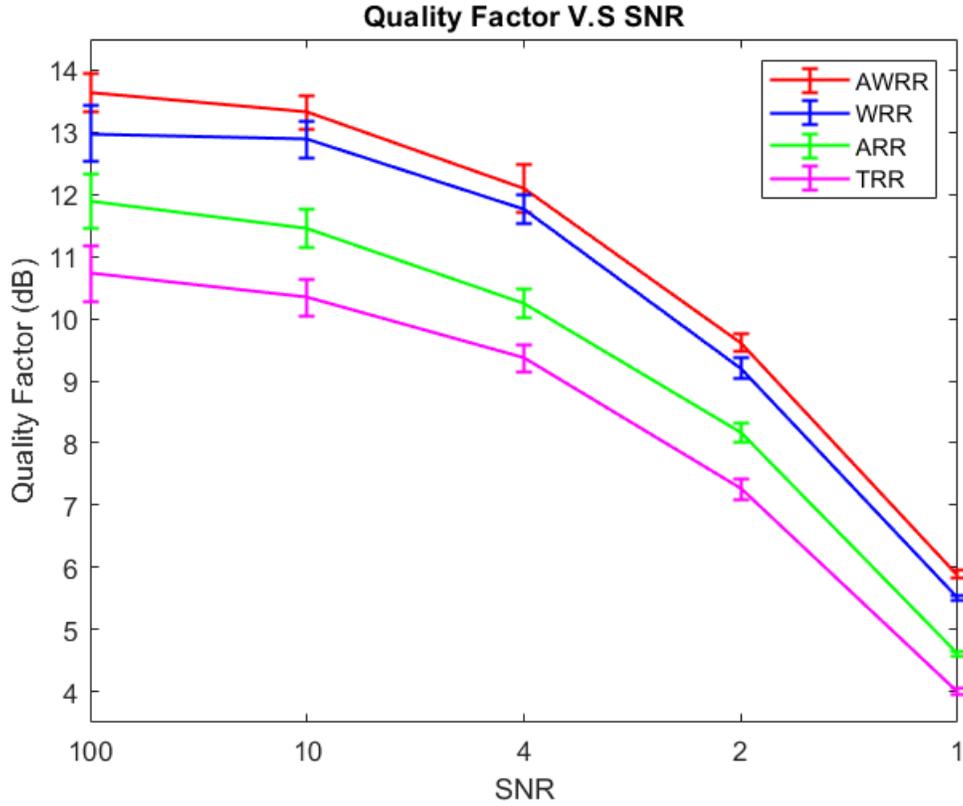


Figure 5.10: The mean and standard error of the reconstruction's quality versus SNR for different methods. These results are obtained by running each algorithm on 20 realizations of the different data set for a signal-to-noise ratio with the gap ratio = 50%, and rank=9 for the predefined rank methods.

Table 5.2 represents the values of the mean and standard deviation of the output quality factor of each method for each signal to noise ratio inputs.

Table 5.2: Mean and standard deviation of the quality of interpolation for each method for different SNR inputs.

Input SNR	Output Quality Factor (dB)							
	AWRR		WRR		ARR		TRR	
	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD
100	13.64	0.41	12.98	0.41	11.90	0.44	10.72	0.30
10	13.32	0.42	12.89	0.42	11.44	0.30	10.34	0.28
4	12.10	0.26	11.77	0.28	10.25	0.23	9.36	0.38
2	9.61	0.26	9.20	0.26	8.16	0.16	7.26	0.15
1	5.88	0.09	5.51	0.09	4.61	0.05	3.40	0.07

5.3.3 Optimally damped rank-reduction method

In the previous section, we studied the performance of the AWRR algorithm compared to the predefined rank-reduction method in interpolation and denoising of 3-D data. In this section, we compare AWRR as an adaptive and data-driven rank-reduction method with a "nearly-adaptive" method proposed by Chen et al. (2016). Their technique first derives the weighting operator, then based on the original weighting operator a damping operator is obtained to decrease the residual noise after the optimal weighting. In this method, the weighting operator is adjusted with an optimal damping operator. Moreover, regardless of the number of the linear events, the rank is selected larger. This "nearly adaptive" algorithm is less sensitive to the rank parameter also reduces the residual noise in the rank-reduction step. Due to this advantage, one can choose a large rank and preserve the signal perfectly while remaining fewer residual errors. The algorithm can be summarized as below.

The optimally damped rank-reduction (ORR) method considers a predefined rank= D , which is usually selected large enough to deal with complicated data and local windows. Then the estimated signal is calculated as follows:

$$\hat{\mathbf{M}} = \mathbf{U}_D \mathbf{\Sigma}_D \mathbf{V}_D^H, \quad (5.16)$$

where D is the selected rank. There is a damping factor as below:

$$\boldsymbol{\nu} = \hat{\delta}^Q (\boldsymbol{\Sigma}_D)^{-Q}, \quad (5.17)$$

where $\hat{\delta}$ represents the maximum element of $\boldsymbol{\Sigma}_D$, and Q is the damping factor.

Using the weighting operator as $\hat{\mathbf{W}} = \text{diag}(\hat{w}_1, \hat{w}_2, \dots, \hat{w}_D)$, and the damping operator from equation 5.17 substituting in equation 5.16, the estimated signal applying the ORR method will be:

$$\hat{\mathbf{M}} = \mathbf{U}_D (\mathbf{I} - \boldsymbol{\nu}) \hat{\mathbf{W}} \boldsymbol{\Sigma}_D \mathbf{V}_D^H, \quad (5.18)$$

where \mathbf{I} is a unit matrix, $\hat{\mathbf{M}}$ denotes the recovered block Hankel matrix by the ORR method. By using the iterative algorithm, data can be interpolated.

The method is insensitive to the rank so, one can choose a large rank and preserve the signal completely with remaining fewer residual errors. One of the drawbacks of this method is its computational cost that makes the algorithm less desirable, particularly for data with more than two spatial dimensions. Because computing the SVD using a large rank can be extremely time-consuming for large-scale problems.

We test the sensitivity of the ORR method assuming that we do not have prior information about the structural complexity of the data and thus the proper rank for data. We evaluate the efficiency of the ORR by selecting the rank=7 to 18. Figure 5.11 shows the diagram of the interpolation quality with respect to the selected rank for the range of 7 to 18. By increasing the selected rank the output quality factor is increasing. The quality factor reaches its maximum value for the selected rank=12 while the curve drops for the ranks larger than 12. However, the quality factor is in the range of 9 to 9.2 (dB) for the chosen rank 10-15. Accordingly, the method is not fully insensitive to the rank unless we select the large rank in a reasonable range.

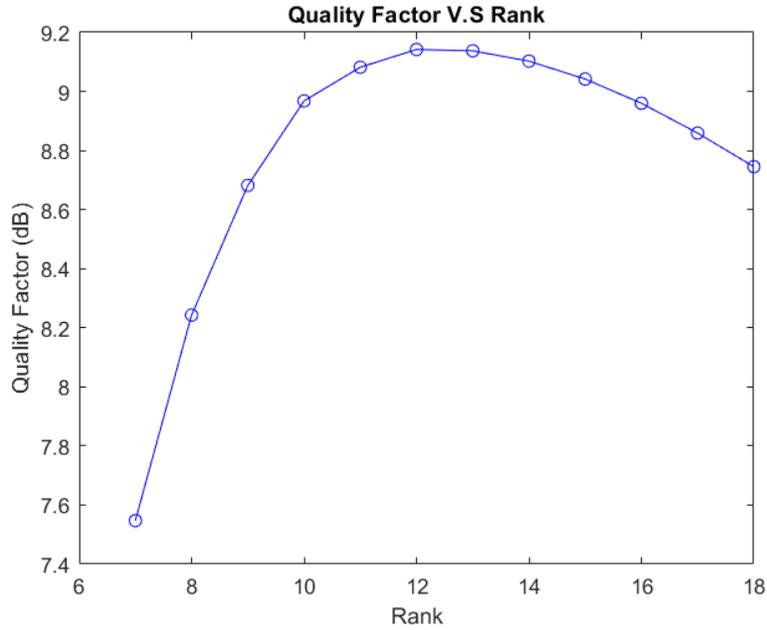


Figure 5.11: Diagram of the output Quality of interpolation using ORR with respect to the selected rank.

To compare the efficiency of AWRR and ORR, a data set with additive random noise $SNR = 2$ is tested for different percentages of missing traces. The selected rank for the ORR method is chosen 12. The parameters such as the number of iterations and the processing window size are selected the same for both methods. Figure 5.12 shows the output quality factor of AWRR and the ORR method with respect to the missing traces percentage. The blue line with the ‘o’ marker is the result of the ORR method and the red line with the ‘*’ marker is the result of the AWRR method.

