Rank-constrained Multidimensional Reconstruction of Seismic Data

by

Fernanda Carozzi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Geophysics

Department of Physics
University of Alberta

Abstract

Our ability to generate accurate images of the Earth interior relies on three steps: data acquisition, data preconditioning, and imaging/inversion. Imaging and inversion methods adopt mathematical physics principles associated with wave propagation phenomena to transform the seismic observations acquired on the surface of the Earth subsurface images. Preconditioning techniques aim to remove coherent and incoherent noise, equalize the energy of the source, and solve problems resulting from inadequate data acquisition.

This thesis focuses on developing data preconditioning methods for multidimensional seismic volumes. I propose new data completion algorithms based on reduced-rank filtering for matrices and multilinear arrays (tensors). These techniques execute signal enhancement and data reconstruction simultaneously.

My research explores unsolved questions of reduced-rank reconstruction methods, including alleviating irregular data sampling and reconstructing data contaminated with outliers. To provide solutions to these questions, I propose two new algorithms. The first algorithm, based on the Multichannel Singular Spectrum Analysis (MSSA) algorithm, is named Interpolated MSSA (I-MSSA). Unlike classical MSSA methods, I-MSSA can honour true spatial coordinates of an irregularly sampled dataset. The second algorithm reconstructs five-dimensional seismic volumes based on robust tensor completion principles. To do so, I explore robust loss functions that attenuate the influence of outliers in the reconstruction.

Preface

A version of the work in chapter 3 of this thesis was published as Carozzi, F. and Sacchi, M.D., 2019. Robust tensor-completion algorithm for 5D seismic-data reconstruction. *Geophysics, 84(2)*, pp.V97-V109.

A version of the work in chapter 5 of this thesis was published as Carozzi, F. and Sacchi, M.D., 2021. Interpolated Multichannel Singular Spectrum Analysis (I-MSSA): a reconstruction method that honors true trace coordinates. *Geophysics, 86(1)*, pp.V55-V70.

I was responsible for designing the algorithms, programming the code, generating the numerical examples and writing the manuscript. Dr. M.D. Sacchi was the supervisory author and was involved with concept formulation and manuscript editing.

To my grandparents,
who have always instilled in me the importance of education.

Acknowledgements

I want to express my most profound appreciation to my supervisor, Professor Mauricio D. Sacchi, for his support and continuous guidance. Throughout my Ph.D., he has always pushed me to consider meaningful research questions, allowing for much growth and discovery. His invaluable expertise, ability to connect topics and unique style of simplifying problems are nurturing and inspiring. I would like to thank my committee members and professors for their support and guidance over the years. I am also grateful for my colleagues at the Signal Analysis and Imaging group. Every member of this group, past and present, contributed enormously to my research and development during the program.

I am forever thankful to my friends at the University of Alberta for the constant support and encouragement. I am also grateful to my family for their patience and motivation in my career pursuits over the years. Finally, I thank my husband, Agustin. He gave me the freedom to follow my dreams, believed in me when I doubted and brought me back to reality when I went astray. I will forever be grateful for your unconditional love and encouragement.

Contents

Appendices

List of Tables

List of Figures

List of abbreviations and symbols

$\mathbf{x}$	Vector
$\mathbf{A}$	Matrix
$\mathbf{A}^T$	Transpose of a matrix
$\mathbf{A}^*$	Hermitian transpose of a matrix
$\mathbf{A}^{-1}$	Inverse of a matrix
$\mathbf{A}^\dagger$	Pseudoinverse of a matrix
$\mathcal{X}$	Tensor
$\{s\}$	Set of elements
X	Fourier transform of $\mathbf{x}$
Ω	Operator
M	Dimension of an array
x	Coefficient
x_i	Elements of a vector
a_{ij}	Elements of a matrix
x_{ijkl}	Elements of a tensor
$\sum_i$	Summation over i-th element
$\lvert \cdot \rvert$	Absolute value
$\lVert \cdot \rVert$	Vector norm
$\langle \cdot \rangle$	Inner product
$\odot$	Outer product
$\circ$	Hadamard or element-wise product
$\hat{\mu}$	Estimate of μ
$f\text{-}\mathbf{x}$	Frequency-space domain
$f\text{-}\mathbf{k}$	Frequency-wavenumber domain
$t\text{-}\mathbf{x}$	Time-space domain
I-MSSA	Interpolated multi-channel singular spectrum analysis
PGD	Projected gradient descent algorithm
PMF	Parallel matrix factorization algorithm
`S/N`	Signal-to-noise ratio
SVD	Singular value decomposition

CHAPTER 1

Introduction

The seismic method uses principles of wave propagation phenomena to obtain an interpretable image of the subsurface. The experiment consists of three main steps. The initial stage is the acquisition of the seismic data. An active source triggers a signal with a controlled pulse of energy. This energy propagates into the subsurface of the earth in the form of seismic waves, transmitting and reflecting energy. Reflected waves are measured by arrays of detectors and processed to obtain the seismic volume. The depth range of the structures to investigate determines the seismic experiments and the possible applications. The near-surface seismic data experiment, required for geotechnical investigations, environmental assessments, and shallow resource exploration, covers up to 1 km in depth. Seismic methods for hydrocarbon exploration, or exploration seismology, targets a depth range of 1 to 10 km and requires expensive acquisition systems with source-detector distances of the order of 10 km. Deeper seismic data, covering up to 100 km depth, can image Earth's crust and often the Moho. The investigation depth defines the specific acquisition, processing, and interpretation tools to be used.

Exploration seismology can target onshore and offshore geological objectives. Since this environment differs significantly, acquisition strategies must be adapted for each case (Vermeer, 1990). Onshore acquisition of seismic data includes shot points with vibroseis or dynamite sources. Fixed geophone stations record the signal on the surface of the earth. These receivers are electromechanical transducers that record the waves impinging on them and convert ground displacement into electrical signals. The signal is digitized and registered as a time series called a seismic trace or seismogram. If seismic sources and receivers are deployed on a line, the experiment results in a 2D profile. We often refer to this particular layout as a 2D seismic acquisition. Modern seismic exploration methods deploy areal arrays of receivers and sources to collect the so-called 3D seismic acquisition. The latter allows capturing seismic responses from 3D geological structures. Offshore acquisitions also use an exploding energy source, usually an air-gun towed by a vessel. The cable or streamer also carries hydrophones or channels. Another popular technique for offshore acquisition is to lay arrays of geophones and hydrophones on the seabed. The method is called Ocean Bottom Seismic (OBS) and finds broad applications in reservoir monitoring.

In both the onshore and the marine environments, seismic data includes signals and noise. For our purposes, signals are seismic reflections that one needs to process to form an interpretable image of the subsurface. Marine datasets, for example, are often affected by multiples or energy that reflects multiple times. Multiples are a consequence of the high contrast in the elastic properties between the sea and the seabed and, for practical purposes, are considered coherent noise that one must eliminate. This thesis addresses mainly processing problems that affect onshore seismic surveys. In onshore data, ground-roll (surface waves) and direct waves contaminate seismic reflections. Typical processing flows start with aggressive denoising methods to eliminate these types of waves.

A standard processing sequence includes the following steps (Yilmaz, 2001): static corrections, ground-roll attenuation, deconvolution, stacking, and migration. Deconvolution improves the temporal resolution of the recorded events by collapsing the signature of the seismic source. Stacking exploits the redundancy of sources and receivers to estimate an initial image of the subsurface. Migration techniques adopt wave propagation principles to estimate subsurface images with their structures in the correct spatial position. Denoising and imaging techniques have stringent sampling requirements. For instance, many processing techniques and migration methods are designed for data acquired on a regular grid of sources and receivers. They also have stringent requirements in terms of Nyquist sampling. For instance, ground-roll becomes challenging to filter when receiver sampling does not honour Nyquist spatial sampling criteria. When acquisition constraints are not correctly observed, reconstruction algorithms can be adopted to reconstruct the recorded seismic wavefield.

1.1 Seismic Data Reconstruction

The seismic wavefield is a continuous signal of time and space. However, the signal acquired by the receivers is a discrete-time series. To fully sample the wavefield, the acquisition design should cover a regular spatial 4D grid. Thus, each sample depends on five coordinates: time, source location (s_x, s_y), and receiver location (r_x, r_y). Such layout defines the offset, midpoint, and azimuth of traces, as indicated in Figure 1.1. A relevant concept related to homogeneous coverage in seismic is the definition of fold. The fold measures the redundancy of data for each common midpoint. Irregularities in fold of coverage and offset and midpoint distribution can result in a low-quality image of the subsurface.

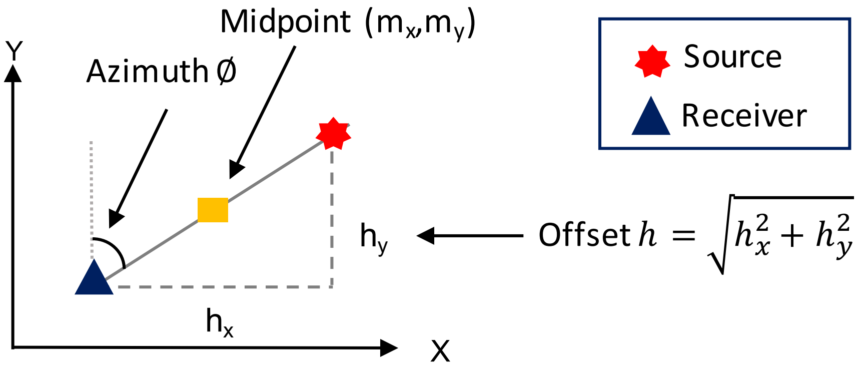

Figure 1.1: Schematic representation of the seismic acquisition geometry, plane view.

Recovering a continuous band-limited signal from its discrete samples requires a sampling rate at least twice as high as its highest frequency. Otherwise, the signal is aliased; the high frequencies of the input signal take the identity of a lower frequency (Oppenheim et al., 2001; Sheriff, 2002). In time, the seismic signal usually fulfills the Nyquist criterion because time-dependent signals are sampled by hardware that can produce the proper sampling. On the contrary, the spatial sampling is typically irregular or insufficient in offset, azimuth and fold. Sources and receivers are deployed by people working in the field. Therefore, it is more challenging to control spatial sampling than temporal sampling. The effect, resulting from economic and geographical constraints, is most noticeable in onshore acquisitions (Ely et al., 2015). However, inadequate sampling also affects marine data because of cable feathering and bad trace editing (Xu et al., 2005).

Reconstruction of seismic data is relevant to many processes, including surface-related multiple removal (SMRE), amplitude versus offset (AVO) and amplitude versus azimuth (AVA) analysis, removal of acquisition footprint and migration. The following paragraphs discuss the connections.

The basic model in seismic processing assumes only primary reflections. Then, the effective removal of multiples is critical for adequately imaging and interpreting offshore fields (Weglein, 1995). Multiples generate due to high reflectivity contrasts at water bottom, and top and base of salt surfaces. However, the leading cause of considerable multiple energy is the reflectivity of the surface (Verschuur et al., 1992). To properly remove the complex multiples in these environments, SRME requires high density and wide azimuth data. Conventional acquisition geometries have sparse source locations and narrow crossline aperture (Lin et al., 2005). Therefore, reconstruction becomes a necessity for multiple removal algorithms.

AVO and AVA analysis study the amplitude variations in the traces with source-receiver distance or angle. AVO/AVA effects have particular relevance in seismic exploration as they are direct hydrocarbon indicators. However, noisy and irregularly sampled traces might distort their effect (Liu, 2004). Since acquiring data with sufficient sampling in offset and azimuth presents logistic and economic challenges, reconstruction gains particular relevance. By improving the offset and azimuth coverage and reducing noise, AVO/AVA effects can highlight velocity and density changes, providing a valuable reservoir characterization tool.

The acquisition footprint is a linear grid pattern on 3D seismic time slices or horizontal amplitude maps on shallow times. These patterns, resulting from suboptimal spatial sampling, correlate with the acquisition geometry affecting the real amplitude and lateral continuity of the seismic events (Al-Bannagi et al., 2005). Reconstruction of seismic traces reduces the acquisition footprint, allowing for better interpretation of the near-surface properties (Trickett et al., 2010; Canning and Gardner, 1998).

Prestack migration also benefits from seismic reconstruction. For example, in Kirchoff migration, each trace is added to the migrated volume by spreading the data along impulse response curves. Ideally, overlapping impulse responses form the correct answer and cancel elsewhere. Cancellation does not happen, in general, if there are irregular variations in the midpoint, offset, azimuth, and velocity sampling

(Gardner and Canning, 1994). Wave equation migration (Claerbout, 1971) also requires common-shot data, with the receiver array covering the entire domain of the survey. Sparse sampling and irregular azimuth coverage result in strong artifacts. Seismic interpolation provides a practical means to obtain proper migrated images.

Given the importance of proper trace recovery, there have been many contributions to seismic data reconstruction and interpolation. The following sections review past work on seismic reconstruction.

1.1.1 Review of previous work

The sampling of onshore seismic data can be inadequate for two reasons. Data might be spatially aliased, or seismic traces might be irregularly acquired. In any case, seismic reconstruction is crucial to image the subsurface properly. Reconstruction algorithms attempt to generate missing traces by using a deterministic spatial interpolation approach. Essentially, algorithms attain reconstruction by assuming that the seismic wavefield is simple enough to be represented as a finite number of basis functions. In other words, on a small window, the seismic signal is a superposition of plane waves (Stanton and Sacchi, 2013). Given the relevance of the field, many authors have developed techniques to address the problem. A broad classification of methods recognizes three categories, algorithms based on the wave equation, algorithms based on signal processing tools, and algorithms based on machine learning. The algorithms based on signal processing tools use reconstruction via domain transforms, prediction error filters, and rank-reduction methods. Table 1.1 shows a diagram emphasizing signal processing tools used in seismic reconstruction. The following paragraphs briefly introduce the most salient techniques in each category.

Wave-equation-based methods

A typical approach in wave-equation-based methods considers the linearization of the Born approximation. The approximation relates data to reflectivity via the subsurface parameters, introduced as a linear operator. The method uses the observations to invert the linear operator and estimate the subsurface reflectivity. From the resulting reflectivity, one can model the data at new spatial positions. The inverse problem is ill-posed and requires the inclusion of regularization methods. In

Transforms	Prediction Filters	Rank Reduction	
		Matrix	Tensor
Fourier	FX	FXY Eigenimage	HOSVD
Radon	FK	SSA/Cadzow	Nuclear Norm
Curvelet	TX	MSSA	Tucker
		Nuclear Norm	Tensorial SVD
		Factorization	PMF

Table 1.1: Signal processing-based seismic reconstruction algorithms

addition, the linearized Born operator requires the subsurface velocity model, which is seldom available. The methods are attractive because they preserve the observed trace position and adopt wave propagation physics to reconstruct the seismic wavefield. Methods proposed by Ronen (1987); Nemeth et al. (1999); Stolt (2002); Fomel (2003); Malcolm et al. (2005), and Kaplan et al. (2010) fall within this category.

Fourier methods

Reconstruction methods based on signal processing techniques do not require a subsurface velocity model and are the workhorse for reconstructing massive datasets in the seismic processing industry. Reconstruction via signal processing often uses linear transforms to represent the available data by coefficients in the transformed domain. The problem is underdetermined and requires a constrained inversion. Once the coefficients are estimated, they are used to synthesize data at new spatial locations. The methods are computationally efficient, easy to use, and usually achieve denoising of the signal. The methods in this category typically consider the Fourier transform, but other transforms as Radon (Kabir and Verschuur, 1995; Trad et al., 2002) and Curvelet (Hennenfent and Herrmann, 2006; Herrmann et al., 2008; Naghizadeh and Sacchi, 2010a) are often explored.

Seismic data reconstruction methods based on multidimensional Fourier transforms gained popularity in the past decades. Duijndam et al. (1999) consider reconstruction from irregular coordinates via a least-squares approach. The algorithm calculates the signal in the regularly sampled Fourier domain from irregularly sampled data. To do so, the authors employ a nonuniform discrete Fourier transform (NDFT) to estimate the Fourier spectrum of the band-limited signal. The algorithm allows for gaps in data and is an alternative to binning and stacking given a reasonable bandwidth and a typical stack fold. Liu and Sacchi (2004) approximate the reconstructed traces with a minimum weighted wavenumber-domain norm. The algorithm, Minimum Weighted Norm Interpolation (MWNI), uses the smoothed periodogram to approximate the matrix of weights required to impose simplicity (sparsity) in the wavenumber Fourier coefficients. The method is computationally more efficient than NDFT as it relies on the Fast Fourier Transform (FFT), but it assumes regularly sampled or binned input data. Antileakage Fourier Transform (ALFT) (Xu et al., 2005) reconsiders the problem of irregularly sampled data and recognize that the Fourier bases are non-orthogonal in an irregular grid. The non-orthogonality of Fourier bases results in spectral leakage, which is energy from one Fourier coefficient leaking into others. To attenuate the effect, the authors use a greedy algorithm to estimate the Fourier coefficients that honour the data. The method also handles mild levels of aliasing. In a different vein, Zwartjes and Gisolf (2006) consider the nonuniform fast Fourier transform (NFFT) to reconstruct an irregularly sampled wavefield. The algorithm includes a non-quadratic function in the model, forcing a sparse solution. This new function stabilizes the inversion and improves reconstruction through gaps by weighting the large model parameters less than the smaller ones. To solve the Fourier coefficients, the authors use a preconditioned IRLS algorithm (Trad et al., 2003).

Abma and Kabir (2006) introduced Projection onto Convex Sets (POCS) for seismic data reconstruction. POCS relies on an iterative scheme that applies a threshold to the transformed data. The algorithm progressively lowers the threshold to include weaker events that might have been ignored or poorly reconstructed in the first iterations. The method reconstructs irregularly missing samples on a regular grid. Other authors extended POCS usage to include antialiasing filters (Gao et al., 2013) and investigate different thresholding schemes. Stanton et al. (2015), for example, demonstrate that soft thresholding combined with debiasing improves results over

Algorithm	Irregular coordinates	Aliasing	Fast Computation
Fourier Inversion with NDFT	✓		
MWNI			✓
ALFT	✓		
POCS			✓
Prediction Filters		✓	

Table 1.2: Salient properties of reconstruction signal-processing-based algorithms

hard thresholding. Finally, POCS can also reconstruct multicomponent data via its quaternionic extension (Stanton and Sacchi, 2011). The method relies on the spectral overlap of components in the frequency-wavenumber domain to improve the reconstruction quality over independent component-wise reconstruction.

Methods considering sparse coefficients are closely related to the field of compressive sensing. The aim is to recover the complete wavefield from randomized incoherent sampling using sparsifying transforms and sparsity-promotion recovery (Herrmann and Hennenfent, 2008; Herrmann, 2010). The sought solution is the one that leverages a sparse spectrum. By taking the inverse Fourier transform of the vector that solves the sparsity promoting problem, one recovers the reconstructed data. Li et al. (2012) define an interpolated restriction operator with coefficients given by a local Lagrange interpolation polynomial. Following a similar idea, Jiang et al. (2017) extend POCS by including a mapping operator in the iterative scheme. The mapping is a weighted summation from neighbouring traces considering the linear, cubic, sinc, Lagrange or spline interpolator. Table 1.2 describes salient properties of reconstruction signal-processing-based algorithms.

Prediction filter methods

Fourier methods are considered a non-parametric approach to spectrum estimation and data reconstruction. The parametric alternative is via prediction filtering, which uses autoregressive modelling to estimate spatial signals. The algorithms can operate in the time-space domain (Claerbout, 1992; Crawley et al., 1999), in the frequency-space domain (Spitz, 1990; Porsani, 1999; Naghizadeh and Sacchi, 2010b), or the frequency-wavenumber domain (Gulunay, 2003). In general, prediction filtering deals with interpolating regularly sampled data onto a high-density regular grid, including interpolation beyond aliasing. In other words, the technique upsamples the data in a grid rather than completing irregularly missing patterns. Prediction filter-based algorithms assume that the signal is a superposition of plane waves. Then, the signal is predictable in space. In the reconstruction problem, one estimates the coefficients of the filter from the low-frequency components of the signal to interpolate the high-frequency aliased events.

Rank-reduction methods

Rank-reduction-based denoising and reconstruction methods also assume that data are predictable in space. The data predictability results in low-rank arrays in the f-x domain. Methods in this category recover irregularly missing samples on a regular grid. Therefore, binning is usually required before the reconstruction. The following section focuses on rank-reduction reconstruction algorithms in more detail.

Machine learning-based methods

Finally, recent advances in seismic data reconstruction correspond to methods included under the machine learning category. These methods differ from previous algorithms in that they do not depend on a model but a training dataset. Jia and Ma (2017) reconstruct undersampled data or missing traces via a support vector regression. Mandelli et al. (2018) train an autoencoder to fill the missing samples in a corrupted prestack seismic image. Greiner et al. (2019) use a convolutional neural network for upscaling a non-aliased wavefield. The authors study, in particular, the interpolation of near-offset traces. Continuing the development of convolutional neural networks, Shi et al. (2019) introduce a random initialization that improves the quality of the reconstruction compared to other methods. Finally, Wang et al. (2019) implement a residual learning network (ResNets) to reconstruct dense seismic data beyond aliasing.

1.1.2 Rank reduction in seismic reconstruction

Rank-reduction methods can enhance the signal-to-noise ration (`S/N`) and reconstruct the seismic data. A broad classification of the methods divides them into two groups. The first group applies rank reduction to seismic data embedded in matrices and the second to seismic data embedded in tensors or multilinear arrays.

Initial contributions in the low-rank field were related to the seismic data denoising. Early efforts include eigenimage filtering in the f-x and f-x-y domain (Freire and Ulrych, 1988; Trickett, 2003). Eigenimage filtering explores the fact that coherent energy maps onto their first eigenimages. The method consists of replacing spatial frequency slices with the sum of the first few eigenimages via the Truncated Singular Value Decomposition (TSVD). Next, Cadzow filtering was introduced to seismic data by embedding the frequency slices in a Hankel matrix before calculating the TSVD (Trickett and Burroughs, 2009; Burroughs and Trickett, 2009). Combinations of eigenimage and Cadzow filter were also explored (Trickett and Burroughs, 2009). Simultaneously, Sacchi (2009) introduces the Singular Spectrum Analysis (SSA) algorithm for seismic data denoising. Figure 1.2 shows the workflow of the algorithm.

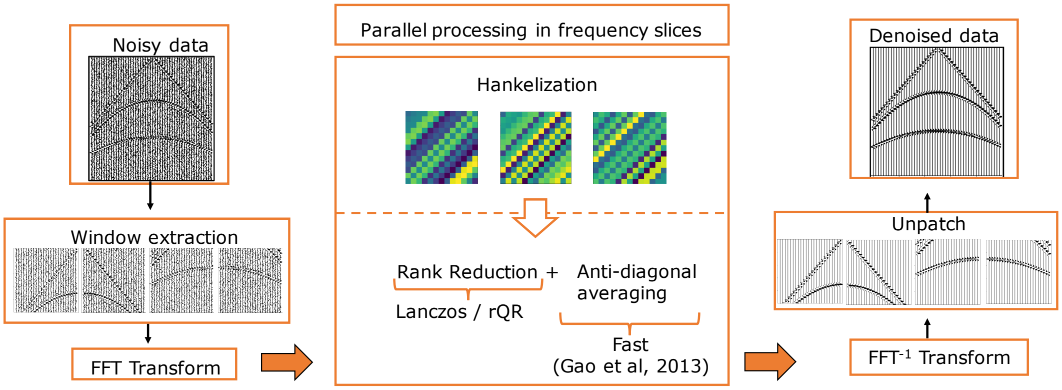

Figure 1.2: Workflow of the SSA algorithm.

Oropeza and Sacchi (2011) advanced rank-reduction methods to simultaneous denoising and interpolation of 3D seismic data. The authors introduce Multichannel

Singular Spectrum Analysis (MSSA), an SSA-based algorithm that allows reconstruction via a POCS-like imputation scheme. MSSA typically considers a linear decrease of the imputation parameter resulting in gradual reinsertion of the estimated filtered data at each iteration. A drawback of the MSSA method is the computational cost of the TSVD on the Hankel matrix. Randomized techniques for rank-reduction, including randomized SVD and QR decomposition, extend the capabilities of the algorithm (Oropeza and Sacchi, 2010; Cheng et al., 2019). A different computational efficiency approach studies a factorization strategy that avoids the SVD (Aravkin et al., 2014; Kumar et al., 2015; Chen and Sacchi, 2015).

The MSSA method was modified to cope with prestack 5D reconstruction by forming block Hankel matrices (Trickett et al., 2010). Several acceleration methods followed based on the Lanczos bidiagonalization technique (Gao et al., 2011, 2013; Cheng et al., 2019). Fast solvers utilize fast matrix-times-vector multiplications of circulant matrices employing FFT properties (Gao et al., 2011). The technique significantly reduces the computational cost of the iterative Lanczos bidiagonalization algorithm and permits the reconstruction of massive volumes of data. Finally, Sacchi et al. (2017) extend the SSA algorithm to multicomponent or 3C seismic data. Multicomponent seismic data samples the three components (one vertical and two horizontal) of the wavefield. Vector SSA embeds the vector measurements into a Hankel matrix, where each element is the 3C field. Bahia and Sacchi (2019) follow a different approach and represent the vector-field as quaternions. The authors use the quaternionic rank-reduction method to solve the problem.

This thesis evaluates a shortcoming of the original MSSA reconstruction algorithm. Even though seismic data are irregularly sampled in the spatial domain, the MSSA method assumes a regular distribution of traces. The standard approach is to map seismic traces to a regular grid before reconstruction. The process, called binning, introduces errors in the amplitude and phase of the traces. López et al. (2016) considered non-uniform Fourier transform to reduce the impact of binning on the final image. In this thesis, I consider an inverse problem that minimizes the residuals between observations and synthesizes data in the original spatial coordinates. The MSSA filter is applied as a projection operator to the desired reconstructed data in the regular grid. The problem reduces to optimizing an objective function via the projected gradient descent method.

The full seismic wavefield has four spatial dimensions in addition to time. In such a high dimensionality, the low-rank properties of the data remain valid. By embedding the data into tensors, reconstruction algorithms benefit from the data redundancy and relations in the different dimensions. Therefore, tensor completion methods operate directly on the multilinear array rather than on data embedded into Hankel or block Hankel matrices. Kreimer and Sacchi (2012) introduce tensorial algebra into the seismic reconstruction problem. The authors recognize that a low-rank tensor represents the ideal (fully sampled and noise-free) seismic data. The algorithm resorts to the High-Order-SVD (HOSVD) algorithm to solve the inverse problem. Kreimer et al. (2013) adopt the nuclear norm minimization to circumvent rank selection. Trickett et al. (2013) introduce the Hankel tensor completion algorithm. The algorithm first forms a tensor from a frequency slice of the data. Then, it applies tensor completion using the canonic polyadic decomposition and ignoring missing elements. Finally, it recovers the reconstructed frequency slice by averaging over every tensor element in which each frequency slice value was originally placed. Da Silva and Herrmann (2015) exploit the Hierarchical Tucker decomposition. Ely et al. (2015) introduce tensorial SVD and Popa et al. (2020) improve the computational time by exploiting the symmetry properties of the Fourier transform. Finally, Gao et al. (2015, 2017) introduce the Parallel Matrix Factorization algorithm (PMF) to the seismic reconstruction problem. Chapter 5 of this thesis presents a robust implementation of PMF to reconstruct seismic data with non-Gaussian noise. The following section describes the importance and the challenges of including non-Gaussian noise in reconstruction algorithms.

1.2 Erratic noise in seismic data

Seismic traces are composed of reflection information contaminated by noise. By adopting an additive noise model, one can represent the observed data as

$$\mathbf{U} = \mathbf{D} + \mathbf{N}\,,$$

where $\mathbf{U}$ is a discrete-time series containing the seismic data or traces, $\mathbf{D}$ is the signal or reflection information, and $\mathbf{N}$ is the noise. The seismic signal can be represented by a superposition of linear events in a small time-space window (Sacchi, 2009; Trickett et al., 2010). On the contrary, the noise is so complex that an accurate description is extremely complicated (Oppenheim et al., 2001). Then, one often models the noise as a stochastic series.

Seismic noise is either incoherent or coherent regarding the correlation with subsequent samples (Yilmaz, 2001). The first type, incoherent noise, typically follows a Gaussian distribution. In onshore seismic acquisitions, sources of random noise are the wind, instrumental noise and surface activities. The second type, coherent noise, includes noise that is correlated either in space or time. The noise that is coherent in space originates typically in the seismic source. Ground roll, mainly composed of Rayleigh waves, is the primary type of coherent linear noise in onshore data. To filter ground roll, one can use f-k-filters or bandpass filters (Yilmaz, 2001). Noise can also be coherent in time. Temporally coherent noise corrupts seismic traces with an erratic or impulsive component, usually not generated by the seismic source. Power lines, misfired shots, and poor coupling of geophones are frequent sources of high-amplitude noise. Expressions of high-amplitude noise in seismic data are spikes, bursts and noisy traces (Anderson and McMechan, 1989). Figure 1.3 shows two field gathers pervaded by impulsive noise.

In seismic data processing, many algorithms rely on the least-squares inversion to improve the subsurface image. In other words, the best image results from minimizing the squares of the residuals between the observed seismic traces and the corresponding model. However, this approach considers a Gaussian distribution of errors. As seismic data usually presents erratic components or non-Gaussian noise, the least-squares approach is not optimal. This thesis considers reconstruction algorithms that are robust to the presence of erratic or high-amplitude errors. The robust approach replaces the l_2 norm with norms that down weight the erratic errors. Conventional robust norms are the l_1 norm, the Huber norm, the bi-weight function, and the Geman-McClure function. This thesis considers a generalized function that includes a flexible parameter to adjust the solution robustness. Setting the parameter to specific values recovers the traditional robust norms.

1.3 Main motivation of this work

The main motivation of this thesis is to utilize the redundancy of multidimensional datasets to reconstruct regularly and irregularly missing data. To accomplish this, I establish constrained inverse problems and consider the algorithm that fits the

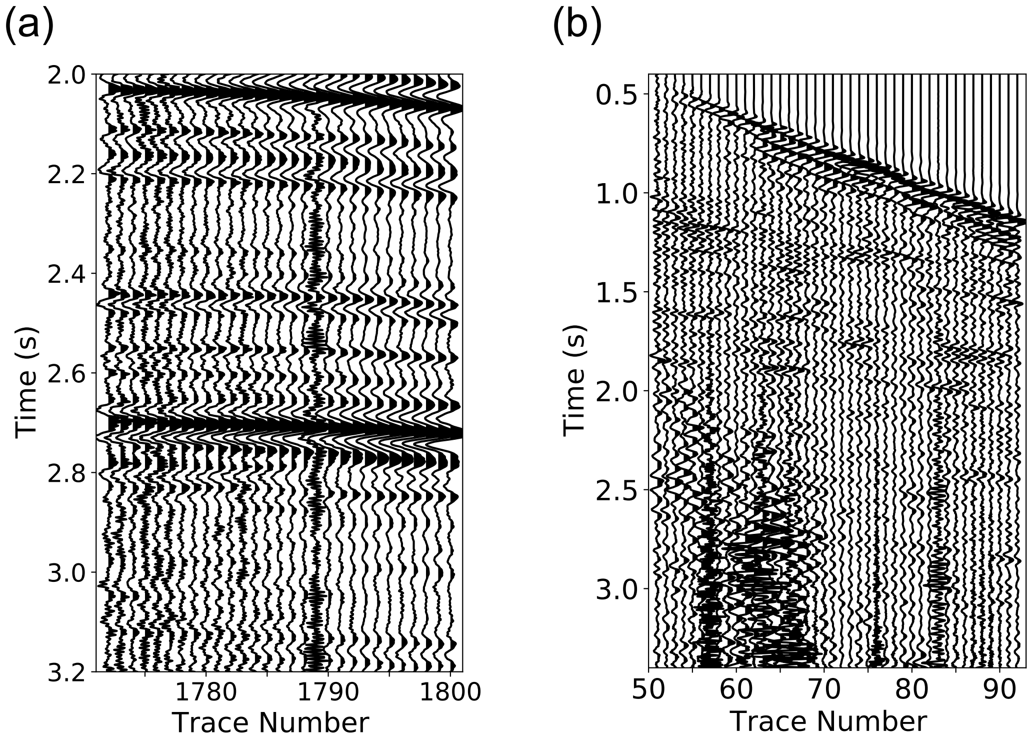

Figure 1.3: Prestack seismic data are frequently pervaded by high-amplitude noise. (a) Teapot dome, RMOTC and the U.S. Department of Energy, SEG Open Data. (b) Stratton 3D , Bureau of Economic Geology, University of Texas, Austin, SEG Open data.

needs. The objective is twofold. First, to design methods that consider the presence of unknown noise distributions by including robust norms. Second, to allow for the reconstruction of irregularly sampled data. Both goals aim to reduce the computational efforts of standard algorithms. The thesis considers the particular application of reconstruction and noise attenuation of seismic wavefields.

1.4 Contributions

The contributions of this thesis can be summarized as follows:

- Introduce Interpolated Multichannel Singular Spectrum Analysis (IMSSA) to reconstruct irregularly sampled data from their real coordinates. I propose reconstruction through a mapping operator that honours the observed locations. I adopt the projected gradient method to approximate the low-rank constraint.

- Evaluate the effect of robust norms in the seismic wavefield. Extend the Parallel Matrix Factorization (PMF) method to account for the presence of non-Gaussian noise.
- Investigate computational tools to reduce the cost of the algorithms as randomized sampling. I also explore reconstruction via cross-spread gathers to reduce the computational burden of windowing the data.

1.5 Outline of the thesis

The thesis is organized as follows

- Chapter 2 formally introduces seismic reconstruction as a rank-reduction recovery problem. The chapter covers an overview of the notation and definitions. Besides, it introduces algorithms for matrix and tensor reconstruction.
- Chapter 3 presents the Interpolated Multichannel Singular Spectrum Analysis algorithm (IMSSA). The proposed method reconstructs seismic data randomly decimated from irregular locations. I propose an inversion scheme that minimizes the misfit between predicted and observed data in the frequency-space domain. The predicted data are mapped to the observed coordinates via an interpolator operator. The inverse problem is solved via the projected gradient method. The projection is imposed via the SSA filter. I show synthetic and field data to test the performance of the algorithms. Field data are processed in cross-spread gathers to favour the natural windowing of the data. This results in reduced processing efforts.
- Chapter 4 explores robust norms in seismic reconstruction. The chapter reviews the applications of robust inversion in geophysics. Following, it introduces M-estimators, the robust analogue of the maximum likelihood method. Next, I propose a robust approach to the reconstruction method. The algorithm minimizes the discrepancies between the predicted data and the observation but replaces the l_2 approach with a generalized robust norm. The

generalized norm regulates its shape via a continuous parameter. The solution is obtained via the projected gradient method. I evaluate synthetic and field tests. Synthetic tests allow examining the solution as a function of the free parameter that defines the convexity of the norm. I concluded that datasets with a high percentage of non-Gaussian noise require non-convex norms for proper reconstruction. However, seismic data typically presents lower percentages of high-amplitude errors. Therefore, robust monotonic norms provide acceptable results.

- In chapter 5, I adapt the Parallel Matrix Factorization method for robust reconstruction of seismic data. I present a new scheme that computes a weighted approximation of the data. The weight depends on the robust norm and the amplitude of the error. The method leads to a two-step algorithm. The first step approximates the data via an imputation algorithm, considering the weights. The second step computes the rank-reduced approximation of the unfoldings of the tensor. I consider synthetic and field examples. Field examples consider a full seismic dataset. I implement 5D slide windowing to approximate the result. Besides, rank-reduction is obtained via randomized QR decomposition, which relaxes the selection of the rank.
- Chapter 6 contains the conclusions of this thesis given by the contributions and limitations of my research.

CHAPTER 2

Rank Reduction

He who refuses to do arithmetic is doomed to talk nonsense.
John McCarthy

The matrix and tensor completion problems intend to estimate an array from its partially observed entries. The problem is underdetermined and, therefore, it does not have a unique solution. However, one can find a solution by imposing additional constraints. In particular, one can assume that the dataset is of low-rank. Thus, a small percentage of correlated samples enable the recovery of the complete array. The completion problem has seen much development in theory and applications in computer vision, machine learning, data mining and collaborative filtering, bioinformatics, control, and signal processing (Zhou et al., 2014; Candès and Plan, 2010).

In seismic data processing, continuous wavefields are adequately sampled in time, but not in their spatial coordinates. Spatial undersampling is a result of economic and physical constraints. In some cases, one can model the seismic wavefield as a limited number of complex exponentials (Stanton and Sacchi, 2013), and show that this model leads to, for instance, Hankel matrices that are low rank. The latter is the essence of reduced-rank methods such as MSSA filtering and Cadzow filtering. This chapter intends to present mathematical tools and simple models that will allow us to develop methods for reduced-rank filtering operating on Hankel forms and tensors.

2.1 Representing seismic data with low-rank arrays

To gain intuition on why one can adopt rank-reduction for denoising and data reconstruction, we first study a simple model. Consider a seismic signal in a small window. The signal can be modelled as an event with constant dip, where the dip or ray-parameter is denoted by p. In the t-x (time-space) domain, such a signal can be represented by a delayed function

$$s(t,x) = w(t-px)\,,$$

where t and x represent time and space, and $w(t)$ is a band-limited wavelet. We can use the Fourier transform to map time to frequency $s(t,x) \to S(\omega,x)$, where $S(\omega,x)$ is often referred as the signal in the f-x (frequency-space) domain. Applying Fourier shift theorem it is easy to show

$$S(\omega,x) = W(\omega)\exp^{-i\omega px}\,,$$

where $\omega = 2\pi f$ is the temporal frequency in radians per seconds and f is frequency in Hz.

We analyze one frequency ω_0, and assume regularly sampled data. The following equation reflects the data for one channel

$$S_n = W\exp^{-i\alpha n}\,, \tag{2.1}$$

where $x_n = n\Delta x$ and $\alpha = \omega_0 p\Delta x$. The spatial distance between channels is Δx. As it is clear that the model is valid for any frequency, I drop the dependency on temporal frequencies to avoid notational clutter. Let us now consider the waveform in the adjacent channel

$$S_{n+1} = W\exp^{-i\alpha(n+1)}\;. \tag{2.2}$$

One can infer a recursive expression between channels given by combining equations 2.1 and 2.2

$$S_{n+1} = S_n P\,, \tag{2.3}$$

where $P = \exp^{-i\alpha}$. That is, the signal is predictable in space with a recursion of order one.

Let us now consider N_x consecutive equally spaced records of the waveform. The trajectory matrix is given by

$$\mathbf{H} = \begin{pmatrix} S_1 & S_2 & \cdots & S_K \\ S_2 & S_3 & \cdots & S_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{L_x} & S_{L_x+1} & \cdots & S_{N_x} \end{pmatrix},$$

where selecting $L_x = \lfloor N_x/2 \rfloor + 1$ and $K = N_x - L_x + 1$ results in a square Hankel matrix when N_x is odd (Trickett, 2008). The symbol $\lfloor \rfloor$ indicates the integer part of the argument. Considering the predictability of the signal in space given by equation 2.3, it is easy to show the following

$$\mathbf{H} = \begin{pmatrix} S_1 & S_1P & \cdots & S_1P^{K-1} \\ S_2 & S_2P & \cdots & S_2P^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{L_x} & S_{L_x}P & \cdots & S_{L_x}P^{K-1}. \end{pmatrix} \tag{2.4}$$

We conclude that the rank of the trajectory matrix **H** with one constant dip equals one. Figure 2.1 shows a 2D seismic section with one dipping linear event and the singular values of the Hankel matrix for frequency index 20. In the presence of uncorrelated noise or missing samples, the rank of the trajectory matrix increases.

Appendix A presents a generalization of the concept to multiple events and higher dimensions. The following sections introduce the matrix and tensor algebra required for the reconstruction problem.

2.2 Matrix completion

The completion problem consists of finding a matrix that best approximates a matrix of incomplete and corrupted data. The underdetermined problem has no unique solution. However, if the unknown matrix is of approximately low rank, it is possible to find a solution by exploiting reduced-rank filtering techniques. This section has the following objectives. First, introduce fundamental concepts as the rank of a matrix and matrix decomposition. Following, define the matrix completion problem. Finally, review popular algorithms to solve the problem.

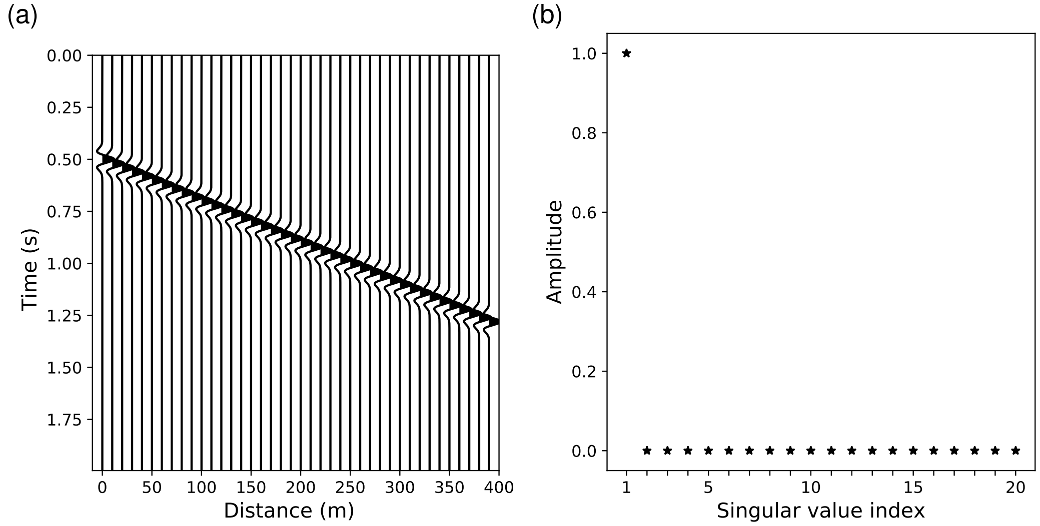

Figure 2.1: Seismic data is low rank. (a) 2D seismic section with one dipping linear event. (b) The rank of the Hankel matrix of a frequency slice equals one.

2.2.1 Preliminaries

Let us consider a square matrix $\mathbf{A}$ of dimensions $M \times M$, and a general vector $\mathbf{x}$ of length M. In general, the multiplication $\mathbf{Ax}$ results in a new vector pointing in a different direction from $\mathbf{x}$. However, there is a subset of exceptional vectors that multiplied by $\mathbf{A}$ do not change direction. That is,

$$\mathbf{Ax} = \lambda \mathbf{x} .$$

Vectors with such properties are called eigenvectors of matrix $\mathbf{A}$. The coefficient λ is called an eigenvalue and describes the stretching or shrinking of the vector $\mathbf{x}$ after the multiplication.

