Optimal seismic acquisition design, reconstruction and denoising

by

Yi Guo

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

Seismic acquisition constitutes a significant economic commitment, accounting for up to 80% of the overall cost of seismic exploration. This cost is intrinsically linked to the quantity of deployed sensors and sources, each carrying its own set of expenses related to acquisition, deployment, and maintenance. With increasing demands to shrink seismic acquisition expenses, minimize ecological impacts, and adhere to health, safety, and environmental (HSE) guidelines, innovative methodologies like compressive sensing (CS) seismic acquisition have emerged. These techniques aim at reducing the number of sensors while still fulfilling the aforementioned objectives. However, practical field constraints can hamper the random sampling strategies typically employed by CS.

Addressing the challenges of elevated operational costs, diminished efficiency, and low sampling density in seismic field acquisition, this dissertation concentrates on the formulation of optimal, cost-effective seismic acquisition layouts to address the challenge of maintaining data quality while utilizing fewer sensors, as compared to conventional high-density surveys. Within this context, the thesis elaborates on two central paradigms — CS seismic acquisition and optimal sparse sensing seismic acquisition — each applicable under different conditions of prior information availability and field operational constraints.

Chapter 2 discusses the CS theory and its application in seismic acquisition, considering data-free survey design. In the realm of data-driven methods, this work embarks on a nuanced examination of diverse strategies for optimization, encapsulated in Chapters 3 and 4. An over-complete pre-trained basis library is also incorporated for data-driven scenarios, facilitating computationally efficient and straightforward data reconstructions. Chapter

5 involves using a data-free objective function for seismic acquisition design. The optimal survey design strategies range from optimal sparse sensing employing QR-column pivoting to more advanced methodologies rooted in reinforcement learning (RL) and deep reinforcement learning (DRL). Noteworthy contributions include the development of a novel theoretical framework for synchronous spatiotemporal compression through design in CS, resulting in a theoretical acquisition cost reduction exceeding 50%. Additionally, an RL-based optimal acquisition design algorithm is introduced, achieving a theoretical reduction in acquisition costs exceeding 65% via the optimization of field sampling points. Particular emphasis is placed on RL and DRL techniques, which are operationalized through the formulation of a Markov decision process (MDP) model for sensor placement decisions.

Empirical validations across diverse application scenarios — including ocean bottom node (OBN) survey, simultaneous source acquisition, time-lapse studies for carbon dioxide storage monitoring, and vibroseis route design — indicate the high efficacy of the proposed methodologies in significantly reducing acquisition costs. The innovative approaches proposed herein hold transformative potential not only for the realm of seismic acquisition but also extend to other domains of geophysical exploration.

Preface

Part of the version of the work in chapter 2 of this thesis has been published in a journal paper: Lin R., Y. Guo, F. Carozzi and M. D. Sacchi, 2022, Simultaneous deblending and source reconstruction for compressive 3D simultaneous-source acquisition data via Interpolated MSSA (I-MSSA): *Geophysics*, **87**(6), V559-V570.

A version of the work in chapter 3 of this thesis has been published in conference abstracts: Y. Guo, R. Lin, and M. D. Sacchi, 2020, Sensor placement optimization for seismic data acquisition and shot reconstruction: *Geoconvention*, Y. Guo and M. D. Sacchi, 2020, Data-driven time-lapse acquisition design via optimal receiver-source placement and reconstruction: *SEG Technical Program Expanded Abstracts*, 66-70, and Y. Guo, R. Lin, and M. D. Sacchi, 2022, Seismic noise attenuation via learned and analytical basis functions: *Second International Meeting for Applied Geoscience & Energy*, 2867-2871.

A version of the work in chapter 4 of this thesis has been published in a journal paper: Y. Guo, R. Lin, and M. D. Sacchi, 2023, Optimal seismic sensor placement based on reinforcement learning approach: An example of OBN acquisition design: *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-12.

For the publications above, I was responsible for designing, developing and programming the algorithms, preparing data examples and the written portions. Dr. Mauricio Sacchi was the supervisory author and was involved in concept formulation and manuscript editing.

To my husband, Rongzhi Lin.

Acknowledgements

I extend my deepest gratitude to my supervisor, Professor Mauricio D. Sacchi, whose unwavering guidance, mentorship, and patience have shaped my scholarly journey. His consistent encouragement has invigorated my quest for meaningful research, complemented by invaluable feedback. Words fail to encapsulate the depth of my appreciation for his ceaseless encouragement, insightful critiques, and compassionate support, particularly during the unprecedented challenges posed by the pandemic and the birth of my child. His guidance has not merely illuminated my research trajectory but has also cast light upon my life's broader landscape.

I extend heartfelt appreciation to my examination committee members, Professor Jeff Gu and Professor Vadim Kravchinsky. Their constructive critiques, intellectual engagement, and consistent encouragement have greatly enriched my doctoral experience. Their involvement has been a privilege, and I am exceedingly fortunate to have had their expertise steer my academic endeavors.

A profound thank you goes to my colleagues in the Signal Analysis and Imaging Group (SAIG) at the University of Alberta. The collective intellectual camaraderie and personal support among the SAIG members have enlivened my academic sojourn, forging a sense of belonging amongst such remarkable geophysicists. Notable mentions include Wenlei Gao, Breno Figuereido Bahia, Fernanda Carozzi, Kristian Torres Bautista, Joaquin Acedo, Alejandro Quiaro, Atila Saraiva Q Soares, Dawei Liu, Hongling Chen, Xuan Ke, Shukui Zhang, Weichen Zhan, and many others for fostering a collegial and supportive environment.

My intellectual foundation has been further fortified by my course instructors, Dr. Jeff Gu, Dr. Aksel Hallin, Dr. Claire Currie, and Dr. Mauricio Sacchi. Their pedagogical contributions have provided the essential scaffolding upon which my research is built.

To my friends and fellow students—Jingchuan Wang, Wenhan Sun, Huifeng Zhang, Weiheng Geng, Yue Wang, Yunfeng Chen, Liuxi Chen, Ruijia Wang, Xiaowen Liu, Yiru Zhou, Tianyang Li, Tongzhen Guo, and others—I owe a bounty of thanks for making my stay at the University of Alberta both colorful and fulfilling.

Last but by no means least, my deepest gratitude is reserved for my family—my parents, parents-in-law, relatives, and especially my cousin, cousin-in-law, aunt, uncle, and grandmother—for their unwavering love and sage advice. Their steadfast support and guidance have been the backbone of my endeavor. A special accolade is reserved for my husband, Rongzhi Lin, whose boundless understanding, companionship, and encouragement have been my steadfast pillars. His presence in my life has been a constant source of strength and inspiration. Finally, my cherished daughter, Ella Lin, although very young, has already imbued my life with immeasurable joy and motivation—I look forward to the inspiration you will undoubtedly continue to provide.