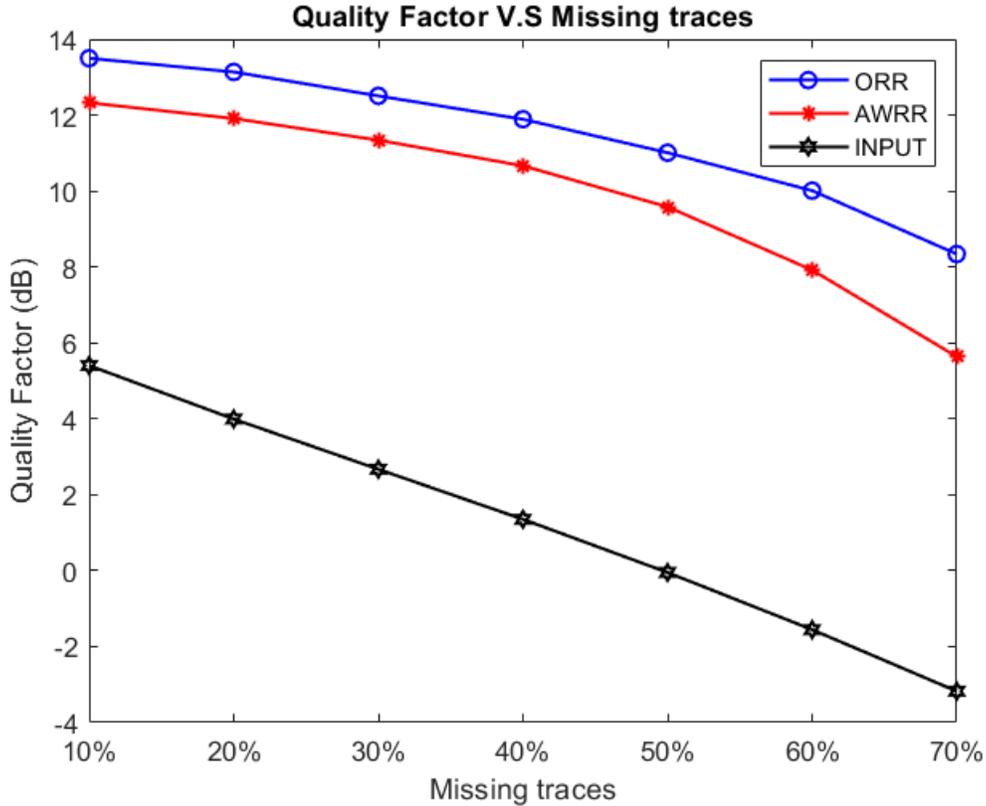


Figure 5.12: The quality of interpolation versus gap percent for ORR, and AWRR.

Figure 5.13 refers to the result of applying ORR and AWRR to the data in figure 5.6, which is contaminated by random noise. The local processing window is set to 23 traces in the inline direction and 11 in the crossline direction for each method, with half a window overlapping in each direction. Figure 5.13 (a) is the initial data before noise contamination and killing traces arranged into a 2-D matrix. Figure 5.13 (a) shows input data with $SNR = 2$ and 60% missing traces. Figure 5.7 (c) is the result of applying ORR, figure 5.13 (e) displays the result of applying AWRR. 5.13 (d) and (f) indicate residual error results subtracting the clean and complete data from the recovered data of figure 5.7 (c) and (e), respectively. We can see that both methods are successful in recovering data; however, the residual errors of the ORR method are less than the AWRR method. The input data quality factor is $QF = -1.44 \text{ dB}$. The outputs quality factor for the ORR and AWRR methods are $QF = 11.12$

and 9.44 dB , respectively.

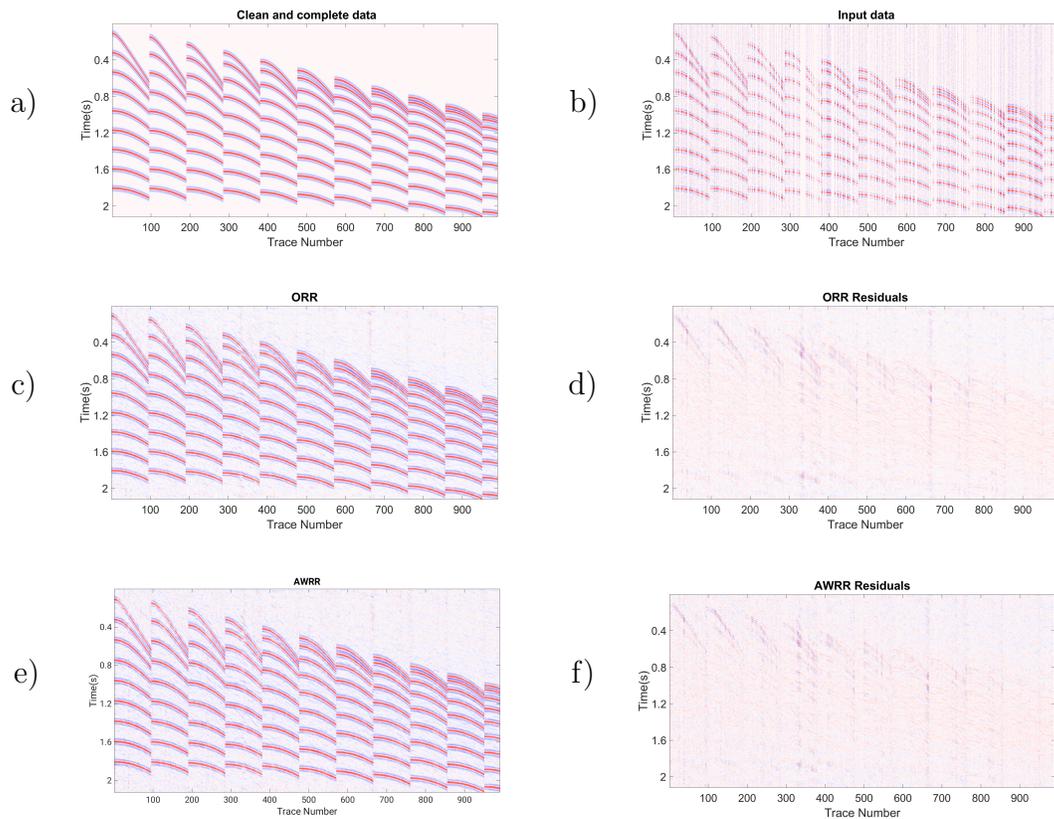


Figure 5.13: Comparison of the interpolation results for ORR method. a) Clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered data by ORR (output $QF = 11.12$ dB), d) residual errors for ORR, e) interpolated data by AWRR (output $QF = 9.44$ dB), f) residual errors for AWRR.

When zooming the data, the comparison of the two methods is more noticeable. Figure 5.14 corresponds to the zooming area of the time 0.4 (s) to 10.2 (s), and the trace number 400 to 460 of data in figure 5.14. We can see that both ORR and AWRR are very powerful in reconstructing the data; however, the ORR's result is smoother than AWRR.