We now consider $\mathbf{A}$, a rectangular matrix of dimensions $M \times N$. The singular value decomposition (SVD) of $\mathbf{A}$ is defined as

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= \Sigma_{i=1}^{n} u_i \sigma_i v_i^T , \end{aligned} \tag{2.5}$$

where $\mathbf{U}$ is an $M \times M$ matrix, $\mathbf{V}$ is an $N \times N$ matrix, and σ is a diagonal matrix of dimension $M \times N$. The columns of $\mathbf{U}$ are called the left singular vectors of $\mathbf{A}$, the columns of $\mathbf{V}$ are the right singular vectors of $\mathbf{A}$, and the elements of σ are called the singular values of $\mathbf{A}$. Matrices $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns, $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

The SVD of $\mathbf{A}$ is strongly connected to the eigenvalue decomposition of the symmetric semidefinite matrices $\mathbf{A}\mathbf{A}^T$ as $\mathbf{A}\mathbf{A}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\mathbf{U}^T$. The columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^T$, the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$, and the diagonal elements of σ are the square roots of the non-zero eigenvalues of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$.

The rank of a matrix describes the number of linearly independent rows or columns of a 2D array. SVD is a rank-revealing decomposition in the sense that a matrix of rank r has only the first r elements of the main diagonal in $\boldsymbol{\Sigma}$ different from zero. From the SVD, a rank-one matrix $\mathbf{A}_1$ equals the outer product of two vectors, $\mathbf{A}_1 = \mathbf{u} \otimes \mathbf{v}$, and a rank r matrix equals the summation of rank-one matrices as $\mathbf{A}_r = \sigma_1\mathbf{u}_1 \otimes \mathbf{v}_1 + \cdots + \sigma_r\mathbf{u}_r \otimes \mathbf{v}_r$.

The SVD also allows for the definition of a pseudoinverse of matrix $\mathbf{A}$ as

$$\mathbf{A}^\dagger = \sum_{i=1}^{r} \mathbf{v}_i \sigma_i^{-1} \mathbf{u}_i^T \, . \tag{2.6}$$

The solution to a linear problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ can be sensitive to perturbations in $\mathbf{A}$. One can measure the sensitivity of the problem via the condition number

$$\texttt{cond}\,(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^\dagger\|_2 = \sigma_1/\sigma_r \, , \tag{2.7}$$

where σ_1 is the largest singular value, σ_r is the smallest one, and $\|\cdot\|_2$ represents the p=2 matrix norm, with $\|\mathbf{A}\|_2 = \sigma_{max}(\mathbf{A})$. A large condition number implies that $\mathbf{A}$ is nearly rank-deficient, that is, the system of linear equations is ill-possed and requires regularization to find a solution.

2.2.2 Low-Rank Matrix Approximation

Given a matrix $\mathbf{Y}$ of size $M \times N$ and rank R, one can estimate the best approximation of $\mathbf{Y}$ via a matrix $\mathbf{X}$ of rank r, with $r < R$. The problem is

$$\min \|\mathbf{Y} - \mathbf{X}\|_F^2 \text{ subject to } \operatorname{rank}(\mathbf{X}) = r\,, \tag{2.8}$$

where $\|\mathbf{Y} - \mathbf{X}\|_F^2 = \sum_{m,n} |Y_{mn} - X_{mn}|^2$ is the square of the standard Frobenius norm.

The low-rank approximation problem has a solution given by the truncated Singular Value Decomposition of the observed matrix (Eckart and Young, 1936)

$$\mathbf{X} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T\,. \tag{2.9}$$

Matrix $\mathbf{X}$ requires only $r(M+N)$ coefficients in contrast to the original $M \times N$ coefficients used for the rank R matrix. This characteristic is an indication that it might be possible to recover a low-rank matrix from relatively few measurements. Figure 2.2 presents the decomposition schematically.

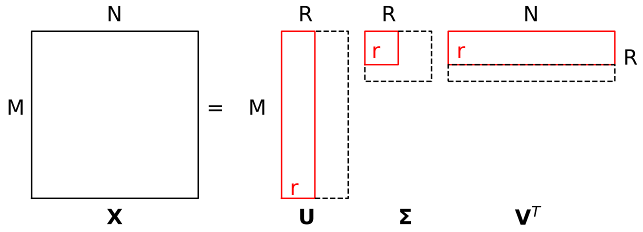

Figure 2.2: Matrix rank reduction via truncated SVD. Matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ can be approximated by $r(M+N)$ coefficients.

2.2.3 Low-rank Matrix Completion

Let us now consider the recovery of a low-rank matrix $\mathbf{X}$ of size $M \times M$ from a subset of its entries. The problem is

$$\min \operatorname{rank}(\mathbf{X})\,, \text{s.t.}\, \Omega(\mathbf{X}) = \mathbf{Y}\,, \tag{2.10}$$

where $\mathbf{X}$ is the complete, low-rank matrix of interest, Ω is a sampling operator returning a subset of the entries, and $\mathbf{Y}$ are the observations. However, the observations are also contaminated by noise, $\mathbf{Y} = \Omega(\mathbf{X}) + \mathbf{N}$. Then, the problem is to find the best low-rank approximation to $\mathbf{Y}$, or

$$\min \operatorname{rank}(\mathbf{X})\,, \text{s.t.}\ \|\Omega(\mathbf{X}) - \mathbf{Y}\|_F^2 \leq \epsilon\,. \tag{2.11}$$

This objective function describes the low-rank matrix completion problem.

The nuclear norm

Minimizing the rank of a matrix corresponds to minimizing the l_0 norm of its singular values (Donoho, 2006). That is,

$$\min \operatorname{rank}(\mathbf{X}) = \min \|\sigma_i\|_0 = \min \sum_{i=1}^{R} |\sigma_i|^0\,,$$

where σ_i are the singular values of $\mathbf{X}$, and the l_0 norm minimizes the number of non-zero entries of a given vector. This definition exposes the non-convexity of the matrix completion problem (Equation 2.11). The tightest relaxation of the l_0 norm is the l_1 norm. In the matrix completion problem, this convex relaxation results in the nuclear norm, or trace norm $\|\mathbf{X}\|_* = \operatorname{trace}\left(\sqrt{\mathbf{X}^*\mathbf{X}}\right) = \sum_{i=1}^{R} \sigma_i$. The problem is now tractable.

Candès and Recht (2009) proved that given an $M \times N$ matrix, the number of available entries m should be greater or equal to $CL^{1.2}R\log L$ to recover the desired matrix. In the expression, C is a positive numerical constant, and $L = \max(M, N)$, and R is the rank. Besides, there should be no empty rows or columns. To prevent this, the singular values of the matrix need to be sufficiently spread. The minimum number of entries needed to recover a matrix exactly was later reviewed by several authors (Candès and Plan, 2010; Keshavan et al., 2010; Recht, 2011), concluding that most low-rank matrices can be recovered even if they have a very low cardinality.

By using Lagrange formulation, one can express the reconstruction problem as an unconstrained convex optimization problem

$$\arg\min_{\mathbf{X}} \|\Omega(\mathbf{X}) - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{X}\|_*\,. \tag{2.12}$$

The equation can be solved using algorithms from the proximal family, as the singular value thresholding algorithm (SVT) (Cai et al., 2010), the fast iterative shrinkage thresholding algorithm (FISTA) (Beck and Teboulle, 2009), and the alternating direction method of multipliers algorithm (ADMM) (Parikh et al., 2014).

Fazel et al. (2013) analyze the solution to Equation 2.12, where the matrix requires to be of low rank and with Hankel or Toeplitz structure. The solution is obtained via ADMM and gradient projection methods.

In general, non-convex regularization methods present better performance than nuclear norm minimization (Nie et al., 2012; Lai et al., 2013; Wen et al., 2018). An alternative approach is to minimize the l_q norm of the singular values with $0 < q < 1$ (Marjanovic and Solo, 2012). This measure of the singular values is called the Schatten-q norm. To solve this non-convex problem, one can use a proximal descent scheme. The matrix completion problem via nuclear norm or Schatten-p norm is generally stable but requires the repeated computation of the SVD to obtain the singular values of the matrix. The process can be computationally intensive for large datasets.

Matrix factorization

When the matrix of interest is large, as for seismic data, storage and computation demand becomes an issue. A common approach is to consider $\mathbf{A} \sim \mathbf{L}\mathbf{R}^T$, where the $M \times r$ and $N \times r$ matrices are also low-rank, restricting the storage demand in all the iterations. The approach is validated by the following property of the nuclear norm

$$\|\mathbf{X}\|_* = \underset{\mathbf{L},\mathbf{R}}{\arg\min} \frac{1}{2}\left(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2\right) \text{ , s.t.}\, \mathbf{X} = \mathbf{L}\mathbf{R}^T \, . \tag{2.13}$$

Then, one can express Equation 2.12 by

$$\underset{\mathbf{L},\mathbf{R}}{\arg\min} \|\Omega\mathbf{L}\mathbf{R}^T - \mathbf{Y}\|_2^2 + \frac{\lambda_1}{2}\|\mathbf{L}\|_F^2 + \frac{\lambda_2}{2}\|\mathbf{R}\|_F^2 \, . \tag{2.14}$$

The misfit term in the cost function is a product of two unknowns. Then, the formulation is non-convex. The problem can be solved via alternating minimization (Recht et al., 2010). The method iteratively fixes each unknown while solving the other one, resulting in two convex subproblems. The iterations are

$$\begin{aligned} \mathbf{R}_{k+1} &= \underset{\mathbf{R}}{\arg\min} \|\Omega \mathbf{L}_k \mathbf{R}^T - \mathbf{Y}\| \\ \mathbf{L}_{k+1} &= \underset{\mathbf{L}}{\arg\min} \|\Omega \mathbf{L} \mathbf{R}_{k+1}^T - \mathbf{Y}\| \,. \end{aligned} \tag{2.15}$$

The method outperforms nuclear norm minimization as the SVD is not calculated. On the other hand, the underlying rank needs to be predefined, which can be challenging in particular applications.

The Projected Gradient Descent

In general, finding a solution to equation 2.11 is NP-hard (Candès and Recht, 2009). A possible relaxation of the problem assumes a priori knowledge of the rank

$$\underset{\mathbf{X}}{\arg\min} \|\Omega(\mathbf{X}) - \mathbf{Y}\|_F^2 \;\; \text{s.t.}\, \text{rank}(\mathbf{X}) \leq r \,. \tag{2.16}$$

The Projected Gradient Descent, which results in the Singular Value Projection algorithm (Jain et al., 2010) and the Iterative Hard Thresholding algorithm (Gavish and Donoho, 2014), solve the optimization problem. Both algorithms are very similar, with a two-step iterative scheme. Each iteration is

$$\begin{aligned} \tilde{\mathbf{X}}_{k+1} &= \mathbf{X}_k - \gamma_k \Omega^* \left(\Omega(\mathbf{X}_k) - \mathbf{Y}\right) \\ \mathbf{X}_{k+1} &= \text{ProjectRank}_r \left(\tilde{\mathbf{X}}_{k+1}\right) \,. \end{aligned} \tag{2.17}$$

The operator called ProjectRank computes components of the SVD up to rank r. If $r \ll \min(M, N)$, this algorithm is faster than computing the full SVD.

This thesis presents an implementation of the Projected Gradient Descent algorithm for seismic reconstruction. The algorithm considers the ProjectedRank operator as a rank minimization over a Hankel matrix containing the values of the sampled wavefield. Chapter 3 presents the mathematical considerations together with synthetic and field results.

2.3 Tensor Completion

Tensors were introduced at the end of the 19th century with the development of the differential calculus and homogeneous polynomials (Comon, 2014; Cichocki et al., 2015). Tensor decomposition techniques have seen major research development since their introduction. In particular, Tucker (1964) and Carroll and Chang (1970) established the Tucker decomposition and Canonical Polyadic Decomposition (CPD) in psychometrics, while Harshman (1970) introduced the concept into linguistics. Many applications can also be found in fields such as chemometrics (Bro, 1997; Smilde et al., 2005), data mining (Acar et al., 2010; Mørup, 2011), graph analysis (Kolda, 2006), and neuroscience (Mørup et al., 2006; Lee et al., 2007). In the field of signal processing, tensors have been used in audio processing, image processing, video processing, machine learning, and biomedical applications, among other fields (Cichocki et al., 2015). In geophysics, tensor algebra was adopted for seismic data reconstruction (Kreimer and Sacchi, 2012; Kreimer et al., 2013; Ely et al., 2015; Gao et al., 2015). The problem is posed as optimizing a cost function that minimizes a misfit in conjunction with a low-rank constraint.

The following sections cover the basics of tensorial algebra and the low-rank tensor recovery problem.

2.3.1 Preliminaries

A tensor $\mathcal{X}$ is a multidimensional array. The dimension of the array is named the order of the tensor or the number of ways or modes. A first-order tensor is a vector $\mathbf{x}$, a second-order tensor is a matrix $\mathbf{X}$, and a third-order tensor or higher is a high-order tensor $\mathcal{X}$. Figure 2.3 shows a diagram of a third-order tensor.

The seismic wavefield has four spatial dimensions. Data depends on five dimensions: time, source and receiver x and y coordinates. We stress that we always prefer to work in the frequency-space domain. In other words, the seismic volume in the frequency-space domain for one single frequency corresponds to a 4D volume that can be embedded in a four-order tensor. Then, one can focus on fourth-order tensors. The elements of a fourth-order tensor are $x_{i,j,k,l}$. One can also gather subarrays by fixing, for example, all the elements of a mode $x_{i,j,:,l}$. A slice is a two-dimensional section of a tensor obtained by fixing all the indices except two, $x_{:,:,k,l}$. Similarly, a

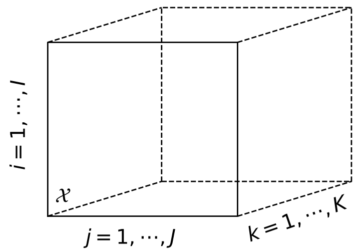

Figure 2.3: Schematic representation of a third-order tensor.

fiber is a one-dimensional subarray, the analogue of a column or row to a matrix, obtained by fixing all the indices except for one $x_{:,j,k,l}$.

One can also reorder the elements of an Nth-order tensor into a matrix. The process is called *unfolding*, *flattening*, or *matricization* and we indicate it as $\mathbf{X}_{(n)} = \text{fold}_{(n)}(\mathcal{X})$. There are different methods to unfold a matrix. This thesis considers the mode-n matricization (Kolda and Bader, 2009). The unfolding rearranges the mode-n fibers to be the columns of the matrix. Figure 2.4 shows the process schematically. In general, the specific permutations of columns are not critical, but they should be consistent throughout the calculations. Similarly, one could vectorize a tensor.

The inner product of two tensors of the same size $\mathcal{X}$, $\mathcal{Y}$ equals the sum of the products of their entries

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} x_{i,j,k,l}\, y_{i,j,k,l}\,. \tag{2.18}$$

Vector spaces with scalar products facilitate the definition of a norm. The tensor Frobenius norm is the square root of the sum of the squares of the elements

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} x_{i,j,k,l}^2}\,. \tag{2.19}$$

With the previous definitions, one concludes that $\langle \mathcal{X}, \mathcal{X} \rangle = \|\mathcal{X}\|_F^2$. The inner product and the Frobenius tensor norm are a high-order generalization of the definitions concerning matrices.

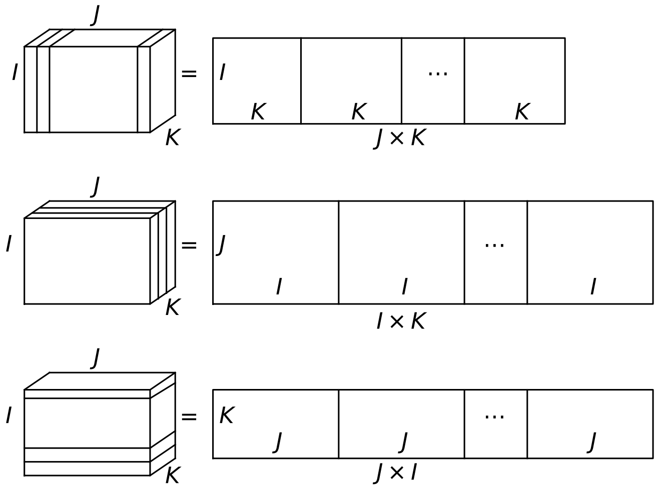

Figure 2.4: A third-order tensor can be unfolded in three different matrices or modes.

Tensors, like matrices and vectors, are equipped with multiple definitions of multiplication. The mode-n product, for example, considers the multiplication of a tensor by a matrix, or vector, in mode n. In the case of a matrix, each mode-n fiber is multiplied by the matrix $\mathbf{U}$. As an example, the mode-3 product of a fourth-order tensor $\mathcal{X} \in \mathbb{R}^{I,J,K,L}$ with a matrix $\mathbf{U} \in \mathbb{R}^{M\times K}$ results in the tensor $\mathcal{Z}$ of size $I \times J \times M \times L$. The element i, j, m, l of tensor $\mathcal{Z}$ is

$$z_{i,j,m,l} = (\mathcal{X} \times_3 \mathbf{U})_{i,j,m,l} = \sum_{k=1}^{K} x_{i,j,k,l}\, u_{m,k}\,. \tag{2.20}$$

Finally, if an Nth-order tensor can be decomposed as the outer product of N vectors

$$\mathcal{A} = \mathbf{b}^{(1)} \otimes \cdots \otimes \mathbf{b}^{(N)}\,, \tag{2.21}$$

the tensor is of rank one. In the definition, $\mathbf{b}^n$ represents a vector $\in \mathbb{R}^{I_n}$, and $\otimes$ is the standard outer product of vectors. This definition is similar to the matrix rank. However, there are substantial differences. The next section delves into this concept and considers two main models or tensor decompositions.

2.3.2 Tensor Decompositions and rank

Tensor decompositions give a concise representation of the underlying structure of the data, revealing the dimension of the subspace. This section describes two tensorial decompositions and the tensor rank associated with them.

The Canonical Polyadic Decomposition

An Nth-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \cdots I_N}$ is expressed in its Polyadic Decomposition (PD) when described as a linear combination of rank-1 tensors (Kolda, 2006; Comon, 2014; Cichocki et al., 2015; de Morais Goulart, 2016; Sidiropoulos et al., 2017). Mathematically, the Polyadic Decomposition is

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \mathbf{b}_r^{(1)} \otimes \ldots \otimes \mathbf{b}_r^{(N)} , \tag{2.22}$$

where λ_r is a normalizing scalar, and $\mathbf{b}^n$ represents a vector $\in \mathbb{R}^{I_n}$ (Figure 2.5). When describing tensor $\mathcal{X}$ with the minimum possible R, one presents its Canonical Polyadic Decomposition (CPD). The rank of the tensor $\mathcal{X}$ equals the value of R for which the CPD holds exactly.

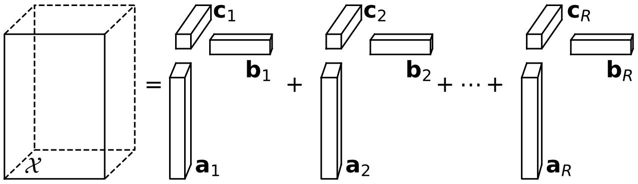

Figure 2.5: Canonical Polyadic Decomposition of a third-order tensor of rank R.

The definition of tensor rank is a counterpart to the definition of matrix rank. However, there are substantial differences to address. First, only one combination of rank-one tensors sums to $\mathcal{X}$, except for the elementary indeterminacies caused by scaling and permutation (Kolda and Bader, 2009). This uniqueness of the CPD is not valid for matrices. Besides, the tensor rank depends on the underlying field and can exceed the smallest dimension of the array.

Given the nature of the definition, defining the rank of a given tensor is an NP-hard problem (Fortnow, 2009). As the CPD is unique for tensors, one could eventually fit multiple CP decompositions with alternative values of R until one fits the data as expected. However, when the data are noisy, the fitting approach cannot determine the rank in any case. Then, the rank should be fixed with a priori knowledge of the data.

There are complementary definitions of the tensor rank. The border rank is the minimum number of rank-one tensors that approximate the tensor of interest within a given tolerance (Comon, 2014). The maximum rank is the largest attainable rank. Finally, the typical rank is any rank that occurs with a probability greater than zero. The typical and maximal ranks are identical for matrices but not for tensors. There may be more than one typical rank in $\mathbb{R}$, while there is always one typical rank in $\mathbb{C}$.

The best rank-k approximation of a given matrix is the solution of the truncated SVD (Eckart and Young, 1936). Contrarily, for higher-order tensors, summing the first k factors of the CP decomposition of a rank R tensor does not yield the best rank-k approximation of the tensor. In other words, the components of a CP decomposition should be calculated simultaneously instead of sequentially. Finally, the best rank-k approximation of a tensor might not exist. These arrays are called degenerate tensors. In such cases, it is useful to consider the border rank.

The Tucker Decomposition

The Tucker Decomposition describes an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \dots I_N}$ as the composition of a core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \dots R_N}$ multiplied by factor matrices in each mode (Kolda, 2006; Comon, 2014; Cichocki et al., 2015; de Morais Goulart, 2016; Sidiropoulos et al., 2017). The matrices are defined as $B^{(n)} = [\mathbf{b}_1^{(n)}, \dots, \mathbf{b}_{R_n}^{(n)}] \in \mathbb{R}^{I_n \times R_n}$, $n = 1, 2, \dots, N$. The decomposition is

$$\mathcal{X} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \dots \sum_{r_N=1}^{R_N} \mathcal{G}_{r_1 \dots r_N} \left(\mathbf{b}_{r_1}^{(1)} \circ \dots \circ \mathbf{b}_{r_N}^{(N)} \right) . \tag{2.23}$$

The core tensor $\mathcal{G}$ is a compressed version of the initial tensor $\mathcal{X}$, especially when $R_1, R_2, \dots, R_N$ are smaller than $I_1, I_2, \dots, I_n$. Figure 2.6 illustrates the decomposition.

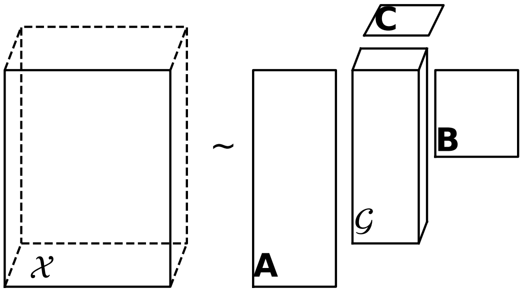

Figure 2.6: Tucker decomposition of a third-order tensor.

In contrast to the CPD, the Tucker decomposition is not unique. Then, it is common practice to choose transformations that restrict the problem. In particular, one can simplify the core, aiming for the highest number of zero elements, which reduces interactions among elements. The extreme case in which $\mathcal{G}$ is a fully diagonal matrix is coincident with the polyadic decomposition.

The High-Order SVD (HOSVD) is the Tucker decomposition in which the factor matrices are orthogonal (De Lathauwer et al., 2000; Kreimer and Sacchi, 2012). In the HOSVD decomposition, the matrices $B^{(1)}, B^{(2)}, \ldots, B^{(N)}$ contain the left singular vectors of the matricization of the tensor $\mathcal{X}$. Besides, the core tensor plays a role similar to that of the matrix of singular values in the SVD. Then, one can reduce the rank of the approximated tensor by truncating the core tensor $\mathcal{G}$.

Previously, we related the rank of the tensor to the CPD. In a similar vein, we define the n-Rank or multilinear rank of a tensor referencing the Tucker decomposition. The n-Rank is the vector whose components are the column rank of the unfoldings of the tensor $\mathcal{X}$. The Tucker decomposition of a tensor of n-Rank equal to $(R_1, R_2, \ldots, R_N)$ is exact. On the contrary, if $(R_1, R_2, \ldots, R_N)$ is less than the n-Rank, the Tucker decomposition provides only an approximation. In contrast to the SVD, the Tucker decomposition is not the best approximation from a least-squares perspective. Nevertheless, every tensor has a best at most rank-k Tucker approximation, even if it is near optimally computed via the High Order SVD.

2.3.3 Low-Rank Tensor Completion

Tensor completion is the problem of recovering a tensor with partially observed elements. The problem deals with obtaining tensor models to enable reconstruction. As a high-dimension generalization of matrices and vectors, tensors exploit redundancy in all the modes. Such redundancy allows us to characterize the complex structures of higher-order data. However, as in the matrix case, the low-rank constraint is necessary to restrict the degrees of freedom of the missing entries.

Mathematically, one can describe the low-rank tensor completion problem as

$$\min \operatorname{rank}(\mathcal{X}) \ , \ \text{s.t}\, \Omega(\mathcal{X}) = \mathcal{Y} \,, \tag{2.24}$$

where $\mathcal{X}$ is the underlying tensor, $\mathcal{Y}$ is the observed one, Ω is the sampling operator that keeps the entries of $\mathcal{X}$ that were observed in $\mathcal{Y}$ and zeros out the ones that were not observed.

The rank function results in a non-convex objective function for the completion problem. Besides, the multiplicity of definitions of rank makes this problem more complicated than its matrix completion counterpart. For example, the tensor rank is an NP-hard problem, and tensor degeneracy is widespread. Therefore, the n-rank is the most frequently used definition of the low-rankness of a tensor.

The Nuclear Norm

As described in the low-rank matrix completion section, the Nuclear Norm is the tightest convex relaxation of the rank of a matrix. Generalizing from this concept, Liu et al. (2012) define the nuclear norm for tensors

$$\|\mathcal{X}\|_* = \sum_{n=1}^{N} \|\mathbf{X}_{(n)}\|_* \,, \tag{2.25}$$

with $\mathbf{X}_{(n)}$ representing the Nth-unfolding of the tensor. The nuclear norm of a tensor is the combination of the trace norms of all matrices unfolded in each mode. Now, the model for tensor decomposition is

$$\min \sum_{n=1}^{N} \|\mathbf{X}_{(n)}\|_* \,, \ \text{s.t.}\, \Omega(\mathcal{X}) = \mathcal{Y} \,. \tag{2.26}$$

Although the matrices in the problem cannot be optimized independently, the convex relaxation allows for a solution of the tensor reconstruction problem without predefining the tensor rank. To solve the problem, Liu et al. (2012) propose optimizing via the block coordinate descent. Gandy et al. (2011) introduce the ADMM, the standard approach in convex tensor completion.

Tensor Factorization

Based on ideas from the weighted least squares method, Tomasi and Bro (2005) present two methods to complete tensors using the PD model. The first approach considers an alternately projection optimization method to estimate the parameters, followed by an imputation scheme

$$\mathcal{X} = \Omega\left(\mathcal{Y}\right) + (1-\Omega)\hat{\mathcal{X}}\,. \tag{2.27}$$

The operator Ω is the sampling operator. Then, $1-\Omega$ is the complement set of the observed samples. The algorithm is derived as an Expectation-Maximization iterative method. The algorithm is easy to implement, but it might converge to a local minimum with increasing missing values.

The second algorithm considers building the model solely with the observed data by minimizing the misfit component-wise

$$J = \sum_{i=1}^{N} \|\mathcal{X}_i - \mathcal{Y}_i\|_F^2\,. \tag{2.28}$$

The problem is solved via a Gauss-Newton algorithm and proves to be computationally efficient for high missing ratios.

Different authors consider similar algorithms in the context of the Tucker Decomposition. This model proves to be more effective than CP as the core tensor captures the interactions between components. Andersson and Bro (1998) and Walczak and Massart (2001) combine the High-Order Orthogonal Iteration algorithm with an imputation scheme. Karatzoglou et al. (2010) consider the optimization approach in combination with the HOSVD. The problem is solved via stochastic gradient descent.

Of particular interest in this thesis is the Parallel Matrix Factorization algorithm (Xu et al., 2015). The method applies low-rank matrix factorization to each unfolding of the tensor and updates the matrix factors alternately. The approach is non-convex, and the size of the matrix factors must be specified in the algorithm. However, the method is faster than the HOSVD (Gao et al., 2015). To recover the tensor $\mathcal{Z}$, PMF solves an optimization problem

$$\min_{\mathcal{Z},\mathcal{X},\mathcal{Y}} \sum_{n=1}^{N} \frac{\alpha_n}{2}\|\mathbf{X}_n\mathbf{Y}_n - \mathbf{Z}_{(n)}\|_F^2\,, \text{ s.t}\,\Omega(\mathcal{Z}) = \mathcal{B}\,, \tag{2.29}$$

where α_n are the weights and satisfy that their sum equals 1. The constraint enforces consistency with observations $\mathcal{B}$. One applies the low-rank approximation to each unfolding by finding matrices $\mathbf{X}_n \in \mathbb{R}^{I_n \times r_n}$ and $\mathbf{Y}_n \in \mathbb{R}^r_n \times \prod_{j\neq nI_j}$, such that $\mathbf{Z}_{(n)} \sim \mathbf{X}_n\mathbf{Y}_n$ for $n = 1, 2, \ldots, N$, where r_n is the estimated rank for the given unfolding. The problem is solved by alternately updating $\mathbf{X}$, $\mathbf{Y}$, and $\mathcal{Z}$. The method avoids using the SVD in its rank reduction step. Chapter 5 of this thesis presents a robust implementation of PMF for seismic reconstruction in the presence of erratic noise.

CHAPTER 3

Interpolated Multichannel Singular Spectrum Analysis (I-MSSA)

So many secrets in the universe to be discovered
Ample samples, yet the Signal remains to be recovered
Farokh Marvasti

The goal of seismic data reconstruction is to simultaneously denoise and regularize field data (Duijndam et al., 1999; Trad, 2009). This work focuses on the MSSA reconstruction method, also called the Cadzow reconstruction method (Trickett, 2008; Sacchi, 2009; Oropeza and Sacchi, 2011).

MSSA was conceived to denoise and reconstruct data deployed on a regular grid. In other words, missing observations are assumed to be irregularly distributed on a regular grid (Oropeza and Sacchi, 2011). In general, seismic data are irregularly sampled in the spatial domain. The standard approach is to assign seismic traces to a regular grid via a process called binning. Binning assigns traces of arbitrary coordinates to the nearest neighbor nodes of the desired regular output grid. The binning process introduces errors in the amplitude and phase of the traces. These errors can be significant for complex data sets with steeply dipping events. I circumvent binning via a new algorithm, the Interpolated MSSA (I-MSSA) method.

The method connects off-the-grid observations to the desired gridded data via a mapping operator. In essence, I solve an inverse problem that minimizes the residuals between observations and synthesized data in the original spatial coordinates. I use the Projected Gradient Descent Method (Cheng and Sacchi, 2016; Bolduc et al., 2017) to find a solution. The algorithm consists of two steps. The first step uses the steepest descent method to estimate the gridded data that honours off-the-grid observations. The second step guarantees convergence to a solution by applying the MSSA filter to the gridded data. The MSSA filter is applied as a projection operator to the desired reconstructed data in the regular grid. In essence, this chapter presents a modified MSSA method to warrant a reconstruction that honours field coordinates.

This chapter follows the subsequent structure. First, I introduce the notation adopted and provide preliminary definitions. I continue with a description of the I-MSSA algorithm. Finally, I test the proposed method via synthetic and field data. I also provide an application where 3D prestack data corresponding to an orthogonal survey is fully reconstructed using cross-spread gathers. I use I-MSSA to restore each subset individually. The output is a complete seismic volume described in a regular CMP grid.

3.1 Method

3.1.1 Preliminaries

The focus of this chapter is on the reconstruction of seismic volumes that depend on two spatial dimensions. However, the method can be generalized to volumes that depend on three or four spatial dimensions. Let us consider a 3D volume of seismic observations denoted by $u(t, r_x, r_y)$, where the variable t indicates time, and r_x and r_y correspond to, for instance, receiver coordinates of a common shot gather. Similarly, r_x and r_y could correspond to inline and crossline coordinates of a cross-spread gather, respectively.

The observed data, $u(t, r_x, r_y)$, can be represented in the frequency-space domain via $U(\omega, r_x, r_y)$ where ω indicates temporal frequency. Without loss of generality, I ignore the variable ω and understand that the proposed process is carried out for all frequencies. I also denote $\mathbf{r}_i = (r_{x_i}, r_{y_i})$ the coordinate of the trace i such that the observed data can be represented as $U(\mathbf{r}_i),\ i = 1, \ldots, N_u$, where N_u is the total number of available traces. I describe the data in the observation grid, $U(\mathbf{r}_i)$, as a linear combination of observations in the desired regular grid

$$U(\mathbf{r}_i) = \sum_{k \in \mathcal{N}_i} W_k \, D(\boldsymbol{\xi}_k) \,, \tag{3.1}$$

where $D(\boldsymbol{\xi}_k)$ represents the desired observations in the regular grid. Likewise, $\mathcal{N}_i$ represents the indexes of the regular grid coordinates surrounding the coordinate $\mathbf{r}_i$, and the coefficients W_k are the associated weights computed as a function of the distance between $\mathbf{r}_i = (r_{x_i}, r_{y_i})$ and $\boldsymbol{\xi}_k = (\xi_{x_k}, \xi_{y_k})$. Last, the regular grid coordinates are

$$\xi_{x_k} = \xi_{x_0} + (k-1)\,\Delta\xi_x \,, \quad k = 1, \ldots, N_x \,, \tag{3.2}$$

$$\xi_{y_k} = \xi_{y_0} + (k-1)\,\Delta\xi_y \,, \quad k = 1, \ldots, N_y \,, \tag{3.3}$$

where x_0, y_0 are the coordinates of the first grid point and $\Delta\xi_x$ and $\Delta\xi_y$ are the x and y grid intervals, respectively. Equation 3.1 represents an operator that maps observations in the desired output grid to the off-the-grid observations, $U(\mathbf{r}_i)$.

A similar expression for the whole data is as follows

$$\mathbf{U} = \mathcal{W}\,\mathbf{D} \,, \tag{3.4}$$

where $\mathbf{U}$ denotes the observed data as a vector of size $N_u \times 1$ containing the observations. Similarly, $\mathbf{D}$ is the desired data on the regular grid, which is an $N_x \times N_y$ matrix of unknown coefficients, and $\mathcal{W}$ is the interpolation operator connecting observations to gridded data. The adjoint of the interpolation operator is the operator $\mathcal{W}^*$. Both $\mathcal{W}$ and $\mathcal{W}^*$ are implicitly implemented as linear operators rather than matrices. Figure 3.1 compares the process of linear interpolation to the assignation of values via nearest neighbor interpolation. In Figure 3.1a, the value in ξ_8 is assigned to the location $\mathbf{r}$, that is, the nearest neighbor grid coordinate. In contrast, in Figure 3.1b, the value $\mathbf{r}$ is calculated via a linear combination of the observation in adjacent coordinates ξ_8, ξ_9, ξ_{13}, and ξ_{14}. Figure 3.1b corresponds to a bilinear interpolator (Press et al., 2007). For the numerical experiments, I adopt a bilinear interpolator and a sinc interpolator tapered with a Kaiser window to limit its spatial support (Fomel, 2001). Appendix B provides pseudo-codes for the bilinear, and the Kaiser window tapered sinc interpolators.

Equation 3.4 cannot be solved directly because the operator $\mathcal{W}$ is non-invertible. I

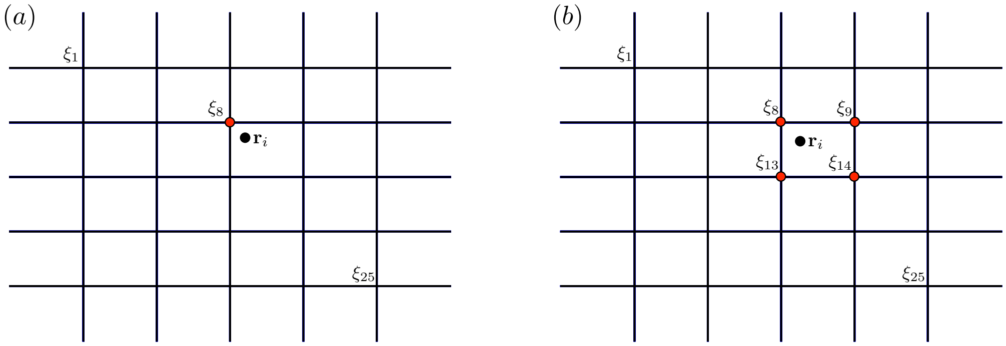

Figure 3.1: (a) Nearest neighbor interpolation. (b) Linear interpolation.

adopt the MSSA filter in the form of a coherence pass projection, which acts as a constraint on $\mathbf{D}$ to solve equation 3.4.

3.1.2 The MSSA filter and MSSA reconstruction

Before introducing the I-MSSA algorithm, I briefly describe the MSSA method following the work of Oropeza and Sacchi (2011). The classical MSSA filter can be applied directly to regularly sampled data, indicated as $\mathbf{D}$. We remind the reader that $\mathbf{D}$ corresponds to data in the f-x-y domain, $D(\omega, \xi_x, \xi_y)$. This volume contains observations regularly distributed on a grid described by equations 3.2 and 3.3. Therefore, for a given frequency ω, $\mathbf{D}$ is a matrix of size $N_x \times N_y$. As a denoising tool, the MSSA filter consists of three steps:

1. From the 2D spatial data matrix $\mathbf{D}$ at a given frequency ω, form a level-2 Block Hankel matrix $\mathbf{H}$. The Block Hankel matrix of size $N_1 \times N_2$ should be approximately square. A typical strategy is to consider $N_1 = L_x \times L_y$ and $N_2 = K_x \times K_y$, where $L_x = \lfloor \frac{N_x}{2} \rfloor + 1$, $K_x = N_x - L_x + 1$, $L_y = \lfloor \frac{N_y}{2} \rfloor + 1$ and $K_y = N_y - L_y + 1$. The symbol $\lfloor \rfloor$ indicates the integer part of the argument (Trickett, 2008; Oropeza and Sacchi, 2011). If one denotes the Hankelization operator by $\mathcal{H}$, then
$$\mathbf{H}(\mathbf{D}) = \mathcal{H}\,\mathbf{D}\,.$$

2. Apply rank reduction to the matrix $\mathbf{H}(\mathbf{D})$ to obtain its low-rank approximation. This operation can be carried out via the Singular Value Decomposition,

or by efficient strategies based on Lanczos bidiagonalization (Gao et al., 2013), or randomization techniques for rank reduction (Oropeza and Sacchi, 2011; Cheng and Sacchi, 2016). By symbolizing the rank reduction operator as $\mathcal{R}$, the low-rank approximation is given by

$$\hat{\mathbf{H}} = \mathcal{R}\,\mathbf{H}(\mathbf{D})\,.$$

3. Finally, apply block anti-diagonal averaging $\mathcal{A}$ to recover the filtered data

$$\hat{\mathbf{D}} = \mathcal{A}\,\hat{\mathbf{H}}\,.$$

Then, I synthesized the MSSA filter using the three operators above, which, after concatenation, can be expressed as one projection, $\mathcal{P} = \mathcal{A}\mathcal{R}\mathcal{H}$. Hence, one can summarize the MSSA filter via the following expression

$$\begin{aligned} \hat{\mathbf{D}} &= \mathcal{A}\,\mathcal{R}\,\mathcal{H}\,\mathbf{D} \\ &= \mathcal{P}\,\mathbf{D}\ . \end{aligned} \tag{3.5}$$

If the data set consists of p plane waves, the rank of the level-2 Hankel matrix is p (Hua, 1992; Yang and Hua, 1996). In general, random noise and missing observations increase the rank of $\mathbf{H}$, and rank reduction can be used to re-establish its ideal rank. Reconstruction of seismic data via MSSA requires the inclusion of a sampling operator $\mathbf{S}$, which, for the 2D case, is an $N_x \times N_y$ matrix with elements given by

$$S_{ij} = \begin{cases} 1 & \text{if grid point } i,j \text{ is occupied by a trace} \\ 0 & \text{if grid point } i,j \text{ is empty}\,. \end{cases} \tag{3.6}$$

If the observed data in the regular grid is $\mathbf{D}^{obs}$, then

$$\mathbf{D}^{obs} = \mathbf{S} \circ \mathbf{D}\,, \tag{3.7}$$

where the symbol $\circ$ indicates element-wise multiplication, also known as the Hadamard product. The MSSA reconstruction method corresponds to the following iteration

$$\mathbf{D}^{\nu} = \alpha\mathbf{D}^{obs} + (\mathbf{1} - \alpha\mathbf{S}) \circ \mathcal{P}\mathbf{D}^{\nu-1}\,. \tag{3.8}$$

The scalar $\alpha \in (0,1]$ is the reinsertion parameter (Oropeza and Sacchi, 2011), ν denotes iteration number, and $\mathbf{1}$ is the $N_x \times N_y$ all-ones matrix. For data sets with a high signal-to-noise ratio, we prefer $\alpha = 1$, which means full reinsertion of observations into the reconstructed solution. Classical MSSA interpolation assumes that observed and reconstructed data occupy the same regular grid. In the following section, I propose I-MSSA, a method for coping with off-the-grid observations.

3.1.3 The I-MSSA via the PGD method

So far, I have described the reconstruction problem as one where one recovers $\mathbf{D}$ from $\mathbf{U}$ (equation 3.1). I replace equation 3.1 by $\mathbf{U} \approx \mathcal{W}\mathbf{D}$ to emphasize the presence of noise in the observations $\mathbf{U}$. In this case, our goal is to optimize the cost function that minimizes the difference between the observations $\mathbf{U}$ and the modeled data $\mathcal{W}\mathbf{D}$. I also select the l_2 error norm to minimize

$$J = \|\mathbf{U} - \mathcal{W}\,\mathbf{D}\|_2^2 \,. \tag{3.9}$$

Equation 3.9 cannot be solved because $\mathcal{W}$ is non-invertible. Hence, the problem requires a constraint to guarantee convergence to a solution. For instance, following Xu et al. (2015), one can assume that the data Hankel matrix $\mathbf{H}(\mathbf{D})$ is of low rank. Equation 3.9 and a low-rank constraint for $\mathbf{H}(\mathbf{D})$ can be combined into a regularized cost function of the form

$$J_R = \|\mathbf{U} - \mathcal{W}\,\mathbf{D}\|_2^2 + \mu\|\mathbf{H}(\mathbf{D}) - \mathbf{A}\mathbf{B}\|_2^2 \,, \tag{3.10}$$

where the $N_1 \times N_2$ level-2 Block Hankel matrix of the desired data is approximated by factors $\mathbf{A}$ and $\mathbf{B}$ of size $N_1 \times p$ and $p \times N_2$, respectively. The problem requires solving for the unknowns $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ simultaneously. In this thesis, we explore a less complicated algorithm via the application of the Projected Gradient Descent (PGD) method (Bertsekas, 1996; Iusem, 2003; Cheng and Sacchi, 2016; Bolduc et al., 2017; Peters et al., 2019). We propose solving

$$\min J \quad \text{subject to } \mathbf{D} \in \mathcal{S} \,, \tag{3.11}$$

where $\mathcal{S}$ is the set of solutions $\mathbf{Z}$, such that $\mathbf{H}(\mathbf{Z})$ is low-rank. The Projected Gradient Descent algorithm can be expressed by an iteration that contains two stages. First, a step in the gradient descending direction of J. Then, a projection onto the set $\mathcal{S}$

$$\mathbf{D}^{\nu} = \mathcal{P}\left[\,\mathbf{D}^{\nu-1} - s^{\nu}\nabla J(\mathbf{D}^{\nu-1})\right], \tag{3.12}$$

where s^{ν} is the step size at iteration ν, and $\mathcal{P}$ is the previously described MSSA filter.