Contents

List of Tables

List of Figures

List of abbreviations and symbols

ADMM	Alternating Direction Method of Multipliers
AI	Artificial Intelligence
ALFT	Antileakage Fourier Transform
ANN	Artificial Neural Network
BP	Basis Pursuit
CCUS	CO_2 Capture, Utilization and Storage
CNN	Convolutional Neural Network
CoSaMP	Compressed Sampling Matching Pursuit
CO_2	Carbon Dioxide
CRG	Common Receiver Gather
CS	Compressive Sensing
CSG	Common Source Gather
Dim	Dimension
DNN	Deep Neural Network
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
EM	Electromagnetic
EPOCS	Extended Projection Onto Convex Sets
FFT	Fast Fourier transform
FISTA	Fast Iterative Shrinkage Thresholding Algorithm
FWI	Full-Waveform Inversion

GA	Genetic Algorithm
GPR	Ground Penetrating Radar
HSE	Health, Safety, and Environment
I-MSSA	Interpolated-Multichannel Singular Spectrum Analysis
MC	Mutual Coherence
MDP	Markov Decision Process
ML	Machine Learning
MSE	Mean Squared Error
MSSA	Multichannel Singular Spectrum Analysis
MT	Magnetotellurics
MWNI	Minimum Weighted Norm Interpolation
NP	Nondeterministic Polynomial
NOUS	The Nonuniform Optimal Sampling
OBN	Ocean Botton Node
OMP	Orthogonal Matching Pursuit
OVT	Offset Vector Tile
PCA	Principal Component Analysis
PEF	Prediction-Error Filter
POCS	Projection Onto Convex Sets
POD	Proper Orthogonal Decomposition
PPO	Proximal Policy Optimization
RE	Reconstruction Error
RI	Restricted Isometry
RL	Reinforcement Learning
ROM	Reduced Order Models
RTM	Reverse Time Migration
SA	Simulated Annealing
SNR	Signal-to-Noise Ratio
SRME	Surface-Related Multiple Elimination

SVD	Singular Value Decomposition
TD	Temporal Difference
VSP	Vertical Seismic Profiling
WEM	Wave-Equation Migration

$\otimes$	Kronecker product
$*$	Vector/Matrix multiplication
β^-, β^+	Hyperparameter
d	Scalar
d_{ij}	Elements of a matrix
$\mathbf{d}, \hat{\mathbf{d}}$	Vector
$\mathbf{D}, \hat{\mathbf{D}}$	Matrix
$\mathcal{D}$	Tensor
$\boldsymbol{\mathcal{F}}$	Fourier transform
$\boldsymbol{\mathcal{F}}^{-1}$	Inverse Fourier transform
$\mathbb{R}$	Data set
$\mathscr{T}$	Sampling/Extraction operator
$\|\|\cdot\|\|_0$	ℓ_0 norm
$\|\|\cdot\|\|_1$	ℓ_1 norm
$\|\|\cdot\|\|_2^2$	ℓ_2 norm
$\|\|\cdot\|\|_p^p$	ℓ_p norm
$\|\|\cdot\|\|_F^2$	The Frobenius norm
$\|\cdot\|$	Absolute value
μ, μ_1, μ_2	Trade-off parameter
$\mathscr{W}$	The interpolation operator
$\boldsymbol{\mathcal{W}}$	The windowing operator (Extract local window data)
$\boldsymbol{\mathcal{W}}^*$	The adjoint operator of $\boldsymbol{\mathcal{W}}$ (Synthesize local window data)
x_i	Elements of a vector
$\xi(x_k, y_k)$	Coordinates of irregular-grid point
$\eta(\hat{x}_i, \hat{y}_j)$	Coordinates of regular-grid point

CHAPTER 1

Introduction

1.1 General background

In the realm of Earth sciences, geophysics stands as an important discipline that endeavors to elucidate the intricacies of Earth's interior. This is primarily achieved through the deployment of numerical methodologies, predominantly anchored in the fundamental laws of physics. Rather than relying on direct observations, geophysicists predominantly employ indirect measurements to retrieve unseen subsurface properties (Tarantola, 2005). Over the years, sophisticated techniques has been developed, allowing geophysicists to quantitatively evaluate these enigmatic subsurface geological formations and properties (Kearey et al., 2002; Lowrie and Fichtner, 2020). In general, subsurface structures and properties are inferred from measurements recorded on Earth's surface.

Among all the geophysical methods, seismic exploration is a key technique for imaging geological formations and identifying potential reservoirs of hydrocarbon accumulations. Artificially generated seismic disturbances propagate waves into the Earth's subsurface, typically from either explosive sources or mechanical vibratory mechanisms. These seismic waves subsequently undergo reflection at various geological interfaces. Ascending back to the Earth's surface, the reflected waves are captured by arrays of sensors[1]. Though rich in information, these raw data require advanced, computationally demanding signal processing and inversion algorithms to accurately delineate the subsurface geological structures.

While the primary focus of this dissertation lies in seismic methodologies, it is worth noting that another category of geophysical approaches—commonly referred to as potential field

[1] In the context of seismic data acquisition, sensors typically refer to electromagnetic transducers strategically positioned on the Earth's surface, or hydrophones situated beneath the oceanic surface.

methods (Kearey et al., 2002)—exists but is not the subject of investigation in this thesis. These potential field techniques hinge on the perturbation-induced distortions of potential fields within the subsurface, attributable to variations in specific physical properties. For instance, gravity prospecting methods quantify distortions in the gravitational potential field, which are linked to the heterogeneity in subsurface density distributions. Analogously, electrical and magnetic methods capture distortions in their respective fields related to subsurface fluctuations in parameters such as resistivity and magnetic susceptibility. Other valuable geophysical techniques such as electromagnetic (EM), magnetotellurics (MT), and ground penetrating radar (GPR), among others, also play a crucial role, each contributing significantly to the broader field of geophysical exploration (Kearey et al., 2002; Lowrie and Fichtner, 2020).

The workflow in applied seismology can be conceptually divided into four critical stages. The initial stage, termed as the *Data Acquisition* stage, encompasses the strategic planning and deployment of seismic sources and receivers, followed by the collection of field data. Subsequently, the *Data Processing* stage is embarked upon, wherein techniques are employed to enhance the signal-to-noise ratio (SNR) and conform the acquired data to a uniformly spaced, dense grid—prerequisites for further processing and imaging. This leads to the *Imaging* stage, where preprocessed seismic data are mapped into subsurface images through the application of wave-equation-based algorithms. The final stage is the *Interpretation* stage, during which these subsurface images are synergistically analyzed with geological data to delineate prospective exploration targets. It should be noted that the financial expenditure on seismic data acquisition alone constitutes approximately more than 80% of the overall costs across these stages.