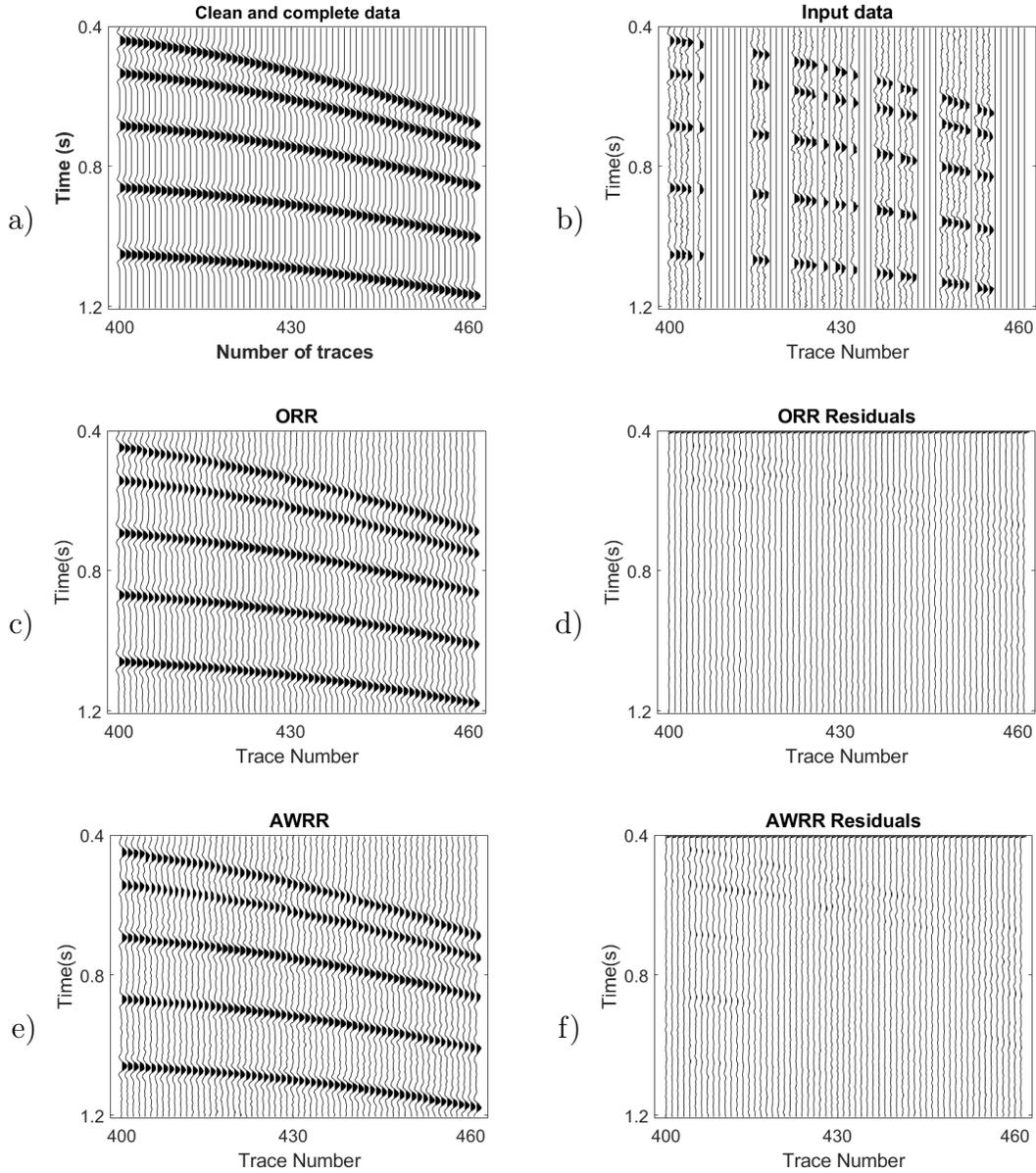


Figure 5.14: Zoomed images of figure 5.13 from the time 0.4 s to 1.2 s and trace number 400 to 460. a) Clean and complete data, b) data with $SNR = 2$ and 50% missing traces, c) recovered data by ORR, d) residual errors for ORR, e) interpolated data by AWRR, f) residual errors for AWRR.

5.3.4 Real field data

This experiment tests the efficiency of the methods on a shot gathers of a 3-D field data set. It is easy to inspect traces in shot/receiver gather displays for poor receivers or any

bad shots, which are the logistic constraints during the seismic survey. Regarding the input data, the best results are obtained with NMO-corrected data (Trad, 2009). However, in this experiment, we applied the MSSA interpolation before NMO correction to see the effect of the algorithm on the curvature.

Figure 5.15 refers to the acquisition coordinates of the shot gather. The red star represents the location of the shot, and the black dot indicates the receiver's location. Figure 5.15 (a) shows the initial distribution of the traces in a shot gather. For this experiment, 41% of the traces are killed. Figure 5.15 (b) illustrates the geometry of input traces after killing 41% of them.

This test is applying the MSSA algorithm using the AWRR and TRR methods in the rank-reduction step. Figure 5.16 (a) shows the input cube of data, which presents the missing traces. Figure 5.16 (b) shows the result of the interpolation using TRR in the rank-reduction step, and figure 5.16 (c) shows the result after using AWRR. For both tests, the number of iterations and processing window remain unchanged. For the TRR approach, the predefined rank is set to $\text{rank} = 10$.

Figure 5.17 represents the cube of data arranged into a 2-D matrix. Figure 5.17 (a) is the input data. Figure 5.16 (b) represents the result of applying TRR. Figure 5.17 (c) shows the result of applying AWRR. Figure 5.18 corresponds to a patch of data from the time 1.35 (s) to 1.75 (s) and the trace number 160 to 190. Figure 5.18 represents the cube of data rearranged in a 2-D matrix. Figure 5.18 (a) is the input data. Figure 5.18 (b) represents the result of applying TRR. Figure 5.18 (c) shows the result of applying AWRR.

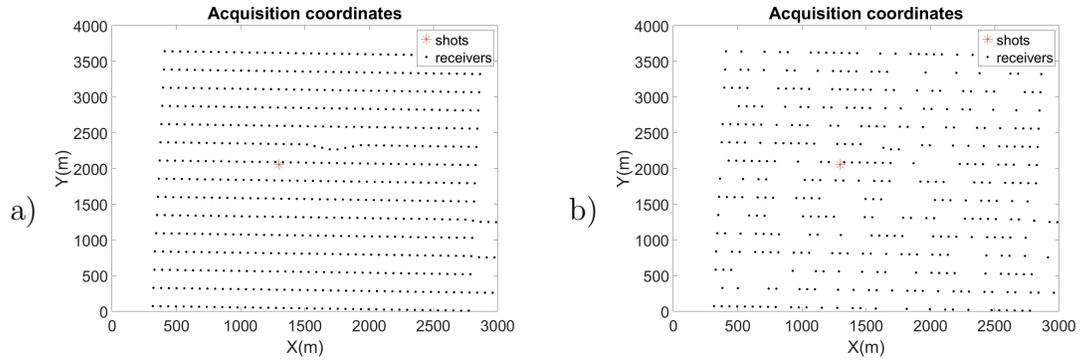


Figure 5.15: Geometry of the field data. (a) Initial distribution of the traces. (b) Distribution of traces after killing 41% of them.

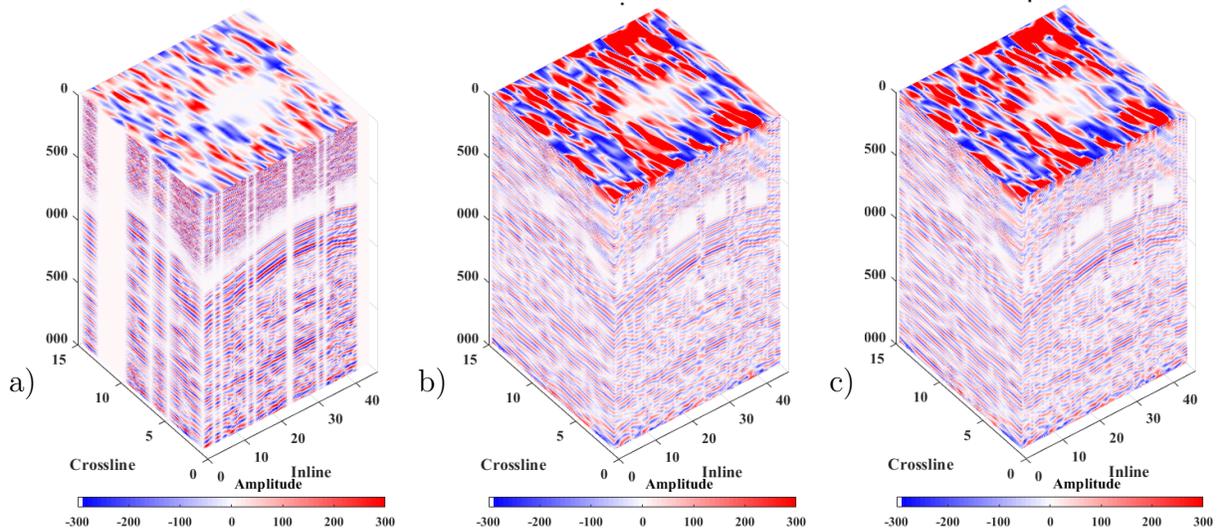


Figure 5.16: Cube of 3-D field data. a) The input data. b) Interpolation result of applying TRR. c) Interpolation result of applying AWRR.