The gradient contains the interaction between off-the-grid data and the data in the desired regular grid. After replacing ∇J in equation 3.12, we have

$$\mathbf{D}^{\nu} = \mathcal{P}\left[\,\mathbf{D}^{\nu-1} - s^{\nu}\mathcal{W}^{*}(\mathcal{W}\,\mathbf{D}^{\nu-1} - \mathbf{U})\right]. \tag{3.13}$$

The term $\mathcal{W}\,\mathbf{D}^{\nu-1} - \mathbf{U}$ defines off-the-grid observation error. Next, the adjoint interpolator $\mathcal{W}^{*}$ maps the error back to the regular desired grid. The fitting goal guarantees that the observed trace coordinates are honoured while the MSSA filter is applied to the reconstructed data at each iteration.

Algorithm 1 shows the final expression for the proposed I-MSSA method. As a stopping criterion, we consider the normalized Frobenius norm of the gradient of the cost function J. The algorithm stops when either $\|\nabla J^{\nu}\|/\|\nabla J^{1}\| < \eta$ or a maximum number of iterations `Max_Iter` is reached.

Algorithm 1 `I-MSSA` via Projected Gradient Descent Algorithm

```
Input: $\mathbf{U}$ (off-the-grid data), $p$ (rank), $\eta$, Max_Iter
$i = 1$
while $\|\nabla J^{i}\|/\|\nabla J^{1}\| \leq \eta$ and $i \leq$ Max_Iter do
    Descent direction: $\nabla J = \mathcal{W}^{*}(\mathcal{W}\mathbf{D} - \mathbf{U})$
    Step size: $s =$ Backtracking($\mathbf{D}, \nabla J$)
    Update: $\mathbf{D} = \mathbf{D} - s\nabla J$
    Projection: $\mathbf{D} =$ MSSA($\mathbf{D}, p$)
    $i \leftarrow i + 1$
end while
```

Backtracking line search

To find the step size in equation 3.13, we used Backtracking Line Search (Nocedal and Wright, 2006). The algorithm starts with an initial estimation of the step size s_0 and iteratively shrinks the step size until satisfying Armijo's condition (Nocedal and

Wright, 2006). Other strategies could have been adopted for adaptive line search, including those discussed by Iusem (2003) and, more recently, by Bolduc et al. (2017). For our particular application, backtracking was preferred because it leads to a Projected Gradient Descent algorithm that converges in a few iterations, typically about 8-10 iterations. Algorithm 2 shows the expression for the backtracking line search used in I-MSSA.

Algorithm 2 `Backtracking`

```
Input: $s_0 > 0$ (initial large step), $\rho \in (0,1)$, $c \in (0,1/2)$
$s = s_0$
while $J(\mathbf{D} - s\nabla J) \leq J(\mathbf{D}) - cs\|\nabla J\|^2$ do
    $s \leftarrow \rho s$
end while
$s^i = s$
```

3.2 Examples

3.2.1 Synthetic Examples

To evaluate I-MSSA, we present a series of synthetic examples. We quantify the quality of the reconstruction via the signal-to-noise ratio of the output, which is defined as follows

$$\texttt{S/N}_o = 10\log_{10}\frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{rec} - \mathbf{d}_{true}\|_2^2}\ , \tag{3.14}$$

where $\|\cdot\|_2^2$ denotes the squared l_2 norm of a vector, $\mathbf{d}_{true}$ is the vectorized ideal volume without decimation or noise contamination, and $\mathbf{d}_{rec}$ is the recovered data after reconstruction. The reader should note that, even though the reconstruction is calculated in the frequency domain, $\texttt{S/N}_o$ is calculated in the time-space domain. We also define the signal-to-noise ratio of the input data via

$$\texttt{S/N}_i[dB] = 10\log_{10}\frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{obs} - \mathbf{d}_{true}\|_2^2}\,, \tag{3.15}$$

which is also the ratio of the power of the clean signal to the power of the noise. To add noise to clean synthetic observations, $\mathbf{d}_{obs} = \mathbf{d}_{true} + \mathbf{n}$, we adopt Gaussian random noise with a variance calculated to yield the desired $\texttt{S/N}_i$.

Linear events

The first example considers a 3D volume with two dipping linear events. The initial grid consists of 25 × 125 grid points separated by 8 m in both x and y directions. After defining the initial grid, the positions were jittered and randomly decimated, to produce the observed data coordinates. The output desired grid has a spacing of 25 m in both x and y directions with a total of 10 × 50 traces. The observed traces account for 40% of the desired number of output traces. We show the initial grid, the coordinates of the observations, and the desired grid in Figure 3.2. For the I-MSSA reconstruction, we consider an ideal rank $p = 2$. For the convergence of the reconstruction, we examine the difference between the normalized gradients of two consecutive iterations. To minimize the number of iterations, we define $\eta = 0.05$ for the bilinear interpolator and $\eta = 0.02$ for the windowed sinc interpolator.

The range of frequencies in which we calculate the reconstruction is 1-80 Hz. The step size was calculated using an inexact line search via backtracking. Heuristically, we found that the optimal parameters for the backtracking algorithm are $c = 0.001$, $\rho = 0.8$, and the initial large step size $s_0 = 5$. For better analysis, we consider two interpolators in the I-MSSA algorithm, a bilinear interpolation operator, and a Kaiser-windowed sinc interpolation operator (See Appendix B for details).

We first consider input data free of random additive noise ($\mathtt{S/N}_i = \infty$). Figure 3.3 displays 3D volumes corresponding to this test. Figure 3.3a shows the observed data after binning, and Figure 3.3b shows the result of applying conventional MSSA after binning. Figure 3.3c shows the reconstruction via I-MSSA with bilinear interpolation, and Figure 3.3d depicts the reconstruction via I-MSSA with sinc interpolation. Finally, Figures 3.3e-3.3g show the residuals between the ideal volume and the reconstructions in Figure 3.3b-3.3d, respectively. The reconstruction via MSSA, with an ideal rank equal to 2, results in $\mathtt{S/N}_o = 18.5$ dB. Conversely, the reconstruction via I-MSSA with the bilinear interpolation results in $\mathtt{S/N}_o = 24.4$ dB. The reconstruction via I-MSSA with the sinc interpolator yields $\mathtt{S/N}_o = 38.9$ dB. The difference in $\mathtt{S/N}_o$ between the MSSA and the I-MSSA reconstruction is a direct consequence of errors introduced by binning in the MSSA method. However, the difference in reconstruction quality between our two implementations of I-MSSA is related to the error in the interpolators. In general, for simple examples, the sinc interpolator yields superior results compared to those obtained with the linear interpolator.

However, as we will see in the following tests, the difference is less significant when working with examples that depart from an ideal superposition of linear events.

Figure 3.4 shows the 2D slices of the data in Figure 3.3. Figure 3.4a shows the observed data after binning. Figure 3.4b shows the reconstruction via conventional MSSA, after binning. Figure 3.4c presents a slice of the volume obtained after I-MSSA reconstruction implemented with the bilinear interpolation. The results of I-MSSA with sinc interpolation are in Figure 3.4d. In all cases, we plot every third trace. Figure 3.4e shows two traces extracted from the MSSA volume. We colored in red the difference between ideal traces and reconstructed traces. Figures 3.4f and 3.4g show the reconstruction via I-MSSA with bilinear and sinc interpolation. We also colored in red the difference between these results and the ideal traces.

Finally, in Figure 3.5, we highlight the difference between the ideal volume and the processed outputs for one trace for the volumes displayed by Figures 3.3a-3.3d, respectively. In all cases, we plot with dash lines the ideal trace for reference. In red is the trace obtained after binning. The figure highlights the time shift introduced by binning. These time shifts affect the input volume that is provided to the MSSA algorithm. In blue, we also portray the reconstructed trace with MSSA after binning. MSSA can cope with the time shifts caused by binning but does not fully compensate for them. In green, we highlight the reconstruction via I-MSSA with bilinear interpolation. In magenta, the reconstruction via I-MSSA with sinc interpolation. I-MSSA applies reconstruction directly with input data in their actual coordinates without having to use binning. Therefore, it does not introduce time shifts in the input volume, obtaining reconstructed volumes of higher $\texttt{S/N}_o$. From the figures, we conclude that I-MSSA, with either the bilinear interpolator or the sinc interpolator, yields better results than MSSA.

In our next example, we add Gaussian noise to the volume of linear events with $\texttt{S/N}_i = 1.2$ dB. We show reconstruction results obtained via MSSA and I-MSSA in Figures 3.6 and 3.7. The reconstruction via MSSA, with an ideal rank $p = 2$, yields $\texttt{S/N}_o = 13.5$ dB. The reconstruction via I-MSSA yields $\texttt{S/N}_o = 15.6$ dB with bilinear interpolation, and $\texttt{S/N}_o = 15.9$ dB with sinc interpolation.

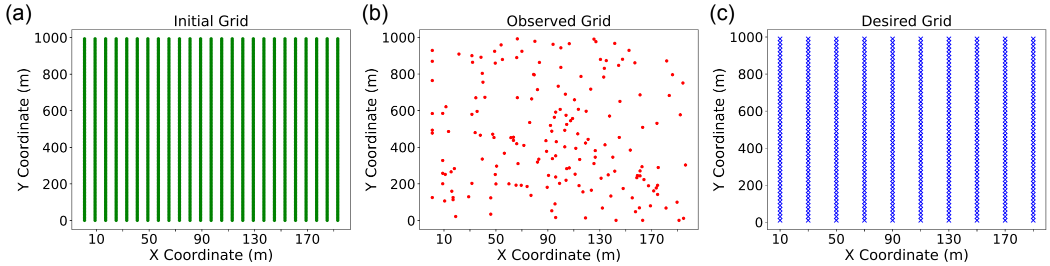

Figure 3.2: Coordinate systems considered in the synthetic examples. (a) Initial grid. (b) Coordinates of the observations. (c) Desired output grid.

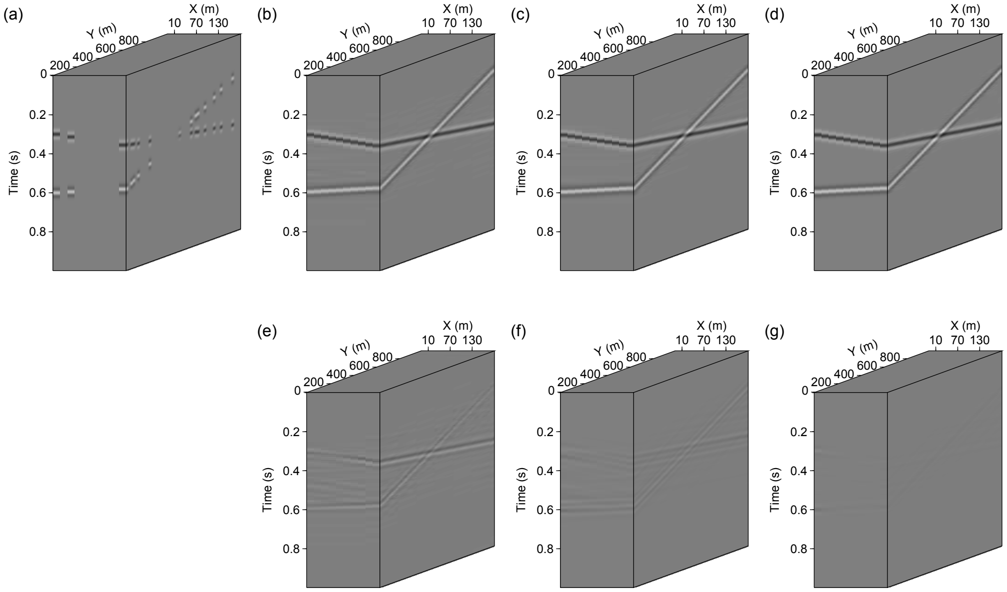

Figure 3.3: Reconstruction of a randomly decimated volume with linear events and $\mathtt{S/N}_i = \infty$. (a) Data volume after binning. (b) Result of the 3D MSSA reconstruction. (c) Reconstruction calculated via the I-MSSA algorithm with bilinear interpolation and (d) sinc interpolation. (e) Residuals between the ideal volume and the reconstructed result in (b) with resulting $\mathtt{S/N}_o = 18.5$ dB. (f) Residuals between the ideal volume and the reconstructed volume in (c) with resulting $\mathtt{S/N}_o = 24.4$ dB. (g) Residuals between the ideal volume and the reconstructed volume in (d) with resulting $\mathtt{S/N}_o = 38.9$ dB.

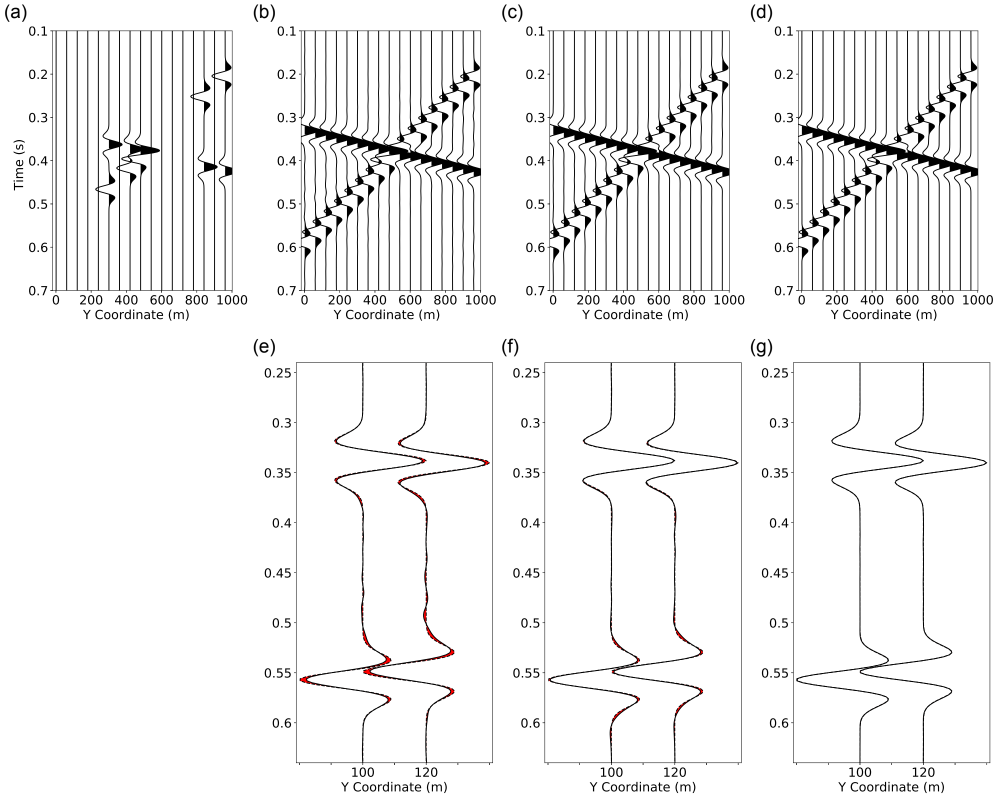

Figure 3.4: An x-slice of the volume in Figure 3.3. (a) Binned volume. (b) Reconstruction via MSSA from the binned volume. The reconstruction yields $\texttt{S/N}_o = 18.5$ dB. (c) Reconstruction via I-MSSA from irregular traces with bilinear interpolation. The reconstruction yields $\texttt{S/N}_o = 24.4$ dB. (d) Reconstruction via I-MSSA from irregular traces with sinc interpolation. The reconstruction yields $\texttt{S/N}_o = 38.9$ dB. Figures (a) to (d) show one every third trace for better visualization. Sections (e) to (g) show a detail of the results in (b) to (d), respectively. Highlighted in red is the difference between the reconstructed traces and the ideal ones.

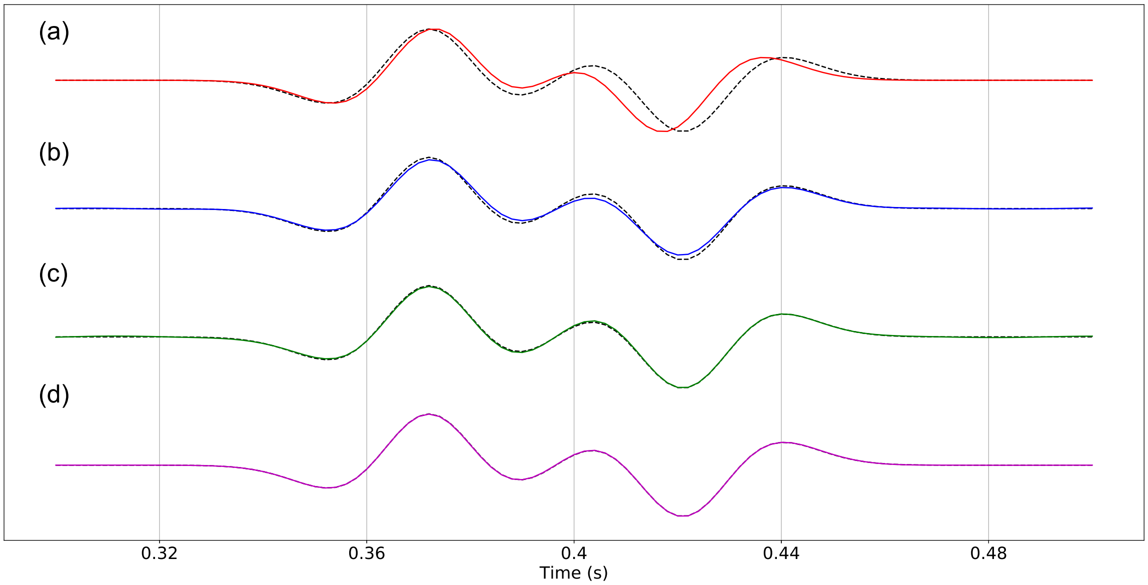

Figure 3.5: One trace from Figures 3.3(a) to (d). In black dashed line is the ideal trace. (a) In red is the binned trace. This figure shows the time shift in the traces induced by binning. (b) In blue, the reconstructed trace with MSSA after binning. MSSA can cope with the time shifts induced by binning but does not fully compensate for the errors. (c) In green, the reconstruction via I-MSSA with bilinear interpolation. (d) In magenta, the reconstruction via I-MSSA with sinc interpolation. I-MSSA considers reconstruction from the observed traces, without binning. Therefore, it does not include time shifts in the input volume, obtaining reconstructed volumes with higher $\mathtt{S/N}_o$.

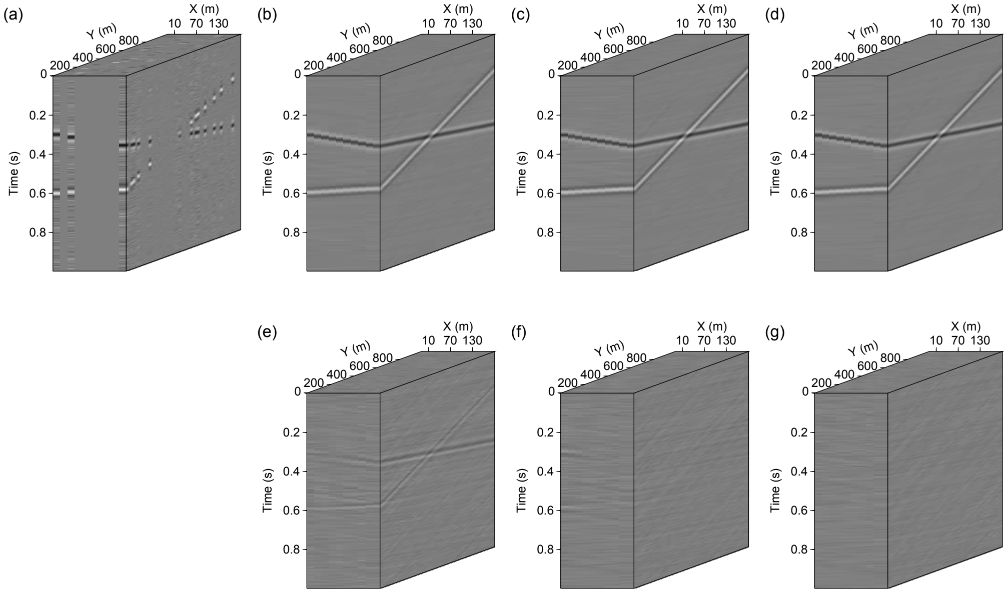

Figure 3.6: Reconstruction of a randomly decimated volume with linear events and Gaussian noise. $\mathtt{S/N}_i = 1.2$ dB. (a) Data volume after binning. (b) MSSA reconstruction. (c) Reconstructed volume obtained via the I-MSSA algorithm with bilinear interpolation, and (d) with sinc interpolation. (e) Residuals between the ideal volume and reconstructed volume in (b) with resulting $\mathtt{S/N}_o = 13.5$ dB. (f) Residuals between the ideal volume and reconstructed volume in (c) with resulting $\mathtt{S/N}_o = 15.6$ dB. (g) Residuals between the ideal volume and reconstructed volume in (d) with resulting $\mathtt{S/N}_o = 15.9$ dB.

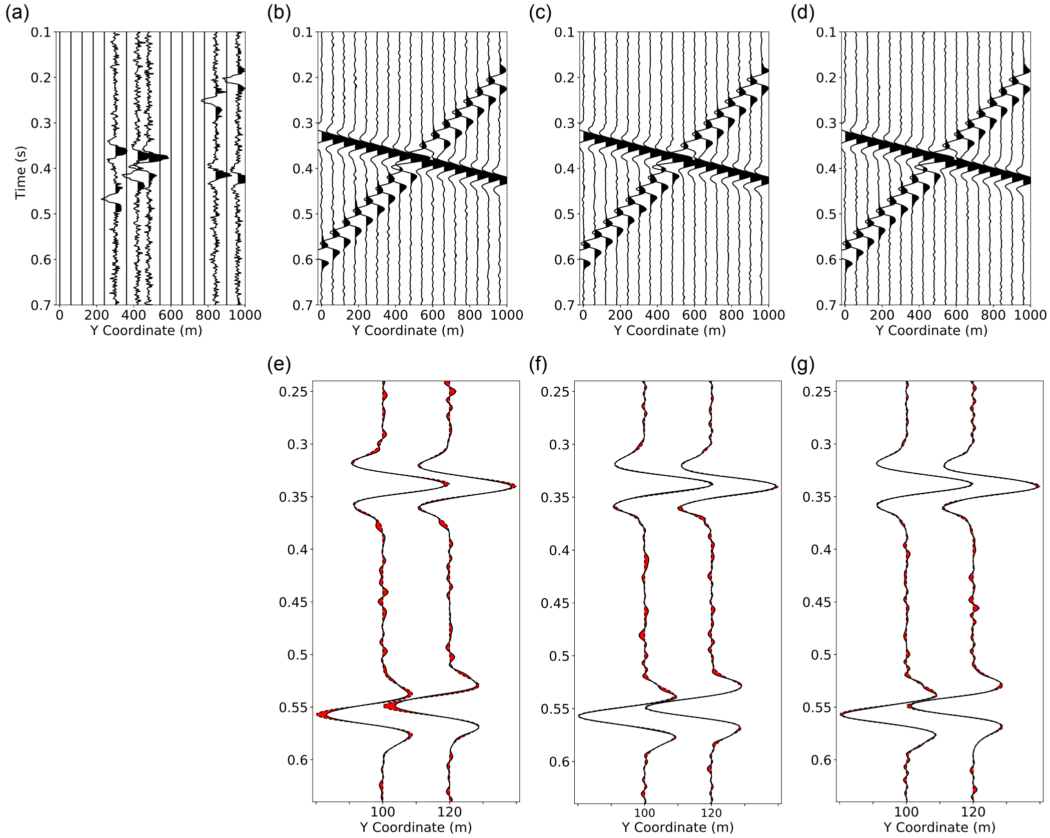

Figure 3.7: Reconstruction of a randomly decimated volume with linear events and Gaussian noise, $\texttt{S/N}_i = 1.2$ dB. This figure shows x-slices of the 3D volume in Figure 3.6. Figures (a) to (d) show one every third trace. (a) Binned volume. (b) Reconstruction via MSSA from the binned volume. The reconstruction yields $\texttt{S/N}_o = 13.5$ dB. (c) Reconstruction via I-MSSA from irregular traces with bilinear interpolation. The reconstruction yields $\texttt{S/N}_o = 15.6$ dB. (d) Reconstruction via I-MSSA from irregular traces with sinc interpolation. The reconstruction yields $\texttt{S/N}_o = 15.9$ dB. (e) to (g) show a detail of the results in (b) to (d), respectively. Highlighted in red is the difference between the reconstructed traces and the ideal ones.

Linear and parabolic events

We also test the method on a synthetic section that contains one dipping linear event and two parabolic events. We use this example to test the technique in a scenario slightly less ideal than the previous example. The dip of the parabolic event varies in space, violating the constant dip assumption made by MSSA. The initial grid is composed of 25×125 grid points. The distance between grid points is 8 m for both x and y-directions. The grid is decimated and jittered to model the observed coordinates. The output desired grid presents 10×50 grid points spaced every 25 m in both x and y-directions. The observed traces represent 40% of the number of desired traces in the regular output grid. Figure 3.2 shows the initial, observed, and output coordinates.

The reconstruction via I-MSSA adopts a rank $p = 6$. The range of frequencies included in the test is 1-80 Hz. The stopping criteria and step size selection for I-MSSA is the same used for the example with linear events. We first consider the case where the input decimated data does not contain additive noise ($\texttt{S/N}_i = \infty$). The results for this example are shown in Figure 3.8. Figure 3.8a shows the data volume after trace binning. Figure 3.8b shows the reconstruction via the MSSA method, with $\texttt{S/N}_o = 9$ dB. Figure 3.8c shows the reconstruction via I-MSSA using bilinear interpolation. This result yields $\texttt{S/N}_o = 10.38$ dB. Figure 3.8d shows the reconstruction with sinc interpolation, resulting in $\texttt{S/N}_o = 10.9$ dB. Figures 3.8e-3.8g show the respective residuals. Figure 3.9 shows the 2D slices extracted from the volumes portrayed in Figure 3.8.

Figure 3.10 describes the experiment with added Gaussian noise of $\texttt{S/N}_i = 1.2$ dB. Figure 3.10a shows the binned traces. Figure 3.10b shows the traces after applying conventional MSSA, with a resulting $\texttt{S/N}_o = 5.8$ dB. Figure 3.10c shows the reconstruction via I-MSSA with bilinear interpolation. This result corresponds to $\texttt{S/N}_o = 7.65$ dB. Figure 3.10d shows the reconstruction via I-MSSA with sinc interpolation. This result corresponds to $\texttt{S/N}_o = 7.8$ dB. Figure 3.10e-3.10g show the respective residuals. Figure 3.11 shows 2D slices extracted from Figure 3.10.

As expected, the resulting $\texttt{S/N}_o$ values for these examples are not as favorable as those obtained for the experiments with linear events. The MSSA method assumes linear events in the t-x domain, a condition that is not satisfied in this example. Nevertheless, these experiments show that reconstruction via I-MSSA can also cope

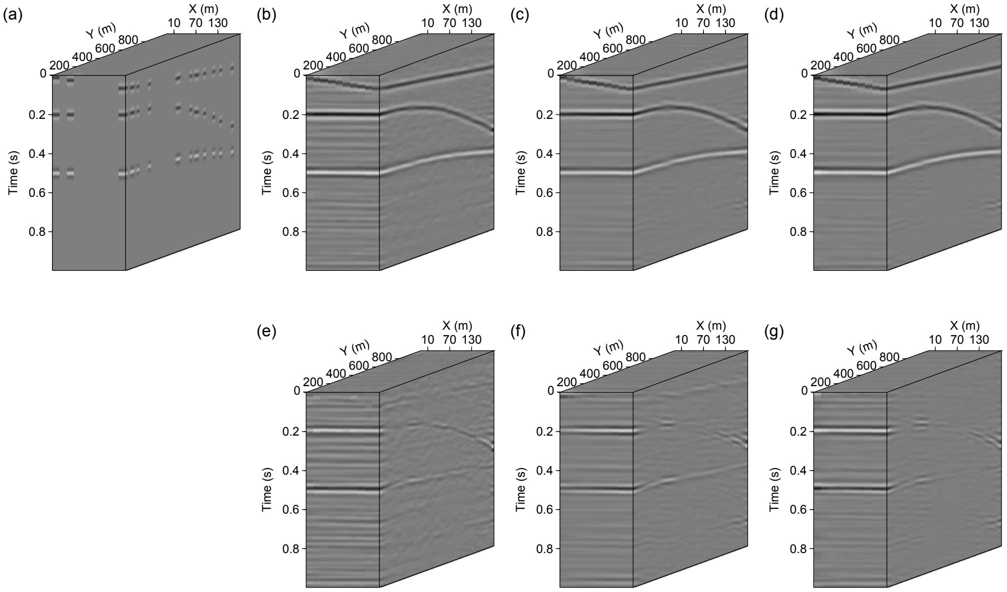

Figure 3.8: Reconstruction of a randomly decimated volume with a combination of linear and parabolic events. For this example, $\mathtt{S/N}_i = \infty$. (a) Data volume after binning. (b) Reconstruction with the MSSA method. (c) Reconstruction calculated via I-MSSA with bilinear interpolation and (d) sinc interpolation. (e) Residuals between the ideal volume and the reconstructed volume in (b) ($\mathtt{S/N}_o = 9$ dB). (f) Residuals between the ideal volume and the reconstructed volume in (c) ($\mathtt{S/N}_o = 10.38$ dB). (g) Residuals between the ideal volume and the reconstructed volume in (d) ($\mathtt{S/N}_o = 10.9$ dB).

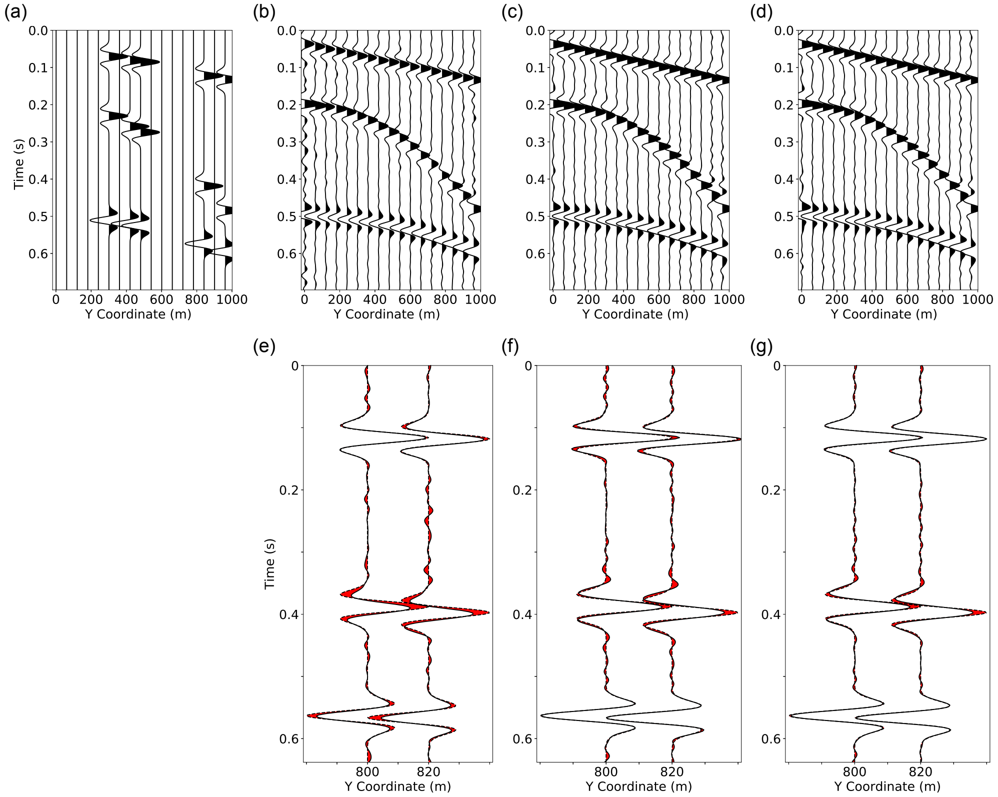

Figure 3.9: Reconstruction of a randomly decimated volume with linear and parabolic events. This figure shows x-slices of the 3D visualization in Figure 3.8. Figures (a) to (d) show one every third trace for better visualization. (a) Binned volume. (b) Reconstruction via MSSA from the binned volume. The reconstruction yields $\mathtt{S/N}_o = 9$ dB. (c) Reconstruction via I-MSSA from irregular traces with bilinear interpolation. The reconstruction yields $\mathtt{S/N}_o = 10.38$ dB. (d) Reconstruction via I-MSSA from irregular traces with sinc interpolation. The reconstruction yields $\mathtt{S/N}_o = 10.9$ dB. Sections (e) to (g) show a detail of the results in (b) to (d), respectively. Highlighted in red is the difference between the reconstructed traces and the ideal ones.

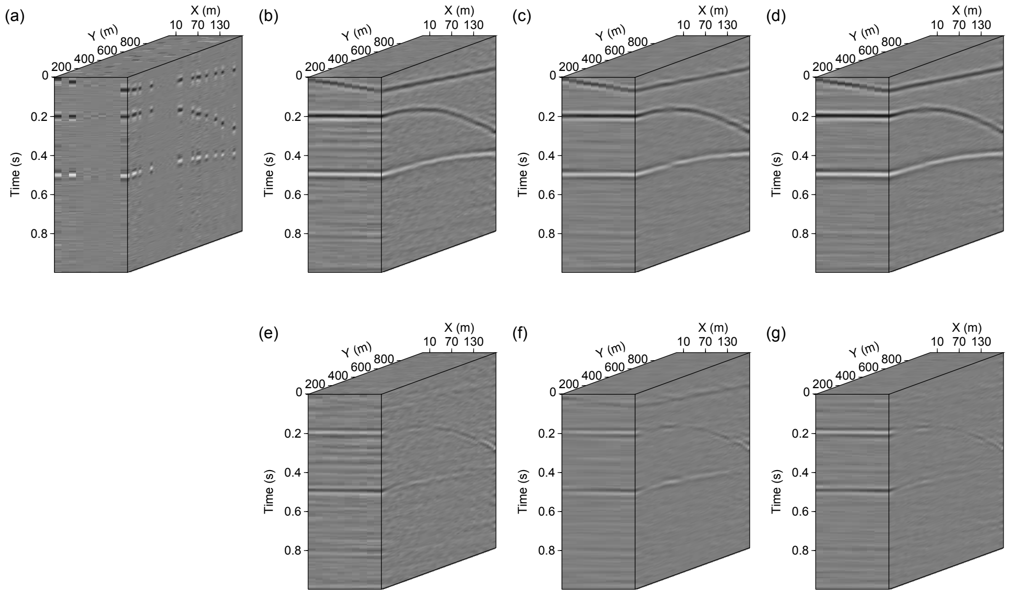

Figure 3.10: Reconstruction of a randomly decimated volume with linear and parabolic events. The volume includes Gaussian noise with $\texttt{S/N}_i = 1.2$ dB. (a) Data volume after binning. (b) Result of 3D MSSA reconstruction. (c) Reconstruction calculated via I-MSSA with bilinear interpolation, and (d) sinc interpolation. (e) Residuals between ideal volume and MSSA reconstructed result in (b). $\texttt{S/N}_o = 5.8$ dB. (f) Residuals between ideal volume and reconstructed result in (c). $\texttt{S/N}_o = 7.65$ dB. (g) Residuals between ideal volume and reconstructed result in (d). $\texttt{S/N}_o = 7.8$ dB.

with moderate spatially variant dips.

Convergence of I-MSSA

We analyze the convergence of the proposed algorithm using results obtained from the four synthetic tests. First, we remind the reader that we have omitted the temporal frequency ω from our expressions to simplify notation, but it is clear that the cost J to minimize in equations 9-11 is a function of ω, $J(\omega)$. We evaluate the scalar $\rho = \sum_{\omega} \|\nabla J(\omega)\|_F^2$ and display it versus iteration to analyze the convergence of I-MSSA. Figure 3.12 shows ρ^k/ρ^1 where k is iteration number. We point out that in our algorithm, the stopping criterion is applied individually to each frequency. Figure 3.12 shows the global behavior of the gradient for all frequencies to summarize

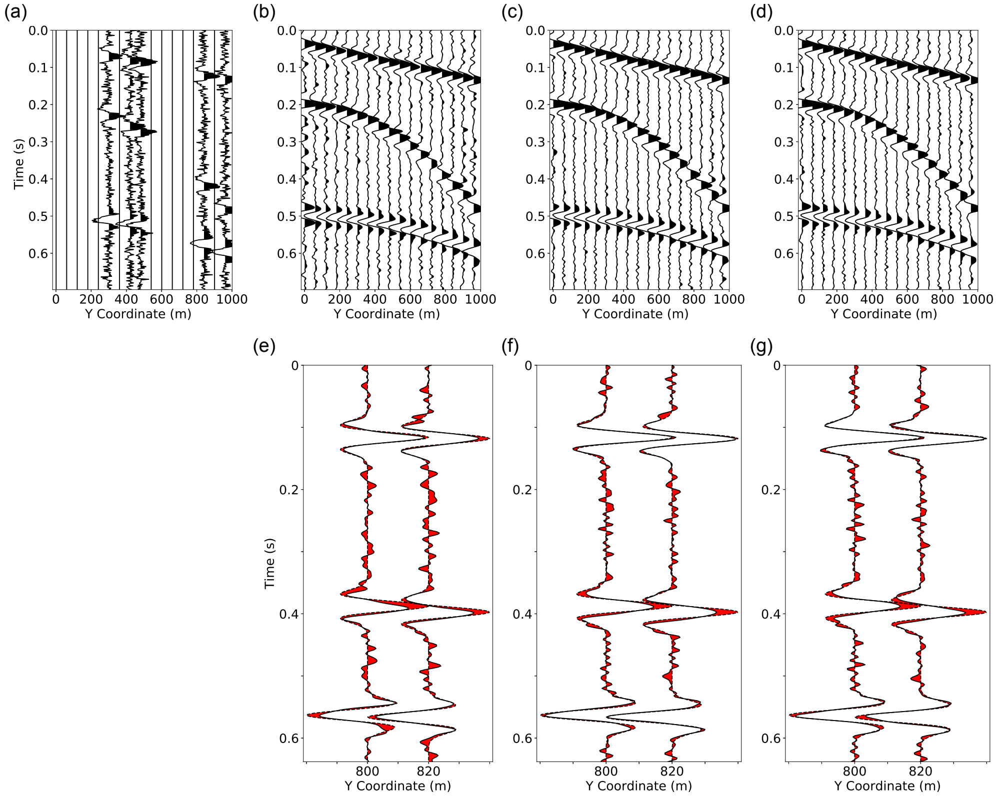

Figure 3.11: Reconstruction of a randomly decimated volume with linear and parabolic events, with Gaussian noise of $\texttt{S/N}_i = 1.2$ dB. This figure shows x-slices of the 3D visualization in Figure 3.10. Figures (a) to (d) show one every third trace for better visualization. (a) Binned volume. (b) Reconstruction via MSSA from the binned volume. The reconstruction yields $\texttt{S/N}_o = 5.8$ dB. (c) Reconstruction via I-MSSA from irregular traces with bilinear interpolation. The reconstruction yields $\texttt{S/N}_o = 7.65$ dB. (d) Reconstruction via I-MSSA from irregular traces with sinc interpolation. The reconstruction yields $\texttt{S/N}_o = 7.8$ dB. Sections (e) to (g) show a detail of the results in (b) to (d), respectively. Highlighted in red is the difference between the reconstructed traces and the ideal ones.

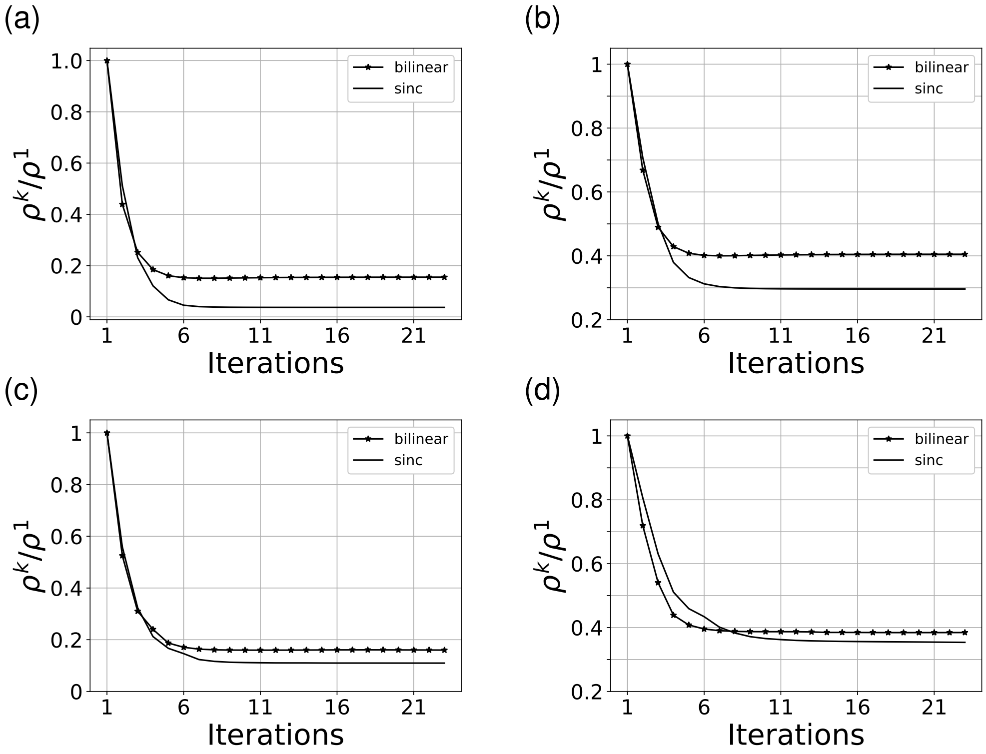

Figure 3.12: Convergence of the I-MSSA algorithm. This diagram shows the normalized, averaged over frequency, squared Frobenius norm of the gradient of the cost function versus iteration. (a) Linear events. (b) Linear events with Gaussian noise. (c) Linear and parabolic events. (d) Linear and parabolic events with Gaussian noise.

results in one diagram.

Figure 3.13 shows the quality of the reconstruction given by $\mathtt{S/N}_o$ versus iteration until convergence. These figures show that the Projected Gradient Descent algorithm with backtracking line-search always converges in fewer iterations than the MSSA algorithm. Besides, I-MSSA with bilinear interpolation requires about 5-7 iterations until convergence, while windowed-sinc interpolation requires about 8-12 iterations. Then, an interpolator such as I-MSSA that considers reconstruction from real coordinates, without binning, results in an improved reconstruction of seismic traces.