The focal point of my research pertains to the realm of seismic exploration, specifically concentrating on aspects related to seismic data acquisition and the subsequent data reconstruction and denoising during the processing stage. In the ensuing sections, I provide a detailed explanation of the seismic reflection method and subsequently delve into discussions on diverse strategies involved in the seismic acquisition design.

1.1.1 The seismic exploration method

Energy stands as the backbone of a nation's sustainable economic and societal progression, being deeply intertwined with people's day-to-day lives. Within the realm of the petroleum industry, exploration geophysics holds considerable importance. Among its various tools, seismic exploration emerges as the main technique in oil and gas geophysical exploration. It offers pivotal insights for oil and gas reservoir development through stages encompassing field data acquisition, processing, imaging, and interpretation. However, the industry is undergoing a transformative shift, especially in the face of the transition to the green economy.

Oil companies are tasked with meeting the dual objectives of reducing costs while improving seismic data quality. As seismic acquisition represents one of the most significant capital investments in the oil and gas geophysical exploration process, it is at the forefront of this revolutionary change. Companies are increasingly compelled to explore innovative seismic acquisition methods that not only fulfill the high-data-density requirements but also do so in a cost-effective manner.

The seismic exploration technique plays a critical role in identifying, exploiting, and monitoring resources such as hydrocarbon reservoirs and in environmental and geotechnical assessments of the near-surface (Yilmaz, 2021). The seismic reflection method employs sources to generate elastic waves propagating into the Earth's subsurface, and these waves undergo transmission and reflection when encountering geological interfaces that separate materials with varying elastic properties and densities. A network of receivers captures the upward-reflected waves. The target subsurface area is illuminated with these waves, and the resultant data is processed numerically to produce detailed images. Traditionally, the seismic exploration method encompasses four stages:

1. **Data acquisition:** During this initial stage, parameter selection is conducted with precision, ensuring compliance with the sampling criteria established by the Nyquist-Shannon theorem. As detailed in Vermeer (1990, 2012), this process can occur on land or at sea. For terrestrial data gathering, sources such as vibroseis or dynamite are employed, and the resultant signals are captured by stationary geophone stations on the Earth's surface. Offshore acquisition typically involves using an air gun, an explosive energy source towed by a vessel, along with a cable or streamer equipped with hydrophones. Moreover, Ocean Bottom Node (OBN) techniques are frequently utilized for offshore data collection. These involve deploying sensor arrays on the ocean floor, a practice that has become integral to reservoir monitoring. Additionally, arranging seismic sources and receivers in a linear format yields a 2D seismic profile, commonly known as 2D seismic acquisition. In contrast, contemporary techniques frequently employ spatial arrays of receivers and sources for 3D seismic acquisition, which is instrumental in capturing the seismic responses from three-dimensional geological structures. Overall, the deployment of receivers and sources can be categorized into three primary methodologies: traditional, compressive sensing, and optimal sparse sensing seismic acquisitions. Each method offers distinct approaches to capturing seismic data, catering to different requirements and objectives in seismic exploration.

2. **Data processing:** This stage aims to rectify inconsistencies originating from the spatial positioning of the sources and receivers. Techniques for incoherent and coherent noise reduction, including advanced denoising methods investigated in my doctoral research, are applied. These methods are elaborated upon in Chapters 2 and 3 and

in Appendix B. Additionally, various reconstruction techniques, employing a range of algorithms and approaches, are used to fill in the gaps in the collected data. The objective is to maintain the fidelity and resolution of the data while making it suitable for more sophisticated processing and interpretation. Data reconstruction methods, another focus of my doctorate, are also included at this stage.

3. **Imaging:** Imaging is primarily used for outlining subsurface structural boundaries. Early imaging techniques were grounded in the Huygens principle and the Kirchhoff integral theorem (Gray et al., 2001). Modern imaging methods, however, have evolved to incorporate either one-way or two-way acoustic or elastic wave equations, which are linearized using the Born approximation, as discussed in Sava and Hill (2009). These advanced imaging techniques not only demarcate structural boundaries but also provide insights into the properties of rock materials, which are vital for the subsequent stage of interpretation.

4. **Interpretation:** In this concluding stage, there is a thorough interpretation of the subsurface images obtained from the imaging process. However, it's important to note that these images alone do not reveal the underlying geological processes, such as the development of structural features, the routes of fluid migration, or the mechanisms of reservoir accumulation. To effectively pinpoint areas rich in hydrocarbons, seismic imagery must be integrated with geological knowledge. This comprehensive interpretation is further enhanced by incorporating regional geological data, analysis of core samples, and properties of formations derived from well logs, as detailed in Brown (2011), which ensures accurate understanding of the subsurface geological structures.

My research is primarily concentrated on reducing the cost associated with seismic exploration, encompassing facets such as minimizing acquisition time, employing compressed spatial sampling density, designing optimal sampling locations, and undertaking associated data processing. One of the principal contributions of my work is the integration of optimal acquisition theory to fine-tune sampling points. This approach can potentially significantly compress both the time and financial resources required for field seismic acquisition. Moreover, the *Ecoseis* concept, introduced by Naghizadeh et al. (2023), presents a notable strategy that underscores the importance of conducting seismic explorations in an environmentally conscious manner. This approach prioritizes the reduction of ecological impact while simultaneously upholding the quality and efficiency of the data collected. In line with a commitment to sustainability, my research also concentrates on seismic acquisition methods with a minimal environmental footprint. This focus is especially pertinent as we navigate the shift toward renewable energy sources and anticipate advancements in Carbon Capture, Utilization, and Storage (CCUS) technologies.

1.1.2 Conventional seismic acquisition and problems

During the seismic data acquisition process, discrete data sets are gathered. These data sets are representations of continuous, finite-energy wavefields and can span up to five dimensions, including time. The acquisition concerns an analog spatiotemporal wavefield with time in the order of seconds and length in the order of kilometers, and the sampling intervals are of the order of milliseconds and meters.

Seismic exploration methods often gather vast amounts of data mined for information during processing. While this strategy has proven highly effective, pursuing higher-resolution images in more complex regions of the Earth exposes inherent flaws in current standard workflows. The primary challenge is the "curse of dimensionality" (Herrmann et al., 2012), exemplified by Nyquist's sampling, which places an increasing burden on the acquisition and processing systems, especially as the scale and the required resolution of the survey areas expand.