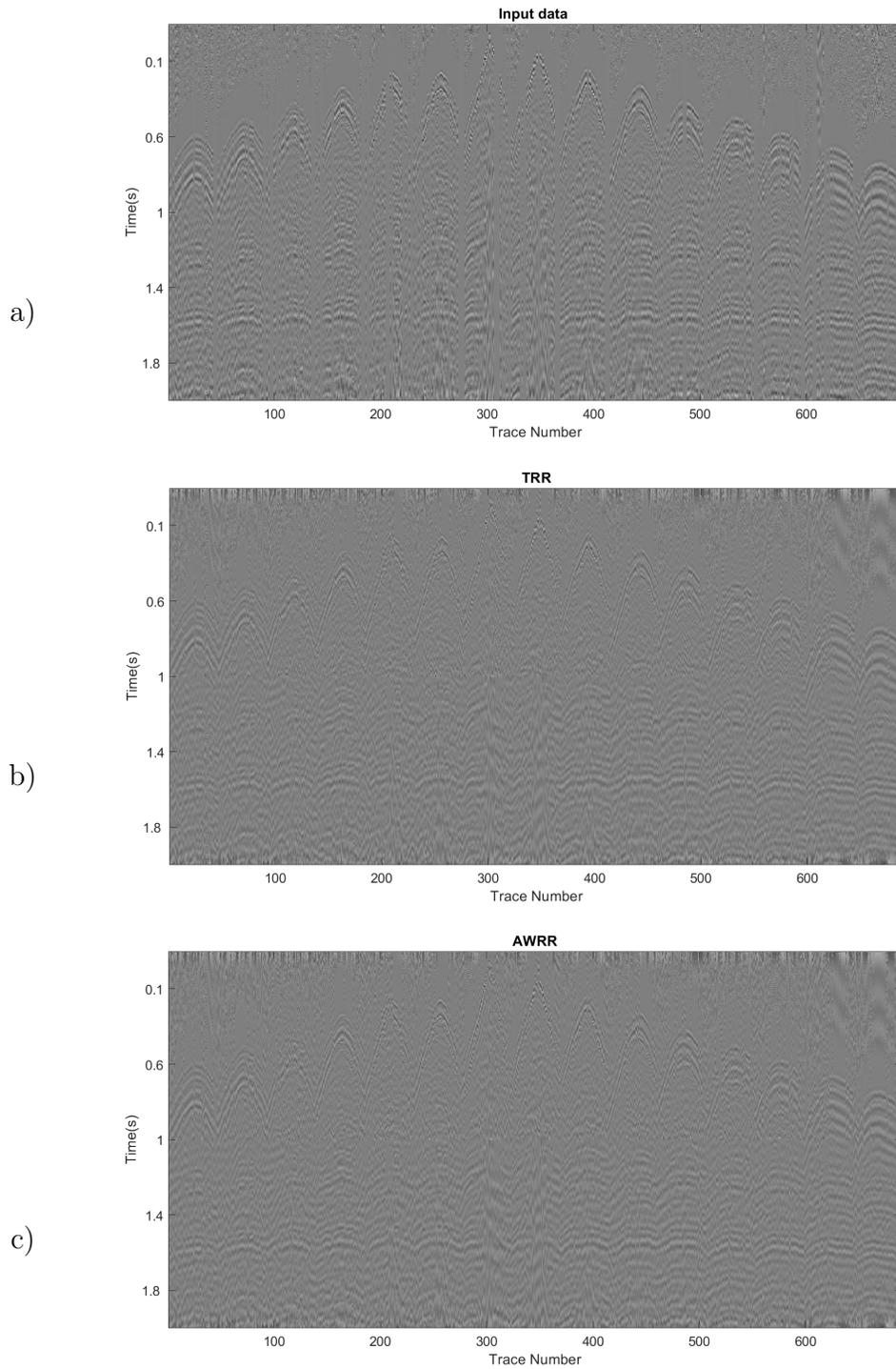


Figure 5.17: a) Input real data, b) interpolation result of applying TRR approach, c) interpolation result of applying AWRR method.

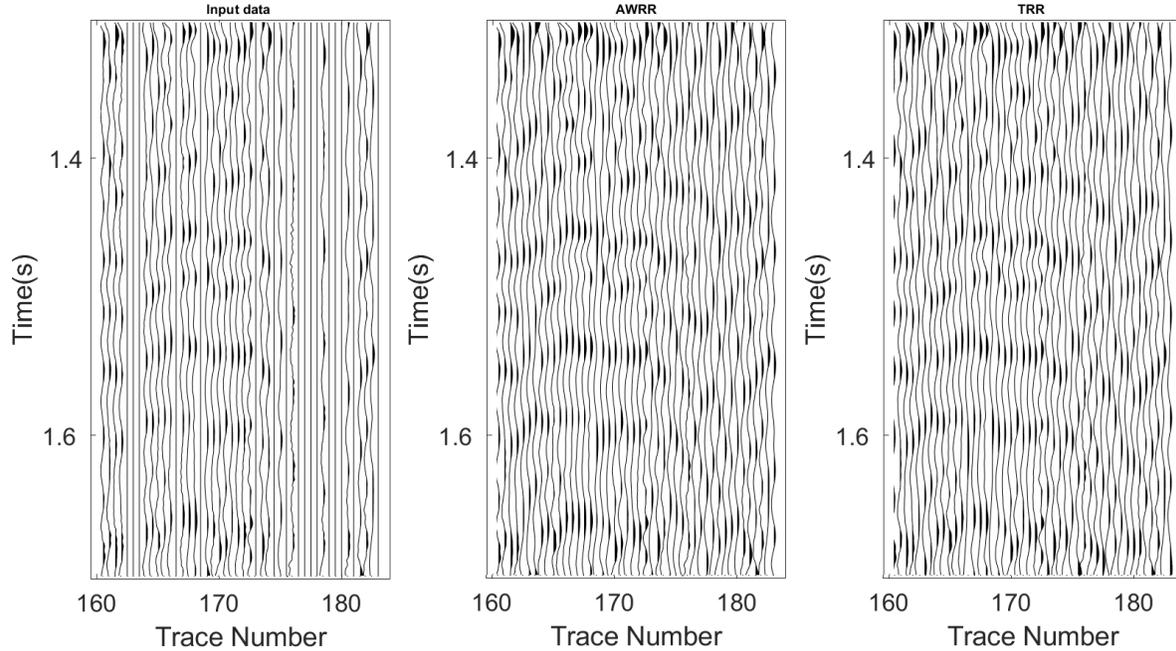


Figure 5.18: Zooming area of figure 5.17 from time 1.35 (s) to 1.75 (s) and trace number 160 to 190. a) input data, b) TRR result, c) AWRR result

5.4 Discussions

In this chapter, I compared different methods of rank-reduction in the MSSA algorithm to interpolate and denoise 3-D seismic data. I proposed the AWRR method, which is a parameter-free method. AWRR is based on a cascade weighting operator combined with the adaptive rank-reduction method. It is adaptive in selecting the rank of the Hankel matrices or the block Hankel matrices. Moreover, because of the weighting operator, AWRR can cause less residual noise.

AWRR is compared with the other rank-reduction methods the results showed that AWRR is more effective than the other methods especially in the case of low signal-to-noise ratio and when the data is sparse. The recovered block Hankel matrices in the frequency slices using AWRR are smoother and contain more details than the other rank-reduction methods.

The AWRR method also was compared with a “nearly adaptive” rank reduction method

called ORR. The results showed better quality of interpolation for ORR compared to WRR. However, the ORR method is not completely data-driven, and it is a computationally expensive algorithm.

Chapter 6

SSA comparison with Minimum Weighted Norm Interpolation

6.1 The binning approach

Most seismic surveys are irregularly sampled. Even if shots and receivers are laid out on a regular grid, offsets and azimuths are not, and midpoints will often be deviated from the bin centers and scattered across the CMP grid. Many interpolation methods need to set seismic traces into a regular grid. Fitting irregularly sampled recorded data to regular grids required careful algorithms because can lead to jittering of events and therefore loss of high frequency information in interpolation algorithms. A common approach to reduce binning errors is reducing bin intervals but that also increases the number of unknowns in the problem making optimization more difficult (Trad, 2009). Binning allocates a cluster of traces of adjacent coordinates to the nearest neighbour point of the desired regular output grid.

Finding an optimal binning size is important since coarse binning produces amplitude and phase errors of the traces specially for dippy events. Fine binning enlarges the size of the input data into the algorithm which makes it difficult to resolve the problem. It is worth noting that some algorithms work faster in a regular grid with more traces than the irregular

grid that contains few traces, therefore, binning size is a trade-off parameter for precision and complexity Trad (2008).

Liu and Sacchi (2004) proposed Minimum Weighted Norm Interpolation (MWNI) to reconstruct spatially band-limited signals. It is implemented via the fast Fourier transform. MWNI controls the bandwidth of the signal and the spectral shape of the reconstructed data by including prior spectral weights. Also, it can interpolate data in fine grid that leads to minimizing binning errors.

In this chapter, we examine the MWNI interpolation method and compare the findings with SSA.

6.2 Theory

In previous chapters, we discussed the Singular Spectrum Analysis (SSA) algorithm as a rank-reduction method of seismic data interpolation. We can summarize the main steps of SSA in 6 steps:

- Transforming data from $t - x$ or $t - x - y$ to $f - x$ or $f - x - y$.
- Generating a Hankel matrix (for 2-D data or a block Hankel matrix for 3-D data) in each slice of frequency.
- Decomposition of the Hankel matrix in its singular spectrum via singular value decomposition.
- Rank reduction of the Hankel matrix.
- Averaging in the Hankel matrix anti-diagonals.
- Inverse Fourier transform to return to the time domain.