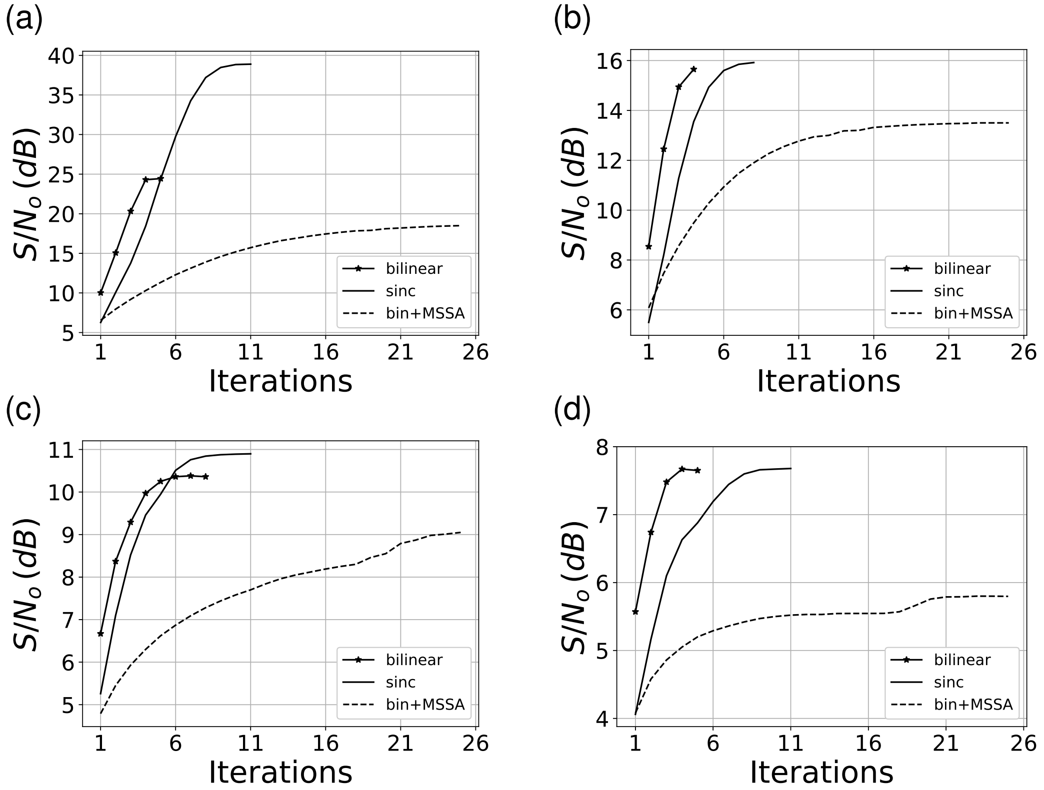

Figure 3.13: Signal-to-noise ratio for MSSA, and I-MSSA considering a bilinear interpolation operator and a sinc interpolation operator. 60% of the traces are decimated. (a) Linear events. (b) Linear events with Gaussian noise. (c) Linear and parabolic events. (d) Linear and parabolic events with Gaussian noise.

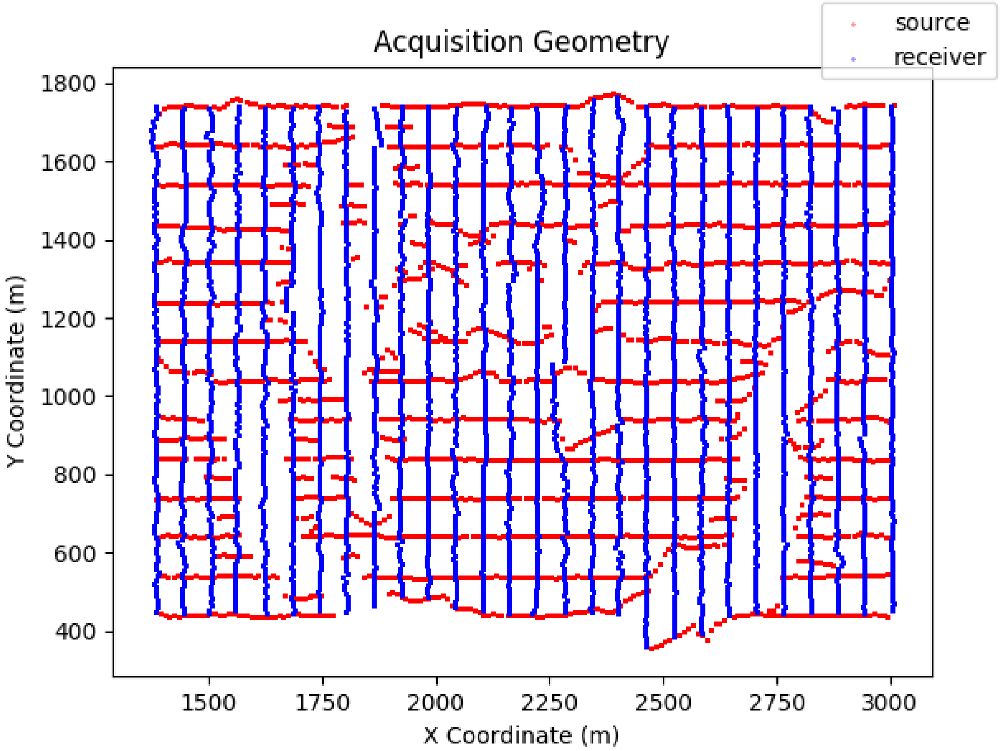

Figure 3.14: Survey acquisition geometry. Red dots indicate the location of the sources. Blue dots represent the location of the receivers.

3.2.2 Field data tests

We considered a 3D onshore seismic data set acquired in the Western Canadian Sedimentary Basin. The data correspond to an orthogonal survey. Figure 3.14 delineates the acquisition geometry. The red dots represent the sources' location, and the blue dots indicate the receivers.

Reconstruction of a shot gather

We extracted a shot gather to test the reconstruction capabilities of I-MSSA. We randomly decimated 60% of the traces and reconstructed them with the proposed method. Finally, we compared the data and the reconstructed result.

Figure 3.15 shows the location of the active source and receivers. Figure 3.16a shows the distribution of the receiver stations for the shot gather. The mean distance between receiver lines is 60 m, and the mean distance between receivers in the same line is 10 m. Figure 3.16b shows the location of traces after decimation, and after reconstruction in Figure 3.16c. The second row of the image depicts an x-slice of the corresponding volume. Figure 3.16d shows the observed traces for a central line, and Figure 3.16e shows the results after the reconstruction with I-MSSA. For the reconstruction, we considered a rank $p = 20$. Figure 3.16f depicts the residuals between the observed data and the reconstruction via I-MSSA.

The residuals in Figure 3.16f show insignificant signal leakage or artifacts. The I-MSSA method also tolerates events that are not strictly linear. As analyzed in the synthetic examples, I-MSSA can accommodate slightly curved events without affecting the final solution.

Reconstruction of cross-spread gathers

We now consider the reconstruction of the orthogonal seismic survey in the CMP domain. To process the data, we organized the traces in cross-spread gathers. Then, for each cross-spread gather, we calculated actual CMP locations and reconstructed the traces via I-MSSA into a regular grid of CMPs. Once all the gathers were reconstructed, we stacked traces that belong to the same regularized CMP coordinate. In this workflow, we reconstruct the full wavefield by processing 3D subvolumes (cross-spreads) of the survey.

Figure 3.17 shows the geometry for one cross-spread gather, and the area in magenta highlights the CMP coverage for that particular gather. Figure 3.18 shows a diagram of the proposed workflow.

Our tests compare three possible approaches to the regularization problem. The first one considers the regularization of the data via binning, without reconstruction. The second one consists of two steps: binning followed by reconstruction of the data via classical MSSA. Finally, we show the results obtained via I-MSSA. It is worth mentioning that the parameters of each algorithm were optimized heuristically to provide a minimum amount of signal leakage in error panels that we compute for quality control. Moreover, we point out that the same output grid was used for MSSA and I-MSSA. The output CMP coordinates correspond to a grid of 5 m in

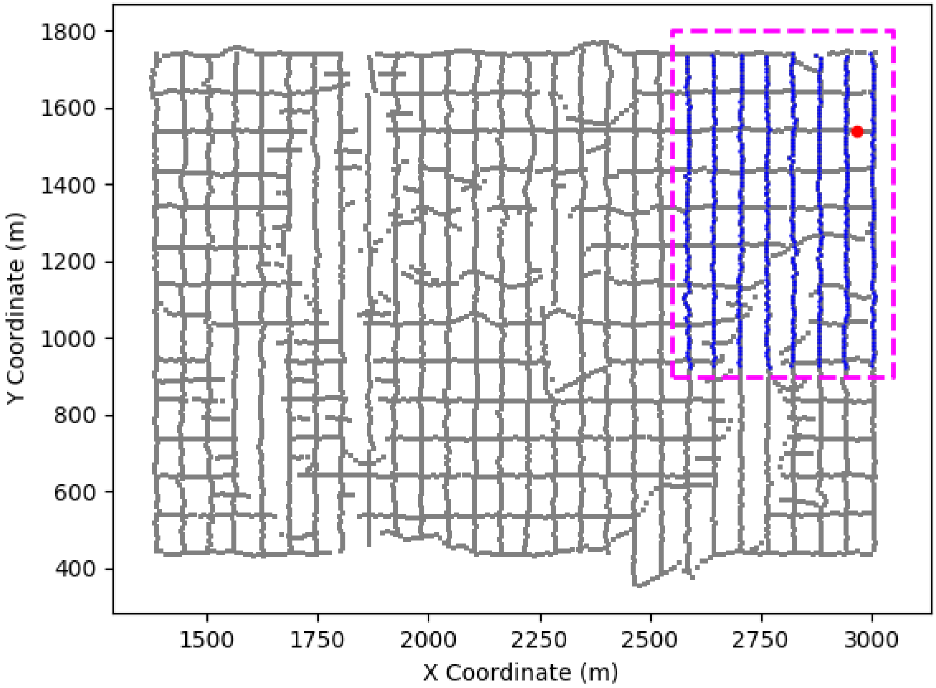

Figure 3.15: Geometry for one-shot gather. The magenta rectangle delineates the area of interest. In red, the location of the source, and in blue, the coordinates of the receivers. Dots in grey represent sources and receivers that are not considered for the reconstruction.

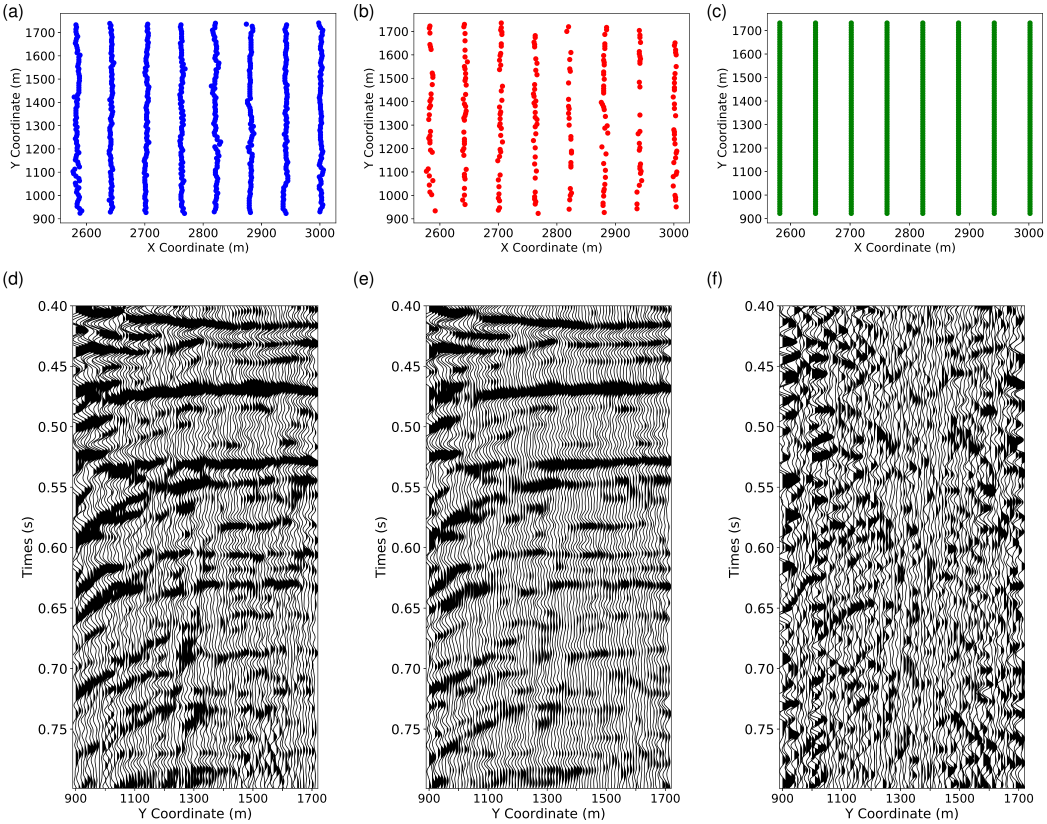

Figure 3.16: Reconstruction of a randomly decimated volume of a shot gather from the Western Canadian Sedimentary Basin. (a) Observed grid. (b) Randomly decimated grid. (c) Output desired grid. (d) Crossline of observed traces. (e) I-MSSA reconstruction. (f) Residuals between regularized observed traces and I-MSSA reconstruction.

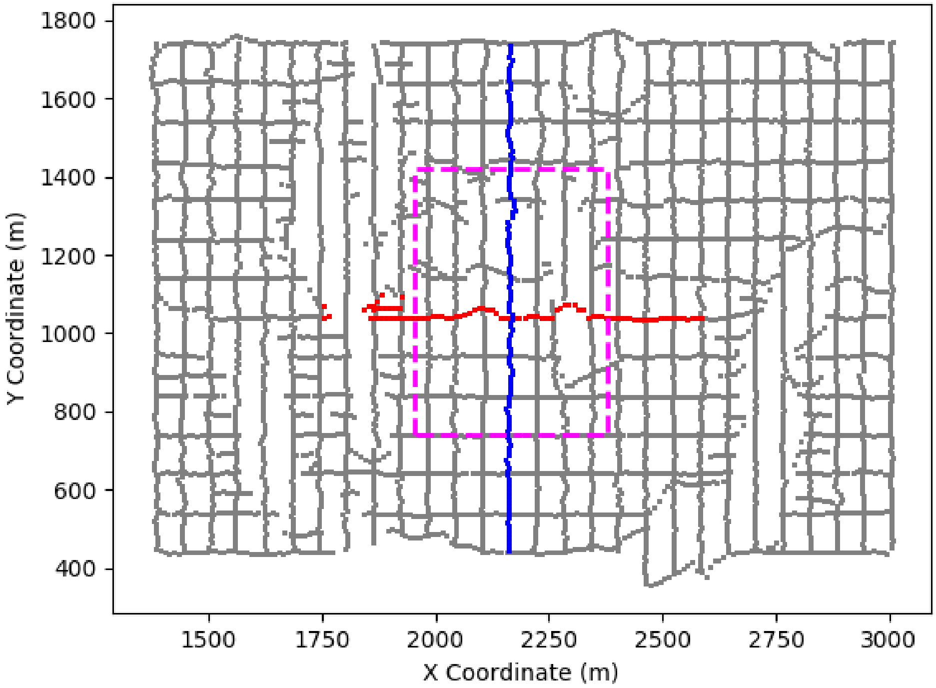

Figure 3.17: Geometry for one cross-spread. In red, the location of the sources. Receivers are indicated in blue. Dots in grey represent sources and receivers that do not belong to this cross-spread. The magenta rectangle delineates the CMP coverage for this cross-spread.

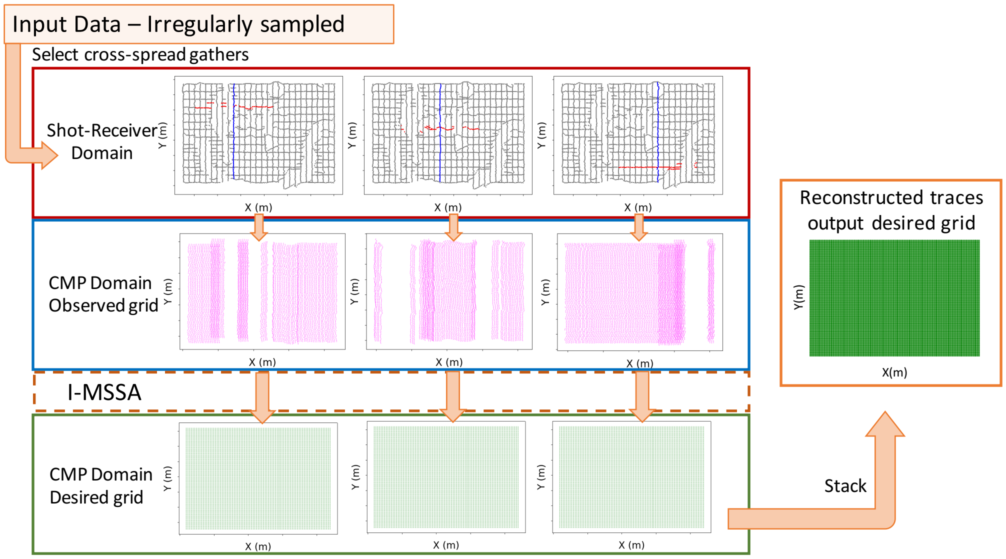

Figure 3.18: Workflow for the reconstruction of cross-spread gathers in the CMP domain. First, we organize the data into cross-spread gathers. Second, we define the observed grid as the true CMP locations of the traces. Third, we apply I-MSSA reconstruction independently for each set of traces belonging to one cross-spread. The flow yields CMP positions that are distributed in the regular desired output grid. Finally, we can stack traces with matching CMP coordinates.

both x and y-directions. We identify the first and second spatial dimensions with common midpoint x (CMPx) and common midpoint y (CMPy).”

Figure 3.19 compares CMPx slice number 20 of one cross-spread for the different methods. Figure 3.19a shows the raw data after binning, without reconstruction. Figure 3.19b shows the result of binning and reconstructing the volume via MSSA. Finally, Figure 3.19c shows the corresponding slices after I-MSSA. Figure 3.20 shows a detail of the previous CMPx sections. Figures 3.20a-3.20c show the observed traces in red. Figure 3.20a shows the binned data. Figure 3.20b shows the result after MSSA reconstruction. Figure 3.20c shows the results after I-MSSA reconstruction. This figure reveals the effect of binning versus reconstruction with interpolation. The misalignment of events in the binned volume results in a reduced quality of the reconstruction via MSSA compared to the I-MSSA method.

Figure 3.21 shows sections of the same volume for 5 consecutive slices. The arrow on top of CMPx number 34 marks the presence of a fault. The discontinuity is noticeable in the reconstruction obtained via I-MSSA.

Figure 3.22 shows time slices for two different cross-spread gathers. The first row presents the time slice at $t = 0.4$ s for cross-spread number 60, and the second row shows the time slice at $t = 0.7$ s for cross-spread number 100. Figures 3.22a and 3.22d show the binned data without reconstruction. Figures 3.22b and 3.22e show the reconstructed data after binning and MSSA. Figures 3.22c and 3.22f show the result after I-MSSA reconstruction. This figure confirms that the reconstruction via I-MSSA results in more continuous events that resemble geologically plausible results. The improved quality of the reconstruction using the I-MSSA algorithm is particularly evident when analyzing the artifacts in Figure 3.22b and 3.22e. Such artifacts are a consequence of using the binned volume as an input for the reconstruction algorithm. On the contrary, they are attenuated in the reconstructed volume via the I-MSSA method, where the input volume includes traces in their actual location.

Finally, Figure 3.23 shows a comparison of the amplitude spectrum before and after reconstruction. Figure 3.23a shows the results for MSSA, and Figure 3.23b shows the result for I-MSSA. We conclude that both methods reconstruct the data without introducing significant changes in the amplitude spectrum.

Figure 3.24 shows a crossline of the stacked cube for the binned data, binned data

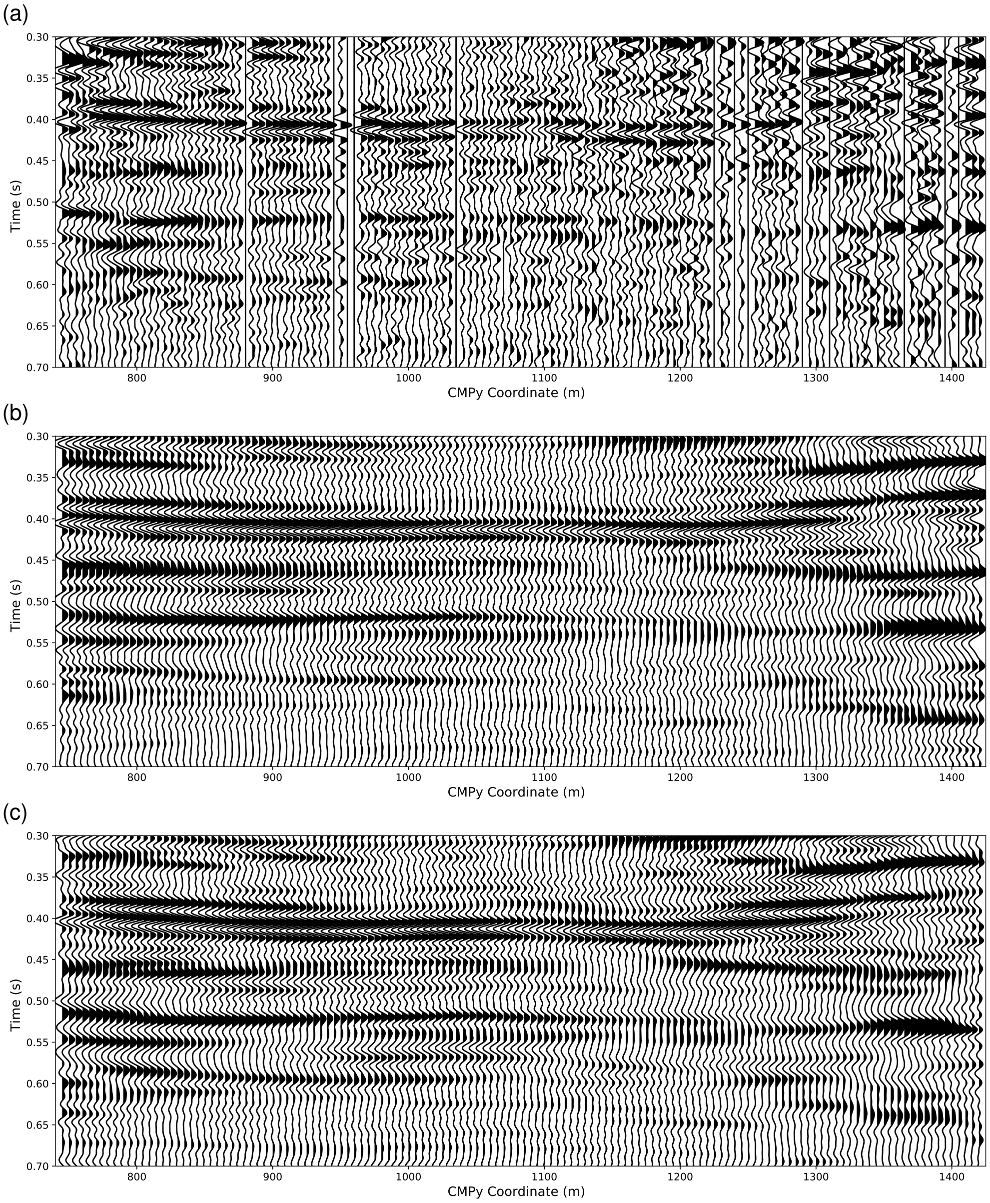

Figure 3.19: Reconstructed field seismic data for fixed CMPx= 20 in cross-spread 60. (a) Binned data, (b) binned data with SSA reconstruction, and (c) I-MSSA reconstruction.

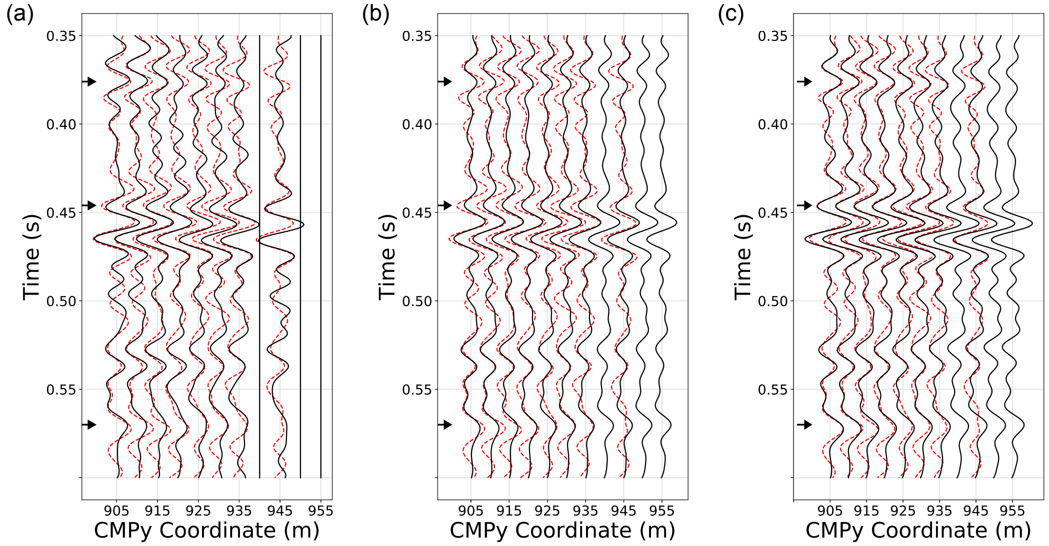

Figure 3.20: Detailed analysis of Figure 3.19. (a) Binned data. (b) Reconstruction via SSA. (c) Reconstruction via I-MSSA. All the figures show the observed traces in red, for reference. The arrows underline the events that are most affected by binning.

with MSSA reconstruction, and I-MSSA traces. We point out that the stack of the data before reconstruction was properly normalized by fold. Although the stacked volumes obtained via MSSA and I-MSSA are very similar, the improvement in the prestack volumes could be beneficial for attribute extraction and quantitative interpretation.

3.3 Discussion

Multichannel Singular Spectrum Analysis is an accurate and flexible reconstruction algorithm. Nonetheless, it requires regular spatial sampling of the input data. Then, binning is necessary to accommodate traces to the grid where one carries out the MSSA reconstruction. Optimal binning in complex data sets could result in errors in the input volume provided to the reconstruction algorithm. Also, some traces might be discarded or averaged when more than one trace falls in the same bin. I-MSSA, on the other hand, uses all the traces with their correct spatial coordinates.

The field examples show the effect of the exclusion of some traces from the recon-

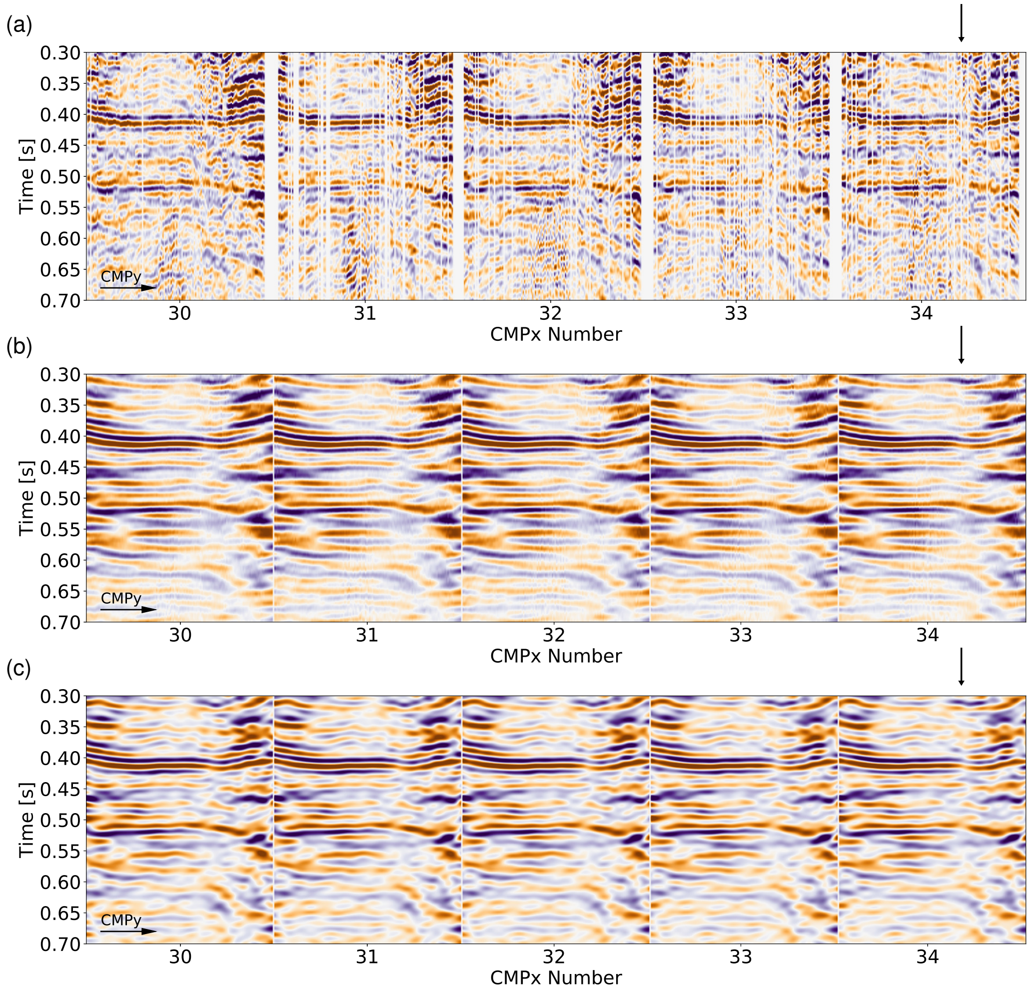

Figure 3.21: Reconstructed field seismic data in cross-spread 60. (a) Binned data for CMPx 30 to 34. (b) Binned data with MSSA reconstruction. (c) I-MSSA reconstruction. The arrow indicates the presence of a fault. The discontinuity is visible in the results obtained via I-MSSA.

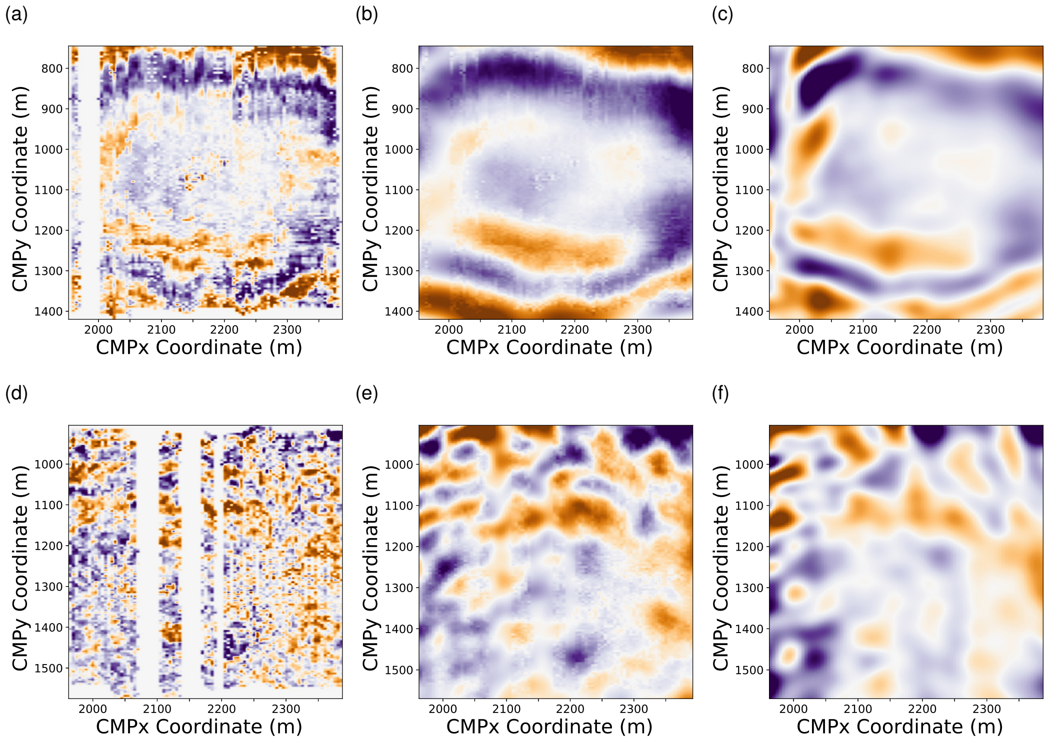

Figure 3.22: Time slices of the reconstructed seismic wavefield. Figures (a), (b), and (c) are binned data, MSSA reconstruction, and I-MSSA reconstruction for cross-spread 60 at time 0.4 s, respectively. Figures (d), (e), and (f) are binned data, MSSA reconstruction, and I-MSSA reconstruction for cross-spread 100 at time 0.7 s, respectively.

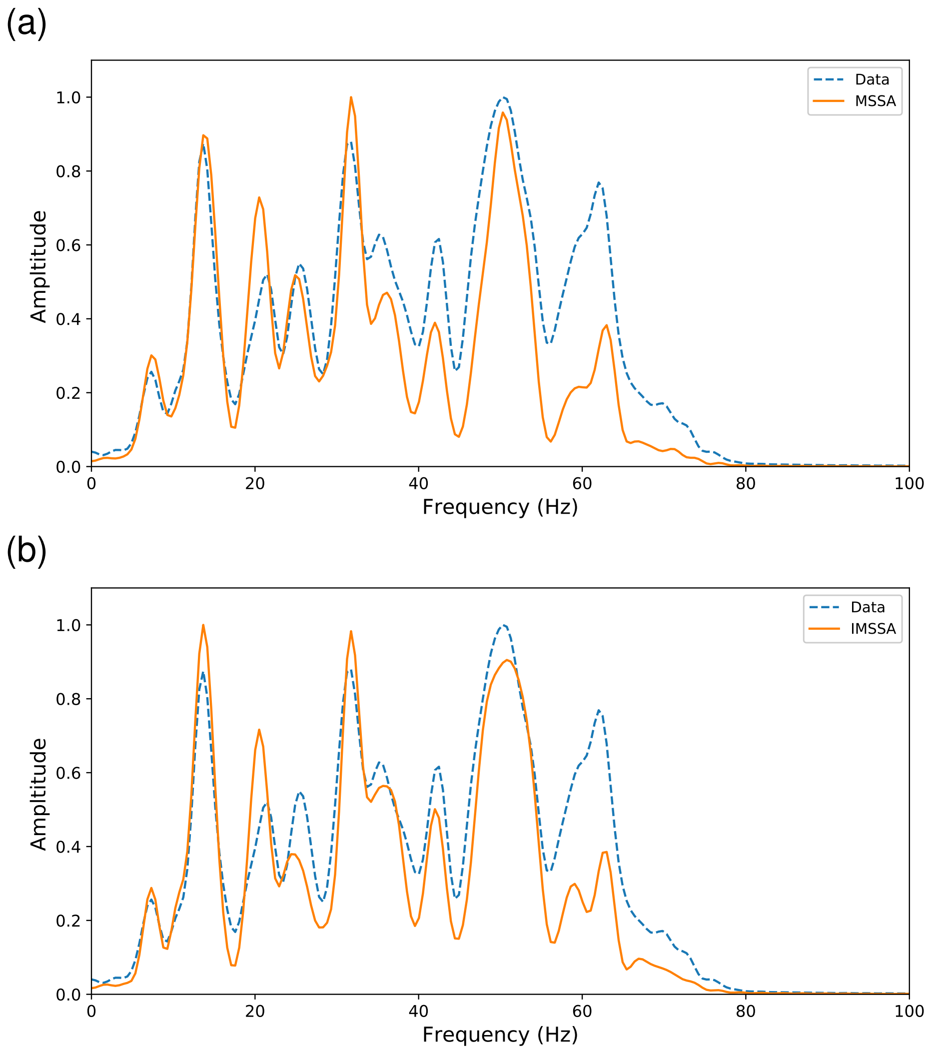

Figure 3.23: Comparison of amplitude spectrum for the data before and after reconstruction with (a) MSSA, and (b) I-MSSA algorithms.

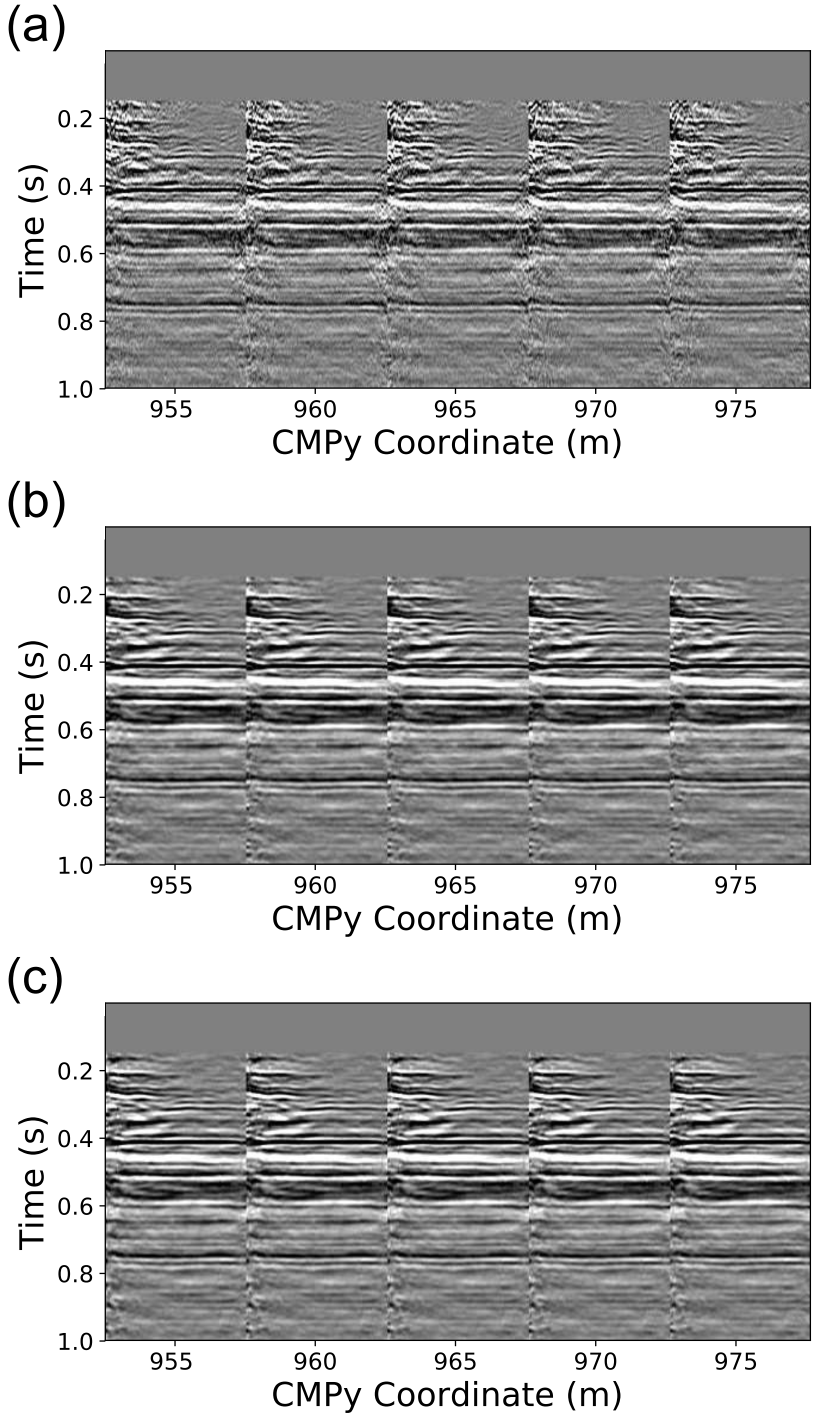

Figure 3.24: Stack for CMPy coordinate 955 to 975 from all cross-spreads. (a) Binned data, (b) binned with SSA reconstruction, and (c) I-MSSA reconstruction.

struction process. As I-MSSA considers the complete data set, it delivers reconstructed volumes with more continuous reflections. This chapter explores reconstruction in the cross-spread gathers domain. The efficiency of cross-spread gathers for prestack denoising and reconstruction has already been studied. For instance, Poole and Herrmann (2007), and Calvert et al. (2008) describe FXY deconvolution in cross-spread gathers, and Shaw et al. (2007) and Vermeer (1998) analyze 3D f-k filtering, among other methods for data enhancement. Similar to the conclusions of the previous authors, we observe that the spatial continuity of events in cross-spread gathers makes it an excellent domain for reduced-rank filtering and reconstruction techniques such as I-MSSA.

3.4 Conclusions

This chapter considers a modification of the MSSA algorithm to cope with irregularly sampled (off-the-grid) traces. The MSSA algorithm was initially proposed to denoise and reconstruct seismic traces organized on regular grids. A typical flow for MSSA involves applying trace binning to move observed coordinate positions to gridded positions. This process, usually called binning, is inherent to the MSSA algorithm that requires forming Hankel or Block Hankel matrices from regularly sampled data. The modification proposed in my work leads to a new algorithm that honours observed trace coordinates. We named the algorithm I-MSSA (Interpolated MSSA). The method minimizes an error function with a constraint given in the form of a projection. In this chapter, I consider the projection as the MSSA coherence pass filter. Then, what seemed a complicated optimization problem was reduced to a simple iterative algorithm where we adopt the Projected Gradient Descent method. The I-MSSA method can obtain solutions that honour exact data coordinates in a few iterations (typically eight to ten iterations).

Synthetic examples did not show significant gains in reconstruction quality when we compared I-MSSA to the conventional MSSA method applied to data after binning. The latter should not prevent users from adopting I-MSSA for prestack seismic data reconstruction. I-MSSA avoids constructing intermediate data volumes, often populated by empty bins, as usually done with classical MSSA reconstruction and with some Fourier reconstruction methods such as Minimum Weighted Norm Interpolation.

When examining the field data experiments, I noticed improvements in prestack volumes reconstructed via I-MSSA over those reconstructed via MSSA. The time slices of the individually reconstructed cross-spread gathers highlight these improvements. In the figures, I observed an increased continuity of events and attenuation of acquisition footprint noise when one adopts I-MSSA. However, I recognize that poststack volumes obtained by MSSA after binning and I-MSSA are very similar. The same conclusion is reported in the literature on prestack seismic reconstruction.

Despite achieving slightly notable improvements, I-MSSA via the Projected Gradient Descent method offers a novel way of posing reduce-rank denoising and reconstruction in cases where one wants to honour exact data coordinates. I-MSSA also addresses a pending problem in the literature on MSSA reconstruction, considering data deployed on irregular locations.

CHAPTER 4

Robust norms in seismic reconstruction

> *When a traveler reaches a fork in the road,*
> *the l_1-norm tells him to take either one way or the other,*
> *but the l_2-norm instructs him to head off into the bushes.*
> John F. Claerbout and Francis Muir

Chapter 3 describes the matrix completion technique as an inverse problem. The solution is the minimum of an objective function with a low-rank constraint, where the objective function measures the squared residuals between the model and the observed data. The approach corresponds to the classical l_2 norm criterion inversion. Such a model results in a convenient mathematical formulation of the problem, which justifies the widespread use of l_2 losses in geophysics.