In traditional seismic acquisition frameworks, adherence to the Nyquist-Shannon sampling theorem (Oppenheim, 1999) is imperative for discretizing data in both time and space. This often necessitates a dense, and consequently expensive, arrangement of sources and receivers on a regular or quasi-regular grid (Monk, 2020). The rationale for such a deployment takes into account various financial, operational, and regulatory considerations, which can differ in real-world scenarios. The quest for high-resolution subsurface images places significant demands for data density, thereby escalating the costs associated with increased numbers of sources and receivers. In such contexts, the imperative for cost-efficiency becomes accentuated. Moreover, these operations inherently involve specific health, safety, and environmental (HSE) risks in the field. Given these conflicting requirements, traditional seismic acquisition methods increasingly appear suboptimal. It becomes essential, therefore, to explore more resource-efficient approaches to seismic data acquisition, leveraging computational capabilities to reconstruct densely and regularly sampled data from undersampled datasets.

Instead of relying on conventional seismic acquisition techniques in the field, exploring optimal acquisition strategies can offer substantial and significant advantages for seismic exploration and production. This approach not only holds immense potential for enhancing economic growth but also plays a critical role in safeguarding national energy security. In contrast to traditional seismic surveys, which prioritize regular and dense spatial sampling, this dissertation aims to relax these sampling constraints to improve cost-effectiveness and minimize environmental impact. Specifically, the research focuses on determining optimal locations for sources and receivers using fewer numbers than conventional methods would require. The central question is how to design an acquisition setup with fewer sources and

receivers that still provide data of sufficient quality for reliable reconstruction, comparable to what would be obtained through a regular, dense (notably more expensive) survey.

1.2 Acquisition design: an overview

In the seismic exploration industry, the acquisition of high-quality seismic data is crucial for accurately identifying and characterizing subsurface geological features. Traditional seismic acquisition design methods rely on human expertise and trial-and-error approaches, which can be time-consuming, costly, and may not necessarily result in the optimal survey design.

Figure 1.1 shows three possible directions for low-cost seismic acquisition considering acquisition time, data compression, and strategic sampling design. For brevity, the single source to multi-source can be realized by simultaneous source acquisition, and the acquisition time is reduced by the factor of the number of sources that work simultaneously. The transition from the regular to the irregular grid is discussed in section 2.3.2, and part of the sampling points can be randomly removed, which is a compression both in time and space. The last case concerns random sampling to optimal sampling, which uses fewer sampling points for low-cost acquisition.

This section briefly illustrates the technology of compressive sensing (CS), machine learning (ML), optimal sparse sensing, and their adaptation in seismic acquisition design. Then, the connection between seismic acquisition and reconstruction is explained. Different sampling schemes are also provided for comparison.

1.2.1 Compressive sensing seismic acquisition

Conventional seismic acquisition relies on the Nyquist sampling theorem, which mandates a high-density, equidistant, and, consequently, high-cost data collection approach. In contrast, compressive sensing (CS) (Baraniuk, 2007; Candes and Tao, 2006; Candes et al., 2006) has emerged as a groundbreaking methodology that challenges the constraints imposed by the Nyquist theorem. CS essentially focuses on recovering un-acquired signals through random sampling coupled with sparse transformations. CS provides a novel approach to seismic data acquisition and processing, specifically designed for datasets that are traditionally viewed as undersampled.

CS has been successfully adopted in seismic acquisition. For instance, in the harsh winter conditions of the North Sea, ConocoPhillips successfully applied an irregular network based on CS theory to enhance data collection efficiency while mitigating the adverse impacts of

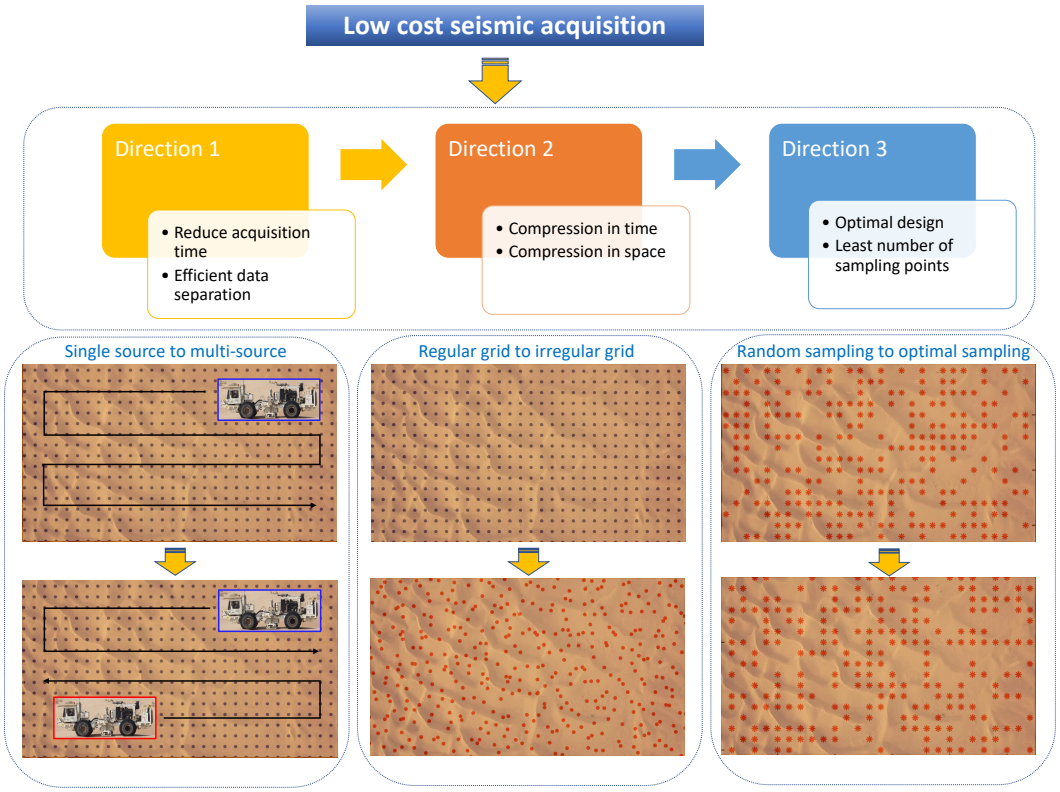

Figure 1.1: Possible approaches for low-cost seismic acquisition.

challenging offshore climates (Mosher et al., 2017). This approach enabled safe, swift, high-quality ocean bottom node (OBN) acquisition. Moreover, through judicious survey design that did not compromise imaging quality, costly nodes were minimized, saving millions of dollars in acquisition costs. This reduced expenditures and furnished practical insights for the broader adoption of this acquisition technology in future operations. In another example, ConocoPhillips executed a non-regular onshore seismic survey in Alaska's North Slope (Mosher et al., 2017). The survey met high-quality, high-density data requirements while being completed within a condensed timeframe of 30 days. Additionally, the fewer sampling points meant that the environmental impact on the permafrost regions was significantly reduced, leading to substantial cost savings. These successful case studies are transformative, altering the traditional mode of seismic acquisition and paving the way for the evolution of low-cost, high-quality seismic acquisition technologies (Brown et al., 2017; Mosher et al., 2017).