By using an iterative algorithm, we can use SSA to interpolate seismic data. The iterative algorithm is as follows:

$$\tilde{\mathbf{S}}_{n+1}(\omega_m) = \alpha_n \mathbf{S}_{obs} + (\mathbf{I} - \alpha_n \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{H} \mathcal{F}(\mathbf{S}_n(\omega_m)) + (\mathbf{I} - \mathbf{R}) \mathcal{F}^{-1} \mathcal{A} \mathcal{R} \mathcal{H} \mathcal{F}(\mathbf{S}_n(\omega_m)), n = 1, 2, \dots, N \quad (6.1)$$

where ω_m is the frequency that SSA is applied on the signals, \mathbf{I} is the matrix of ones, $\alpha_n \in (0, 1]$ is the weight factor and linearly changes with the number of iterations, \mathbf{S}_{obs} denotes the observed data, \mathcal{F} indicates the Fourier transform, and \mathcal{F}^{-1} shows the inverse Fourier transform. \mathcal{H} points to the Hankelization operator, \mathcal{R} reveals rank reduction operator, and \mathcal{A} gives the averaging anti-diagonal operator (Oropeza and Sacchi, 2011).

We studied different methods of rank-reduction for the step of rank-reduction that can enhance the interpolation's results. We proposed an adaptive weighting rank-reduction method that is data-driven and can select the rank of the Hankel matrix properly. The SSA method assumes a regular distribution of traces and suffers from critical deficiencies in interpolating regularly decimated traces.

MWNI works in the Fourier domain and the interpolation algorithm is implemented for each frequency on the spatial dimensions. Let us consider a 2-D seismic data in the time domain. which can be shown as:

$$S(x, t) = W(t - px) \quad (6.2)$$

where x defines the space, t is time, p is dip and $w(t)$ indicates the wavelet. First data are taken to the $f - x$ domain by a Fourier transform.

$$S(x, \omega) = A(\omega) e^{-i\omega px} \quad (6.3)$$

where ω is the temporal frequency and A is the amplitude. $S_\omega = [s_1, s_2, \dots, s_{N_x}]^H$. The complete data and the observed data are connected by a linear system:

$$\mathbf{d} = L\mathbf{S}. \quad (6.4)$$

Assume \mathbf{R} as a sampling operator that is a matrix of ones and zeros. As an example of an observed data with $N_x = 9$ and some missing traces like $\mathbf{S}_\omega = [s_1, 0, s_3, 0, s_5, s_6, s_7, 0, 0]^H$, we can show:

$$\begin{pmatrix} s_1 \\ s_3 \\ s_5 \\ s_6 \\ s_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} (s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9)^H \quad (6.5)$$

Solving equation 6.4 leads to an undetermined system of equations. Among all the possible solutions, MWNI chooses a solution which minimizes a model norm. The inversion can be reduced to solving the constrained minimization problem:

minimizing $\|\mathbf{S}\|_p^2$ subjects to $\mathbf{d} = L\mathbf{S}$, where $\|\cdot\|_p^2$ indicates a weighted norm. In the presence of noise, the above constraints can be modified as: Minimize $\|\mathbf{W}_s\mathbf{S}\|_p^p$ subject to $\|\mathbf{W}_d(\mathbf{d} - L\mathbf{S})\|_q^q = \phi_d$, where p and q indicate the different norms used to estimate the size of the quantities involved, ϕ_d is the estimation of noise plus a residual due to the failure of the model, \mathbf{W}_s is a matrix of model weights and \mathbf{W}_d is a matrix of data weights. To obtain the desired solution we should minimize the cost function:

$$(\lambda\mathbf{W}_s^H\mathbf{W}_s + L^H\mathbf{W}_d^H\mathbf{W}_dL)\mathbf{S} = L^H\mathbf{W}_d^H\mathbf{W}_d\mathbf{d} \quad (6.6)$$

where λ is a trade-off parameter. By fixing the model to some previous estimation (like the spatial spectra from the previous temporal frequency) and applying a linear minimization

by conjugate gradient algorithm (CG) equation 6.4 is solved. To prevent a potential zero division (for a zero model) we can apply right preconditioning:

$$\mathbf{d} = L\mathbf{W}_s^{-1}\tilde{\mathbf{W}}_s\mathbf{S} \quad (6.7)$$

with $\mathbf{W}_s\mathbf{S}$ being a new model $\tilde{\mathbf{S}}$, and L_s^{-1} being a new operator \tilde{L} . The optimization problem leads to: Minimize $\|\tilde{\mathbf{S}}\|_p^p$ subject to

$$\|\mathbf{W}_d(\mathbf{d} - L\mathbf{W}_s^{-1}\tilde{\mathbf{S}})\|_q^q = \phi_d \quad (6.8)$$

The following system is the result of the minimization of the cost function of equation 6.8:

$$(\lambda\mathbf{I} + \mathbf{W}_s^{-H}L^H\mathbf{W}_d^H\mathbf{W}_dL\mathbf{W}_s^{-1})\tilde{\mathbf{S}} = \mathbf{W}_s^H \quad (6.9)$$

Equation 6.9 is solved by setting the trade-off parameter to 0 and letting the number of internal iterations in the conjugate gradient play the role of regularizer (Trad, 2003). The FFT algorithm assumes data are regular so binning is needed before applying the algorithms for interpolation (Liu and Sacchi, 2004; Trad, 2009).

6.2.1 Synthetic data

In this section, we set a regularization test for a 2-D synthetic shot gather, containing 2 hyperbolic and one linear events. The input traces were located at irregular offset, where 30% of which was randomly decimated. In this test, the offset was regularized with different binning sizes (figure 6.1). Figure 6.1 (a) and (b) shows the input data in the $t - x$ and $f - k$ domain. Figure 6.1 (c) and (d) indicates the complete and clean data in $t - x$ and $f - k$.

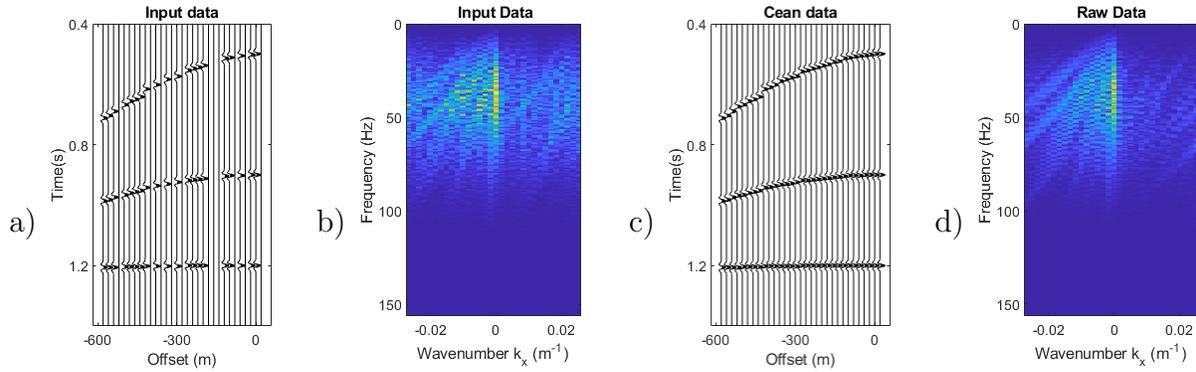


Figure 6.1: 2-D data with irregular offset. a) Input data with 30% randomly missing traces. b) $f - k$ spectrum of the input data. c) Desired data. d) $f - k$ spectrum of the desired data.

Figure 6.2 shows the geometry distribution of the input data where the ‘o’ indicates the true coordinates of the traces and ‘.’ refers to the regularized grid of traces with different binning sizes. Figure 6.2 (a), (b), (c), and (d) show the regularized grid for binning sizes 18, 9, 6, and 4.5 (m), respectively. The best match between the true traces locations and the regularized grid refers to the finest bin size, which was 4.5 (m).