However, the l_2 norm assumes that the measurement uncertainties follow a Gaussian distribution. In reality, this supposition is only approximately correct, as a small portion of the observations usually presents atypical deviations from the general pattern. Unfortunately, the l_2 criterion is highly sensitive to the presence of gross errors, even if they represent a slim percentage of the overall measurements.

Gaussian and erratic noise frequently pervade seismic data. Erratic noise is mainly present in onshore prestack data originating from misfired shots, poor coupling between the geophone and the surface, surface activity and others. This chapter aims to evaluate the impact of high amplitude noise in seismic reconstruction. I propose a robust alternative to reconstruction via an inverse problem that measures the misfit with a generalized loss function. The scheme iteratively estimates a robust approximation and evaluates a projection into the subset of possible solutions via the SSA filter.

This chapter has the following structure. First, it introduces M-estimators, the robust analogue of the maximum likelihood method. As a particular example of an M-estimator, I evaluate a generalized loss function that regulates its shape via a continuous parameter. The chapter continues with a review of the role of robust statistics in geophysics. Following, I propose a robust reconstruction approach based on the SSA filter. The innovation is the inclusion of the generalized loss function in the misfit. In this implementation, the minimum solution is calculated via the Projected Gradient Descent method (PGD). Finally, this chapter presents synthetic and real examples to evaluate the performance of the algorithm.

4.1 M-estimators

The data collected in many fields are contaminated by noise. Therefore, estimating the signal from the noisy measurements becomes a relevant topic in signal processing (Zhao et al., 2018). A classic approach to the problem considers the maximum likelihood method. The method captures the model under which the data has the highest probability of occurring (Brandt, 2014; Liano, 1996). If the samples have a normal distribution, their mean returns an optimal estimate of the real value (Maronna et al., 2019).

In reality, observations usually present *outliers*, atypical measurements that are well separated from the bulk of the data (Maronna et al., 2019). Examples of outliers are interference in radar and sonar systems or impulsive noise in mobile communications (Zoubir et al., 2012). A set of measurements with outliers results in a distribution with heavier tails than the ideal normal distribution (Tarantola, 2005; Amundsen, 1991; Claerbout and Muir, 1973; Scales and Gersztenkorn, 1988). In this case, classic estimates are suboptimal and fail to provide good fits to the data.

One approach to estimating a signal from sets of measurements with outliers is detection and deletion. There are many outlier detection methods, but they might not be practical for reconstruction purposes (Branham, 1986). In general, it is also true that for seismic data processing, detection and deletion methods are used, which, in general, eliminate a complete seismic trace rather than portions of it. On the other hand, the work presented in this chapter does not attempt to eliminate whole traces from processing but to weight them properly during a robust reconstruction process.

The generalization of maximum likelihood estimates for non-Gaussian distributions is called M-estimators (Carrillo et al., 2016). M-estimators are defined by the cost function of an optimization problem

$$\hat{\mu} = \min_{\mu} \sum \rho\left(\mu, x_i\right) \tag{4.1}$$

or by the implicit equation

$$\sum \psi\left(\mu, x_i\right) = 0\,, \tag{4.2}$$

where ψ is the derivative of the loss with respect to the unknown.

The robustness of M-estimators results from down weighting outliers outside a defined interval. Moreover, robust estimators show approximately optimal performance under slight perturbations of the model. Popular M-estimators include objective functions considering the l_1 norm (Lee et al., 2006), the l_1/l_2 norm (Bube and Langan, 1997), the Geman-McClure error criterion (Geman and McClure, 1985) and the Huber norm (Carrillo et al., 2016). More recently, Barron (2019) proposed a general function that includes most typical norms and pseudo-norms. The following paragraphs introduce the M-estimators mentioned above.

Median estimator

Let us consider a set of observations $x = \{x_1, \ldots, x_n\}$ of a given property of interest μ. One can assume that the observations are affected by noise via the linear model

$$x_i = \mu + u_i\,,\ i = 1, \ldots, n\,, \tag{4.3}$$

where the noise u_i in each observation is independent and identically distributed. The goal is to calculate an estimated $\hat{\mu}$ of the real value μ from the set of noisy observations x_i. This problem is called the location parameter estimation problem.

An approach to the problem is to minimize the sum of the squared residuals

$$\hat{\mu} = \min_{\mu} \sum_{i=1}^{n} (\mu - x_i)^2 \, . \tag{4.4}$$

One equates the partial derivative of the sum with respect to the unknown to zero, which results in an analytical expression for the estimate

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \, . \tag{4.5}$$

The estimator $\hat{\mu}$ is the sample mean or the arithmetic average of the data. If the observations have a normal distribution, the sample mean is the maximum likelihood estimator. It is clear from the definition that a single observation that is deviated from the bulk of the data has an unbounded influence on the mean. Then, the estimate is not robust to outliers.

Instead of minimizing the sum of the squared errors, one could minimize the summed absolute values

$$\hat{\mu} = \min_{\mu} \sum_{i=1}^{n} |\mu - x_i| \, . \tag{4.6}$$

If we set the partial derivative with respect to the unknown to zero, it results in

$$0 = \sum_{i=1}^{n} \mathrm{sgn}\left(\mu - x_i\right) \, . \tag{4.7}$$

Equation 4.7 defines the median of the samples. It sets $\hat{\mu}$ such that there is an equal number of observations exceeding and below the estimate. The median represents the maximum likelihood estimator of samples with a zero-mean Laplacian distribution. In the presence of an outlier, the median is not affected as much as the mean. In this sense, this is a robust estimator.

Considering the location parameter estimation problem and the definition of M-estimators (Equations 4.1 and 4.6), the loss function is $\rho(\mu - x_i) = |\mu - x_i|$, the l_1 norm (Lee et al., 2006). Its derivative, the ψ function, is $\psi(\mu - x_i) = \mathtt{sgn}\,(\mu - x_i)$. In the linear regression context, this solution is called the least-absolute deviation (LAD). The estimator is convex, piece-wise continuous, and non-differentiable in the origin. Then, it cannot be used by methods that require derivatives.

l_p estimator

The l_p function represents a family of estimators that include the mean and the median. Consider the set of observations $x = \{x_1, \ldots, x_n\}$. The estimator $\hat{\mu}$ is the minimum of the objective function

$$\hat{\mu} = \min_{\mu} \sum_{i=1}^{n} \frac{1}{\sigma_i^p} |\mu - x_i|^p \,, \tag{4.8}$$

where σ is a scale parameter, and $p > 0$ controls the shape of the modelled distribution. When $p = 2$, the l_p estimator yields the weighted mean for a Gaussian distribution. Conversely, $p = 1$ results in the weighted median estimator, or l_1 norm. If $0 < p < 1$, the model assumes very impulsive samples (Carrillo et al., 2016).

Note that, if $p > 1$, the l_p norm is convex and continuous everywhere. If $p = 1$, the function is piece-wise continuous but remains convex. When $0 < p < 1$, the norm is piece-wise continuous and non-convex.

The l_p estimator represents the maximum likelihood estimator of samples with zero-centred generalized Gaussian distributions and provides particularly robust estimations for impulsive noise applications. The lower the value of p, the more heavy-tailed is the distribution.

Myriad estimator

Let us now assume that the samples present a Cauchy distribution. If we consider the maximum likelihood method, it estimates the property of interest as

$$\hat{\mu} = \min_{\hat{\mu}} \sum_{i=1}^{n} \log[\sigma^2 + (\mu - x_i)^2] \,. \tag{4.9}$$

The equation defines the Least Lorentzian Squares criterion, and the solution is the myriad estimator. The scale parameter σ defines the robustness properties of the Lorentzian norm. The Lorentzian norm behaves as an l_2 cost function for variations that are small compared to $1/\sigma$. The function is convex in the interval $-1/\sigma \leq x_i \leq 1/\sigma$ and log-concave outside it. Further, the function is everywhere continuous and differentiable.

Huber estimator

M-estimates are defined by Equation 4.1 and do not necessarily represent a given distribution (Maronna et al., 2019). For example, the Huber estimator (Bube and Langan, 1997) combines the l_2 and l_1 norm in the cost function. Considering the definition of M-estimators, one can set ρ to the Huber's function

$$\rho\left(c,x\right)=\begin{cases} 1/2x^2; & |x|\le c \\ c|x|-1/2c^2; & |x|>c\,. \end{cases} \tag{4.10}$$

The Huber function is piece-wise continuous, convex, and differentiable. The tunning constant c modifies the degree of robustness of the estimator. Huber's function provides strong theoretical guarantees for the convergence of optimization problems. As a result, it is the most popular norm in robust statistics.

From Huber's function, one obtains the Huber estimator as

$$\psi\left(c,x\right)=\begin{cases} x & ;|x|\le c \\ c\,\mathrm{sgn}(x) & ;|x|>c\,. \end{cases} \tag{4.11}$$

The M-estimators in the limit cases when $c\to\infty$ and $c\to 0$ are the mean and the median.

Bi-square estimator

A different set of robust functions includes those with non-monotonic ψ functions. In this case, the loss function increases more slowly than Huber's function for large deviations. The bi-weight family of functions proposed by Beaton and Tukey (1974) is an example of such estimators

$$\rho\left(c,x\right)=\begin{cases} 1/6c^2\left\{1-\left[1-\left(x/c\right)^2\right]^3\right\} & ;|x|\le c \\ 1 & ;|x|>c\,, \end{cases} \tag{4.12}$$

where c is a tuning constant. The derivative of the loss function is

$$\psi(c,x)=\begin{cases} x\left[1-\left(\frac{x}{c}\right)^2\right]^2 & ;|x|\le c \\ 0 & ;|x|>c\,. \end{cases} \tag{4.13}$$

Note that ψ is everywhere differentiable and vanishes outside $[-c, c]$. Those M-estimators with ψ functions that tend to zero at infinity are called *redescending* estimators (Maronna et al., 2019). Redescending M-estimators offer an increase in robustness toward large outliers as they completely reject their effect on the estimator and fit the clean measurements. Other examples of redescending estimators are the Geman-McClure and the Welsch criterion.

A generalized M-estimator

It is common practice in robust statistics to interchange M-estimators when designing a model. Even though the characteristic provides flexibility to the approach, tunning specific parameters and evaluating convergence requires time and expertise. To overcome the hassle, Barron (2019) introduces a generalized M-estimator that encompasses most features of traditional robust estimators. The generalized objective function controls the robustness of the solution via a shaping parameter $\alpha \in \mathbb{R}$. The loss function emerges from the generalized Charbonnier loss (Sun et al., 2010; Barron, 2019). For practical implementations, the generalized loss function is

$$\rho\left(x, \alpha, c\right) = \frac{b}{d}\left(\left(\frac{\left(\frac{x}{c}\right)^2}{b} + 1\right)^{(d/2)} - 1\right), \tag{4.14}$$

where $b = |\alpha - 2| + \epsilon$, $d = \begin{cases} \alpha + \epsilon & \alpha \geq 0 \\ \alpha - \epsilon & \alpha < 0 \end{cases}$, $\epsilon = 10^{-5}$ is a small scalar to avoid singularities, and $c > 0$ is a scale parameter that controls the size of the loss quadratic bowl near $x = 0$. The inclusion of parameters b and d circumvents singularities and indeterminations.

In equation 4.14, when $\alpha = 2$, $\rho \to l_2$, which results in a non-robust estimator. For $\alpha = 1$, the loss is a smoothed form of Huber estimator, typically known as l_1/l_2 or pseudo-Huber norm. If $\alpha = 0$, equation 4.14 approximates Cauchy estimator. Setting $\alpha = -2$ results in the Geman-McClure estimator (Geman and McClure, 1985). In the limit, when $\alpha \to$-∞, the generalized norm approximates the Welsch estimator (Dennis Jr and Welsch, 1978). Table 4.1 presents the correspondence between traditional misfit criteria and the generalized measure of misfit for selected values of α.

Shape Parameter α	Criterion	$\rho(v = x/c)$
2	l_2	$\frac{1}{2}\lvert v\rvert^2$
1	l_1/l_2	$\sqrt{1+\lvert v\rvert^2}-1$
0	Cauchy	$\ln\left(1+\frac{1}{2}\lvert v\rvert^2\right)$
-2	Geman-McClure	$2\lvert v\rvert^2/\left(\lvert v\rvert^2+4\right)$
→-∞	Welsch	$1-\exp\left(-\frac{1}{2}\lvert v\rvert^2\right)$

Table 4.1: Identification of typical misfit criteria with the generalized loss function, for selected values of α.

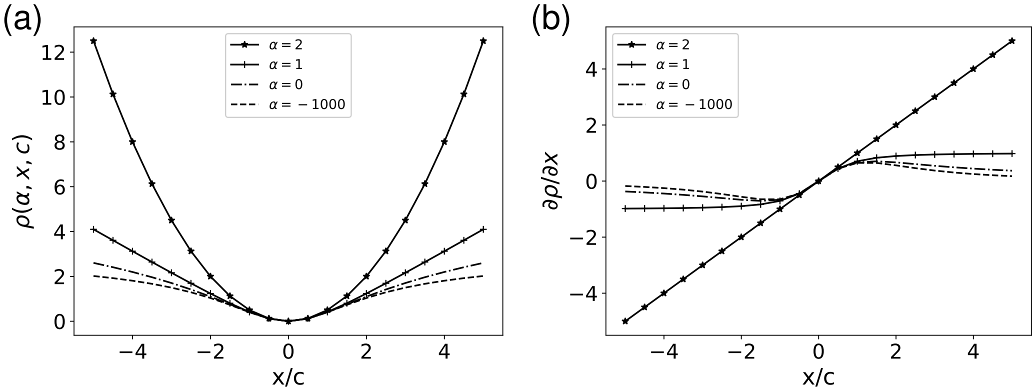

Figure 4.1: Robust metrics for several values of α. (a) Functional ρ . (b) Influence function.

Figure 4.1a shows the objective function for different values of α, revealing the nature of the robustness of M-estimators. The l_2 norm assigns the highest values to the most significant deviations. In other words, the cost function increases its value faster when $x \to \infty$. Conversely, robust estimators down weight the assignments to higher deviations.

Figure 4.1b depicts gradients or influence functions, which enable a possible classification of norms (Holland and Welsch, 1977). The influence function of the non-robust l_2 norm ($\alpha = 2$) is linear, increasing the influence of outliers on the results. As α decreases in equation 4.14, the effect of significant residuals on the gradient also decreases. If $\alpha = 1$, all outliers have the same weight, and the gradient is monotonic. Norms with monotonic influence functions are the l_1 norm, the hybrid l_1/l_2 norm, and the Huber norm. If $\alpha < 1$, the influence function decreases as the residuals increase. The effect of the outliers on the solution diminishes as its magnitude increases. These norms lead to redescending estimators. If the influence function is asymptotic to zero for atypical errors, the norm is a soft redescending function. Examples of these functions are the Cauchy norm and the Geman-McClure function. Finally, if the influence function is equal to zero for sufficiently large outliers, then the function is a hard redescending estimator. A popular hard redescending function in geophysics is the bi-weight norm (Beaton and Tukey, 1974).

The following sections describe the contribution of robust statistics in geophysics and present a robust seismic reconstruction method.

4.2 Robust statistics and its geophysical applications

Seismic traces are composed of reflection information contaminated by noise. Figure 4.2 presents an analysis of the density distribution of the samples corresponding to two noisy synthetic gathers. Figure 4.2a considers random noise with a signal-to-noise ratio of 1 dB. Plotted in red is the probability density function of a Gaussian or normal distribution. From the figure, one can conclude that the random component of the noise follows a Gaussian distribution. Figure 4.2b shows the histogram of traces affected by random and erratic noise. The histogram follows a heavier-tailed distribution than the Gaussian one, plotted in red. Surveys with heavy tail distributions require robust estimators.

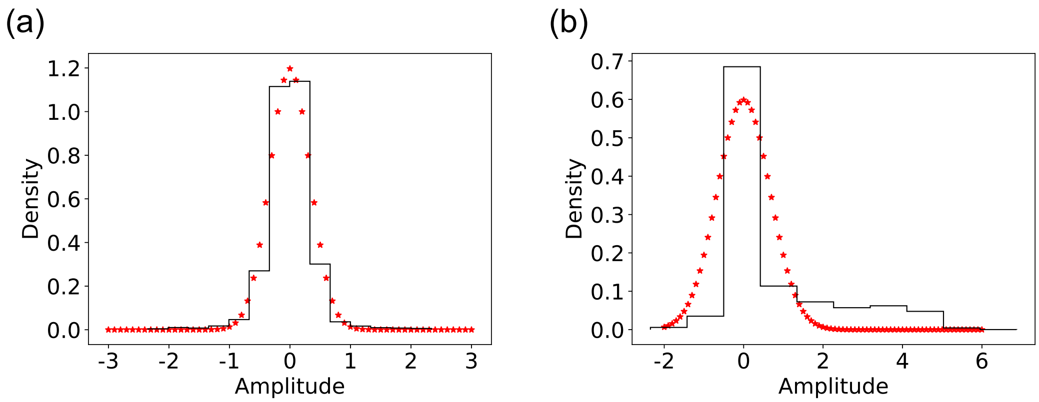

Figure 4.2: Density distribution of errors. (a) Random errors fit a Gaussian distribution. (b) Random and erratic errors cannnot fit the model due to the heavier tail of the samples.

In seismic data processing, many algorithms rely on the least-squares inversion to improve subsurface imaging. As seismic data usually presents erratic components, the least-squares approach is not optimal. Claerbout and Muir (1973) introduce the theory of robust inversion to seismic data. The authors discuss the design of predictive deconvolution filters for the denoising of multiples. In this context, the multiples represent the predictable component, and the primaries are the unpredictable component of the signal. Predictive deconvolution can separate the predictable component from the reflection series. By considering an absolute value criterion instead of the classic least-squares, the result is less affected by the unpredictable signal, and the primary reflections stand out. Taylor et al. (1979) study wavelet estimation via deconvolution. Their work considers an l_1 norm that reduces the shortcoming of the least-squares method. The method was later generalized to the l_p norms (Debeye and Van Riel, 1990; Gholami and Sacchi, 2012). However, it is important to stress that Taylor et al. (1979) and Debeye and Van Riel (1990) were specifically concerned with the retrieval of sparse models and not with robust inversion. The problem of retrieval of sparse models and robust estimation have strong commonalities because, in both cases, the resulting model or resulting residuals are sparse.

Travel-time tomography aims to reconstruct a velocity model of the subsurface from observed reflection travel times. It has critical applications in geophysics, but it is a computationally intensive inverse problem. Scales and Gersztenkorn (1988) frame the problem within the robust inverse theory and consider a solution via conjugate gradients and Iteratively Reweighted Least Squares (IRLS). Bube and Langan (1997) adopt an l_1/l_2 norm, which is computationally more tractable than the l_1 norm.

In the velocity analysis problem, Guitton and Symes (2003) estimate the velocity model parameters from the common midpoint gathers contaminated with non-Gaussian noise. The authors use the Huber norm combined with the L-BFGS algorithm, a quasi-Newton optimizer, to solve the inverse problem. Ji (2006) solves the robust inverse problem via a modified conjugate gradient method that incorporates weights to guide the gradient vector at each iteration. Ji (2012) includes a redescending M-estimator to the problem. The solution is obtained via IRLS with conjugate gradients.

A relatively new application to robust inversion is the Full Waveform Inversion (FWI) problem, a data-fitting procedure based on full-wavefield modelling. The main application of FWI is in calculating velocity models of the subsurface (Virieux and Operto, 2009). Crase et al. (1990) compare the inversion results with three robust norms, the l_1, the myriad norm, and the hyperbolic secant criterion. The authors use a preconditioned steepest descent to solve the linearized inverse problem. Brossier et al. (2010) report experiments with l_1, Huber, and l_1/l_2 norms.

In simultaneous-source data separation (also called the deblending problem), seismic data acquired with simultaneous sources Ibrahim and Sacchi (2014) can be separated via robust estimation of Radon coefficients. Radon algorithms for this purpose are based on the l_1 error loss and l_1 to promote sparsity of coefficients. In a similar vein, Lin and Sacchi (2020) apply the steepest descent method with a coherence pass operator defined by the robust Radon transform to obtain deblended, denoised data. Bahia et al. (2020) consider a robust Fourier thresholding method by imposing sparsity in the transform that describes the data and robustness on the misfit term.

Classical denoising and reconstruction algorithms cannot handle erratic noise. Trickett et al. (2012) were the first to acknowledge the effects of spatially erratic noise on rank-reduction filters. Chen and Sacchi (2015) present a denoising algorithm based on an objective function that minimizes the residuals between the observed data and an ideal low-rank matrix. The misfit follows Tukey's bi-weight norm. Sternfels et al. (2015) discuss a joint low-rank and sparse inversion convex optimization problem. In this approach, the low-rank data is separated from the sparse erratic noise via the joint minimization of a nuclear norm term and an l_1-norm term, constrained by the fit to the incomplete data. The problem is solved via ADMM. Cheng et al. (2015) and Li et al. (2019) solve the denoising problem via robust principal component analysis. Robust projection filters have also been proposed (Chen and Sacchi, 2017).

Robust methods play an essential role not only in the solution of geophysical inverse problems and in data preconditioning techniques. In both cases, the main goal is to minimize the influence of erratic noise on the resolution of an inverse problem or in the solution of denoising and data reconstruction problem.

4.3 Robust rank-reduction reconstruction

This chapter aims to estimate a complete, denoised seismic array $\mathbf{D}(t,m,n)\,,m=1\ldots,N_x\,,n=1\ldots,N_y$ from an incomplete set of observations, $\mathbf{U}(t,m,n)$. The observed and ideal data are sampled in the same grid to simplify the problem. The examples consider a 3D dataset, but the algorithm is easily adapted to higher dimensions. The method approximates a solution by optimizing the cost function that minimizes the difference between the data and the ideal traces

$$J=\|\mathbf{U}-\mathcal{S}\mathbf{D}\|_\rho\,, \tag{4.15}$$

where $\mathcal{S}$, the sampling operator, equals 1 if the sample was observed and 0 otherwise. I omitted indices to declutter notation. As I consider an adaptive approach, $\|\cdot\|_\rho$ represents a proper norm for the corresponding dataset. To guarantee convergence, I constraint the results. Hence, the problem is

$$\min J, \text{ subject to } \forall\omega:\mathbf{Z}_\omega\in\mathcal{T}\,, \tag{4.16}$$

where $\mathcal{T}=\{\mathbf{Z}_\omega:\text{rank}(\mathcal{H}[\mathbf{Z}_\omega])=k_\omega\}$. In the equation, $\mathcal{H}$ is the Hankelization operator applied to each frequency slice of $\mathbf{Z}$ and k_ω is the corresponding ideal rank.

Minimization of Equation 4.16 results in the following iterative algorithm

$$\begin{aligned}\mathbf{Z} &= \mathbf{D}^{\nu-1}-s^\nu\mathbf{W}\left(\mathbf{U}-\mathcal{S}\mathbf{D}^{\nu-1}\right)\\ \mathbf{D}^\nu &= \mathcal{P}[\mathbf{Z}]\,,\end{aligned} \tag{4.17}$$

where $\mathcal{P}=\mathcal{F}^{-1}\mathcal{A}\mathcal{R}\mathcal{H}\mathcal{F}$ represents the Singular Spectrum Analysis (SSA) filter applied independently to each frequency slice, $\mathcal{R}$ is the rank reduction operator, and $\mathcal{A}$ is the operator that averages along anti-diagonals (Oropeza and Sacchi, 2011). Finally, $\mathbf{W}$ is a diagonal matrix of weights

$$w(x, \alpha, c) = \left(\frac{|x/c|^2}{b} + 1 \right)^{(d/2-1)} ,$$

where b and d are determined by the selection of α (Equation 4.14).

Hence, the gradient of the objective function ∇J considers a matrix of weights that depends on the error. Such a matrix down weights atypical samples from the optimization with a level of rejection controlled by the parameter α.

Algorithm 3 shows the proposed robust reconstruction algorithm. I initialize the algorithm with the l_2 solution. As a stopping criterion, I consider the normalized Frobenius norm of the gradient of the cost function J. The algorithm stops when either $\|\nabla J^{\nu}\|/\|\nabla J^1\| < \eta$ or a maximum number of iterations `Max_Iter` is reached. To find the step size in equation 4.17, I used an inexact Backtracking Line Search (Nocedal and Wright, 2006).

Algorithm 3 Robust seismic reconstruction via Projected Gradients

```
Input: $\mathbf{u}$ (observed data), $p$ (rank), $\eta$, Max_Iter
$i = 1$
while $\|\nabla J^i\|/\|\nabla J^1\| \leq \eta$ and $i \leq$ Max_Iter do
    Descent direction: $\nabla J_\rho(\rho, \mathbf{D}, \mathbf{U}) = W(\mathcal{S}\mathbf{D} - \mathbf{U})$
    Step size: $s = \texttt{Backtracking}(\mathbf{D}, \nabla J)$
    Update: $\mathbf{Z} = \mathbf{D} - s\nabla J$
    Projection: $\mathbf{D} = \texttt{MSSA}(\mathbf{Z}, p)$
    $i \leftarrow i + 1$
end while
```

4.4 Examples

4.4.1 Synthetic examples

The synthetic examples consider a 3D volume with two linear events and high amplitude noise modelled as a Gaussian mixture. A mixture model considers samples from partitioned populations, where each subpopulation has a particular probability density (Brandt, 1998).

The $\mathtt{S/N}_i$ quantifies the noise in the input as

$$\mathtt{S/N}_i = \frac{\|\mathbf{d}_{true}\|_2^2}{\sigma_N^2}\,, \tag{4.18}$$

where $\mathbf{d}_{true}$ represents the clean signal, and σ_N^2 is the variance of the Gaussian mixture. Considering 0 mean subpopulations, the variance of the Gaussian mixture equals

$$\sigma_N^2 = \sum_{i=1}^{N} p_i * \sigma_i^2\,, \tag{4.19}$$

where N is the number of subpopulations, p_i is the probability of occurrence of a sample of the i-th subpopulation, and σ_i^2 is the variance of the i-th subpopulation.

This section presents two synthetic tests. The first test considers erratic noise randomly distributed in the volume. The erratic noise is a Gaussian mixture composed of two subpopulations with 0 mean. The first subpopulation has a dispersion $\sigma_1 = 0.01$ and a percentage of occurrence in the range $p_1 = [0.95 : 0.75]$. For the second subpopulation, σ_2 varies from 1 to 8. Following, I decimate 10% of the traces to examine denoising and reconstruction simultaneously. The ideal rank for the example is 3, and the step size s is approximated via the backtracking algorithm (Nocedal and Wright, 2006). The output signal-to-noise ratio quantifies the reconstruction quality of the method

$$\mathtt{S/N}_o[dB] = 10 * \log_{10} \frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{rec} - \mathbf{d}_{true}\|_2^2} \tag{4.20}$$

where $\mathbf{d}_{rec}$ represents the recovered data.

Figure 4.3 presents the results of the test. Figure 4.3a evaluates the reconstruction with $\alpha = 1$. The reconstruction was acceptable for a low percentage of erratic noise, with values of $\mathtt{S/N}_o$ above 10 dB. However, when the erratic noise was more frequent or presented higher amplitudes, the monotonic norm did not impose proper weights on the outliers. Figure 4.3b shows the result for $\alpha \to -\infty$. The redescending property of the influence function allowed for improved overall reconstruction. Figure 4.4 shows a x-slice of the studied volumes. Figure 4.4a shows the input data contaminated with high amplitude noise. In this case, $\sigma_2 = 2$ with 15% of erratic

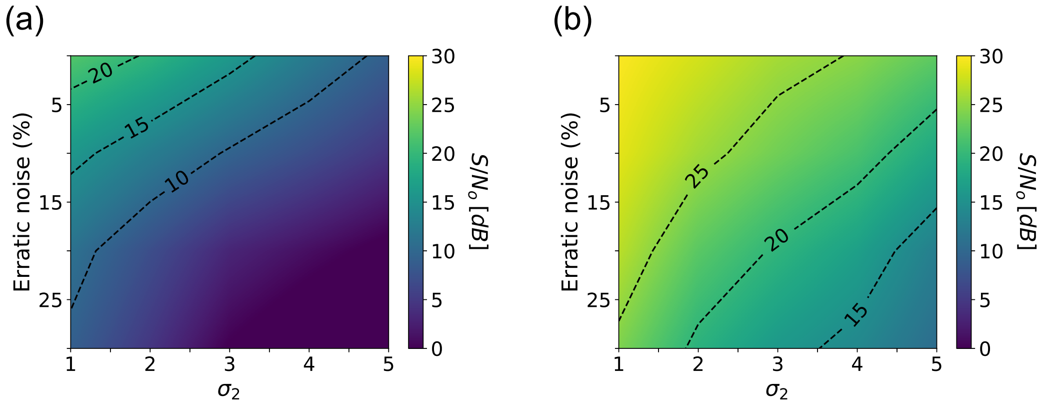

Figure 4.3: `S/N`$_o$ for a seismic volume affected with erratic noise. (a) Reconstruction with $\alpha = 1$. The robust norm results in proper reconstruction only for moderate erratic noise. (b) Reconstruction with α →-∞. A redescending norm results in improved reconstruction for increased presence of high-amplitude erratic noise.

samples. Figure 4.4b shows a slice of the ideal volume. Figures 4.4c and 4.4e show the results for $\alpha = 1$ and α →-∞, respectively. Finally, Figures 4.4d and 4.4f show the residuals between the ideal output traces and the reconstructed volume. As expected, the redescending norm results in overall improved reconstruction quality.

The following analysis considers the effect of the parameter α in the reconstruction. The test utilizes the same seismic volume as before, but with 25% of erratic samples. Figure 4.5 shows the `S/N`$_o$ as a function of parameter α. From the results, I conclude that the redescending influence functions improve the reconstruction capabilities in the presence of high amplitude noise.

The second test considers the presence of a percentage of erratic traces. The erratic noise presents $\sigma_1 = 0.1$ and σ_2 between 1 and 10. We modelled cases with a range of 10% to 80% of noisy traces. Besides, background noise affected all the traces with a `S/N`$_i$ = 1.2 dB and 10% of traces were decimated. Figure 4.6 presents the results of the reconstruction. As in the previous tests, $\alpha = 1$ results in proper reconstruction when the erratic noise is moderate. For traces with high amplitude noise, the algorithm required a redescending norm to obtain accurate results. I conclude that if more than 30% of traces are affected by high amplitude noise, the algorithm requires robust redescending norms for proper reconstruction.

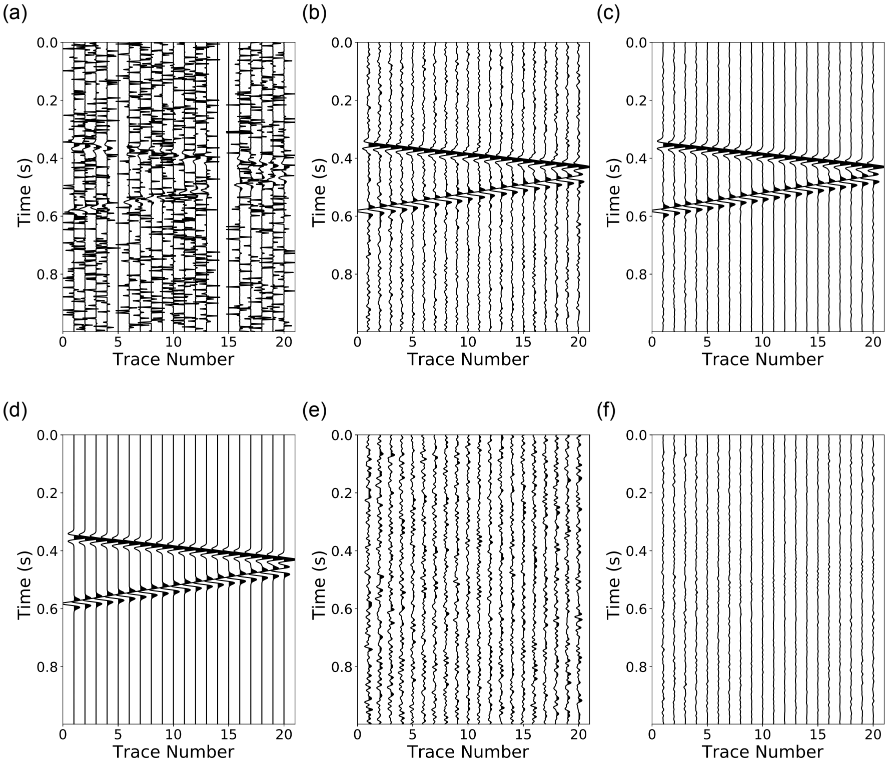

Figure 4.4: X-section of a 3D volume. (a) Input data with erratic noise. (b) Ideal output traces. (c) Reconstruction with $\alpha = 1$. (d) Residuals between ideal output and reconstruction. (e) Reconstruction with $\alpha \rightarrow$-∞. (f) Error between ideal output and reconstruction.

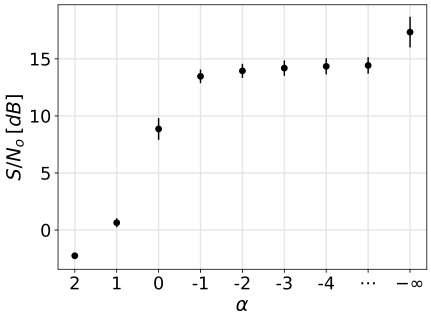

Figure 4.5: Reconstruction of a seismic volume with erratic noise with flexible selection of robustness. The parameter α determines the shape of the influence function, and therefore, its robustness to high amplitude errors. $\alpha = 2$ represents a non-robust solution. Decreasing values of α result in increasingly redescending robust norms.

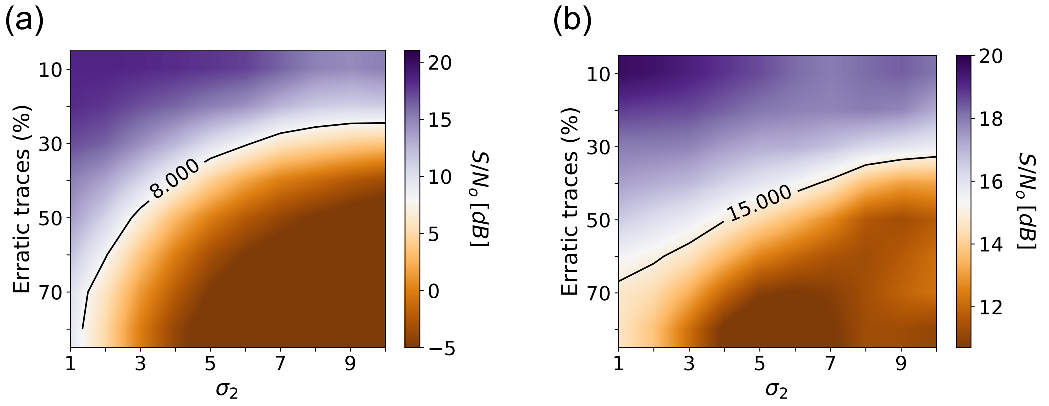

Figure 4.6: `S/N`$_o$ for a seismic volume with traces with erratic traces. (a) Reconstruction with $\alpha = 1$. The robust norm results in proper reconstruction only for a moderate number of erratic traces. (b) Reconstruction with $\alpha \rightarrow$-∞. A redescending norm results in improved reconstruction for increased presence of erratic traces.

Figure 4.7 shows a x-slice of the studied volumes. Figure 4.7a presents the input data contaminated with high amplitude noise, with $\sigma_2 = 9$ and 10% of erratic traces. Figure 4.7b shows a slice of the ideal volume. Figures 4.7c and 4.7d show the results for $\alpha = 1$ and $\alpha \rightarrow$-∞, respectively. Finally, Figures 4.7d and 4.7f show the residuals between the ideal output traces and the reconstructed volume. As expected, the redescending norm results in overall improved reconstruction quality.

4.4.2 Field data tests

The field test considers a 3D onshore data set from a heavy-oil field in the Western Canadian Sedimentary Basin. Sources and receivers correspond to an orthogonal survey. The data is NMO corrected and low-pass filtered. A filter was applied to remove high-frequency noise caused by field operations. I extract one cross-spread gather and add 10% of erratic traces. The reconstruct is on the CMP domain. Figure 4.8 presents the results. Figure 4.8a shows the data contaminated by high-amplitude noise. Figure 4.8b presents the non-robust reconstruction result. This panel shows that a robust loss function is required to reconstruct the data. Figures 4.8c and 4.8d show the robust reconstruction with $\alpha = 1$ and α =-1000, respectively. As the percentage of erratic traces is moderate, both results are acceptable.

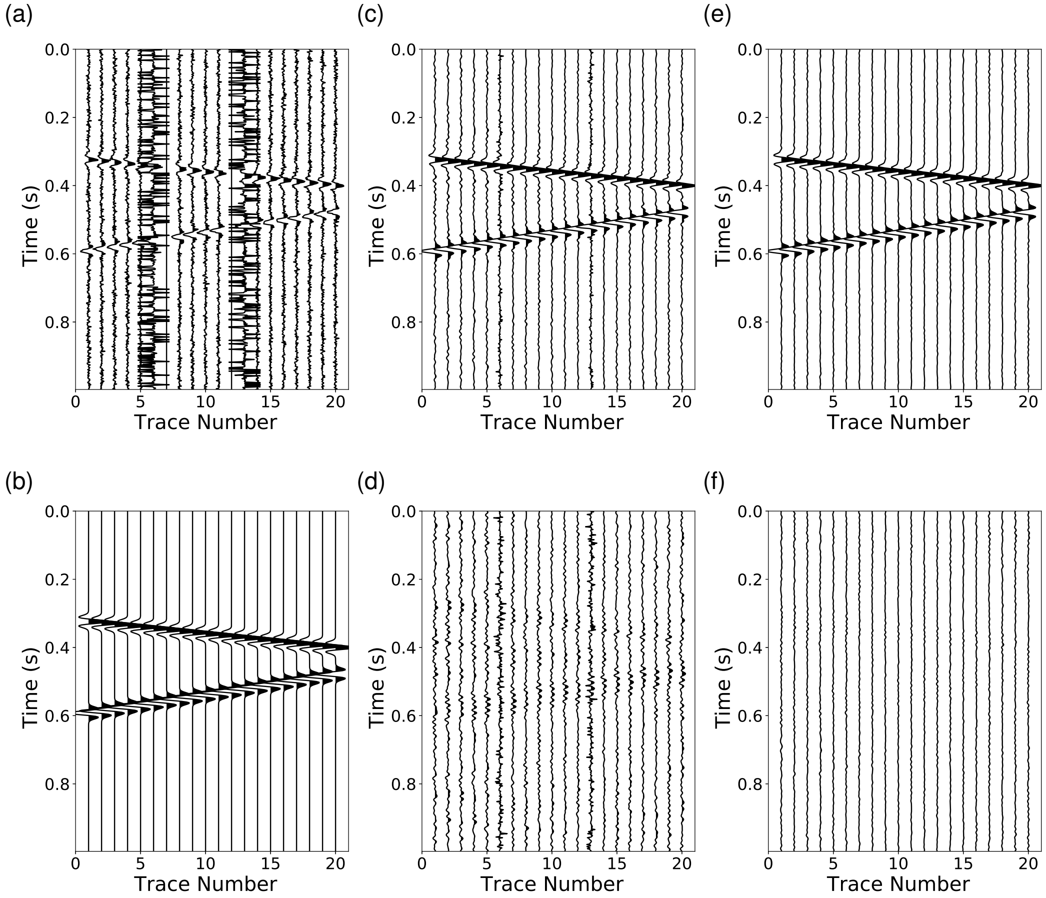

Figure 4.7: X-section of a 3D volume. (a) Input data with erratic traces. (b) Ideal output traces. (c) Reconstruction with $\alpha = 1$. (d) Residuals between ideal output and reconstruction. (e) Reconstruction with $\alpha \rightarrow$-∞. (f) Error between ideal output and reconstruction.

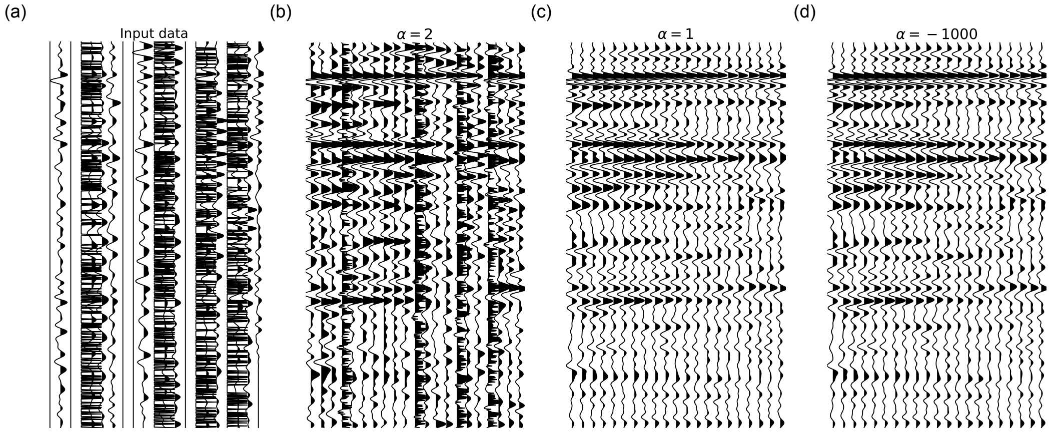

Figure 4.8: Reconstruction of a cross-spread gather. (a) Input data with erratic traces. (b) Non-robust reconstruction (c) Robust reconstruction with $\alpha = 1$. (d) Robust reconstruction with $\alpha \rightarrow$-∞.

4.5 Conclusions

This chapter introduces a robust algorithm for the reconstruction of seismic volumes via MSSA. The objective function of the problem is minimized via the Projected Gradient method. The misfit measures the discrepancy between the recorded data and the model via a generalized cost function controlled by one parameter, α. For $\alpha = 2$, the loss function reduces to the non-robust l_2 norm. Conversely, for $\alpha = 1$, the cost function is equivalent to a robust l_1/l_2 norm. Additionally, redescending estimators are attained for $\alpha \rightarrow$-∞.

The proposed robust MSSA was tested with synthetic and a field example. The $\texttt{S/N}_o$ quantified the quality reconstruction of the results. I conclude that a robust loss function is required for a reliable reconstruction of the data. For an increased presence of outliers, a redescending estimator ($\alpha \rightarrow$-∞) provides better reconstruction results.

CHAPTER 5

Robust Parallel Matrix Factorization

Multidimensional seismic data reconstruction has emerged as a primary topic of research in seismic data processing. Although there exists a large number of algorithms for multidimensional seismic data reconstruction, they often adopt the l_2 norm to measure the discrepancy between observed and reconstructed data. Strictly speaking, these algorithms assume well-behaved noise that ideally follows a Gaussian distribution. When erratic noise contaminates the seismic traces, a 5D reconstruction must adopt a robust criterion to measure the difference between observed and reconstructed data.