1.2.2 Machine learning assisted seismic acquisition

In the swiftly advancing field of machine learning (ML), numerous applications have emerged within exploration geophysics, predominantly focusing on aspects like seismic data processing, imaging, and interpretation. However, the utilization of ML for optimizing seismic data acquisition remains relatively unexplored. There are a few sporadic instances where attempts have been made, such as the work done in 2019 by Blacquiere et al., which incorporated artificial intelligence (AI) through convolutional neural networks (CNNs) and genetic algorithms (GA) to fine-tune acquisition parameters (Blacquiere and Nakayama, 2019).

Reinforcement learning (RL) theory offers a foundational perspective (Sutton and Barto, 2018), deeply anchored in psychological (Romanes, 1883) and neuroscientific (Schultz et al., 1997) views on animal behavior, about how agents can optimize their interaction with the environments. However, for this theory to be effective in real-world scenarios, agents face a challenging mission: extracting efficient environmental representations from high-dimensional sensory data and leveraging this to apply past experiences to new situations (Mnih et al., 2015). Impressively, creatures like humans manage this task by seamlessly merging RL with layered sensory processing systems (Fukushima, 1980; Serre et al., 2005). This synergy is further supported by numerous neural studies showing solid correlations between signals from dopaminergic neurons and the algorithms used in temporal difference RL (Schultz et al., 1997). Leveraging the latest advancements in deep neural network (DNN) training, Mnih et al. (2015) introduced an innovative artificial agent called a deep Q-network (DQN) designed to master effective strategies straight from intricate sensory data through end-to-end RL.

Despite its close affinity with AI, RL remains an underexplored methodology within the context of exploration geophysics. This is particularly curious given the inherently dynamic and decision-centric nature of seismic acquisition, which seems to be a natural fit for RL's optimization capabilities. Thus, Guo et al. (2023) first proposed adopting RL for seismic acquisition design, and the application in the OBN survey demonstrates that fewer sensors at optimal positions can be an alternate solution for conventional acquisition after proper reconstruction.

Other than designing the acquisition layout, ML customarily exploits dominant features in a dataset for tasks such as classification and prediction (Kotsiantis et al., 2007; Osisanwo et al., 2017) to motivate the reconstruction aspect. Often, these prominent features can be discerned through dimensionality reduction methods like proper orthogonal decomposition (POD) (Pinnau, 2008; Lu et al., 2019; Mendible et al., 2020), which is adopted for the pre-learned basis library in section 1.4.

1.2.3 Optimal sparse sensing seismic acquisition

Optimal sparse sensing seismic survey endeavors to delineate the optimal survey layout for seismic acquisition, with a dual objective: diminishing the acquisition costs and minimizing the sampling points used. Despite these reductions, the resulting data is designed to match the quality of traditionally acquired datasets, ensuring its viability for subsequent imaging and interpretation tasks. Solving for the optimal positioning of sources and receivers, based on a specific optimization objective, has been demonstrated to be a nondeterministic polynomial (NP)-hard problem, which necessitates an exhaustive brute-force search across all feasible combinations of source-receiver placements on a designated grid (Brunton et al., 2016; Manohar et al., 2018; Guo and Sacchi, 2020).

In practical scenarios, I dig into the design of real-world acquisition settings using optimal sensing. Figure 1.2 showcase the standard versus optimal arrangements for sources and receivers. In the conventional approach, sources and receivers are uniformly and densely positioned. However, in the optimal setting, sources and receivers are strategically placed only at locations that are deemed optimal. By requiring fewer sources and receivers, not only are equipment costs reduced, but fieldwork durations are also shortened. Thus, the total acquisition expenses are lower than those in conventional surveys.

Considering whether prior data from the target area is available or not, this thesis introduces both data-free and data-driven survey design. In the data-driven realm, an optimized acquisition scheme founded on optimal sensor placement theory is adapted to seismic survey design. Unlike CS, which employs a random sampling approach, this theory aims to refine

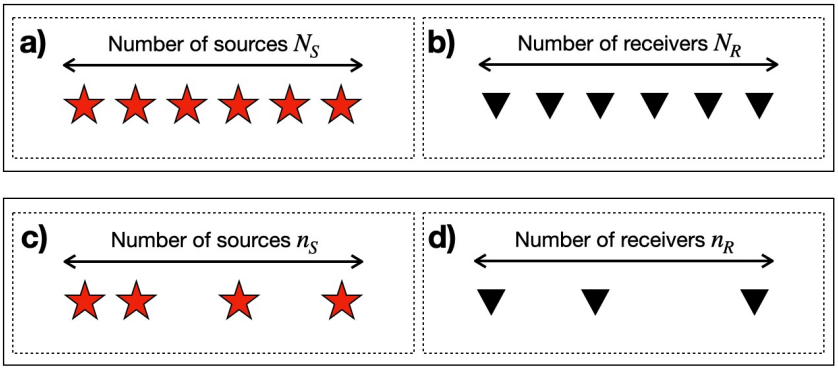

Figure 1.2: Depiction of the source and receiver placements for both the traditional and optimal survey designs. The red stars signify source locations, while the black triangles denote receiver positions. (a) and (b) display the layouts for the source and receiver in the conventional acquisition geometry, respectively. Conversely, (c) and (d) exhibit the layouts for the source and receiver in the optimal acquisition configuration, respectively.

the random sampling distribution into an optimal one. The objective is to achieve data acquisition with even fewer sampling points than what CS typically requires.

The crux of the issue in the optimal design of efficient, cost-effective seismic acquisition boils down to two core questions: How to achieve highly compressed data, and how to reconstruct the compressed data with high fidelity? To address these fundamental queries, optimal seismic data acquisition inherently involves a balancing act, where one has to make critical trade-offs between the number of sampling points and the quality of the reconstructed data. Given the complex nature of this problem, this thesis delves into three theoretical methods for acquisition design:

- Data-driven approaches to acquisition design (corresponding to Chapter 3): This method capitalizes on empirically collected data to inform the acquisition strategy and aims to utilize the least number of sensors.
- Reinforcement learning-based acquisition design (corresponding to Chapter 4): In this approach, reinforcement learning (RL) algorithms are used to navigate the complexities of sensor placement and other acquisition parameters. The goal is to optimize the acquisition process based on real-time feedback dynamically. It aims to use data analytics to identify an efficient and cost-effective design that does not compromise data quality.
- Deep reinforcement learning-based acquisition design (corresponding to Chapter 5): This method extends the capabilities of RL by incorporating the power of deep learning

algorithms. It allows for more nuanced optimization and provides additional layers of complexity, enabling more accurate and efficient seismic data acquisition with the least number of sensors and not compromising data quality.