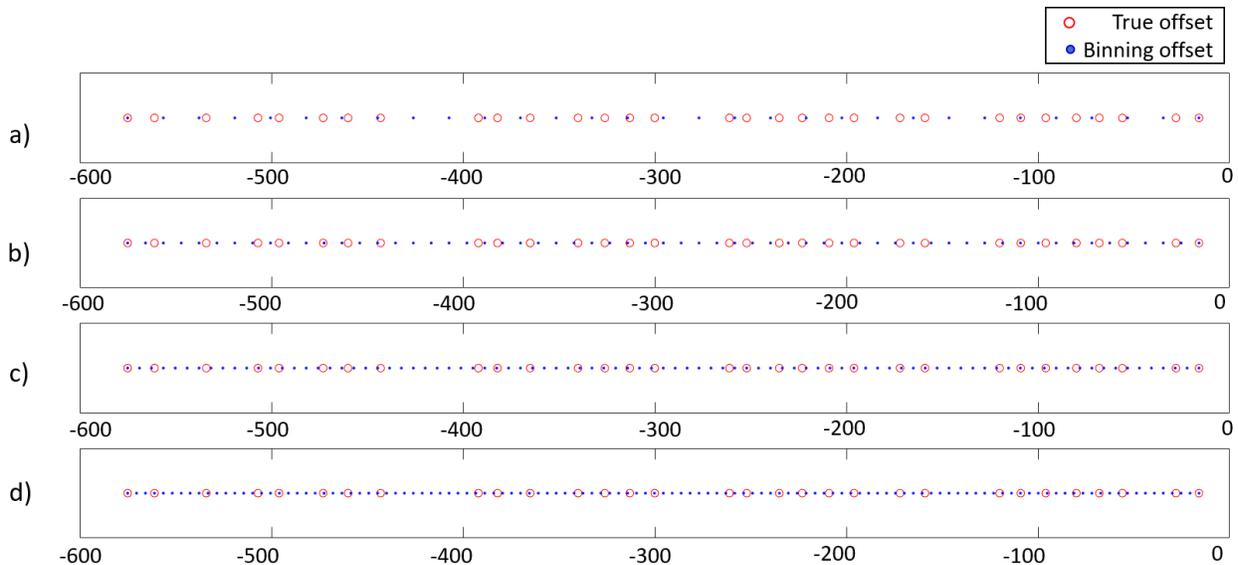


Figure 6.2: Geometry distribution of traces with irregular offset and different binning sizes. a) Bin size=18 (m), b) bin size=9 (m), c) bin size =6 (m), d) bin size=4.5 (m).

Figure 6.3 corresponds to the results of MWNI interpolation for different bin sizes. Fig-

Figure 6.3 (a), (d), (g), and (j) show the regularized and interpolated data for bin cells 18, 9, 6, and 4.5 (m), respectively. Figure 6.3 (b), (e), (h), and (k) show the corresponding interpolated results in their exact coordinates, the output quality of interpolation was calculated $QF = 9, 9.26, 9.35,$ and $10,$ respectively. The residual errors for the different input regularized data are shown in figure 6.3 (c), (f), (i), and (l) respectively. We can see that the flat event was interpolated entirely, whereas the curved events have some vertical errors. By reducing the size of the bin cells, the quality of interpolation was enhanced.

Figure 6.4 corresponds to the $f - k$ spectrum of the result of MWNI interpolation for different binning sizes. Figure 6.4 (a), (b), (c), and (d) indicates the $f - k$ filter when the regular offset was set to 18, 9, 6, and 4.5 (m) respectively. The effect of inserting zero traces regularly is aliasing which was removed after the MWNI interpolation.

Figure 6.5 (a) shows the result of applying SSA to the data in figure 6.1 the output quality was $QF = 8.5$ (dB). Figure 6.5 (b) depicts the results of subtracting interpolated data from the complete data. Figure 6.5 (c) is interpolated data in the $f - k$ domain.

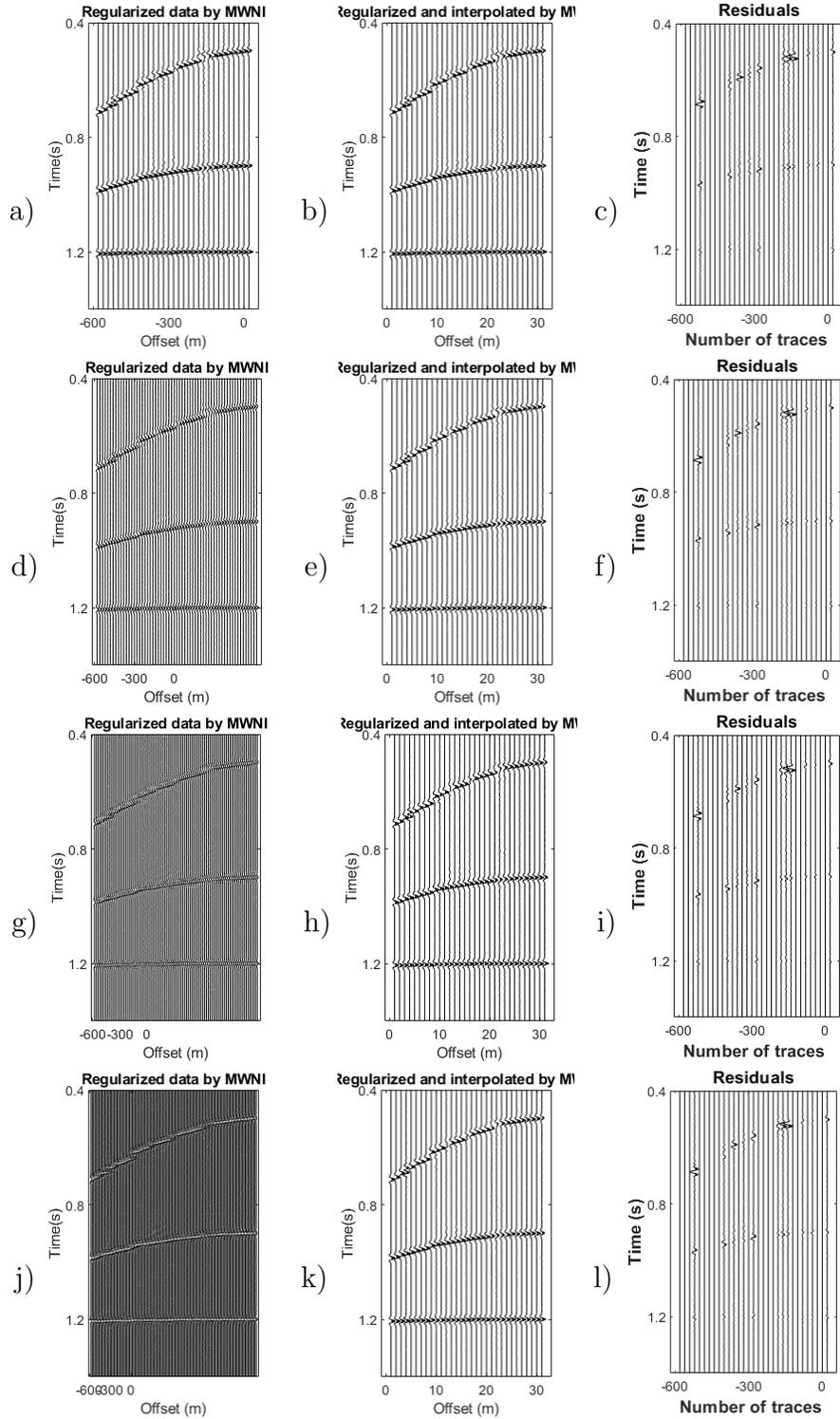


Figure 6.3: Results of MWNI interpolation on a 2-D data. a), d), g), and h) Regularized and interpolated data for different bin size of 18, 9, 6, and 4.5(m), respectively. b), e), h), and k) interpolated traces in their true coordinates. c), f), i), and l) Residual errors.

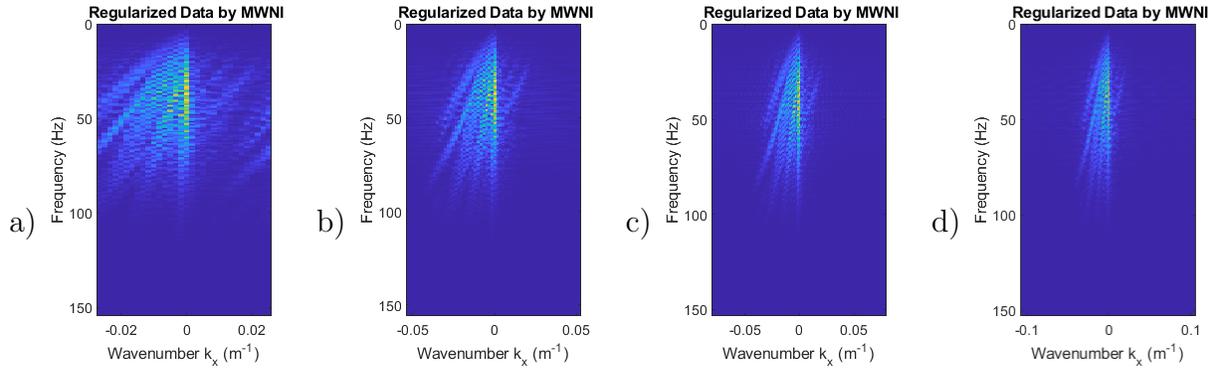


Figure 6.4: $f - k$ spectrum of the MWNI interpolation. a) Bin size=18(m), b) 9(m), c) 6(m), and d) 4.5(m).