The goal is to propose a robust tensor-completion method for the reconstruction of 5D seismic volumes and to evaluate its output via synthetic and field-data examples. I adopt robust measures of misfit to minimize the influence of erratic noise on the reconstructed seismic volume. The main contribution is the addition of a robust misfit function into the Parallel Matrix Factorization (PMF) algorithm (Xu et al., 2015), an SVD-free tensor-completion technique for seismic-data reconstruction. Previous research in the area adopted a nonrobust PMF formulation based on a l_2 measure of fit (Gao et al., 2015, 2017). I also introduce the randomized QR decomposition in the rank-reduction stage of the PMF algorithm (Halko et al., 2011). The randomized QR decomposition simplifies parameter selection (rank) in the PMF algorithm (Chiron et al., 2014; Cheng and Sacchi, 2016).

This chapter is organized as follows. First, it provides a brief overview of the notation and definitions that are used throughout the article. Then, it formulates nonrobust and robust PMF tensor-reconstruction methods. Finally, it presents synthetic and field data examples.

5.1 Method

5.1.1 Prestack seismic volumes are tensors

I consider seismic data in the receiver-source domain. I denote the source coordinates by (s_x, s_y). Similarly, I denote the receiver coordinates by (g_x, g_y). Data acquired via a 3D seismic acquisition layout is often reorganized in offset-midpoint space using the transform

$$\begin{aligned} m_x &= \frac{s_x + g_x}{2}, \qquad & m_y &= \frac{s_y + g_y}{2} \\ h_x &= \frac{s_x - g_x}{2}, \qquad & h_y &= \frac{s_y - g_y}{2} \\ h &= \sqrt{h_x^2 + h_y^2}\,, \qquad & \phi &= \arctan(\frac{h_x}{h_y})\,, \end{aligned}$$

where (m_x, m_y) indicates the inline and crossline midpoint coordinate. The variables h and ϕ denote absolute offset (or offset) and azimuth (in radians), respectively. Similarly, inline and crossline offset are denoted by h_x and h_y, respectively. The data tensor is assembled by the following process. First, I read a seismic trace, and transform it to the frequency domain via the Fourier transform: $d(t, m_x, m_y, h, \phi) \leftrightarrow D(\omega, m_x, m_y, h, \phi)$. In this sense, I eliminate the time dependency obtaining, for one temporal frequency ω, a seismic volume that depends on four spatial coordinates. I also define the 4D desired acquisition geometry on a regular grid as follows

$$\begin{aligned} m'_x &= m_x^{min} + \Delta x\,(i_1 - 1)\,, \quad & i_1 &= 1, \ldots, I_1\,, \\ m'_y &= m_y^{min} + \Delta y\,(i_2 - 1)\,, \quad & i_2 &= 1, \ldots, I_2\,, \\ h' &= h^{min} + \Delta h\,(i_3 - 1)\,, \quad & i_3 &= 1, \ldots, I_3\,, \\ \phi' &= \phi^{min} + \Delta \phi\,(i_4 - 1)\,, \quad & i_4 &= 1, \ldots, I_4\,, \end{aligned}$$

where the coordinates (m'_x, m'_y, h', ϕ') have indexes (i_1, i_2, i_3, i_4) associated to them. The number of midpoints in the inline and crossline directions are I_1 and I_2, respectively. Similarly, the number of offsets and azimuths are given by I_3 and

I_4, respectively. Then, I assign the seismic trace to the closest bin with centre (m'_x, m'_y, h', ϕ') and indices (i_1, i_2, i_3, i_4). I continue reading traces to form the tensor $\mathcal{D}(\omega)_{i_1,i_2,i_3,i_4} \leftarrow D(\omega, m_x, m_y, h, \phi)$. After this process, often referred as binning, the final tensor of observations will contain empty grid points. In the situation that more than one trace lands on a given bin, I use simple averaging to assign one observation per bin. Finally, the tensor-reconstruction algorithm is applied to all frequencies $\omega \in [\omega_{min}, \omega_{max}]$. In other words, the regularization algorithm operates on 4D volumes represented by 4th-order tensors. To gain notational simplicity, I drop the dependency on frequency in the notation. Hence, for the remaining of the article, $\mathcal{D}(\omega)_{i_1,i_2,i_3,i_4} \equiv \mathcal{D}_{i_1,i_2,i_3,i_4}$. Finally, the reconstructed data are transformed back to the time domain. I remind the reader that the proper frequency-domain symmetry must be enforced before such transformation.

As a reminder, for each frequency, the regular data tensor $\mathcal{D}$ can be unfolded in 4 matrices or modes. The matrices $\mathbf{D}_k$, $k = 1, \ldots, 4$ obtained from the ideal noise-free and fully sampled tensor $\mathcal{D}$ are assumed low-rank. In this context, p_k refers to the rank of the kth unfolding, $\mathbf{D}_k$.Then, the tensor rank is defined via the vector (p_1, p_2, p_3, p_4).

Finally, a few words concerning the jargon used in the field of seismic-data reconstruction are in order. We adopt the name 5D reconstruction for the regularization of data that depend on four spatial coordinates and time (Liu et al., 2004; Trad, 2008). In other words, the input patch of data to reconstruct is a 5D volume. However, when the reconstruction is conducted in the frequency-space domain, the regularization algorithm operates on 4D volumes that are represented by 4th-order tensors.

5.1.2 Tensor reconstruction with the l_2-norm (Frobenius) misfit function: Nonrobust PMF method

The examples focus on 5D seismic-data reconstruction, where the completion algorithm is applied to 4th-order tensors in the frequency-space domain. However, the analysis is provided for the general case of tensors of order N. I adopt the following linear model to represent the observed data:

$$\mathcal{P} \circ \mathcal{Z} \approx \mathcal{D}, \tag{5.1}$$

where $\circ$ is the Hadamard's or element-wise product. The Nth-order tensor $\mathcal{Z}$ represents the desired complete data volume. The sampling operator $\mathcal{P}$ is an Nth-order tensor with entries given as follows:

$$\mathcal{P}_{i_1,i_2,i_3,\dots i_N} = \begin{cases} 0 & \text{if bin} \quad i_1, i_2, i_3, \dots i_N \quad \text{is empty} \\ 1 & \text{otherwise}\,. \end{cases} \tag{5.2}$$

Therefore, the problem as a linear system of equations where the goal is to estimate $\mathcal{Z}$, the complete data volume, from $\mathcal{D}$.

Furthermore, I assume Gaussian noise to model the mismatch in expression 5.1. Therefore, to solve for $\mathcal{Z}$, one needs to minimize the squared Frobenius norm of the error $\|\mathcal{E}\|_F^2 = \|\mathcal{P} \circ \mathcal{Z} - \mathcal{D}\|_F^2$ with respect to the unknown tensor $\mathcal{Z}$. The problem is underdetermined, there is an infinite number of solutions that can equally fit the data. The solution of the aforementioned problem requires a regularization constraint. In the PMF method, the regularization term is given by a low-rank constraint. I assume that the unfolded tensors $\mathbf{Z}_{(k)} = \texttt{unfold}_{(k)}\{\mathcal{Z}\}$ can be approximated by low-rank matrices (Xu et al., 2015; Gao et al., 2015). The retrieval of $\mathcal{Z}$ from $\mathcal{D}$ is then posed as a constrained minimization problem where the goal is to minimize the cost function

$$\begin{aligned} J &= J_M + \mu J_R \\ &= \frac{1}{\sigma^2}\|\mathcal{P} \circ \mathcal{Z} - \mathcal{D}\|_F^2 + \mu \sum_{k=1}^{N} \|\texttt{fold}_{(k)}\{\hat{\mathbf{Z}}_{(k)}\} - \mathcal{Z}\|_F^2 & \text{(5.3a)} \\ &= \frac{1}{\sigma^2}\|\mathcal{P} \circ \mathcal{Z} - \mathcal{D}\|_F^2 + \mu \sum_{k=1}^{N} \|\hat{\mathbf{Z}}_{(k)} - \texttt{unfold}_{(k)}\{\mathcal{Z}\}\|_F^2\,. & \text{(5.3b)} \end{aligned}$$

The term J_M is the misfit function. The scalar σ^2 is the variance of the noise. The second term of the objective function, J_R, is the regularization term. The matrices $\hat{\mathbf{Z}}_{(k)}$ denote the rank p_k approximation of the k-mode $\mathbf{Z}_{(k)} = \texttt{unfold}_{(k)}\{\mathcal{Z}\}$. The scalar $\mu > 0$, the trade-off parameter, controls the relative strength of the misfit term versus the low-rank regularization term. I have also written the regularization term in two equivalent expressions (equations 5.3a and 5.3b). The latter facilitates the derivation of the PMF algorithm.

The cost function contains $N + 1$ unknowns: $\mathcal{Z}$ and $\hat{\mathbf{Z}}_k$, $k = 1, \dots, N$. I pose the estimation of $\mathcal{Z}$ from $\mathcal{D}$ via an alternating minimization algorithm. I first solve for

$\mathcal{Z}$ considering known matrices $\hat{\mathbf{Z}}_k, k = 1, \ldots N$. The minimum of J is found by solving

$$\frac{\partial J}{\partial \mathcal{Z}^*} = 0\,, \tag{5.4}$$

where I have adopted complex differentiation to carry out the derivates of the cost function J with respect to unknowns (Brandwood, 1983; Chen and Sacchi, 2015) and obtained

$$\mathcal{Z} = (\mathcal{I} - \alpha\mathcal{P}) \circ \mathcal{C} + \alpha\mathcal{D}\,, \tag{5.5}$$

where $\mathcal{I}$ is an Nth-order tensor with all elements equal to one. The tensor $\mathcal{C}$ is given by

$$\mathcal{C} = \frac{1}{N}\sum_{k=1}^{N} \texttt{fold}_{(k)}\{\hat{\mathbf{Z}}_{(k)}\}\ , \tag{5.6}$$

and α is a parameter given by $\alpha = \frac{1}{1+N\mu\sigma^2}$. It is evident that the problem is nonlinear because equations 5.5 and 5.6 are coupled. In other words, the tensor $\mathcal{C}$ depends on low-rank matrices that are computed from the unknown $\mathcal{Z}$.

One also needs to solve for the unknown matrices $\hat{\mathbf{Z}}_k$ by considering $\mathcal{Z}$ known. The matrices $\hat{\mathbf{Z}}_k, k = 1, \ldots, N$ that minimize equation 5.3b are estimated by invoking the Eckart-Young-Mirsky theorem (Eckart and Young, 1936; Mirsky, 1960). In other words, $\hat{\mathbf{Z}}_k$ is the low-rank approximation of the matrix $\texttt{unfold}_{(k)}\{\mathcal{Z}\}$. The latter could be computed via the Singular Value Decomposition (Eckart and Young, 1936) or by any rank-reduction algorithm capable of approximating (in the Frobenius sense) a matrix by one of lower rank. The rank-reduction stage of the algorithm adopts a randomized QR decomposition (Halko et al., 2011; Chiron et al., 2014; Cheng and Sacchi, 2015, 2016). One can show that the random projection adopted by the randomized QR decomposition provides less tight bounds on the recovery of the original array than SVD (Liberty et al., 2007; Halko et al., 2011). Further detail on randomized QR decomposition is given in Appendix C.

Algorithm 1 summarizes the nonrobust PMF method where `Rank_Reduction` is a user-supplied function to perform the reduced-rank approximation.

5.1.3 Robust PMF method

The preceding section showed the derivation of the PMF tensor-completion iterative algorithm with the l_2 misfit function. The l_2 misfit norm of the error tensor is

Algorithm 4 Low-Rank tensor completion with l_2 misfit

```
Input: $\mathcal{D}, \mathcal{P}, \mu, \sigma^2, p_k, maxiter, tol$
$i = 1, \mathcal{Z}^i = \mathcal{D}, \alpha = \frac{1}{1+N\mu\sigma^2}$
while $\|\mathcal{Z}^{i+1} - \mathcal{Z}^i\|_F^2 \,/\, \|\mathcal{Z}^i\|_F^2 \leq tol \;\&\; i \leq maxiter$ do
    for $k = 1 : N$ do
        $\mathbf{Z}^i_{(k)} = \texttt{unfold}_{(k)}\{\mathcal{Z}^i\}$
        $\hat{\mathbf{Z}}^i_{(k)} = \texttt{Rank\_Reduction}\{\mathbf{Z}^i_{(k)}, p_k\}$
    end for
    $\mathcal{C}^i = \frac{1}{N}\sum_{k=1}^{N} \texttt{fold}_{(k)}\{\hat{\mathbf{Z}}^i_{(k)}\}$
    $\mathcal{Z}^{i+1} = (\mathcal{I} - \alpha\mathcal{P}) \circ \mathcal{C}^i + \alpha\mathcal{D}$
    $i \leftarrow i + 1$
end while
```

nonrobust. This section redefines the misfit by adopting a robust misfit function. I consider the l_1/l_2 norm, a differentiable approximation to the well-known l_1 norm. The l_1/l_2 misfit function (Bube and Langan, 1997; Lee et al., 2006) is

$$\begin{aligned} J_M &= \sum_{i_1,i_2,\ldots,i_N} \rho(\mathcal{E}_{i_1,i_2,\ldots,i_N}) \\ &= \sum_{i_1,i_2,\ldots,i_N} \sqrt{1 + |\mathcal{E}_{i_1,i_2,\ldots,i_N}|^2/\sigma^2}\ , \end{aligned} \tag{5.7}$$

where it is easy to see that $\rho(x) \approx |x|$ when σ is small. The scalar σ^2 is now the scale parameter associated with the distribution that induces the measure of the misfit. If we replace the Frobenius norm in equation 5.3a by equation 5.7 and take derivatives with respect to the unknown variables $\mathcal{Z}$, we obtain the robust PMF algorithm

$$\mathcal{Z} = (\mathcal{I} - \mathcal{A} \circ \mathcal{P}) \circ \mathcal{C} + \mathcal{A} \circ \mathcal{D}\ , \tag{5.8}$$

where $\mathcal{A}$ is a tensor with elements given by

$$\mathcal{A}_{i_1,i_2,i_3,\ldots i_N} = \frac{1}{1 + N\mu\sigma^2\sqrt{1 + |\mathcal{E}_{i_1,i_2,i_3,\ldots i_N}|^2/\sigma^2}}\ , \tag{5.9}$$

with $\mathcal{E} = \mathcal{P} \circ \mathcal{Z} - \mathcal{D}$ and $\mathcal{C}$ is given by equation 5.6. Equivalently to the nonrobust PMF method, I have adopted the randomized QR decomposition to compute the low-rank matrices $\hat{\mathbf{Z}}_{(k)}$ required to estimate $\mathcal{C}$. Algorithm 2 provides the robust PMF algorithm where `Evaluate_Weights` is the function that evaluates the tensor

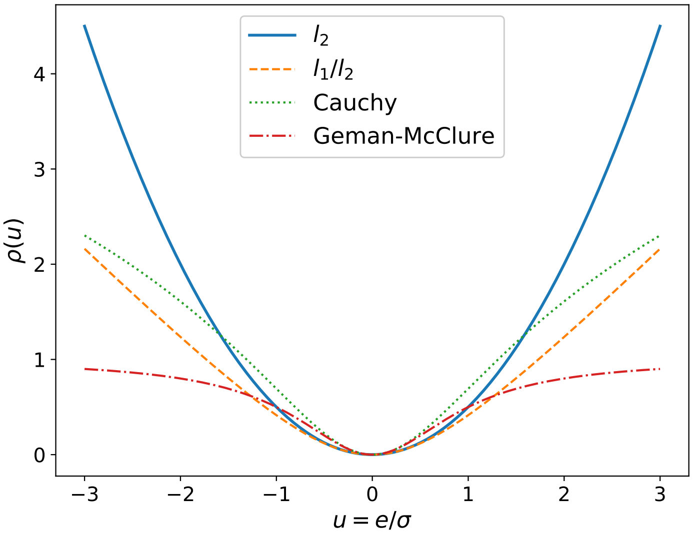

Figure 5.1: Functionals ρ that can be adopted to measure the error discrepancy in the PMF algorithm. The expressions for $\rho(u)$ are given in Table 5.1.

of weights (see Table 5.1).

Notice the similarity between the nonrobust and robust PMF methods. The tensor of weights $\mathcal{A}$ can be replaced by the constant α to obtain the nonrobust PMF algorithm (equation 5.5). In the robust PMF method, the tensor of observations $\mathcal{D}$ is weighted by the tensor $\mathcal{A}$. The weights depend on the magnitude of the error $\mathcal{E}$ evaluated at the preceding step of the iterative algorithm. In essence, the outliers are de-emphasized by the product $\mathcal{A} \circ \mathcal{D}$. The elements of the tensor of weights satisfy $0 < \mathcal{A}_{i_1,i_2,i_3,\dots i_N} < 1$. The condition also applies for the nonrobust case where $0 < \alpha < 1$ for trade-off parameter $\mu > 0$.

In the derivation, I have selected the l_1/l_2 robust measure of the misfit. Similar results can be obtained by adopting the Cauchy (Sacchi and Ulrych, 1995) or the Geman-McClure (Geman and McClure, 1985) criteria. Table 5.1 offers the function $\rho(u)$ and the associated weights $\mathcal{A}$ for the l_2-norm misfit and for the robust l_1/l_2, Cauchy, and Geman-McClure error norms. Figure 5.1 shows the functional $\rho(u)$ for the l_2-norm, the l_1/l_2 norm, and the Cauchy and Geman-McClure criteria.

Estimator	Misfit Criterion	$\rho(v)$	Weights $\mathcal{A}(u)$
Nonrobust	l_2	$\|u\|^2$	$\frac{1}{1+N\mu\sigma^2}$
Robust	l_1/l_2	$\sqrt{1+\|u\|^2}$	$\frac{1}{1+N\mu\sigma^2\sqrt{1+\|u\|^2}}$
	Cauchy	$ln(1+\|u\|^2)$	$\frac{1}{1+N\mu\sigma^2(1+\|u\|^2)}$
	Geman-McClure	$\frac{\|u\|^2}{1+\|u\|^2}$	$\frac{1}{1+N\mu\sigma^2(1+\|u\|^2)^2}$

Table 5.1: Nonrobust and robust misfit criteria adopted for PMF, $u = e/\sigma$.

Algorithm 5 Low-Rank tensor completion with robust misfit

```
Input: $\mathcal{D}, \mathcal{P}, \mu, \sigma, p_k, maxiter, tol$
$\mathcal{Z}^i = \mathcal{D}, i = 1$
while $\|\mathcal{Z}^{i+1} - \mathcal{Z}^i\|_F^2 \,/\, \|\mathcal{Z}^i\|_F^2 \leq tol$ & $i \leq maxiter$ do
    for $k = 1 : N$ do
        $\mathbf{Z}^i_{(k)} = \texttt{unfold}_{(k)}\{\mathcal{Z}^i\}$
        $\hat{\mathbf{Z}}^i_{(k)} = \texttt{Rank\_Reduction}\{\mathbf{Z}^i_{(k)}, p_k\}$
    end for
    $\mathcal{C}^i = \frac{1}{N}\sum_{k=1}^{N} \texttt{fold}_{(k)}\{\hat{\mathbf{Z}}^i_{(k)}\}$
    $\mathcal{E}^i = \mathcal{D} - \mathcal{P} \circ \mathcal{Z}^i$
    $\mathcal{A}^i = \texttt{Evaluate\_Weights}\{\mathcal{E}^i, \mu, \sigma\}$
    $\mathcal{Z}^{i+1} = \big(\mathcal{I} - \mathcal{A}^i \circ \mathcal{P}\big) \circ \mathcal{C}^i + \mathcal{A}^i \circ \mathcal{D}$
    $i \leftarrow i + 1$
end while
```

5.2 Examples

5.2.1 Synthetic data

This section evaluates the proposed robust PMF tensor-reconstruction method and compares the robust solution to the nonrobust PMF solution (Gao et al., 2015). To do so, I created two volumes of spatial data of size 12 × 12 × 12 × 12 traces and 500 time samples with a time-sampling rate of 2 ms. The first volume was composed of a superposition of events with linear moveout. The second volume consisted of events with parabolic traveltime curves. For all the synthetic experiments, I set a convergence tolerance $tol = 0.004$ and a maximum number of iterations $maxiter = 50$ (see Algorithms 1 and 2). The band of frequencies used for reconstruction is 1 to 70 Hz. I evaluate the reconstruction quality via the signal-to-noise ratio of the solution (or output) in decibels:

$$\texttt{S/N}_o[dB] = 10 \log_{10} \frac{\|\mathbf{d}_{true}\|_F^2}{\|\mathbf{d}_{rec} - \mathbf{d}_{true}\|_F^2} \ , \tag{5.10}$$

where $\mathbf{d}_{true}$ and $\mathbf{d}_{rec}$ indicate the true volume prior to decimation and contamination with noise and the reconstructed volume, respectively. Note that, even though the PMF algorithm is executed in the frequency-space domain, the final $\texttt{S/N}_o$ is computed with true and reconstructed seismograms in the time domain.

Volume of linear events

The first volume used to test the robust tensor-completion PMF algorithm is composed of three events with linear moveout. The linear events have been synthesized using three waves with ray parameters in each spatial dimension given by $q_1 = [-0.001, 0.002, -0.003]$ s/m, $q_2 = [0.0005, 0.001, 0.0001]$ s/m, $q_3 = [-0.001, 0.001, 0.001]$ s/m, and $q_4 = [0.002, -0.0005, 0.001]$ s/m. The first and second spatial dimensions are CMPx and CMPy, respectively. Similarly, the third and fourth spatial dimensions are offset and azimuth. The noise is modeled via a Gaussian mixture that corresponds to a $\texttt{S/N}_i = 1.2$ dB. The signal-to-noise ratio, $\texttt{S/N}_i$, is defined by the ratio of the power of the clean signal to the variance of the Gaussian mixture. We opt for a Gaussian mixture to mimic erratic non-Gaussian noise. For this purpose, we mix two series of Gaussian deviates with standard error 0.001 and 10.0 with a probability of occurrence $Pr = 0.9$ for the series with the smallest standard deviation. By this means, 10% of the samples of the simulated noise are large outliers, and the remaining are small additive noise. To further complicate the example, I add traces mimicking sinusoidal noise of central frequency 20 Hz. The sinusoidal traces are randomly distributed in the spatial volume, and they represent 5% of the total volume. The volume corrupted with noise is then randomly decimated by an amount of 60%. The total number of traces before decimation is $12^4 = 20736$. The total number of traces after decimation is about 8300.

The trade-off parameter $\mu = 1$ is chosen by experimenting with different values and by picking the value that generates a minimum amount of signal leakage in the error panel. I use the same trade-off parameter μ for both the l_2 nonrobust PMF and robust PMF solutions. Given that the data is contaminated with non-Gaussian noise, one cannot find a trade-off parameter that properly minimizes leakage when adopting the l_2 norm. For this particular example, I also set $p_k = 7,\ k = 1, 2, 3, 4$ where p_k is the number of random projections of the randomized QR decomposition for mode k (Appendix C).

For the robust PMF method, the parameter σ is selected via the following heuristic expression $\sigma = 0.0001 \times \|\mathcal{E}^1(\omega)\|_F$, where $\mathcal{E}^1(\omega)$ is the error at first iteration. In other words, σ is frequency-dependent. One could have also adopted a test of goodness of fit to determine μ as in Gao et al. (2017). However, developing such a criterion is not an easy task for data contaminated with non-Gaussian noise.

Figure 5.2a shows the data for CMPy 2, 5, 8, and 10, and CMPx in the range 1 to 12. Offset and azimuth are fixed. Figure 5.2b shows the same traces after contamination with noise and decimation. The nonrobust l_2 reconstruction is portrayed by Figure 5.2c. Similarly, the robust PMF reconstruction is provided in Figure 5.2d. The corresponding error panels can be found in Figures 5.2e and f, respectively.

Figure 5.3 shows results for the same volume but offset 2, 6, 9, and 12 and CMPx in the range 1 to 12. In this case, CMPy and azimuth are fixed. Figures 5.3a and b show the ideal traces and the same traces after contamination with noise and decimation. Figures 5.3c and e show the nonrobust reconstruction and the error, respectively. Similarly, the robust PMF reconstruction is provided in Figure 5.3d, and the error in Figure 5.3f.

It is clear from the previous figures that the nonrobust PMF method is unable to reproduce the data accurately, then signal leakage is observed in the error panels (Figures 5.2e and 5.3e). On the other hand, the robust PMF algorithm properly reconstructs the seismic volume. The reconstruction quality for the nonrobust PMF algorithm is $\texttt{S/N}_o = 7$ dB, whereas the reconstruction quality for the robust PMF algorithm is $\texttt{S/N}_o = 14$ dB.

Volume of parabolic events

Finally, the synthetic experiment is repeated for parabolic events. In this case, the data has four parabolic events with residual move-out at the last trace of each dimension given by $q_1 = [0.08, 0.02, 0.12, -0.04]$ s, $q_2 = [0.04, 0.02, 0.04, -0.06]$ s, $q_3 = [0.04, 0.01, 0.016, -0.02]$ s, and $q_4 = [0.01, 0.04, 0.02, -0.08]$ s. Erratic noise was generated using the parameters described in the preceding example, and the data is decimated by 60%. I also adopt $\mu = 0.3$ for the nonrobust and the robust algorithm. Last, in both cases, I adopt $p_k = 8$, $k = 1, 2, 3, 4$ where p_k is the number of random projections of the randomized QR decomposition for mode k.

Figure 5.4a shows the data for CMPy 2, 3, 6, and 12 and CMPx in the range 1 to 12. Offset and azimuth are fixed. Figure 5.4b shows the same traces after contamination with noise and decimation. The nonrobust l_2 reconstruction is portrayed by Figure 5.4c. Similarly, the robust PMF reconstruction is displayed in Figure 5.4d. The reconstruction-error volumes are portrayed in Figures 5.4e and f.

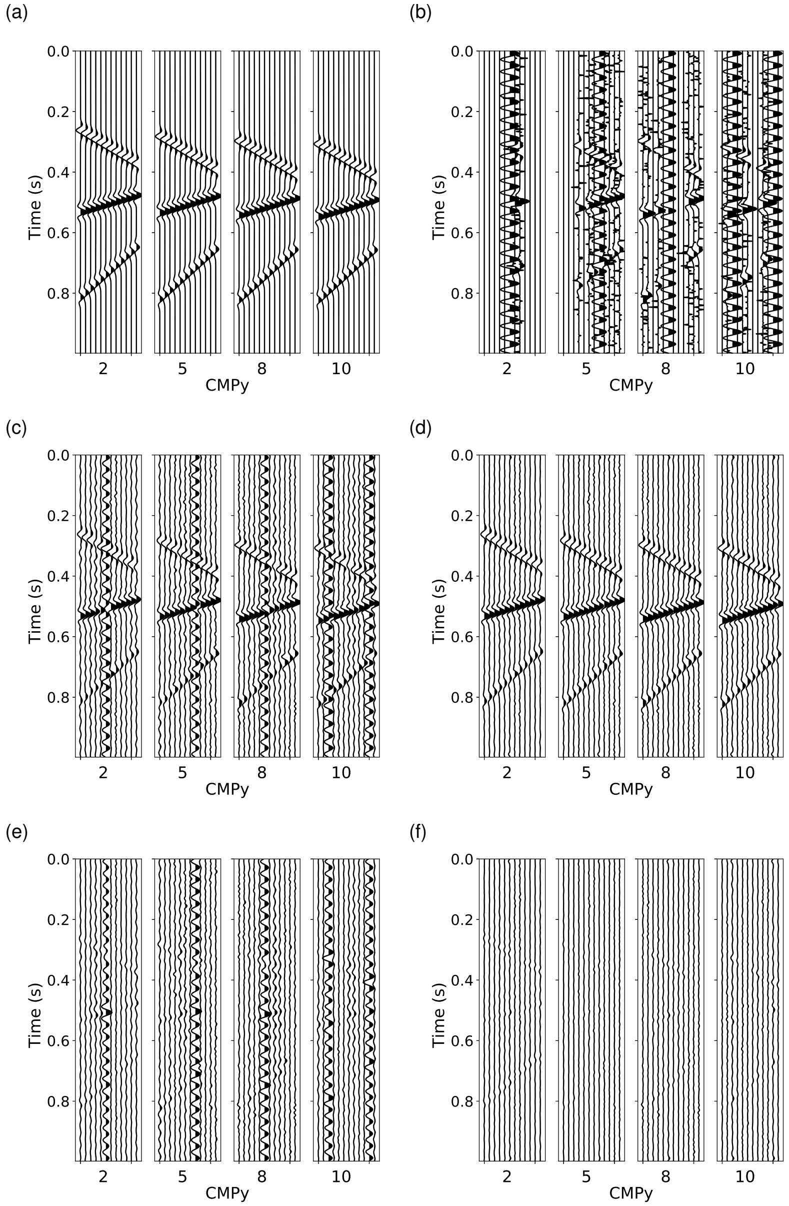

Figure 5.2: Reconstruction of linear events with data contaminated with erratic noise, and sinusoidal noise. (a) Original data. (b) Data corrupted with erratic noise and sinusoidal noise and decimated by 60%. (c) Nonrobust l_2 PMF reconstruction. $\texttt{S/N}_o = 7$ dB. (d) Robust PMF reconstruction. The l_1/l_2 norm was adopted as a robust measure of misfit. $\texttt{S/N}_o = 14$ dB. (e) Error associated with the nonrobust solution. (f) Error associated with the robust solution.

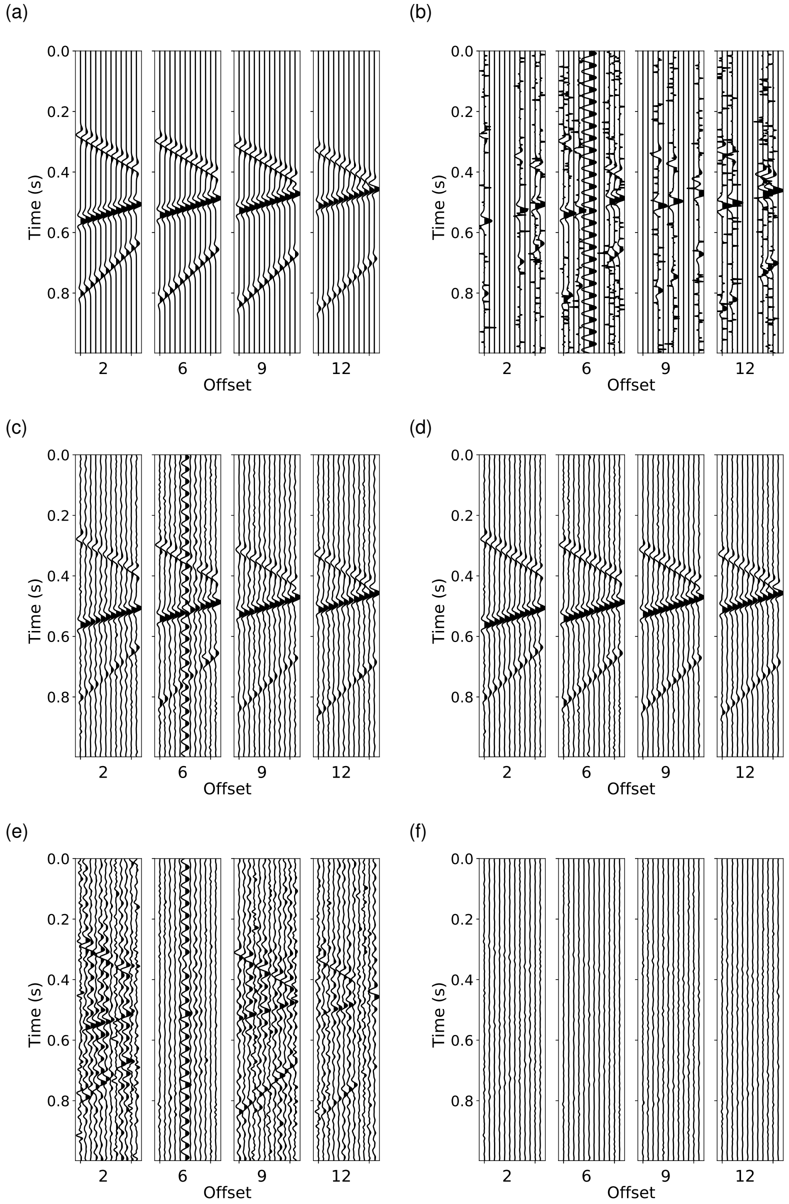

Figure 5.3: Reconstruction of linear events with data contaminated with erratic noise, and sinusoidal noise. (a) Original data. (b) Data corrupted with erratic noise and sinusoidal noise and decimated by 60%. (c) Nonrobust l_2 PMF reconstruction. $\texttt{S/N}_o = 7$ dB. (d) Robust PMF reconstruction. The l_1/l_2 norm was adopted as a robust measure of misfit. $\texttt{S/N}_o = 14$ dB. (e) Error associated with the nonrobust solution. (f) Error associated with the robust solution.

Figure 5.5 shows the same 5D volume, but for offsets 2, 3, 9, and 12, and CMPx in the range 1 to 12. In this case, CMPy and azimuth are fixed. Figures 5.5a and b show the ideal and noisy traces. Figures 5.5c and e show the nonrobust reconstruction and the corresponding errors. Finally, Figures 5.5d and f show the robust reconstruction with the respective error panel.

From the previous figures, it is clear that the nonrobust PMF method is unable to reconstruct the data correctly, and a signal leakage is observed in the error panel (Figures 5.4e and 5.5e. For the nonrobust PMF algorithm, the best reconstruction quality that one can attain is $\mathtt{S/N}_o = 16$ dB. The reconstruction quality for the robust PMF method is $\mathtt{S/N}_o = 20$ dB.

Tests are also carried out with Cauchy and Geman-McClure error criteria (Table 1). Results obtained with these criteria are similar to those obtained via the l_1/l_2 misfit.

5.2.2 Field-data test

I also test the nonrobust and robust PMF algorithms on a 5D onshore seismic dataset. The data was acquired to monitor a heavy-oil field in Alberta, Canada. Sources and receivers were distributed following an orthogonal survey. Their locations are shown in Figure 5.6. The NMO correction was applied to the data prior to reconstruction to avoid spectral wrapping in the frequency-wavenumber domain (Gao et al., 2015). Moreover, a low-pass filter is applied to remove high-frequency noise caused by field operations. The cutoff frequency of the low-pass filter is 100 Hz. The data was binned into a 4D spatial tensor by considering the following geometry:

$\Delta x = 5$m, total number midpoints in the inline direction = 275

$\Delta y = 5$m, total number midpoints in the crossline direction = 161

$\Delta h = 100$m, number of offset sectors $I_3 = 6$

Offset sector 1, $h[m] \in [50, 150)$
Offset sector 2, $h[m] \in [150, 250)$
Offset sector 3, $h[m] \in [250, 350)$
Offset sector 4, $h[m] \in [350, 450)$

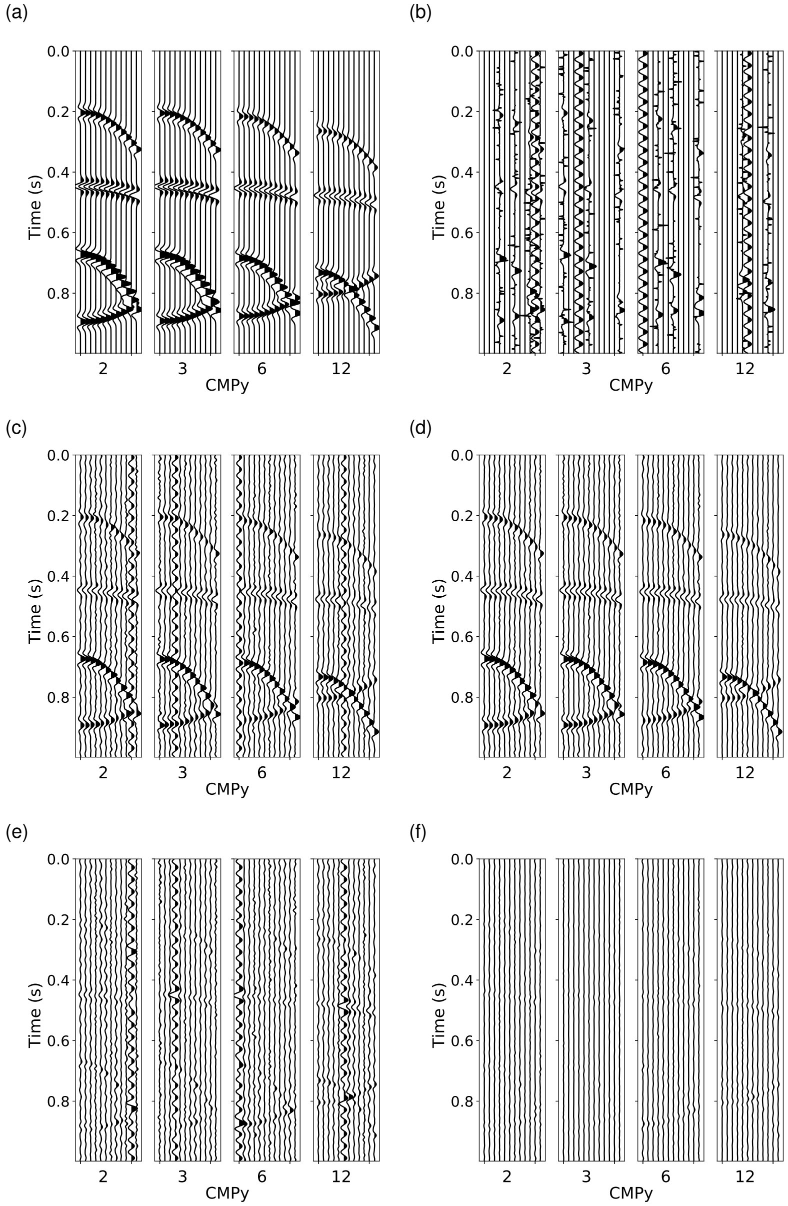

Figure 5.4: Reconstruction of parabolic events with data contaminated with erratic noise and sinusoidal noise. (a) Original data. (b) Data corrupted with erratic noise and sinusoidal noise and decimated by 60%. (c) Nonrobust l_2 PFM reconstruction. $\texttt{S/N}_o = 16$ dB. (d) Robust PMF reconstruction. The l_1/l_2 norm was adopted as a robust measure of misfit. $\texttt{S/N}_o = 20$ dB. (e) Error associated with the nonrobust solution. (f) Error associated with the robust solution.

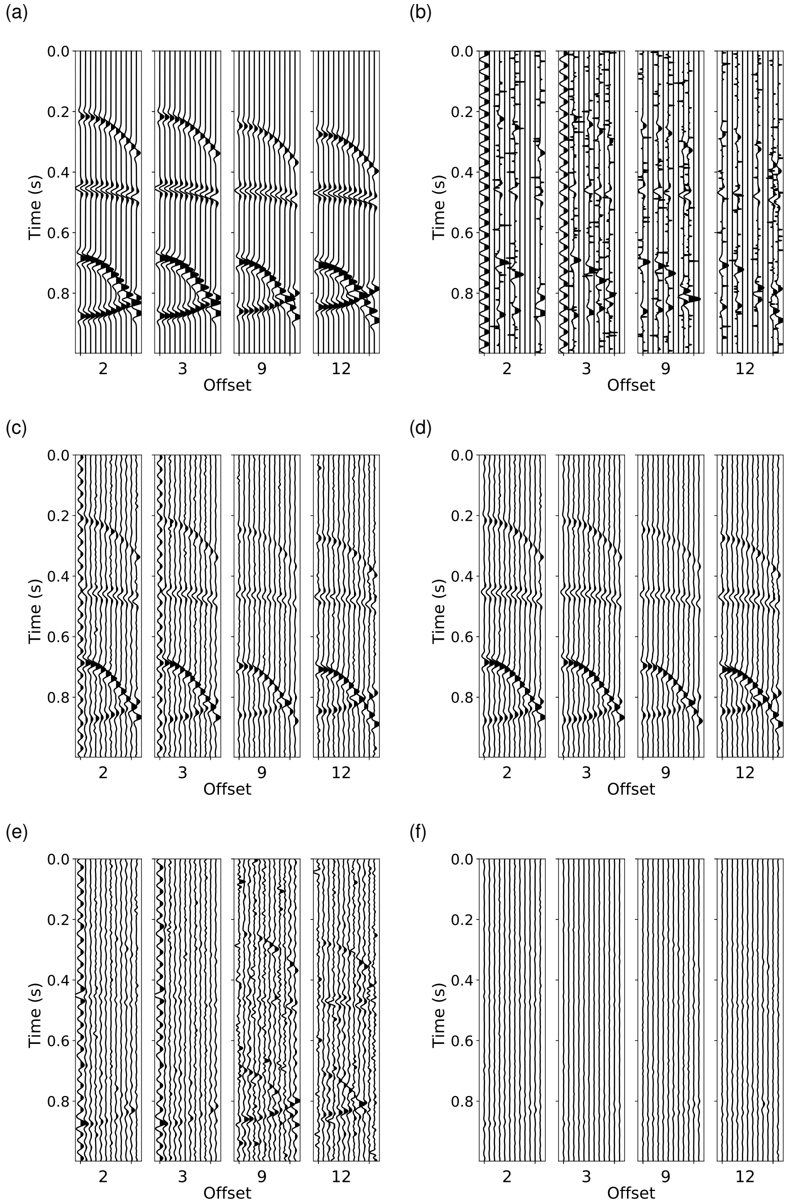

Figure 5.5: Reconstruction of parabolic events with data contaminated with erratic noise and sinusoidal noise. (a) Original data. (b) Data corrupted with erratic noise and sinusoidal noise and decimated by 60%. (c) Nonrobust l_2 PFM reconstruction. $\texttt{S/N}_o = 16$ dB. (d) Robust PMF reconstruction. The l_1/l_2 norm was adopted as a robust measure of misfit. $\texttt{S/N}_o = 20$ dB. (e) Error associated with the nonrobust solution. (f) Error associated with the robust solution.

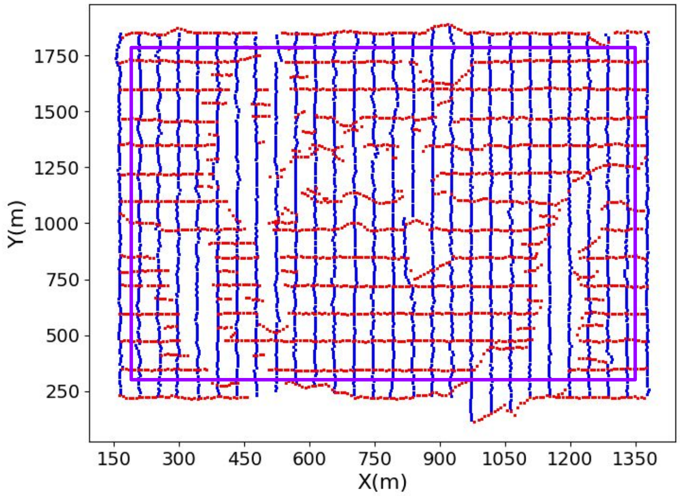

Figure 5.6: Survey acquisition geometry. Sources are ploted in red and receivers in blue. The violet rectangle represents the extent of the CMP bins that were reconstructed in the test.