By pioneering these methods, this study aims to achieve two primary objectives:

- Optimization of acquisition design: Each method is rigorously tested to determine its effectiveness in reducing the number of sampling points needed. The ultimate goal is to establish a new paradigm in seismic data acquisition that is more cost-effective and efficient than existing CS-based methods.
- High-fidelity data reconstruction: Concurrent with the optimization of acquisition design, the research also focuses on developing robust techniques for reconstructing the acquired seismic data. This ensures that the quality of the subsurface images generated is not compromised despite the reduction in sampling points.

The summary map (Figure 1.3) of the research puts forward the proposed three major technical routes and four significant applications of this thesis.

Each idea presents a different balance between the number of sampling points used and the quality of the data reconstructed, providing a comprehensive set of options for optimizing seismic data acquisition. The three technique routes are:

1. Minimize sampling points, flexibility in data quality:
    - Goal: Identify the least number of sampling points needed for acquisition.
    - Premise: Data quality and density requirements are not strict, allowing for greater flexibility.
    - Approach: Optimal sparse sensing illustrated in Chapter 3. The primary focus is on minimizing the number of sampling points used, even if this compromises some level of data quality.
2. Quality-prioritized design with reduced sampling:
    - Goal: Ensure that the quality of the reconstructed data matches that of a conventional survey.
    - Premise: Data quality must be maintained, but the number of sampling points can be adjusted to different requirements.
    - Approach: Optimal design based on RL approach shown in Chapter 4. Maintain high-quality data while seeking to reduce the number of sampling points used compared to traditional methods.

Figure 1.3: Three approaches and four applications proposed in this thesis in the context of low-cost seismic acquisition.

3. High-quality data with minimum sampling points:

    - Goal: Achieve high-quality reconstructed data with the fewest sampling points possible.
    - Premise: This idea is contingent upon the success of the first two routes.
    - Approach: Optimal design based on DRL theory demonstrated in Chapter 5. Strive for the highest standard of optimal acquisition design by combining minimized sampling points with maintained data quality.

Once the methods and algorithms for the aforementioned three approaches are fully developed, I proceed to validate them by applying the optimal acquisition design strategy across four distinct seismic acquisition frameworks:

1. Simultaneous source acquisition: Traditional seismic surveys often require large arrays of sources and sensors, making them expensive and time-consuming. Optimizing the acquisition process for simultaneous sources can reduce the number of required components, thereby reducing both the time and financial investments required (Oropeza and Sacchi, 2011; Li et al., 2013; Mosher et al., 2014b, 2017; Lin and Sacchi, 2020; Lin et al., 2022c). The corresponding examples are shown in Chapter 3.

2. Carbon dioxide geological storage monitoring: Given the urgency to combat climate change, monitoring carbon dioxide (CO_2) storage and migration in geological formations is a priority. An optimized seismic acquisition design will help in quicker and more accurate monitoring of CO_2 storage sites, thereby contributing to the global effort to reduce greenhouse gas emissions (Davis et al., 2003; Arts et al., 2005; Chadwick et al., 2004, 2005, 2009, 2010; Lumley, 2010; Fabriol et al., 2011). The corresponding examples are shown in Chapter 3.

3. Ocean bottom node acquisition: The offshore oil and gas sector often faces numerous challenges, including high costs and logistical complexities in seismic data collection. Applying the optimized acquisition techniques in OBN settings aims to reduce the costs associated with seismic exploration substantially (Mosher et al., 2017; Monk, 2020). The corresponding examples are shown in Chapter 4.

4. Vibroseis route design: Using the least number of sensors and putting them at the optimal positions to find the shortest path moving through all the chosen sensors possess many advantages, i.e., the corresponding acquisition cost is lower and the field working time is reduced, which is essential for HSE concern (Mosher et al., 2017). The related examples are shown in Chapter 5.

In summary, examining the complete research workflow, each element — from the initial scientific questions to the introduction of three distinct technical approaches and, ultimately, the application to four different real-world scenarios — is interconnected and progressively deepens the understanding. Consequently, this thesis not only provides a novel perspective for seismic acquisition design but also serves as a reference that can inspire similar work in other disciplines requiring optimal acquisition designs.

1.3 Different sampling schemes

Due to the ever increasing demands for higher resolution, achieving complete sampling becomes economically and practically infeasible. As a result, data is often sampled below the Nyquist rate, which usually corresponds to periodic undersampling of sources or receivers for the spatial coordinates. Unfortunately, such undersamplings can introduce significant artifacts. To address this problem, recent research efforts have been focused on developing enhanced sampling strategies. These strategies often involve randomizing the spatial placements of sources and receivers.

The theorem of sampling presents a subject matter of significant relevance in the realms of physical sciences and engineering. There are two primary classifications for sampling techniques: uniform and nonuniform. Uniform sampling involves the regular collection of a signal at a consistent interval, which allows for signal reconstruction guided by the renowned Nyquist theorem. A detailed overview of uniform sampling and its characteristics can be found in the work by Unser (2000). On the other hand, nonuniform sampling is categorized by whether the locations of sample collection are predetermined or uncertain, as discussed by Vandewalle et al. (2007). Multichannel sampling is another form that leads to the so-called super-resolution reconstruction techniques (Bertero and Boccacci, 2003; Park et al., 2003).

1.3.1 Random sampling scheme

Random sampling entails the selection of samples from a regular data grid in a stochastic manner. To keep the discussion clear and concise, the focus is on regular sampling with randomly missing data points, i.e., discrete random (under)sampling. Unless otherwise specified, the term *random* is used in the discrete sense. Such a sampling approach can give rise to algorithms capable of near-perfect Fourier reconstruction, particularly when the signal's spectrum is either sparse or confined to a specific band (Naghizadeh and Sacchi, 2010b).

Compared to regular undersampling, random undersampling diffuses coherent aliasing into benign incoherent noise, simplifying the complex task of interpolation to a more straightforward noise reduction problem. Hennenfent and Herrmann (2008a) demonstrated that irregular/random undersampling is not a drawback for particular transform-based interpolation methods. One of the seminal insights of CS is that random projection measurements of the state are inherently incoherent concerning nearly any generic basis. This incoherence is pivotal for the efficient recovery of sparse signals. However, obtaining such random projections of the full state can pose a considerable challenge in physical applications.