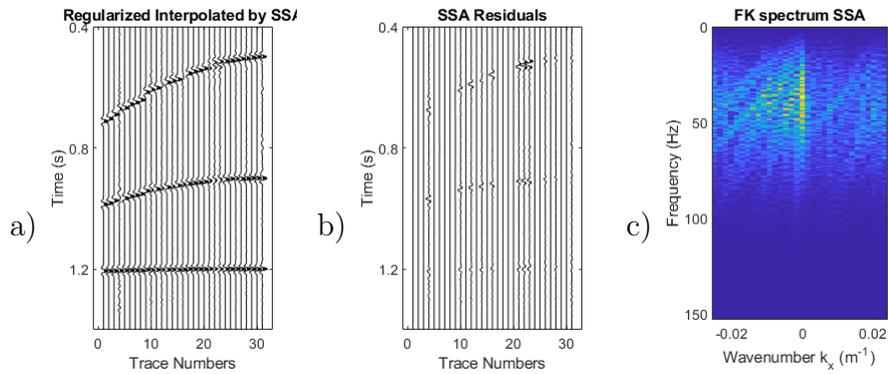


Figure 6.5: Results of SSA interpolation a) Interpolated data in a rough grid of 18(m), b) residual errors, c) $f - k$ spectrum of the interpolated data.

Chapter 7

Conclusions and Recommendations

Seismic data irregularities caused by acquisition constraints, economic limitations, or removal of bad traces, decreases the performance of many processing tools and imaging required for seismic interpretation. Therefore, interpolation of seismic data is a high priority tool in geophysics. In this thesis, we have studied the application of Singular Spectrum Analysis (SSA) with an iterative algorithm as a rank-reduction interpolation method. The primary objective was to improve the results of SSA interpolation by introducing different methods of rank-reduction operators.

In chapter 2, we reviewed the four main steps of the basic SSA algorithm and different uses of SSA for the case of time series, such as smoothing the series, finding the trends, or observing the periodic events. The main steps of the basic SSA algorithm can be summarized in four steps. Embedding of the time series in a trajectory matrix in the form of a Hankel matrix. Decomposition of this Hankel matrix into its singular values and singular vector matrices by SVD. Rank-reduction of the Hankel matrix by making zero the unwanted singular values (Truncated SVD) and recovering the matrix spanned by the main components. Averaging on the anti-diagonals of the restored Hankel matrix to reclaim the time series. This dimensionality reduction of the matrix eliminates unwanted components like noise or

artifacts produced by gaps.

In chapter 3, we studied the application of SSA to 2-D and 3-D seismic data. The main difference for seismic data in comparison with 1-D time series is that the algorithm is applied in the Fourier domain for each frequency slice. Here, we looked at the problem of how to define the rank of the Hankel matrix in each frequency slice when knowing the number of linear events in the processing window. SSA can be expanded to more than two dimensions with some rearranging of the Hankel matrices. Generating a block of Hankel matrices in each frequency slice, Multichannel Singular Spectrum Analysis (MSSA), uses the information of different dimensions of the seismic data simultaneously. Rank-reduction of the block Hankel matrices enhances the results of SSA. We reviewed MSSA for 3-D data using only three dimensions (one gather at the time), but full 3-D data interpolation requires the expansion to 5-D to take into account the data continuity across inlines, crosslines, offsets and azimuths. This requires creating windows with contributions from many shots and receiver gathers simultaneously. This can be done by generating the fourth-order of the Hankel matrix. In this chapter, the implementation of SSA with a POCS-like algorithm was studied. This iterative algorithm recovers amplitudes and performs denoising simultaneously. We examined the 3-D synthetic seismic data for interpolation and denoising by MSSA.

In chapter 4, we considered the problem of rank selection. Finding the optimal rank for rank-reduction in seismic data can be quite challenging and it has an important effect on the interpolation results. On one hand, choosing the rank too large leads to significant remaining noise or incomplete interpolation. On the other hand, underestimation of the rank distorts the signal by removing singular vectors that are necessary for a proper data reconstruction. In this chapter, we proposed an adaptive rank-reduction method by exploring the singular values of the Hankel matrix in each frequency slice. The method chooses the maximum energy ratio between the two following singular values as the criterion to define the optimal

rank. For 3-D data, the block Hankel matrices in higher frequencies are recovered with higher ranks. Consequently, the method selects the second cutoff in the singular values spectrum.

In Chapter 5, we considered the problem of additive random noise. We included a weighting operator in the rank-reduction step to reduce the effect of noise. We proposed a data driven method of rank reduction called Adaptive Weighting Rank-Reduction (AWRR), which leads to a decrease of residual noise in the recovered data. We compared rank reduction methods with and without applying the weighting operator on the block Hankel matrix. As an additional improvement, the result of AWRR in our tests was smoother than when applying other methods. We compared AWRR with a “nearly adaptive” rank reduction method called the Optimally damped Rank-Reduction method (ORR) for different percent of missing traces. Although the results of ORR outperformed AWRR in our tests, its computationally cost was larger.

Finally, in Chapter 6, we briefly discussed the Minimum Norm Weighted Interpolation method and the effect of binning when samples are located at variant distances from each other. This problem affects both MSSA and MWNI, and we made a quick comparison of the results from the two methods in these conditions.

7.1 Future Work

Interpolation of regularly decimated traces with SSA (upsampling), as opposed to gaps or irregular sampling, is an important topic for further research. For this case, it is interesting to inspect carefully the singular vectors of the Hankel matrix. For a data containing one event, the reduced rank Hankel matrix can be generated by the first singular value. Looking at the first column of the left and right singular vectors, we see zeroes representing the missing traces. The first singular vector can be recovered by omitting these zeroes and fitting a

polynomial curve to it. Figure 7.1 shows the result of the application of SSA on regularly decimated traces containing one linear event and their $f - k$ spectrum.

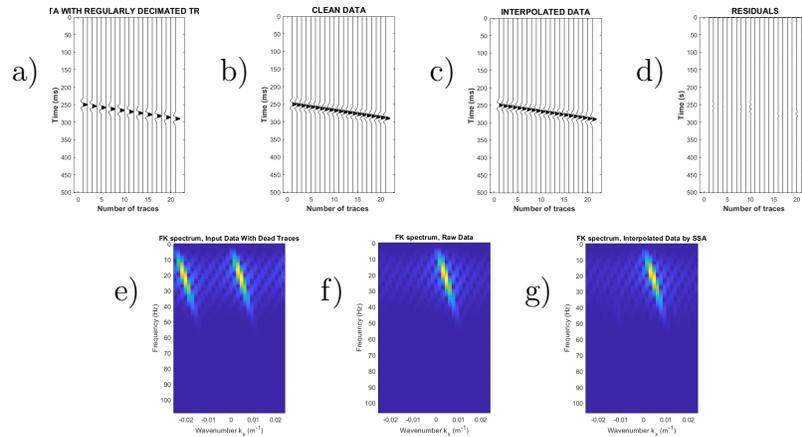


Figure 7.1: Interpolating decimated traces by SSA or upsampling. a) Input data. b) Desired data. c) Interpolated by SSA. d) Residual errors. e) $f - k$ spectrum of input data. f) $f - k$ spectrum of desired data. g) $f - k$ spectrum of interpolates data.

Most interpolation rank-reduction methods such as SSA are SVD-based methods. A SVD decomposition of a Hankel matrix is a very time-expensive operation, especially for block Hankel matrices as used in MSSA. An iterative approach in the interpolation part of SSA makes the algorithm time consuming. To make MSSA with AWRR feasible for higher dimensions, it is recommended to move from the SVD-based method to the SVD-free based rank minimization method.

The SSA approach has many similarities with what is known in machine learning as kernel dimensionality reduction, for example, Kernel PCA. It would be important to understand if the Hankel matrix formation can be formulated in terms of what is called the Kernel-trick to reduce computational costs.

Bibliography

- Abma, R. and Kabir, N. (2006). 3d interpolation of irregular data with a pocs algorithm. *Geophysics*, 71(6):E91–E97.
- Aoki, M. (2013). *State space modeling of time series*. Springer Science & Business Media.
- Benaych-Georges, F. (2009). Rectangular random matrices, related convolution. *Probability Theory and Related Fields*, 144(3-4):471–515.
- Benaych-Georges, F. and Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135.
- Bochner, S., Chandrasekharan, K., and Chandrasekharan, K. (1949). *Fourier transforms*. Number 19. Princeton University Press.
- Bregman, L. M. (1965). The method of successive projection for finding a common point of convex sets. *Sov. Math. Dok.*, 162(3):688–692.
- Broomhead, D. S. and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2-3):217–236.
- Carozzi, F. and Sacchi, M. D. (2021). Interpolated multichannel singular spectrum analysis: A reconstruction method that honors true trace coordinates. *Geophysics*, 86(1):V55–V70.
- Chen, K. and Sacchi, M. D. (2015). Robust reduced-rank filtering for erratic seismic noise attenuation. *Geophysics*, 80(1):V1–V11.