Offset sector 5, $h[m] \in [450, 550)$

Offset sector 6, $h[m] \in [550, 650]$

$\Delta\phi = 45\,\text{degrees}$, number of azimuth sectors $I_4 = 8$

Azimuth sector 1, $\phi[^\circ] \in [0, 45)$

Azimuth sector 2, $\phi[^\circ] \in [45, 90)$

Azimuth sector 3, $\phi[^\circ] \in [90, 135)$

Azimuth sector 4, $\phi[^\circ] \in [135, 180)$

Azimuth sector 5, $\phi[^\circ] \in [180, 225)$

Azimuth sector 6, $\phi[^\circ] \in [225, 270])$

Azimuth sector 7, $\phi[^\circ] \in [270, 315)$

Azimuth sector 8, $\phi[^\circ] \in [315, 360]$.

The fold of the data after binning is shown in Figure 5.7.

I also define overlapping windows or *patches* in space. This technique is used to ensure stationarity of the wavefield. Each patch consists of 45×45 midpoint bins

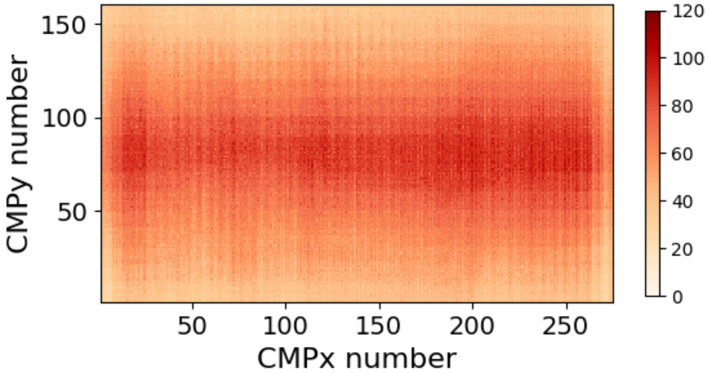

Figure 5.7: Fold map of the data.

with all offset and azimuth sectors. Therefore, I define $I_1 = 45$, $I_2 = 45$, $I_3 = 6$, and $I_4 = 8$ which leads to 4D spatial patches of size $45 \times 45 \times 6 \times 8$. The patches are overlapped by 5 midpoints in both inline and crossline coordinates. This configuration yields a total of 28 patches that cover the area delimited by the rectangle in magenta in Figure 5.6.

The reconstruction is performed on every patch. I use one patch to tune the trade-off parameters and then apply them for the rest of the patches. The frequencies range from 1 Hz to 100 Hz, and there is a maximum number of 50 iterations per frequency and a tolerance $tol = 0.0005$. For the nonrobust PMF reconstruction I adopt $\mu \times \sigma^2 = 1$. For the robust PMF completion I adopt the l_1/l_2 norm with parameter $\mu = 500$, and σ is determined as a function of the error at the first iteration $\sigma = 0.0001 \times \|\mathcal{E}^1(\omega)\|_F$. In both cases, randomized QR decomposition is applied with parameter $p_k = 20, 20, 5, 5$. The computational time for one patch in a desktop computer with an Intel Core i5 processor with a speed of 3.30GHz running Julia in a single processor is 64 seconds. Individual reconstructed 5D patches are re-assembled into the $276 \times 161 \times 6 \times 8$ midpoint-offsets-azimuth volume. Areas of overlap are weighted with a cosine taper to avoid windowing artifacts in the re-assembled volume.

Figures 5.8 to 5.10 show different slices of the volume before and after reconstruction. Figure 5.8 shows the original data and the reconstructed data via the nonrobust and robust PMF algorithms for fixed CMPx 10, offset sector 3 (250-350 m), and azimuth sector 5 (180-225°). This particular figure illustrates the attenuation

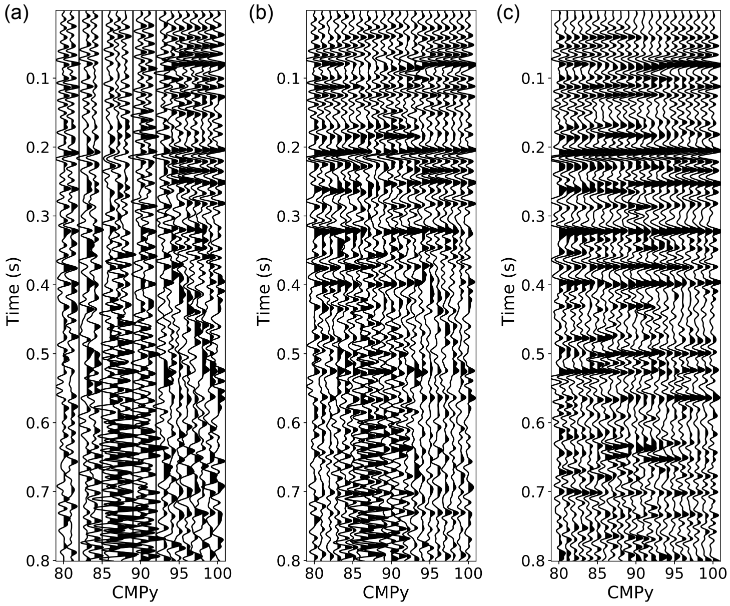

Figure 5.8: Reconstructed field seismic data for fixed CMPx=10, offset sector 3, and azimuth sector 5. (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

of coherent high-frequency noise corresponding to the robust PMF reconstruction (Figure 5.8(c). Figure 5.9 portrays the original data and the nonrobust and robust PMF reconstructions for CMPy 90, offset sector 4 (350-450 m), and azimuth sector 3 (90-135°). In Figure 5.10, I change the aspect ratio of the illustration and provide a 350 ms window portraying the original data and the nonrobust and robust PMF reconstructions for CMPy 80, offset sector 3 (250-350 m), and azimuth sector 5 (180-225°). Again, it is clear that the robust PMF algorithm is not only able to recover the observed traces, but it also attenuates erratic noise such as the oscillatory signals near the position CMPx 150 in Figure 5.10(a).

Figures 5.11 and 5.12 show 3D visualizations of the 5D reconstructed tensor for offset sector 2 (150-250 m), azimuth 4 (135-180°), and offset sector 4 (350-450 m), azimuth 4 (135-180°), respectively. Finally, Figure 5.13 shows the stacked cube for the field data, nonrobust and robust reconstruction via PMF algorithms. The stack of the data before reconstruction was properly normalized by fold. Visual comparison of stacked volumes shows modest improvement on the quality of the seismic, yet the reconstructed cube could result in improved quantitative interpretation and

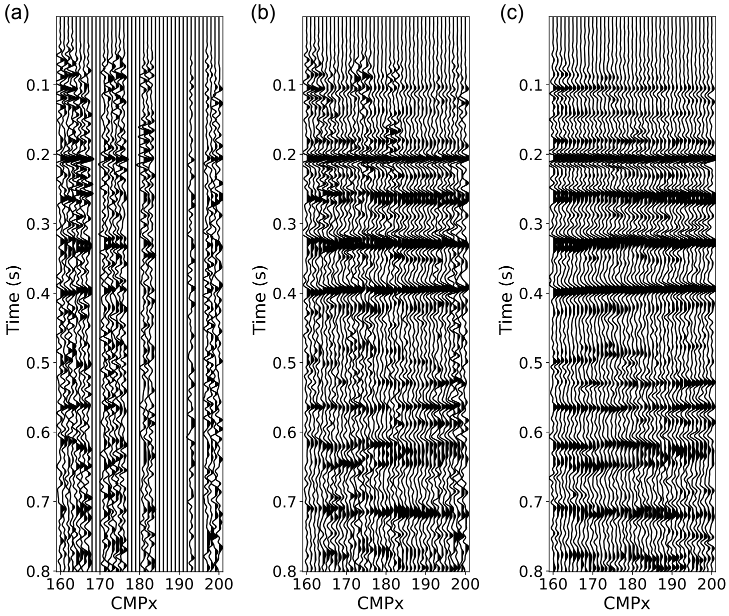

Figure 5.9: Reconstructed field seismic data for fixed CMPy=90, offset sector 4, and azimuth sector 3. (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

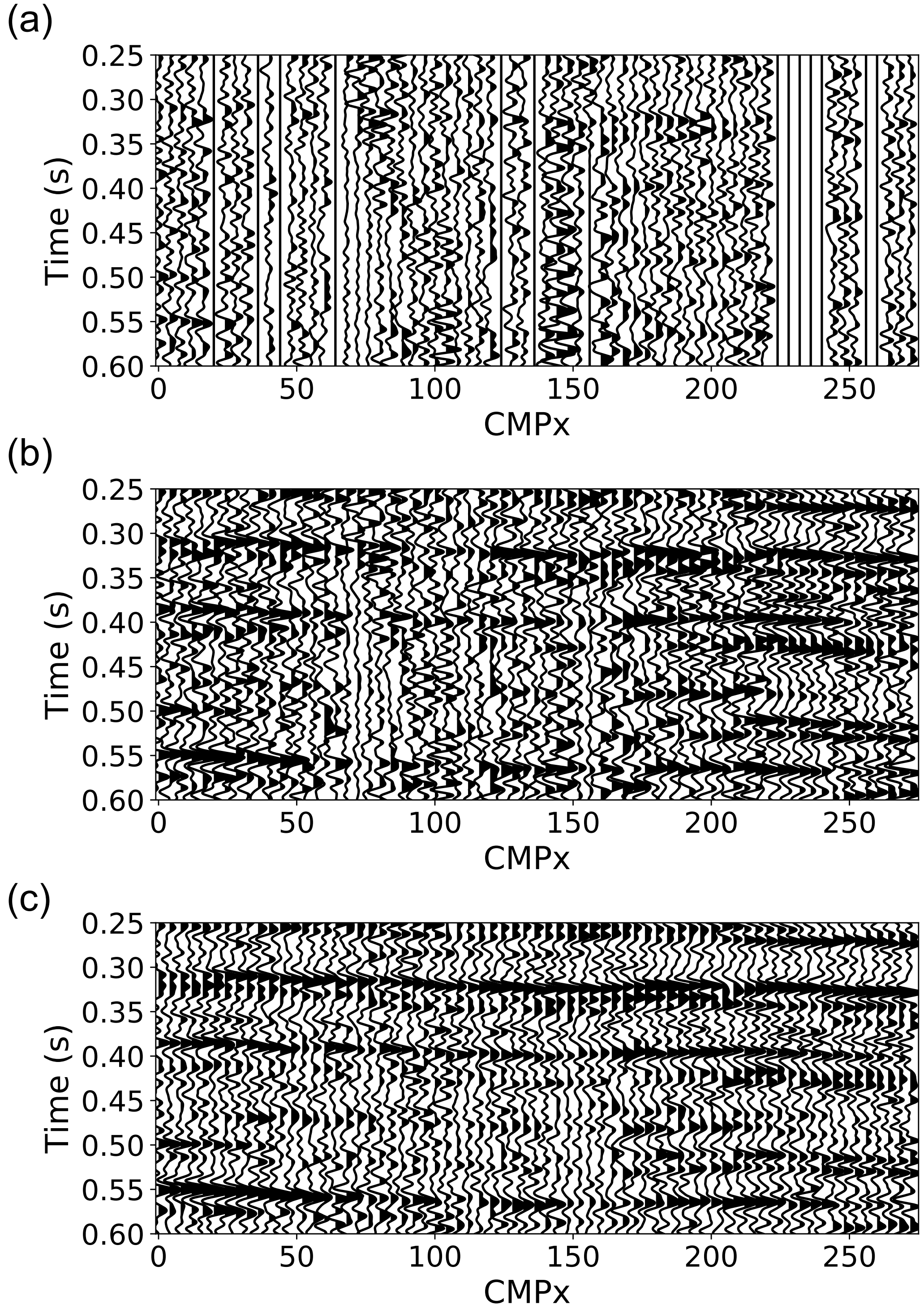

Figure 5.10: Reconstructed field seismic data for fixed CMPy=80, offset sector 3, and azimuth sector 5. (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

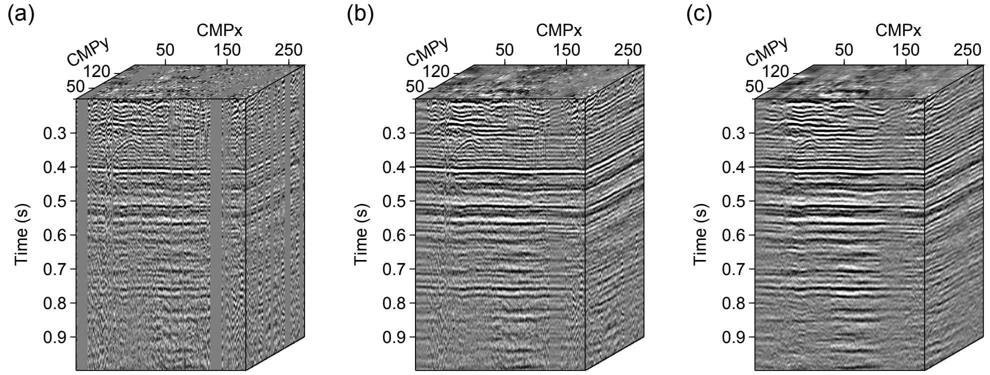

Figure 5.11: Data for offset sector 2 (150-250 m) and azimuth 4 (135-180°). (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

attribute analysis.

5.3 Conclusions

Prestack seismic-data reconstruction can be posed as a tensor-completion problem. Onshore seismic data are frequently corrupted by erratic noise that does not obey any particular distribution. I have adapted the Parallel Matrix Factorization (PMF) method to cope with such erratic noise by incorporating a robust error norm. Such modification considers differences in the magnitude of the reconstruction error, which de-emphasizes *bad observations* and minimizes their influence on the final reconstructed volume. I have presented results using the l_1/l_2 error norm, but similar results could be attained by adopting the Cauchy and Geman-McClure error functionals.

The iterative PMF algorithm consists of two stages. In the first step, a new estimate of the data is calculated via a simple imputation algorithm. In the second step, rank-reduction is applied to matrices obtained via tensor unfolding of the estimated data tensor. In the iterative algorithm, new data obtained by the rank-reduction stage and observations are weighted to generate the current estimate of the reconstructed data. The nonrobust PMF algorithm does not take into account differences in

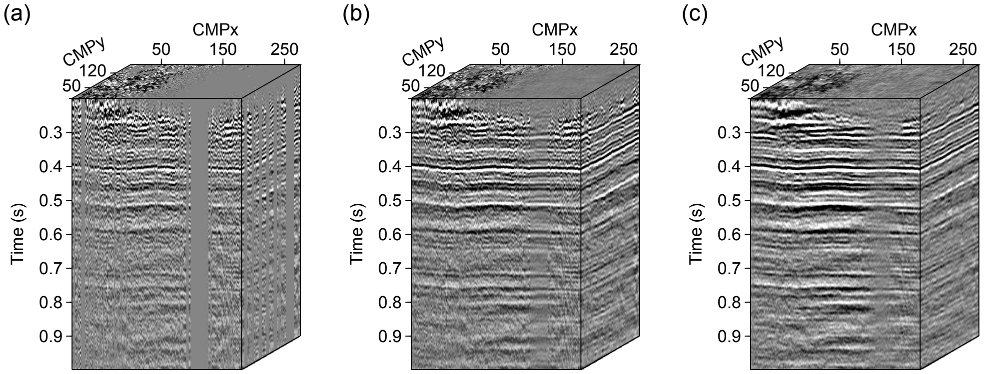

Figure 5.12: Data for offset sector 4 (350-450 m) and azimuth 4 (135-180°). (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

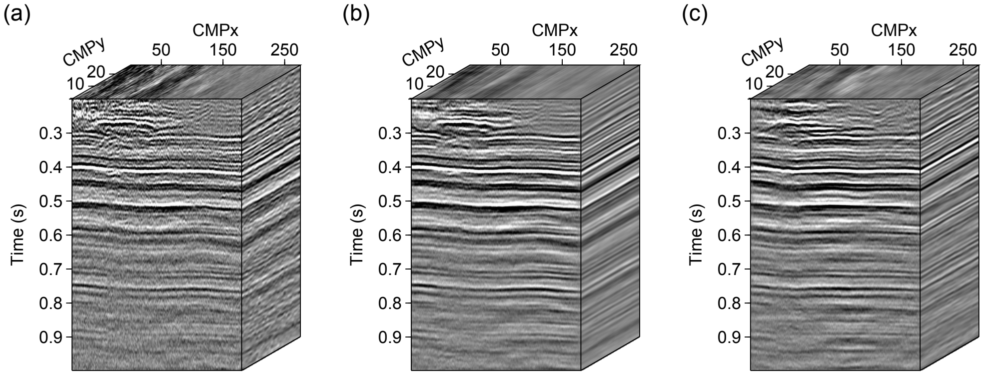

Figure 5.13: Stacks. (a) Original data. (b) Nonrobust PMF reconstruction. (c) Robust PMF reconstruction.

the magnitude of the reconstruction error. Therefore it is inclined to introduce erratic noise back into the reconstructed prestack volume. The proposed robust PMF algorithm, on the other hand, de-emphasizes *bad observations* and minimizes their influence on the final reconstructed volume.

The rank-reduction stage of the PMF algorithm can be implemented by numerous methods. I adopt a randomized QR decomposition as the engine of the rank-reduction stage. The randomized QR decomposition relaxes the precise determination of the best rank of the unfoldings, and, therefore, it adds practicality to the algorithm.

I recognize that numerical experimentation was adopted to drive this research and reach conclusions; this is a route often chosen in prestack seismic reconstruction. Numerical tests were the immediate solution to address the ubiquitous parameter selection problem. Future research entails finding objective criteria for parameter estimation (trade-off parameter, standard deviation, and tensor rank) for nonrobust and robust PMF algorithms.

I also tested the proposed algorithm with real data from the Western Canadian Sedimentary Basin. The field data experiments permit us to conclude that the proposed algorithm can recover unrecorded seismic traces while attenuating erratic noise in recorded traces. In addition, the analysis of stacked sections confirms that the algorithm does not introduce artifacts.

CHAPTER 6

Conclusions

Multidimensional reconstruction can significantly improve seismic data by reproducing the ideal, regularly sampled seismic wavefield one would have desired to acquire. It also allows the reconstruction of data gaps and the correct positioning of seismic traces at the desired output grid. Reconstruction methods are particularly beneficial for onshore seismic data as they homogenize fold in CMP sectors producing datasets that contain a regular distribution of offsets and azimuths. The latter is essential for imaging, AVO/AVA inversion, and the removal of acquisition footprints. These footprints, observed in seismic time slices, can hamper the interpretation of subtle geological features.

Geological interpretation of seismic data evaluates time migrated or depth migrated images that have undergone 5D reconstruction before imaging. Despite the widespread application of reconstruction methods in exploration seismology, many industrial applications of 5D reconstruction consider Fourier inversion methods. Reduced-rank methods as engines for 5D reconstruction are not as popular as methods based on Fourier kernels, albeit been investigated for about ten years. Reduced-rank filtering has been adopted mainly for denoising. This thesis investigates and provides solutions to several shortcomings of reduced-rank reconstruction methods, including irregularly sampled seismic data and the reconstruction of data contaminated with erratic noise.

The classical MSSA assumes input data deployed on regular grids, yet seismic data are usually irregularly sampled in the spatial domain. Although seismic processing flows often tolerate binning (nearest neighbour interpolation) errors, there are situations where an emphasis on accurate input trace coordinates should prevail. I reformulated the MSSA method to honour the real spatial coordinates of seismic traces. The new method, I-MSSA, solves an optimization problem with a rank constraint. The I-MSSA algorithm provides a tool for coping with data regularization when one does not want to apply binning to input datasets. For example, when processing the cross-spread gathers of a seismic volume. Chapter 3 presented a case study with a workflow to process these subvolumes of seismic data.

Another problem that lacked detailed investigation is the reconstruction of data contaminated with a non-Gaussian error distribution or, in general, erratic noise of unknown distribution. Random and erratic noise contaminates onshore seismic data. Bursts of energy appear at isolated locations, making a good portion of the seismic traces unusable. Rather than eliminating traces, I proposed using robust norm minimization of the erratic error combined with rank-constraints to develop a robust SSA method (Chapter 4). The scheme can be extended to other interpolation methods that rely on regular grids, presenting a solution to the off-the-grid processing problem when the engine for denoising or reconstruction assumes regular sampling.

Similarly, in Chapter 5, I introduced a robust tensor completion method to reconstruct data contaminated with erratic noise. The algorithm generalizes the PMF method proposed for 5D data completion to data containing traces contaminated with non-Gaussian noise. The scheme provides a fast method for rank-reduction to apply robust PMF in realistic field data scenarios as those presented in Chapter 5. To do so, I replaced matrix factorization or the SVD (the two methods PMF often used) with a randomized QR decomposition that reduces the computation time of rank-reduction applied to unfolded tensors.

Below is a detail of the contributions and a summary of each chapter.

Chapter 2 reviews the formulation of low-rank approximation and reconstruction via low-rank considerations. The chapter introduces the algebra and some typical formulations of the problem and demonstrates that seismic data can be approximated via low-rank arrays. This chapter provides a comprehensive introduction to a student pursuing research in this area. The latter was my primary goal in adding Chapter 2 to the thesis.

Chapter 3 explores the extension of MSSA to datasets with irregular coordinates. MSSA assumes that seismic data has a regular geometry. In reality, given the complexity of the operations and financial constraints, field data is usually acquired with some level of randomness. Conventional processing workflows include binning as a preprocessing step for seismic reconstruction via MSSA. Binning is not an optimal process as it introduces amplitude and phase errors in the input traces. Chapter 3 introduces I-MSSA, a new reconstruction method that honours the real coordinates of traces. The method formulates reconstruction as an inverse problem, where the misfit includes a mapping operator that relates irregular coordinates to an ideal output grid. For each frequency slice, I-MSSA enforces the low-rank condition via a projection operator implemented as an SSA filter. I tested the method with synthetic and field examples. Despite achieving slightly notable improvements, I-MSSA offers a novel way of solving the rank-reduction reconstruction problem for off-the-grid data. Binning of input datasets constructs intermediate data volumes, discarding or averaging multiple traces that correspond to the same bin and populating with null traces where the bin is empty. Binning also introduces phase and amplitude errors to the dataset. I-MSSA eliminates the increased requirement of computation and memory introduced by binning. More importantly, it considers the complete set of data without discarding or averaging samples. These features of I-MSSA were my primary motivation for pursuing this line of research.

Chapter 4 introduces robust norms in seismic reconstruction via rank-reduction algorithms. The chapter reviews traditional robust norms used in geophysical applications and introduces a generalized function that regulates its shape via a continuous parameter. The generalized function provides flexibility to the inversion. Afterwards, I present a scheme that minimizes robust measures of fit combined with rank-constraint principles. The method allows a robust adaptation of the SSA scheme. This chapter also evaluates the requirements for monotonic or redescending norms in seismic reconstruction. I tested the algorithm via synthetic and field data and concluded that redescending estimators provide adequate reconstruction in the presence of a high percentage of non-Gaussian errors. However, seismic data typically presents low percentages of erratic traces. Therefore, monotonic norms, such as the Huber norm, generally provide acceptable results. Chapter 4 explores foundational concepts in robust statistics and provides a cornerstone for the algorithms presented in the following chapter.

Chapter 5 presents a robust implementation of PMF for seismic reconstruction. PMF reconstructs the signal by approximating the tensor unfoldings to low-rank matrices. The algorithm estimates the solution via an alternate least-squares approach that results in two steps. The first step calculates the data via an imputation algorithm, and the second step applies rank-reduction to the unfoldings of the estimated data tensor. The nonrobust algorithm introduces erratic noise back into the estimated volume in each iteration. I propose a robust PMF that down weights high amplitude observations to limit their influence in the solution. The algorithm results in a similar engine to PMF, but the imputation scheme includes weights defined by the considered robust norm. The scheme also considers a randomized QR decomposition as the rank-reduction engine. The randomized approach alleviates the computational strain and relaxes the determination of the best rank for the unfoldings. I tested the algorithm with synthetic and field data. Tests showed that the algorithm attenuates noise and recovers missing samples. The stacked sections of field data also showed that the algorithm does not introduce artifacts.

6.1 Software libraries

Many of the results presented in this thesis were compiled on my GitHub account github.com/fercarozzi. The libraries are developed in julia language (Bezanson et al., 2017) and perform multidimensional reconstruction of seismic data with the algorithms previously described. In addition, I have used SeismicJulia (Stanton and Sacchi, 2016) to model synthetic experiments and read, write and process field datasets. My contribution to developing SeismicJulia included maintenance, development and documentation of the libraries.

6.2 Recommendations and future research directions

The bulk of my work considers methods that operate in the f-x domain. In other words, the data are initially transformed from the time domain to the frequency domain. Next, processing and inversion are carried on all frequency slices, honouring Fourier symmetries. Finally, after all the frequency slices of the volume are simultaneously processed, the inverse Fourier transform is adopted to transform the results back to the time domain. I have not investigated the application of tensor completion methods to volumes in the time-space domain as this is not a typical approach for seismic data processing. However, I recognize the possibility of obtaining exciting results when processing the data in the time domain with algorithms based on tensor-reconstruction principles.

The previous is not advised for the I-MSSA algorithm as it structures the problem via trajectory matrices. In essence, one extract patches of seismic data and assimilate those patches into Hankel matrices. To demonstrate the predictability of seismic signals in the t-x domain is not straightforward. Therefore, these volumes should not be loosely considered low-rank in the t-x domain.

Another recommendation for future work is to consider the recent advancement of machine learning applications to seismic reconstruction. Some impressive results were obtained after considerable training (Mandelli et al., 2018). Machine learning methods in collaboration with data augmentation techniques will play an essential role in seismic data reconstruction in the future. However, generalizing reconstruction algorithms, or finding models that operate on diverse sets of onshore data, is still an unsolved problem. Also, numerically modelling data that mimics realistic noisy onshore surveys is challenging. The latter might impose severe limitations on building datasets for training. However, this seems to be a contemporary problem in machine learning not restricted to geophysical applications.

Bibliography

Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, no. 6, E91–E97.

Acar, E., D. M. Dunlavy, T. G. Kolda, and M. Mørup, 2010, Scalable tensor factorizations with missing data: SIAM International Conference on Data Mining, 701–712.

Al-Bannagi, M. S., K. Fang, P. G. Kelamis, and G. S. Douglass, 2005, Acquisition footprint suppression via the truncated SVD technique: Case studies from Saudi Arabia: The Leading Edge, **24**, no. 8, 832–834.

Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, no. 12, 2027–2035.

Anderson, R. G., and G. A. McMechan, 1989, Automatic editing of noisy seismic data: Geophysical prospecting, **37**, no. 8, 875–892.

Andersson, C. A., and R. Bro, 1998, Improving the speed of multi-way algorithms: Part I. Tucker 3: Chemometrics and Intelligent Laboratory Systems, **42**, 93–103.

Aravkin, A., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2014, Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation: SIAM Journal on Scientific Computing, **36**, no. 5, S237–S266.

Bahia, B., R. Lin, and M. D. Sacchi, 2020, Iterative deblending with robust Fourier thresholding: SEG Technical Program Expanded Abstracts 2020, 3279–3283.

Bahia, B., and M. Sacchi, 2019, Robust singular spectrum analysis via the bifactored gradient descent algorithm: SEG Technical Program Expanded Abstracts 2019, 4640–4644.

Barron, J. T., 2019, A general and adaptive robust loss function: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4331–4339.

Beaton, A. E., and J. W. Tukey, 1974, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data: Technometrics, **16**, no. 2, 147–185.

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM journal on imaging sciences, **2**, no. 1, 183–202.

Bertsekas, D. P., 1996, Constrained optimization and Lagrange multiplier methods: Athena Scientific.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah, 2017, Julia: A fresh approach to numerical computing: SIAM Review, **59**, no. 1, 65–98.

Bolduc, E., G. C. Knee, E. M. Gauger, and J. Leach, 2017, Projected gradient descent algorithms for quantum state tomography: npj Quantum information, **3**, no. 1, 44.

Brandt, S., 1998, Data analysis: Springer.

———, 2014, The method of Maximum Likelihood, *in* Data Analysis: Springer, 153–173.

Brandwood, D., 1983, A complex gradient operator and its application in adaptive array theory: IEE Proceedings F - Communications, Radar and Signal Processing, IET, 11–16.

Branham, Jr., R. L., 1986, Is robust estimation useful for astronomical data reduction?: Quarterly Journal of the Royal Astronomical Society, **27**, 182–193.

Bro, R., 1997, PARAFAC. Tutorial and applications: Chemometrics and intelligent laboratory systems, **38**, no. 2, 149–172.

Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: Geophysics, **75**, no. 3, R37–R46.

Bube, K. P., and R. L. Langan, 1997, Hybrid l_1/l_2 minimization with applications to tomography: Geophysics, **62**, no. 4, 1183–1195.

Burroughs, L., and S. Trickett, 2009, Prestack rank-reducing noise suppression: Practice: SEG Technical Program Expanded Abstracts 2009, **28**, 3337–3341.

Cai, J.-F., E. J. Candès, and Z. Shen, 2010, A singular value thresholding algorithm for matrix completion: SIAM Journal on Optimization, **20**, no. 4, 1956–1982.

Calvert, A., E. Jenner, R. Jefferson, R. Bloor, N. Adams, R. Ramkhelawan, and C. St. Clair, 2008, Preserving azimuthal velocity information: Experiences with cross-spread noise attenuation and offset vector tile preSTM: SEG Technical Program Expanded Abstracts 2008, 207–211.

Candès, E., and Y. Plan, 2010, Matrix completion with noise: Proceedings of the

IEEE, **98**, no. 6, 925 –936.

Candès, E., and B. Recht, 2009, Exact matrix completion via convex optimization: Foundations of Computational Mathematics, **9**, no. 6, 717–772.

Canning, A., and G. H. Gardner, 1998, Reducing 3-D acquisition footprint for 3-D DMO and 3-D prestack migration: Geophysics, **63**, no. 4, 1177–1183.

Carrillo, R. E., A. B. Ramirez, G. R. Arce, K. E. Barner, and B. M. Sadler, 2016, Robust compressive sensing of sparse signals: a review: EURASIP Journal on Advances in Signal Processing, **2016**, no. 1, 108.

Carroll, J. D., and J.-J. Chang, 1970, Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition: Psychometrika, **35**, no. 3, 283–319.

Chen, K., and M. D. Sacchi, 2015, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, no. 1, V1–V11.

———, 2017, Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation: Geophysical Prospecting, **65**, 650–668.

Cheng, J., K. Chen, and M. D. Sacchi, 2015, Application of Robust Principal Component Analysis (RPCA) to suppress erratic noise in seismic records: SEG Technical Program Expanded Abstracts 2015, 4646–4651.

Cheng, J., M. Sacchi, and J. Gao, 2019, Computational efficient multidimensional singular spectrum analysis for prestack seismic data reconstruction: Geophysics, **84**, no. 2, V111–V119.

Cheng, J., and M. D. Sacchi, 2015, A fast rank-reduction algorithm for 3D deblending via randomized QR decomposition: SEG Technical Program Expanded Abstracts 2015, 3830–3835.

———, 2016, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition: Geophysics, **81**, no. 1, V89–V101.

Chiron, L., M. A. van Agthoven, B. Kieffer, C. Rolando, and M.-A. Delsuc, 2014, Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry: Proceedings of the National Academy of Sciences, **111**, no. 4, 1385–1390.

Cichocki, A., D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, 2015, Tensor decompositions for signal processing applications: From two-way to multiway component analysis: IEEE Signal Processing Magazine, **32**, no. 2, 145–163.

Claerbout, J., 1992, Earth soundings analysis: Processing versus inversion: Black-well Science.

Claerbout, J., and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **38**, 826–844.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, no. 3, 467–481.

Comon, P., 2014, Tensors: a brief introduction: IEEE Signal Processing Magazine, **31**, no. 3, 44–53.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: application to real data: Geophysics, **55**, no. 5, 527–538.

Crawley, S., R. Clapp, and J. Clearbout, 1999, Interpolation with smoothly non-stationary prediction-error filters: SEG Technical Program Expanded Abstracts 1999, 1154–1157.

Da Silva, C., and F. J. Herrmann, 2015, Optimization on the hierarchical Tucker manifold–applications to tensor completion: Linear Algebra and its Applications, **481**, 131–173.

De Lathauwer, L., B. De Moor, and J. Vandewalle, 2000, A multilinear singular value decomposition: SIAM J. Matrix Anal. Appl., **21**, no. 4, 1253–1278.

de Morais Goulart, J. H., 2016, Estimation of structured tensor models and recovery of low-rank tensors: PhD thesis, Université Côte d'Azur.

Debeye, H., and P. Van Riel, 1990, l_p norm deconvolution: Geophysical Prospecting, **38**, no. 4, 381–403.

Dennis Jr, J. E., and R. E. Welsch, 1978, Techniques for nonlinear least squares and robust regression: Communications in Statistics-simulation and Computation, **7**, no. 4, 345–359.

Donoho, D. L., 2006, Compressed sensing: IEEE Transactions on Information Theory, **52**, no. 4, 1289–1306.

Duijndam, A. J. W., M. A. Schonewille, and C. O. H. Hindriks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, no. 2, 524–538.

Eckart, C., and G. Young, 1936, The approximation of one matrix by another of lower rank: Psychometrika, **1**, 211–218.

Ely, G., S. Aeron, N. Hao, and M. E. Kilmer, 2015, 5D seismic data completion and denoising using a novel class of tensor decompositions: Geophysics, **80**, no. 4, V83–V95.

Fazel, M., T. K. Pong, D. Sun, and P. Tseng, 2013, Hankel matrix rank minimization with applications to system identification and realization: SIAM Journal on Matrix Analysis and Applications, **34**, no. 3, 946–977.

Fomel, S., 2001, Three-dimensional seismic data regularization: PhD thesis, Stanford University.

———, 2003, Seismic reflection data interpolation with differential offset and shot continuation: Geophysics, **68**, no. 2, 733–744.

Fortnow, L., 2009, The status of the p versus np problem: Communications of the ACM, **52**, no. 9, 78–86.

Freire, S., and T. Ulrych, 1988, Application of singular value decomposition to vertical seismic profiling: Geophysics, **53**, no. 6, 778–785.

Gandy, S., B. Recht, and I. Yamada, 2011, Tensor completion and low-n-rank tensor recovery via convex optimization: Inverse Problems, **27**, no. 2, 025010.

Gao, J., J. Cheng, and M. D. Sacchi, 2017, Five-dimensional seismic reconstruction using Parallel Square Matrix Factorization: IEEE Transactions on Geoscience and Remote Sensing, **55**, no. 4, 2124–2135.

Gao, J., M. D. Sacchi, and X. Chen, 2011, A fast rank reduction method for the reconstruction of 5D seismic volumes: SEG Technical Program Expanded Abstracts 2011, **30**, no. 1, 3622–3627.

———, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: Geophysics, **78**, no. 1, V21–V30.

Gao, J., A. Stanton, and M. D. Sacchi, 2015, Parallel Matrix Factorization algorithm and its application to 5D seismic reconstruction and denoising: Geophysics, **80**, no. 6, V173–V187.

Gardner, G. H., and A. Canning, 1994, Effects of irregular sampling on 3-D prestack migration: SEG Technical Program Expanded Abstracts 1994, 1553–1556.

Gavish, M., and D. L. Donoho, 2014, The optimal hard threshold for singular values is $4/\sqrt{3}$: IEEE Transactions on Information Theory, **60**, no. 8, 5040–5053.

Geman, S., and D. McClure, 1985, Bayesian image analysis: An application to single photon emission tomography: Proceedings of the American Statistical Association, Statistical Computing Section, 12–18.

Gholami, A., and M. Sacchi, 2012, A fast and automatic sparse deconvolution in the presence of outliers: IEEE Transactions on Geoscience and Remote Sensing, **50**, no. 10, 4105–4116.

Greiner, T. L., O. Kolbjørnsen, J. E. Lie, E. H. Nilsen, A. K. Evensen, and L. Gelius, 2019, Cross-streamer wavefield interpolation using deep convolutional neural network: SEG Technical Program Expanded Abstracts 2019, 2207–2211.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, no. 4, 1310–1319.

Gulunay, N., 2003, Seismic trace interpolation in the Fourier transform domain: Geophysics, **68**, no. 1, 355–369.

Halko, N., P.-G. Martinsson, and J. A. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: SIAM review, **53**, no. 2, 217–288.

Harshman, R. A., 1970, Foundations of the PARAFAC procedure: Models and conditions for an explanatory multi-modal factor analysis: UCLA Working Papers in Phonetics, **16**, 1–84.

Hennenfent, G., and F. J. Herrmann, 2006, Seismic denoising with nonuniformly sampled curvelets: IEEE Trans. on Computing in Science and Engineering, **8**, no. 3, 16–25.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, no. 6, WB173–WB187.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008, Curvelet-based seismic data processing: A multiscale and nonlinear approach: Geophysics, **73**, no. 1, A1–A5.

Holland, P. W., and R. E. Welsch, 1977, Robust regression using iteratively reweighted least-squares: Communications in Statistics-theory and Methods, **6**, no. 9, 813–827.

Hua, Y., 1992, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil: IEEE Transactions on Signal Processing, **40**, no. 9, 2267–2280.

Ibrahim, A., and M. D. Sacchi, 2014, Simultaneous source separation using a robust Radon transform: Geophysics, **79**, no. 1, V1–V11.

Iusem, A. N., 2003, On the convergence properties of the projected gradient method for convex optimization: Computational and Applied Mathematics, **22**, no. 1, 35–52.

Jain, P., R. Meka, and I. S. Dhillon, 2010, Guaranteed rank minimization via singular value projection: Advances in Neural Information Processing Systems, 937–945.

Ji, J., 2006, Cgg method for robust inversion and its application to velocity-stack inversion: Geophysics, **71**, no. 4, R59–R67.

———, 2012, Robust inversion using biweight norm and its application to seismic inversion: Exploration Geophysics, **43**, 70–76.

Jia, Y., and J. Ma, 2017, What can machine learning do for seismic data processing? An interpolation application: Geophysics, **82**, no. 3, V163–V177.

Jiang, T., B. Gong, F. Qiao, Y. Jiang, A. Chen, D. Hren, and Z. Meng, 2017, Compressive seismic reconstruction with extended POCS for arbitrary irregular acquisition: SEG Technical Program Expanded Abstracts 2017, 4272–4277.

Kabir, M. M. N., and D. J. Verschuur, 1995, Restoration of missing offsets by parabolic Radon transform: Geophysical Prospecting, **43**, no. 3, 347–368.

Kaplan, S. T., M. Naghizadeh, and M. D. Sacchi, 2010, Data reconstruction with shot-profile least-squares migration: Geophysics, **75**, no. 6, WB121–WB136.

Karatzoglou, A., X. Amatriain, L. Baltrunas, and N. Oliver, 2010, Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering: Proceedings of the fourth ACM conference on Recommender systems, 79–86.

Keshavan, R. H., A. Montanari, and S. Oh, 2010, Matrix completion from a few entries: IEEE transactions on information theory, **56**, no. 6, 2980–2998.

Kolda, T. G., 2006, Multilinear operators for higher-order decompositions: Technical report, Sandia National Laboratories, Albuquerque, NM and Livermore, CA.

Kolda, T. G., and B. W. Bader, 2009, Tensor decompositions and applications: SIAM review, **51**, no. 3, 455–500.

Kreimer, N., and M. D. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, no. 3, V113–V122.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, no. 6, V273–V284.

Kumar, R., C. Da Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2015, Efficient matrix completion for seismic data reconstruction: Geophysics, **80**, no. 5, V97–V114.

Lai, M.-J., Y. Xu, and W. Yin, 2013, Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization: SIAM Journal on Numerical Analysis, **51**, no. 2, 927–957.

Lee, H., Y.-D. Kim, A. Cichocki, and S. Choi, 2007, Nonnegative tensor factorization for continuous EEG classification: International journal of neural systems, **17**, no. 04, 305–317.

Lee, S.-I., H. Lee, P. Abbeel, and A. Y. Ng, 2006, Efficient l_1 regularized logistic regression: Association for the Advancement of Artificial Intelligence, 401–408.

Li, C., C. C. Mosher, and S. T. Kaplan, 2012, Interpolated compressive sensing for seismic data reconstruction: SEG Technical Program Expanded Abstracts 2012, 1–6.

Li, W., K. Chen, F. Ahmed, and W. Jeong, 2019, Rank revealing and vector optimization methods for adaptive robust denoising: SEG Technical Program Expanded Abstracts 2019, 4695–4699.

Liano, K., 1996, Robust error measure for supervised neural network learning with outliers: IEEE Transactions on Neural Networks, **7**, no. 1, 246–250.

Liberty, E., F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, 2007, Randomized algorithms for the low-rank approximation of matrices: Proceedings of the National Academy of Sciences, **104**, no. 51, 20167–20172.

Lin, D., J. Young, W.-j. Lin, and M. Griffiths, 2005, 3D SRME prediction and subtraction for better imaging: SEG Technical Program Expanded Abstracts 2005, 2088–2091.

Lin, R., and M. Sacchi, 2020, Separation of simultaneous sources acquired with a high blending factor via coherence pass robust Radon operators: Geophysics, **85**, no. 3, 1–98.

Liu, B., 2004, Multi-dimensional reconstruction of seismic data: PhD thesis, University of Alberta.

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, no. 6, 1560–1568.

Liu, B., M. D. Sacchi, and D. Trad, 2004, Simultaneous interpolation of 4 spatial dimensions: 74th Annual International Meeting, Soc. of Expl. Geophys., 2009–2012.

Liu, J., P. Musialski, P. Wonka, and J. Ye, 2012, Tensor completion for estimating missing values in visual data: IEEE transactions on pattern analysis and machine intelligence, **35**, no. 1, 208–220.

López, O., R. Kumar, Ö. Yılmaz, and F. J. Herrmann, 2016, Off-the-grid low-rank matrix recovery and seismic data reconstruction: IEEE Journal of Selected Topics in Signal Processing, **10**, no. 4, 658–671.

Malcolm, A. E., M. V. de Hoop, and J. H. LeRousseau, 2005, The applicability of dip moveout/azimuth moveout in the presence of caustics: Geophysics, **70**, no. 1, S1–S17.

Mandelli, S., F. Borra, V. Lipari, P. Bestagini, A. Sarti, and S. Tubaro, 2018, Seismic data interpolation through convolutional autoencoder: SEG Technical Program Expanded Abstracts 2018, 4101–4105.

Marjanovic, G., and V. Solo, 2012, On l_q optimization and matrix completion: IEEE Transactions on signal processing, **60**, no. 11, 5714–5724.

Maronna, R. A., R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, 2019, Robust statistics: theory and methods (with R): John Wiley & Sons.

Mirsky, L., 1960, Symmetric gauge functions and unitarily invariant norms: The Quarterly Journal of Mathematics, **11**, 50–59.

Mørup, M., 2011, Applications of tensor (multiway array) factorizations and decompositions in data mining: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **1**, no. 1, 24–40.

Mørup, M., L. K. Hansen, J. Parnas, and S. M. Arnfred, 2006, Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization: Technical University of Denmark Technical Report, 1–28.