1.3.2 Jittered sampling scheme

Random undersampling that adheres to a discrete uniform distribution establishes advantageous conditions for a reconstruction method that promotes sparsity in the Fourier domain. Nonetheless, a global transformation like the Fourier transform generally fails to yield a sparse representation of seismic wavefields. These scenarios necessitate a more localized transformation, such as the windowed Fourier or Curvelet transform, as outlined by Zwartjes and Sacchi (2007) and Herrmann and Hennenfent (2008), respectively. Issues emerge when dealing with data gaps [2] exceeding the spatial and temporal scope of the transform elements, as discussed by Trad et al. (2005). Titova et al. (2021) implemented a two-stage sampling approach, where the second stage was specifically designed to address and fill in substantial gaps that were identified following an initial random sampling scheme. As a result, undersampling strategies that do not manage the maximum gap size are deemed less favorable.

Controlling the maximum gap size is an essential practical aspect when reconstructing wavefields using localized sparsifying transforms, but this is an aspect that random undersampling does not inherently manage. To bridge this gap, jittered undersampling, which was pioneered by Hennenfent and Herrmann (2008a) and the related work for seismic data (Leneman, 1966; Dippe and Wold, 1992), merges the advantages of random sampling with the controlled maximum gap size. Jittered undersampling is particularly advantageous for methods that utilize localized transform elements, such as the windowed Fourier or Curvelet transform (Candes and Tao, 2006). These approaches can be sensitive to gaps exceeding the spatial and temporal bounds of the transform elements, a challenge highlighted by Trad et al. (2005). Precisely, the essence of the jittered sampling technique is based on the critical observation that irregularities in sub-Nyquist sampling can be advantageous for nonlinear, sparsity-promoting wavefield-reconstruction algorithms. These irregular sampling patterns

[2] The *gap* is defined as the difference between the spacing of two consecutive sampled traces and the interval of a dense interpolation grid, with ideal sampling exhibiting zero gaps, according to Hennenfent and Herrmann (2008a)

help convert coherent aliasing into incoherent noise that typically has a lesser impact on the reconstruction quality. When sampling follows a discrete random uniform distribution, it doesn't offer the ability to regulate the maximum gap size, which can be problematic for transforms relying on localized elements, as these gaps can disrupt the data continuity necessary for such transforms. This approach is instrumental in developing a robust sparsity-promoting wavefield recovery strategy that adheres to the principles of CS. Under certain conditions, it establishes a conducive recovery condition for seismic wavefield-reconstruction techniques that enforce sparsity in the Fourier domain or related transforms, as evidenced by the research of Sacchi et al. (1998); Trad and Ulrych (1999); Xu et al. (2005); Abma and Kabir (2006) and Zwartjes and Sacchi (2007). An example application can be seen in the work of Moldoveanu (2010), who adapted the jittered sampling concept from Hennenfent and Herrmann (2008a) to overcome the physical constraints of marine data acquisition. They developed a comprehensive random coil-sampling strategy, where multiple vessels operate in unison, moving along randomly centered coil paths, to achieve full coverage.

1.3.3 Optimal sampling scheme

By identifying these innovative sampling methods as examples of CS, one can develop sampling and computational strategies that harness the inherent structure in seismic data. Consequently, we can devise sub-Nyquist sampling strategies, and the efficacy of these methods mainly depends on subsampling techniques that disrupt the periodicity found in traditional sampling methods. The optimal sampling scheme tries to place sensors in the most informative positions. I define three different sampling schemes according to various requirements, corresponding to the three routes in section 1.2.3.

Figure 1.4 shows different undersampling schemes with various randomness. The fine grid of open circles denotes the candidate positions of the sensors. The solid circles correspond to the coarse sampling locations. The first four sampling schemes in Figure 1.4 illustrate sampling with increasing randomness. Four out of five samples are missing, and Hennenfent and Herrmann (2008a) have proven that regular undersampling is the most challenging, while random and optimally jittered samplings according to a discrete uniform distribution are among the most favorable. The basic idea of jittered undersampling is to regularly decimate the interpolation grid and subsequently perturb the coarse-grid sample points on the fine grid. As for random undersampling according to a discrete uniform distribution, in which each location is equally likely to be sampled, a discrete uniform distribution for the perturbation around the coarse-grid points is considered. In particular, if the signal is sufficiently sparse, these schemes lead to a reconstruction as good as dense regular sampling. In comparison, the three optimal undersampling schemes have been well explained in section 1.2.3, considering the trade-off between the number of sensors and the reconstruction results.

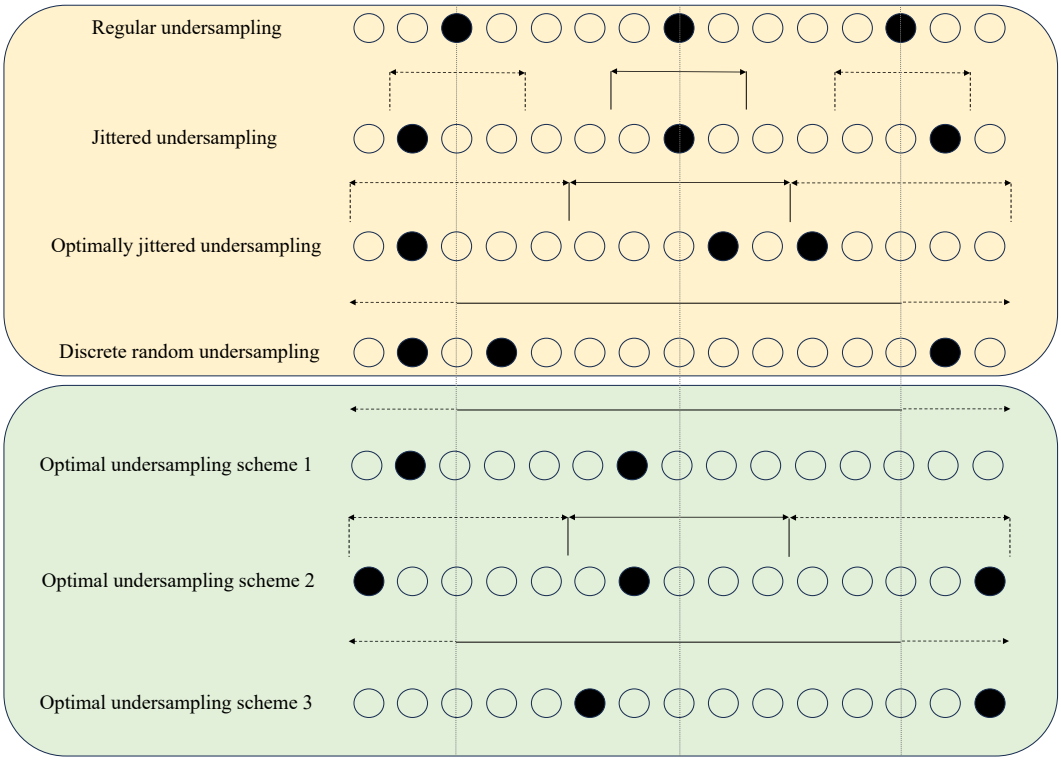

Figure 1.4: Schematic comparison between different undersampling schemes. The circles define the fine grid on which the original signal is alias-free. The solid circles represent the actual sampling points for the different undersampling schemes. The upper four sampling schemes are existing sampling strategies, and the lower three correspond to the three proposed optimal sampling schemes. For the optimally jittered undersampling and the optimal undersampling scheme 2, the maximum acquisition gap is controlled.