- Chen, Y., Zhang, D., Jin, Z., Chen, X., Zu, S., Huang, W., and Gan, S. (2016). Simultaneous denoising and reconstruction of 5-d seismic data via damped rank-reduction method. *Geophysical Journal International*, 206(3):1695–1717.
- Cody, R., Harmouche, J., and Narasimhan, S. (2018). Leak detection in water distribution pipes using singular spectrum analysis. *Urban Water Journal*, 15(7):636–644.
- Duijndam, A., Schonewille, M., and Hindriks, C. (1999). Reconstruction of band-limited signals, irregularly sampled along one spatial direction. *Geophysics*, 64(2):524–538.
- Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.
- Elsner, J. B. and Tsonis, A. A. (2013). *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media.
- Ely, G., Aeron, S., Hao, N., and Kilmer, M. E. (2015). 5d seismic data completion and denoising using a novel class of tensor decompositions. *Geophysics*, 80(4):V83–V95.
- Fomel, S. (2003). Seismic reflection data interpolation with differential offset and shot continuation. *Geophysics*, 68(2):733–744.
- Gao, J., Sacchi, M. D., and Chen, X. (2013). A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions. *Geophysics*, 78(1):V21–V30.
- Gao, J.-J., Chen, X.-H., Li, J.-Y., Liu, G.-C., and Ma, J. (2010). Irregular seismic data reconstruction based on exponential threshold model of pocs method. *Applied Geophysics*, 7(3):229–238.
- Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashov, D., Mann, M., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., et al. (2002). Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1):3–1.

- Golyandina, N. and Korobeynikov, A. (2014). Basic singular spectrum analysis and forecasting with r. *Computational Statistics & Data Analysis*, 71:934–954.
- Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. A. (2001). *Analysis of time series structure: SSA and related techniques*. CRC press.
- Golyandina, N. and Zhigljavsky, A. (2013). *Singular Spectrum Analysis for time series*. Springer Science & Business Media.
- Greiner, T. L., Kolbjørnsen, O., Lie, J. E., Nilsen, E. H., Evensen, A. K., and Gelius, L. (2019). Cross-streamer wavefield interpolation using deep convolutional neural network. In *SEG Technical Program Expanded Abstracts 2019*, pages 2207–2211. Society of Exploration Geophysicists.
- Hassani, H., Soofi, A. S., and Zhigljavsky, A. A. (2010). Predicting daily exchange rate with singular spectrum analysis. *Nonlinear Analysis: Real World Applications*, 11(3):2023–2034.
- Hsieh, W. W. and Wu, A. (2002). Nonlinear multichannel singular spectrum analysis of the tropical pacific climate variability using a neural network approach. *Journal of Geophysical Research: Oceans*, 107(C7):13–1.
- Jin, Y., Wu, X., Chen, J., Han, Z., and Hu, W. (2018). Seismic data denoising by deep-residual networks. In *SEG Technical Program Expanded Abstracts 2018*, pages 4593–4597. Society of Exploration Geophysicists.
- Kreimer, N. and Sacchi, M. D. (2012). A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. *Geophysics*, 77(3):V113–V122.
- Kreimer, N., Stanton, A., and Sacchi, M. D. (2013). Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *Geophysics*, 78(6):V273–V284.

- Kumar, R., Da Silva, C., Akalin, O., Aravkin, A. Y., Mansour, H., Recht, B., and Herrmann, F. J. (2015). Efficient matrix completion for seismic data reconstruction. *Geophysics*, 80(5):V97–V114.
- Liu, B. and Sacchi, M. D. (2004). Minimum weighted norm interpolation of seismic records. *Geophysics*, 69(6):1560–1568.
- Liu, G. and Chen, X. (2018). Seismic data interpolation using frequency-domain complex nonstationary autoregression. *Geophysical Prospecting*, 66(3):478–497.
- Liu, Y. and Fomel, S. (2011). Seismic data interpolation beyond aliasing using regularized nonstationary autoregression. *Geophysics*, 76(5):V69–V77.
- Mandelli, S., Borra, F., Lipari, V., Bestagini, P., Sarti, A., and Tubaro, S. (2018). Seismic data interpolation through convolutional autoencoder. In *SEG Technical Program Expanded Abstracts 2018*, pages 4101–4105. Society of Exploration Geophysicists.
- Mikhailiuk, A. and Faul, A. (2018). Deep learning applied to seismic data interpolation. In *80th EAGE Conference and Exhibition 2018*, volume 2018, pages 1–5. European Association of Geoscientists & Engineers.
- Nadakuditi, R. R. (2014). Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018.
- Naghizadeh, M. and Sacchi, M. (2013). Multidimensional de-aliased cadzow reconstruction of seismic records. *Geophysics*, 78(1):A1–A5.
- Oropeza, V. (2010). The singular spectrum analysis method and its application to seismic data denoising and reconstruction.
- Oropeza, V. and Sacchi, M. (2011). Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76(3):V25–V32.

- Porsani, M. J. (1999). Seismic trace interpolation using half-step prediction filters. *Geophysics*, 64(5):1461–1467.
- Read, P. (1992). Applications of singular systems analysis to ‘baroclinic chaos’. *Physica D: Nonlinear Phenomena*, 58(1-4):455–468.
- Ronen, J. (1987). Wave-equation trace interpolation. *Geophysics*, 52(7):973–984.
- Sacchi, M. (2009). Fx singular spectrum analysis: Cspg cseg cwls convention. *Abstracts, Calgary, Alberta, Canada*, pages 392–395.
- Sacchi, M. D., Ulrych, T. J., and Walker, C. J. (1998). Interpolation and extrapolation using a high-resolution discrete fourier transform. *IEEE Transactions on Signal Processing*, 46(1):31–38.
- Shi, Y., Wu, X., and Fomel, S. (2020). Deep learning parameterization for geophysical inverse problems. In *SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, Beijing, China, 5-7 November 2019*, pages 36–40. Society of Exploration Geophysicists.
- Siahkoochi, A., Kumar, R., and Herrmann, F. (2018). Seismic data reconstruction with generative adversarial networks. In *80th EAGE Conference and Exhibition 2018*, volume 2018, pages 1–5. European Association of Geoscientists & Engineers.
- Spitz, S. (1991). Seismic trace interpolation in the fx domain. *Geophysics*, 56(6):785–794.
- Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer.
- Trad, D. (2003). Interpolation and multiple attenuation with migration operators. *Geophysics*, 68(6):2043–2054.
- Trad, D. (2008). Five dimensional seismic data interpolation. In *SEG Technical Program Expanded Abstracts 2008*, pages 978–982. Society of Exploration Geophysicists.

- Trad, D. (2009). Five-dimensional interpolation: Recovering from acquisition constraints. *Geophysics*, 74(6):V123–V132.
- Trad, D., J. D. and Cheadle, S. (2005). Understanding land data interpolation. In *SEG Technical Program Expanded Abstracts 2005*, pages 2158–2161. Society of Exploration Geophysicists.
- Trickett, S. (2008). F-xy cadzow noise suppression. In *SEG Technical Program Expanded Abstracts 2008*, pages 2586–2590. Society of Exploration Geophysicists.
- Trickett, S. and Burroughs, L. (2009). Prestack rank-reducing noise suppression: Theory. In *SEG Technical Program Expanded Abstracts 2009*, pages 3332–3336. Society of Exploration Geophysicists.
- Van Loan, C. F. et al. (1996). Matrix computations (johns hopkins studies in mathematical sciences).
- Vautard, R., Yiou, P., and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1-4):95–126.
- Wang, B., Zhang, N., Lu, W., and Wang, J. (2019). Deep-learning-based seismic data interpolation: A preliminary result. *Geophysics*, 84(1):V11–V20.
- Wang, B., Zhang, N., Lu, W., Zhang, P., and Geng, J. (2018). Seismic data interpolation using deep learning based residual networks. In *80th EAGE Conference and Exhibition 2018*, volume 2018, pages 1–5. European Association of Geoscientists & Engineers.
- Wang, Y., Cao, J., and Yang, C. (2011). Recovery of seismic wavefields based on compressive sensing by an l_1 -norm constrained trust region method and the piecewise random subsampling. *Geophysical Journal International*, 187(1):199–213.
- Wu, J. and Bai, M. (2018). Adaptive rank-reduction method for seismic data reconstruction. *Journal of Geophysics and Engineering*, 15(4):1688–1703.

Yilmaz, Ö. (2001). *Seismic data analysis*, volume 1. Society of exploration geophysicists
Tulsa, OK.