Naghizadeh, M., and M. D. Sacchi, 2010a, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: Geophysics, **75**, no. 6, WB189–WB202.

———, 2010b, Seismic data reconstruction using multidimensional prediction filters: Geophysical Prospecting, **58**, no. 2, 157–173.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Nie, F., H. Wang, X. Cai, H. Huang, and C. Ding, 2012, Robust matrix completion via joint schatten p-norm and lp-norm minimization: 2012 IEEE 12th International Conference on Data Mining, 566–574.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization 2nd.

Oppenheim, A. V., J. R. Buck, and R. W. Schafer, 2001, Discrete-time signal processing. vol. 2, 2 ed.: Upper Saddle River, NJ: Prentice Hall.

Oropeza, V., and M. D. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32.

Oropeza, V. E., and M. D. Sacchi, 2010, A randomized SVD for Multichannel Singular Spectrum Analysis (MSSA) noise attenuation: SEG Technical Program

Expanded Abstracts 2010, **29**, 3539–3544.

Parikh, N., S. Boyd, et al., 2014, Proximal algorithms: Foundations and Trends in Optimization, **1**, no. 3, 127–239.

Peters, B., B. R. Smithyman, and F. J. Herrmann, 2019, Projection methods and applications for seismic nonlinear inverse problems with multiple constraints: Geophysics, **84**, no. 2, R251–R269.

Poole, G., and P. Herrmann, 2007, Multidimensional data regularization for modern acquisition geometries: SEG Technical Program Expanded Abstracts 2007, 2585–2589.

Popa, J., S. Minkoff, and Y. Lou, 2020, Improving seismic data completion via low-rank tensor optimization: SEG Technical Program Expanded Abstracts 2020, 2774–2778.

Porsani, M., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, no. 5, 1461–1467.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007, Numerical recipes 3rd edition: The art of scientific computing, 3 ed.: Cambridge university press.

Recht, B., 2011, A simpler approach to matrix completion: Journal of Machine Learning Research, **12**, no. 12.

Recht, B., M. Fazel, and P. Parrilo, 2010, Guaranteed minimun-rank solutions of linear matrix equations via nuclear norm minimization: SIAM review, **52**, no. 3, 471–501.

Ronen, J., 1987, Wave-equation trace interpolation: Geophysics, **52**, no. 7, 973–984.

Sacchi, M., 2009, *FX* Singular Spectrum Analysis: CSPG CSEG CWLS Convention, 392–395.

Sacchi, M. D., J. Cheng, and S. Janzen, 2017, Theory and application of vector Singular Spectrum Analysis (SSA) for multicomponent seismic data reconstruction: CSEG Annual meeting.

Sacchi, M. D., and T. J. Ulrych, 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, no. 4, 1169–1177.

Scales, J. A., and A. Gersztenkorn, 1988, Robust methods in inverse theory: Inverse Problems, **4**, 1071–1091.

Shaw, S., C. Mosher, S. Chiu, Y. Shen, J. Howell, and M. Wuenscher, 2007, Examples of 3D noise attenuation processes applied to land and OBC data: Presented at the 69th EAGE Conference and Exhibition-Workshop Package.

Sheriff, R. E., 2002, Encyclopedic dictionary of applied geophysics: Society of exploration geophysicists.

Shi, Y., X. Wu, and S. Fomel, 2019, Deep learning parameterization for geophysical inverse problems: SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, Beijing, China, 36–40.

Sidiropoulos, N. D., L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, 2017, Tensor decomposition for signal processing and machine learning: IEEE Transactions on Signal Processing, **65**, no. 13, 3551–3582.

Smilde, A., R. Bro, and P. Geladi, 2005, Multi-way analysis: applications in the chemical sciences: John Wiley & Sons.

Spitz, S., 1990, 3-D seismic interpolation in the F-XY domain: SEG Technical Program Expanded Abstracts 1990, 1641–1643.

Stanton, A., N. Kreimer, D. Bonar, M. Naghizadeh, and M. Sacchi, 2012, A comparison of 5D reconstruction methods: SEG Technical Program Expanded Abstracts 2012, 1–5.

Stanton, A., and M. Sacchi, 2011, Multicomponent seismic data reconstruction using quaternion Fourier transform and POCS: SEG Technical Program Expanded Abstracts 2011, 1267–1272.

———, 2013, All roads lead to Rome: Predictability, sparsity, rank and pre-stack seismic data reconstruction: CSEG Recorder, **38**, 1919–1928.

Stanton, A., M. Sacchi, R. Abma, and J. A. Stein, 2015, Mitigating artifacts in projection onto convex sets interpolation: SEG Technical Program Expanded Abstracts 2015, 3779–3783.

Stanton, A., and M. D. Sacchi, 2016, Efficient geophysical research in julia: CSEG GeoConvention, 1–3.

Sternfels, R., G. Viguier, R. Gondoin, and D. Le Meur, 2015, Multidimensional simultaneous random plus erratic noise attenuation and interpolation for seismic data by joint low-rank and sparse inversion: Geophysics, **80**, no. 6, WD129–WD141.

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, no. 3, 890–908.

Sun, D., S. Roth, and M. J. Black, 2010, Secrets of optical flow estimation and their principles: 2010 IEEE computer society conference on computer vision and pattern recognition, 2432–2439.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Taylor, H. L., S. C. Banks, and J. F. McCoy, 1979, Deconvolution with the l_1 norm: Geophysics, **44**, no. 1, 39–52.

Tomasi, G., and R. Bro, 2005, PARAFAC and missing values: Chemometrics and Intelligent Laboratory Systems, **75**, no. 2, 163 – 180.

Trad, D., 2008, Five dimensional seismic data interpolation: 78th Annual International Meeting, SEG, Expanded Abstracts, 978–981.

———, 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, no. 6, V123–V132.

Trad, D., T. J. Ulrych, and M. Sacchi, 2003, Latest views of the sparse Radon transform: Geophysics, **68**, no. 1, 386–399.

Trad, D., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with high-resolution time-variant Radon transforms: Geophysics, **67**, no. 2, 644–656.

Trickett, S., 2008, *F-XY* Cadzow noise suppression: SEG Technical Program Expanded Abstracts 2008, **27**, no. 1, 2586–2590.

Trickett, S., and L. Burroughs, 2009, Prestack rank-reducing noise suppression: Theory: SEG Technical Program Expanded Abstracts 2009, **28**, no. 1, 3332–3336.

Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank-reduction filtering for erratic noise: SEG Technical Program Expanded Abstracts 2012, 1–5.

———, 2013, Interpolation using hankel tensor completion: SEG Technical Program Expanded Abstracts 2013, 3634–3638.

Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: SEG Technical Program Expanded Abstracts 2010, **29**, no. 1, 3829–3833.

Trickett, S. R., 2003, *F-XY* eigenimage noise suppression: Geophysics, **68**, no. 2, 751–759.

Tucker, L. R., 1964, The extension of factor analysis to three-dimensional matrices: Contributions to mathematical psychology, 110–127.

Vermeer, G., 1990, Seismic wavefield sampling: Society of Exploration Geophysicsits.

Vermeer, G. J., 1998, 3-D symmetric sampling: Geophysics, **63**, no. 5, 1629–1647.

Verschuur, D. J., A. Berkhout, and C. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, no. 9, 1166–1177.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

Walczak, B., and D. Massart, 2001, Dealing with missing data: part I: Chemometrics and Intelligent Laboratory Systems, **58**, no. 1, 15–27.

Wang, B., N. Zhang, W. Lu, and J. Wang, 2019, Deep-learning-based seismic data interpolation: A preliminary result: Geophysics, **84**, no. 1, V11–V20.

Weglein, A. B., 1995, Multiple attenuation: Recent advances and the road ahead: SEG Technical Program Expanded Abstracts 1995, 1492–1495.

Wen, F., L. Chu, P. Liu, and R. C. Qiu, 2018, A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning: IEEE Access, **6**, 69883–69906.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87–V95.

Xu, Y., R. Hao, W. Yin, and Z. Su, 2015, Parallel matrix factorization for low-rank tensor completion: Inverse Problems and Imaging, **9**, no. 2, 601–624.

Yang, H., and Y. Hua, 1996, On rank of block Hankel matrix for 2-D frequency detection and estimation: IEEE Transactions on Signal Processing, **44**, no. 4, 1046–1048.

Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of exploration geophysicists.

Zhao, Q., Q. Du, X. Gong, and Y. Chen, 2018, Signal-preserving erratic noise attenuation via iterative robust sparsity-promoting filter: IEEE Transactions on Geoscience and Remote Sensing, **56**, no. 6, 3547–3560.

Zhou, X., C. Yang, H. Zhao, and W. Yu, 2014, Low-rank modeling and its applications in image analysis: ACM Computing surveys (CSUR), **47**, no. 2, 1–33.

Zoubir, A. M., V. Koivunen, Y. Chakhchoukh, and M. Muma, 2012, Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts: IEEE Signal Processing Magazine, **29**, no. 4, 61–80.

Zwartjes, P., and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: Geophysics, **71**, no. 6, V171–V186.

APPENDIX A

Low-rank approximation in seismic data

This section describes the mathematical principles that validate the reconstruction of seismic data via low-rank constraints. Chapter 2 shows the mathematical support for a 2D section with one linear dipping event. This appendix generalizes the concept. First, I introduce the 2D case with R linear events with distinct dips. Following, I consider a 3D volume with R linear events. Both cases form a trajectory matrix, or block-Hankel matrices to obtain a low-rank array. Next, I generalize to the multidimensional case. The full seismic wavefield depends on 5 dimensions, four space coordinates and the time coordinate. The principle of low-rankness of block-Hankel matrices is well-founded in the multidimensional case. However, the approach does not fully employ the data redundancy. Therefore, we consider rank-reduction approximation on the data tensor (Kreimer and Sacchi, 2012; Stanton et al., 2012; Stanton and Sacchi, 2013).

2D linear events

Following the scheme from the introduction, we now evaluate a section with R linear events with distinct dips. In frequency space, one can represent the section as a superposition of plane waves

$$S_n(\omega, x) = \sum_{r=1}^{R} W_r(\omega) \exp^{-i\omega p_r n \Delta x} \ .$$

Considering N_x consecutive equally spaced records, and dropping the dependency on frequency

$$\mathbf{S} = \mathbf{Z}\mathbf{W} \tag{A.1}$$

where $\mathbf{S} = (S_1\, S_2\, \cdots\, S_N)^T$, $\mathbf{W} = (W_1\, W_2\, \cdots\, W_R)^T$, and

$$\mathbf{Z} = \begin{pmatrix} \exp^{-i\alpha_1} & \exp^{-i\alpha_2} & \cdots & \exp^{-i\alpha_R} \\ \exp^{-i2\alpha_1} & \exp^{-i2\alpha_2} & \cdots & \exp^{-i2\alpha_R} \\ \vdots & \vdots & \ddots & \vdots \\ \exp^{-iN_x\alpha_1} & \exp^{-iN_x\alpha_2} & \cdots & \exp^{-iN_x\alpha_R} \end{pmatrix},$$

where each column of matrix $\mathbf{Z}$ is called a steering vector with elements $z_{l,m} = \exp^{-il\alpha_m}$, $l = 1, \cdots, N_x$ $m = 1, \cdots, R$, and $\alpha_m = \omega_o p_m \Delta x$. Assuming that the R dips are distinct and the number of events is smaller than the number of traces, the R columns are linearly independent. Then, the matrix is of rank R and any row is a linear combination of R other rows. In particular, row $R+1$ is

$$\begin{pmatrix} \exp^{-i(R+1)\alpha_1} & \exp^{-i(R+1)\alpha_2} & \cdots & \exp^{-i(R+1)\alpha_R} \end{pmatrix} = \\ (P_R\, P_{R-1}\, \cdots\, P_1) \begin{pmatrix} \exp^{-i\alpha_1} & \exp^{-i\alpha_2} & \cdots & \exp^{-i\alpha_k} \\ \exp^{-i2\alpha_1} & \exp^{-i2\alpha_2} & \cdots & \exp^{-i2\alpha_k} \\ \vdots & \vdots & \ddots & \vdots \\ \exp^{-iR\alpha_1} & \exp^{-iR\alpha_2} & \cdots & \exp^{-iR\alpha_k} \end{pmatrix}. \tag{A.2}$$

Multiplying both sides by vector $\mathbf{W}$ results in a linear recursive relation

$$S_{R+1} = (P_R\, P_{R-1}\, \cdots\, P_1) \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_R \end{pmatrix}. \tag{A.3}$$

One can obtain the elements of $(P_R\, P_{R-1}\, \cdots\, P_1)$ from the previous equation if the dips are known.

Let us now consider the trajectory matrix for data **S** with N_x channels

$$\mathbf{H} = \begin{pmatrix} S_1 & \cdots & S_R & S_{R+1} & \cdots & S_K \\ S_2 & \cdots & S_{R+1} & S_{R+2} & \cdots & S_{K+1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \\ S_{L_x} & \cdots & S_{R+L_x-1} & S_{R+L_x} & \cdots & S_{N_x} \end{pmatrix}, \tag{A.4}$$

where L_x and K follow the same criteria as for the one dip case. Considering the recursive relation described in equation A.3,

$$\mathbf{H} = \begin{pmatrix} S_1 & \cdots & S_R & \sum_{r=1}^{R} P_r S_{R+1-r} & \cdots & \sum_{r=1}^{R} P_r S_{K-r} \\ S_2 & \cdots & S_{R+1} & \sum_{r=1}^{R} P_r S_{R+2-r} & \cdots & \sum_{r=1}^{R} P_r S_{K+1-r} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \\ S_{L_x} & \cdots & S_{R+L_x-1} & \sum_{r=1}^{R} P_r S_{R+L_x-r} & \cdots & \sum_{r=1}^{R} P_r S_{N_x-r} \end{pmatrix}. \tag{A.5}$$

The columns $R+1$ to K of the Hankel matrix are a linear combination of the previous R columns. Therefore, the rank of the trajectory matrix equals R, the number of distinct dips in the signal. Figure A.1 shows a seismic section with 3 dipping linear events and the singular values of the corresponding Hankel matrix for frequency index 20.

3D linear events

Considering a 3D seismic volume with R distinct events, one can model the signal as

$$s(t,x,z) = \sum_{r=1}^{R} w(t - p_r x - q_r z)\,, \tag{A.6}$$

where t represents the time domain, and x, z represent the space domain. As in the previous section, $w(t)$ is a pulse. Considering the 1D Fourier transform, from the time domain to the frequency domain, the signal is

$$S(\omega,x,z) = \sum_{r=1}^{R} W_r(\omega) \exp^{-iwp_r x - iwq_r z} \tag{A.7}$$

where $\omega = 2\pi f$ is the temporal frequency. Each term in equation A.7 represents a sinusoidal or monochromatic plane wave, a special solution to the wave equation.

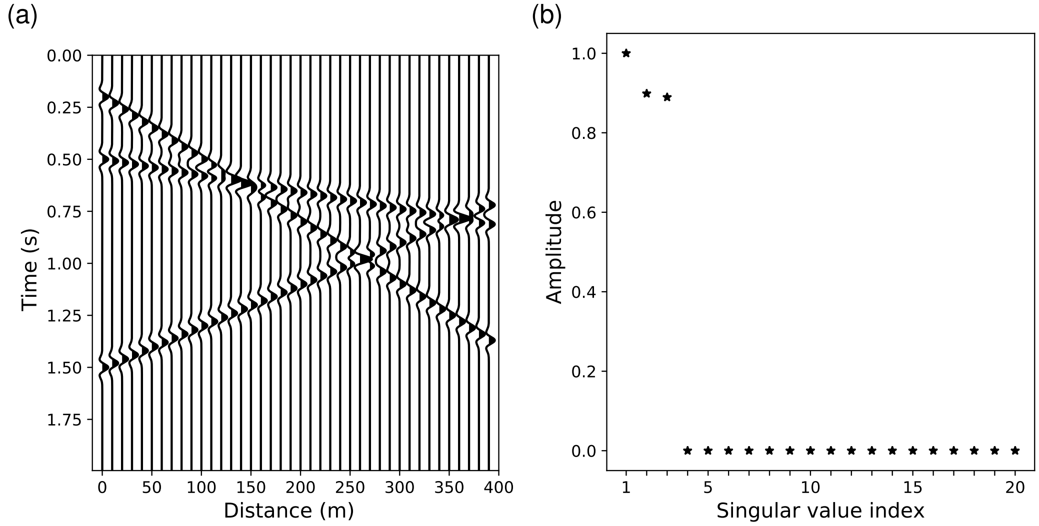

Figure A.1: Seismic data is low rank. (a) 2D seismic section with three dipping linear event. (b) The rank of the corresponding Hankel matrix equals three.

Analyzing one frequency ω_0, and assuming regularly sampled data, the signal at one channel is

$$S_{m,n} = \sum_{r=1}^{R} W_r \exp^{-i\alpha_r m - i\beta_r n} \tag{A.8}$$

where $x_m = m\Delta x$, $z_n = n\Delta z$, $\alpha_r = \omega_0 p_r \Delta x$, and $\beta_r = \omega_0 q_r \Delta z$.

In the case of $M \times N$ channels, one can describe the volume as a matrix

$$\mathbf{S} = \mathbf{YWZ}\,, \tag{A.9}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1R} & \\ y_{21} & y_{22} & \cdots & y_{2R} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{M1} & y_{M2} & \cdots & y_{MR} & \end{pmatrix}$$

$$\mathbf{W} = \mathrm{diag}(W_1, W_2, \cdots, W_R)$$

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1N} & \\ z_{21} & z_{22} & \cdots & z_{2N} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{R1} & z_{R2} & \cdots & z_{RN} & \end{pmatrix},$$

with $y_{mr} = \exp^{-i\alpha_r m}$, and $z_{rn} = \exp^{-i\beta_r n}$. Provided that the number of channels, M and N are larger than the number of distinct events R, $\text{rank}(\mathbf{S}) \leq R$. Besides, $\text{rank}(\mathbf{S}) = R$ if and only if the ranks of $\mathbf{Y}$ and $\mathbf{Z}$ are both equal to R. That is, the events have distinct dips in each direction.

Let us now consider the trajectory matrix for data $\mathbf{S}$

$$\mathbf{H} = \begin{pmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \cdots & \mathbf{S}_{M-K} & \\ \mathbf{S}_2 & \mathbf{S}_3 & \cdots & \mathbf{S}_{M-K+1} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{S}_K & \mathbf{S}_{K+1} & \cdots & \mathbf{S}_M & \end{pmatrix}, \tag{A.10}$$

where

$$\mathbf{S}_m = \begin{pmatrix} S_{m1} & S_{m2} & \cdots & S_{m(N-L)} & \\ S_{m2} & S_{m3} & \cdots & S_{m(N-L+1)} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ S_{mL} & S_{m(L+1)} & \cdots & S_{mN} & \end{pmatrix}. \tag{A.11}$$

The trajectory matrix $\mathbf{H}$ is a $K \times (M-K+1)$ Hankel block matrix, and $\mathbf{S}_m$ is a $L \times (N-L+1)$ Hankel matrix. In the case when $L = 1$ and $K = M$, the block Hankel matrix equals equation A.3. Then, the 2D case is a particular case of the multidimensional analysis.

Similar to equation A.9, one can describe the Hankel block matrix as

$$\mathbf{H} = \mathbf{B}_L \mathbf{W} \mathbf{B}_R\,, \tag{A.12}$$

where $\mathbf{B}_L = \left(\mathbf{Z}_L \; \mathbf{Z}_L\mathbf{Y}_d \; \cdots \; \mathbf{Z}_L\mathbf{Y}_d^{K-R}\right)^T$, $\mathbf{B}_R = \left(\mathbf{Z}_R \; \mathbf{Y}_d\mathbf{Z}_R \; \cdots \; \mathbf{Y}_d^{M-K}\mathbf{Z}_R\right)^T$, $\mathbf{Y}_d = \text{diag}(y_{r1})$,

$$\mathbf{Z}_L = \begin{pmatrix} z_{11} & z_{21} & \cdots & z_{R1} & \\ z_{12} & z_{22} & \cdots & z_{R2} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{1L} & z_{2L} & \cdots & z_{RL} & \end{pmatrix},$$

$$\mathbf{Z}_R = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1(N-L)} & \\ z_{21} & z_{22} & \cdots & z_{2(N-L)} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{R1} & z_{R2} & \cdots & z_{R(N-L)} & \end{pmatrix}.$$

The rank of the Hankel block matrix equals R if and only if the rank of E_R and E_L equal R. Following Yang and Hua (1996) and Hua (1992), the previous condition is valid if $M - R + 1 \geq K \geq R$ and $N - R + 1 \geq L \geq R$, and only if $KL \geq R$ and $(M - K + 1)(N - L + 1) \geq R$. Figure A.2 shows a 3D seismic volume with 3 dipping linear events and the singular values of the corresponding Hankel matrix for frequency index 20.

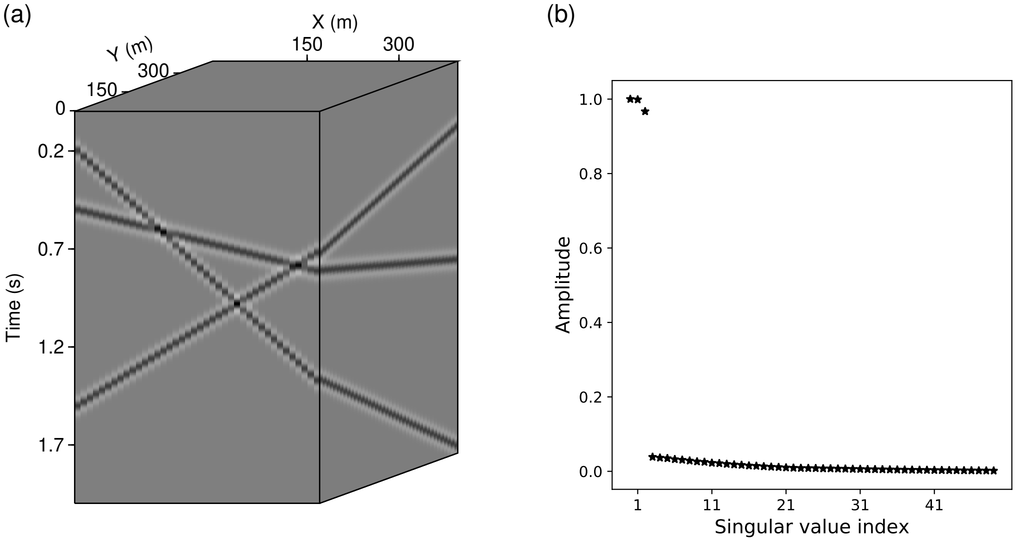

Figure A.2: Seismic data is low rank. (a) 3D seismic section with three dipping linear event. (b) The rank of the corresponding Hankel matrix equals three.

Multidimensional seismic data

The full seismic wavefield has four spatial dimensions in addition to time. One can describe the spatial coordinates combining the locations of the sources and receivers, or midpoints and azimuths. Even though such high-dimensionality presents visualization challenges, the low-rank properties of the data remains valid. Therefore, one can apply the block Hankel matrix principles to reconstruct the full wavefield.

However, rank-reduction on four dimensional data is computationally expensive and does not benefit from the data redundancy and relations in the different dimensions. Multidimensional algebra exploits such capabilities. Therefore, we embed the seismic dataset in a tensor.

A seismic volume with R linear events and four spatial coordinates can be modeled as

$$s(t, x_1, x_2, x_3, x_4) = \sum_{r=1}^{R} w(t - p_{1r}x_1 - p_{2r}x_2 - p_{3r}x_3 - p_{4r}x_4) \tag{A.13}$$

where p_{jr} are the ray parameters for the R events in x_j, $j = 1, 2, 3, 4$. Considering the 1D Fourier transform from time to the frequency domain

$$S(\omega, x_1, x_2, x_3, x_4) = \sum_{r=1}^{R} W_r(\omega) \exp(-i\alpha_r n_1 - i\beta_r n_2 - i\gamma_r n_3 - i\epsilon_r n_4) \tag{A.14}$$

where $W(\omega)$ is the frequency dependent amplitude of the events, ω is the frequency, $\alpha_r = \omega p_{1r}\Delta x_1$, $\beta_r = \omega p_{2r}\Delta x_2$, $\gamma_r = \omega p_{3r}\Delta x_3$, $\epsilon_r = \omega p_{4r}\Delta x_4$, and $x_j = n_j \Delta x_j$, $j = 1, 2, 3, 4$.

As a simplification, one can evaluate the model for one frequency ω_0. Dropping the frequency dependency and adopting $W(\omega_0) = 1$

$$S_{i_1,i_2,i_3,i_4} = \sum_{r=1}^{R} \mathbf{a}_{ri_1} \circ \mathbf{b}_{ri_2} \circ \mathbf{c}_{ri_3} \circ \mathbf{d}_{ri_4} \,, \tag{A.15}$$

where vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ represent $\exp -i\alpha_r i_1$, $\exp -i\beta_r i_2$, $\exp -i\gamma_r i_3$, and $\exp -i\epsilon_r \alpha i_4$, respectively. Equation A.15 resembles a CP decomposition with R terms.

Following Kolda and Bader (2009) and Kreimer et al. (2013), one can form the unfoldings of tensor $\mathcal{S}$ as

$$\begin{aligned}\mathbf{S}^{(1)} &= \mathbf{A}\left(\mathbf{D}\circ\mathbf{C}\circ\mathbf{B}\right)^T\\ \mathbf{S}^{(2)} &= \mathbf{B}\left(\mathbf{D}\circ\mathbf{C}\circ\mathbf{A}\right)^T\\ \mathbf{S}^{(3)} &= \mathbf{C}\left(\mathbf{D}\circ\mathbf{B}\circ\mathbf{A}\right)^T\\ \mathbf{S}^{(4)} &= \mathbf{D}\left(\mathbf{C}\circ\mathbf{B}\circ\mathbf{A}\right)^T\ .\end{aligned} \tag{A.16}$$

Focusing on the mode-1 unfolding,

$$\begin{aligned}\mathbf{S}^{(1)} &= \mathbf{A}\left(\mathbf{D}\circ\mathbf{C}\circ\mathbf{B}\right)^T\\ &= \mathbf{A}\left(\left[\mathbf{d}_1\otimes\mathbf{c}_1\otimes\mathbf{b}_1\cdots\mathbf{d}_R\otimes\mathbf{c}_R\otimes\mathbf{b}_R\right]\right)^T\ .\end{aligned} \tag{A.17}$$

The Kronecker product of $\mathbf{d}_r\otimes\mathbf{c}_r\otimes\mathbf{b}_r$, $r=1,\cdots,R$ is a vector. Therefore, it has rank 1. Assuming that the R events have different dips, then each of the products is linearly independent from the rest and the rank of the matrix equals R. Likewise, matrix $\mathbf{A}$ represents R independent events resulting in rank R. Considering the properties of the unfoldings,

$$\operatorname{rank}\left(\mathbf{S}^{(1)}\right) \leq \min\left(\operatorname{rank}\left(\mathbf{A}\right), \operatorname{rank}\left(\mathbf{U}\right)\right) \tag{A.18}$$

$$= \min(R,R) = R\,, \tag{A.19}$$

where $\mathbf{U} = \left(\left[\mathbf{d}_1\otimes\mathbf{c}_1\otimes\mathbf{b}_1\cdots\mathbf{d}_R\otimes\mathbf{c}_R\otimes\mathbf{b}_R\right]\right)^T$. Then, the rank of the mode-1 unfolding of $\mathcal{S}$ equals the number of unique events. The proof for the other unfoldings follow the same scheme, with a change in the order of the products. Therefore, $\mathcal{S}$ has 4-ranks equal to the number of independent. Figure A.3 shows the singular values corresponding to the unfoldings of a 5D seismic volume with 3 dipping linear events.

Besides, Kreimer et al. (2013) shows that the low-rankness of the tensor holds for events with variable dips, for example, parabolic events. The concept holds for decomposable events, that is, events separable in the spatial directions.

The previous analysis presents the bases of seismic reconstruction via rank-reduction.

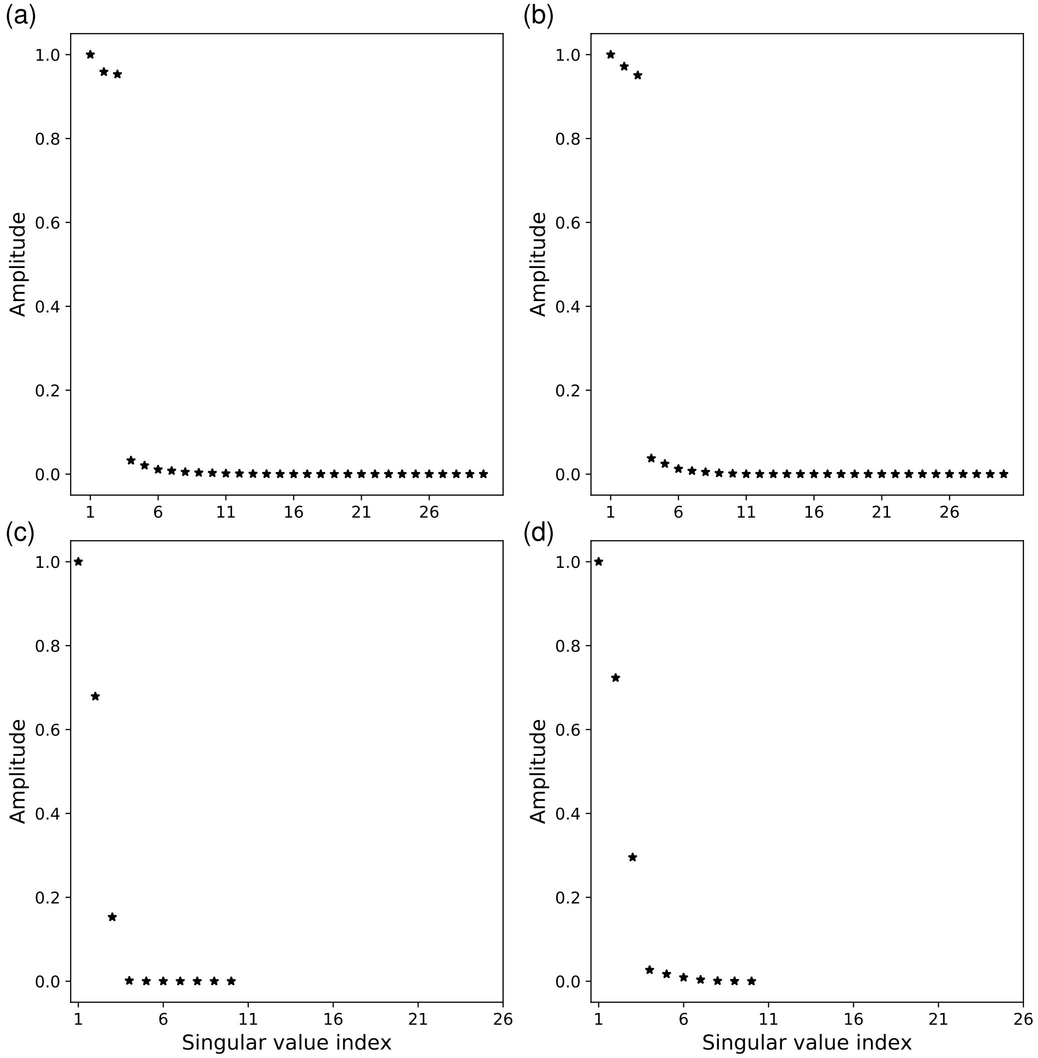

Figure A.3: Seismic data is low rank. Singular values of the unfolding of the modes of a 5D seismic volume.(a) Mode 1 unfolding. (b) Mode 2 unfolding. (c) Mode 3 unfolding. (d) Mode 4 unfolding.

APPENDIX B

Interpolation algorithms

The I-MSSA algorithm recovers a regular seismic volume from a set of observations that are located off-the-grid. To do so, it represents off-the-grid data as a linear combination of observations in the desired regular grid. In particular, we describe two operators, bilinear interpolation and 2D Kaiser window tapered sinc interpolator.

The interpolator estimates a value in an arbitrary location as a linear combination of gridded data. The irregular data in the observation grid is $U(\mathbf{r}_i)$, where $\mathbf{r}_i$ is the observed coordinate of the sample. Likewise, the regular data is given by $D(\boldsymbol{\xi}_k)$, where $\boldsymbol{\xi}$ is a coordinate in the regular mesh.

We assume that the point $\mathbf{r}_i$ is surrounded by tabulated points $\boldsymbol{\xi}_i$ that belong to the regular grid. The data in the observation grid takes the form

$$U(\mathbf{r}_i) = \sum_{k \in \mathcal{N}_i} W_k \, D(\boldsymbol{\xi}_k) \, . \tag{B.1}$$

The above is the forward interpolation operator where several gridded samples $D(\boldsymbol{\xi}_k)$ synthesize one off-the-grid sample via a weighted sum. The adjoint interpolation operator is a spraying operator (Claerbout, 1992) that distributes weighed copies of the off-the-grid sample to grid coordinates

$$\tilde{D}(\boldsymbol{\xi}_k) = W_k U(\mathbf{r}_i) \quad k \in \mathcal{N}_i \, . \tag{B.2}$$

The explicit form of the 2D bilinear interpolation operator is given by

$$\begin{aligned} U(\mathbf{r}_i) &= (1-t)(1-u)D(\xi_{x_1},\xi_{y_1}) + t(1-u)D(\xi_{x_2},\xi_{y_1}) \\ &+ tuD(\xi_{x_2},\xi_{y_2}) + (1-t)uD(\xi_{x_1},\xi_{y_2})\,, \end{aligned} \tag{B.3}$$

where

$$t = (r_x - \xi_{x_1})/\Delta\xi_x\,, \tag{B.4}$$

$$u = (r_y - \xi_{y_1})/\Delta\xi_y)\,, \tag{B.5}$$

where $\Delta\xi_x$ and $\Delta\xi_y$ are the x and y grid intervals, respectively. Therefore, the weights depend on the distance of the observation point to the regular coordinate. The adjoint operator distributes the data value in the four grid points. We describe the forward and adjoint operators in Algorithm 6.

Algorithm 6 Bilinear interpolator operator

```
Input: $\mathbf{r}$, $\boldsymbol{\xi}$
if adj = true, Input = $U$, Output = $\tilde{D}$
if adj = false, Input = $D$, Output = $U$
for $k = 1 : N_u$ do
    $ia \leftarrow \text{floor}((r_{x_k} - \xi_{ox})/\Delta\xi_x) + 1;\; ib \leftarrow\; ia + 1$
    $ja \leftarrow \text{floor}((r_{y_k} - \xi_{oy})/\Delta\xi_y) + 1;\; jb \leftarrow\; ja + 1$
    if $1 < ib \le N_x$ and $1 < jb \le N_y$ then
        $t \leftarrow (r_{x_k} - \xi_{x_{ia}})/\Delta\xi_x$
        $u \leftarrow (r_{y_k} - \xi_{y_{ja}})/\Delta\xi_y$
        if adj = true then
            $\tilde{D}(\xi_{x_{ia}},\xi_{y_{ja}}) += (1-t)\,(1-u)\,U(k)$
            $\tilde{D}(\xi_{x_{ib}},\xi_{y_{ja}}) += t\,(1-u)\,U(k)$
            $\tilde{D}(\xi_{x_{ib}},\xi_{y_{jb}}) += t\,u\,U(k)$
            $\tilde{D}(\xi_{x_{ia}},\xi_{y_{jb}}) += (1-t)\,u\,U(k)$
        else
            $U(k) = (1-t)(1-u)D(\xi_{ia},\xi_{ja}) + t(1-u)D(\xi_{ib},\xi_{ja}) + tuD(\xi_{ib},\xi_{jb}) + (1-t)uD(\xi_{ia},\xi_{jb})$
        end if
    end if
end for
```

A bilinear interpolator is very practical, but also simple. A more accurate interpolator for seismic data processing is the Kaiser window tapered sinc interpolator. The weights of the 2D interpolator are given by sinc function of limited spatial support (Fomel, 2001)

$$W(x) = \text{sinc}(\pi x) \frac{I_0\left(a\sqrt{1-x^2}\right)}{I_0(a)}\,, \tag{B.6}$$

where I_0 represents the zero-order modified Bessel function of the first kind, a is a parameter that needs to be empirically estimated, and x is the argument which is either t or u given by equations B.4 and B.5. Algorithm 7 shows an implementation of the Kaiser window tapered sinc interpolation for the 2D case.

Algorithm 7 2D Kaiser window tapered sinc interpolator

```
Input: $\mathbf{r}$, $\boldsymbol{\xi}$, assume a 7 point operator in $x$ and $y$.
if adj = true, Input = $U$, Output = $\tilde{D}$
if adj = false, Input = $D$, Output = $U$
for $k = 1 : N_u$ do
    $ia \leftarrow \text{floor}((r_{x_k} - \xi_{ox})/\Delta\xi_x) + 1$;
    $ja \leftarrow \text{floor}((r_{y_k} - \xi_{oy})/\Delta\xi_y) + 1$;
    for $i : ia - 3 : ia + 3$ do
        for $j : ja - 3 : ja + 3$ do
            $t \leftarrow (r_{x_k} - \xi_{x_i})/\Delta\xi_x$
            $u \leftarrow (r_{y_k} - \xi_{y_j})/\Delta\xi_y$
            if adj = true then
                $D(\xi_{x_i}, \xi_{y_j}) += U(k)\, W(t)\, W(u)$
            else
                $U(k) += D(\xi_{x_i}, \xi_{y_j})\, W(t)\, W(u)$
            end if
        end for
    end for
end for
```

APPENDIX C

Randomized QR

An essential component of the PMF algorithm is the rank-reduction step. In the original algorithm, rank-reduction of unfolded matrices was implemented via iterative matrix factorization (Xu et al., 2015). The latter is equivalent to applying rank-reduction to unfolded tensors via the SVD. In this appendix, we summarize the randomized QR decomposition adopted for rank-reduction, used throughout our work.

To simplify the notation we denote $\mathbf{Z}$ any of the matrices $\mathbf{Z}_{(k)}$, $k = 1, \ldots, N$. Let us consider $\mathbf{Z}$ of size $N_1 \times N_2$, and $\mathbf{\Omega}$ a matrix of size $N_2 \times p$ with columns formed by random normal vectors of unit variance. The first step of the randomized QR decomposition is to decrease the size of the original matrix $\mathbf{Z}$ by multiplication with the random matrix $\mathbf{\Omega}$ to form a new matrix

$$\mathbf{M} = \mathbf{Z}_{(k)}\mathbf{\Omega} \ . \tag{C.1}$$

Notice that $\mathbf{M}$ is size $N_1 \times p$. Halko et al. (2011) demonstrated that $\mathbf{M}$ captures most of the information in the range of the matrix $\mathbf{Z}$. In other words, the columns of $\mathbf{M}$ span a subspace of the image of the matrix $\mathbf{Z}$. Next, we compute the reduced QR decomposition of $\mathbf{M}$ to obtain an orthonormal basis $\mathbf{Q}$ and an upper-triangular matrix $\mathbf{R}$ such that

$$\mathbf{QR} = \mathbf{M} \ . \tag{C.2}$$

The low-rank approximation of the matrix $\mathbf{Z}$ is computed by applying the projection operator $\mathbf{Q}\mathbf{Q}^H$ to $\mathbf{Z}$:

$$\hat{\mathbf{Z}} = \mathbf{Q}\mathbf{Q}^H\mathbf{Z} \; . \tag{C.3}$$

By adopting the randomized QR algorithm, we gain computational efficiency in comparison to applying SVD or iterative matrix factorization. The efficiency arises from the reduction of the size of the original matrix $\mathbf{Z}$ by the random projection. However, this gain in computational performance is not the primary motivation for adopting the randomized QR decomposition. SVD computes the low-rank approximation closest to $\mathbf{Z}$ in the Frobenius sense. Incorrect knowledge of the actual rank of $\mathbf{Z}$ can lead to a suboptimal reconstruction. On the other hand, the random projection adopted in the randomized QR decomposition provides less tight bounds on the recovery of $\mathbf{Z}$. This concept is shown with an example. In Figure C.1 we show $\texttt{S/N}_o$ for synthetic examples consisting of 5D volumes of parabolic events contaminated with non-Gaussian noise as described in chapter 5. We have computed the $\texttt{S/N}_o$ in decibels as a function of rank and decimation for SVD and randomized QR decomposition for the robust PMF algorithm. Figure C.1 shows that the region where the PMF algorithm is capable of retrieving high-quality solutions is broader for the randomized QR decomposition than for the SVD algorithm. Similar results were also reported by Chiron et al. (2014) and Cheng and Sacchi (2016). In essence, the randomized QR decomposition is less sensitive to the precise knowledge of the rank of the matrix $\mathbf{Z}$ than the SVD decomposition. Last, it is necessary to mention that results associated to rank-reduction via iterative matrix factorization (Xu et al., 2015) are equivalent to those obtained with the SVD.

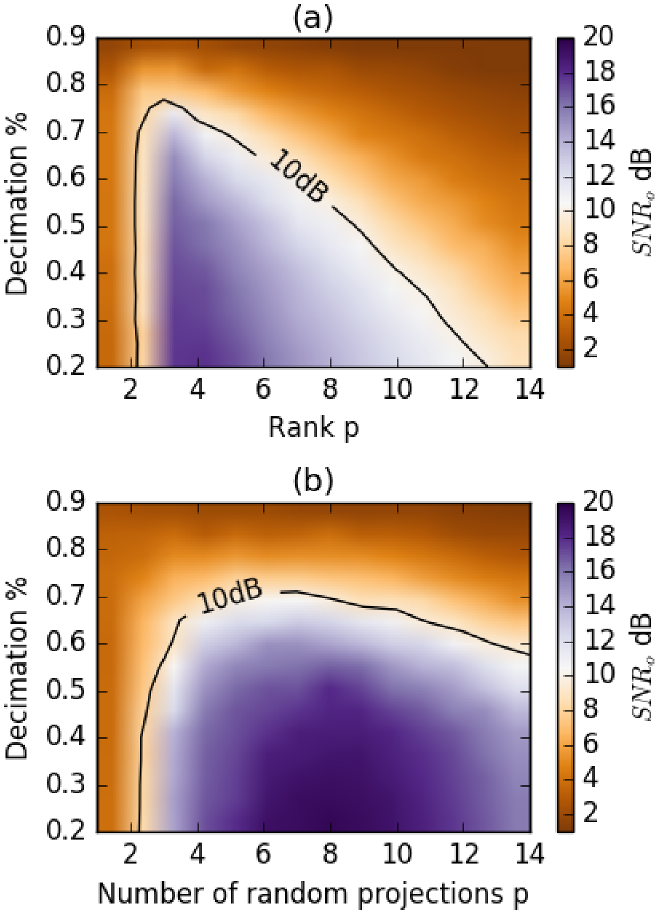

Figure C.1: Reconstruction quality measured via $\mathtt{S/N}_o$ for simulation with varying amount of data decimation and rank or number of random projections. (a) Robust PMF with rank-reduction implemented via SVD; rank p. (b) Robust PMF with rank-reduction implemented via randomized QR algorithm; number of random projections p .