The subsequent section delves into the seismic acquisition design problem, specifically focusing on the strategy of sensor placement. This issue is particularly significant as the choice of sensor locations can greatly influence the quality of the acquired seismic data, as well as the efficiency and cost-effectiveness of the entire seismic survey.

1.4 Problem description

Consider an ideal seismic data cube $\mathcal{D}(t, r, s)$, where t indicates time, r and s denote the receiver and source position, respectively. Assume the time axis is discretized via the expression $t = (it - 1)\,\Delta t, \quad it = 1 \dots N_t$, and similarly, assume a regular distribution of receivers and sources with $r = r_0 + (ir - 1)\Delta r, \quad ir = 1 \dots N_r$ and $s = s_0 + (is - 1)\,\Delta s, \quad is = 1 \dots N_s$, where the integers N_t, N_r, and N_s are the number of time samples, receivers, and sources, respectively. Similarly, Δt corresponds to the temporal sampling interval, whereas Δr and Δs denote receiver and source intervals, respectively. The cube can be represented as a third-order tensor $\mathcal{D}$ of size $N_t \times N_r \times N_s$. An element of the tensor is defined as $[\mathcal{D}]_{it,ir,is}$.

For illustration purposes, consider an optimal receiver problem, the ideal data tensor $\mathcal{D}$ can be reshaped into a matrix $\mathbf{D}$ of size $(N_t \times N_r) \times N_s$ where each column of $\mathbf{D}$ corresponds to a common source gather (CSG). This organization of the data facilitates estimating the basis functions adopted to solve the reconstruction problem, and it also leads to a sampling problem expressed via a simple matrix-times-matrix multiplication. Assume the data is acquired with N_s sources and only K receivers ($K < N_r$) are needed for the optimal design problem. In other words, I consider the case where only a subset of K receivers are deployed in the field, and I question where these K receivers should be deployed to maximize the ability of a particular method to reconstruct the data.

For the optimal receiver problem, the sampling operator acts only on receivers. The receivers are deployed at positions given by the sampling set $\boldsymbol{\nu} = [\nu_1, \nu_2 \dots \nu_K]$. For instance, if the ideal data is composed of $N_r = 8$ receiver positions and $\boldsymbol{\nu} = [1, 2, 5, 6, 8]$ means extracting receivers 1, 2, 5, 6 and 8 ($K = 5$) to form the sampled or observed data. To continue with the analysis and in order to define the sampling operator for this particular problem, the following matrix is also defined

$$\mathbf{T}_r = \begin{bmatrix} \mathbf{T}_{\nu_1} & \mathbf{T}_{\nu_2} & \dots & \mathbf{T}_{\nu_K} \end{bmatrix}^T, \tag{1.1}$$

where vectors $\mathbf{T}_{\nu_i}$ are column vectors of length N_r with a one at element ν_i and zeros elsewhere. No sampling in time is performed, and the whole seismogram is taken for each sampled receiver. This can be easily accommodated by expressing the data sampling matrix

via

$$\mathbf{T} = \mathbf{T}_r \otimes \mathbf{I}_{N_t}, \tag{1.2}$$

where $\mathbf{I}_{N_t}$ is the $N_t \times N_t$ identity matrix and the symbol $\otimes$ stands for Kronecker product. $\mathbf{T}$ is the sampling matrix, a sparse matrix of size $(N_t \times K) \times (N_t \times N_r)$, which extracts K selected seismograms belonging to receivers with sampling positions $\nu_k, k = 1 \dots K$, from the ideal data $\mathbf{D}$. The latter permits the expression of the sampled data, referred to as the observed data for a given sampling set $\boldsymbol{\nu}$, as follows

$$\mathbf{D}_{\boldsymbol{\nu}} = \mathbf{T}\mathbf{D}. \tag{1.3}$$

Notice that $\mathbf{T}$ acts on each column of $\mathbf{D}$, and as stated previously, each column of $\mathbf{D}$ corresponds to a CSG. The reconstruction problem involves estimating $\mathbf{D}$ (the ideal data) from the decimated data $\mathbf{D}_{\boldsymbol{\nu}}$, which is underdetermined because the ideal data has more seismograms than the decimated data. Appendix A provides a simplified illustration of the sampling procedure outlined in this section, serving as a practical example for better understanding.

A CS or optimally designed acquisition followed by an effective reconstruction process ensures the subsequent steps—data processing, imaging, and interpretation—are based on the most reliable and comprehensive dataset available, thereby improving the overall quality and reliability of seismic analyses. Thus, the following describes the connection between acquisition and reconstruction.

1.5 Connection between seismic data acquisition and reconstruction

Seismic data acquisition and reconstruction are intrinsically linked processes. Each influences the other in critical ways, and understanding the connection between them can lead to more efficient and effective seismic surveys.

CS techniques and optimal survey designs aim to capture the most critical seismic information while minimizing costs and computational load. However, these approaches are often designed to capture sparse or sub-Nyquist sampled data. Bednar (1996) argued that no real theoretical requirement exists for regular spatial sampling of seismic data. However, popular multitrace processing algorithms, e.g., surface-related multiple elimination (SRME) (Verschuur et al., 1992) and wave-equation migration (WEM) (Claerbout, 1971), need a dense and regular coverage of the survey area. Nevertheless, field data sets are typically sampled

irregularly or coarsely along one or more spatial coordinates and need to be reconstructed to form a complete and interpretable dataset.

The quality and resolution of seismic data primarily depend on the acquisition stage. The placement of seismic sources and sensors' density all affect the resulting data. Reconstruction algorithms are applied to fill the gaps or improve the signal-to-noise ratio (SNR) if data is not fully acquired. Besides, considering high-density data acquisition can be expensive and time-consuming, especially in challenging environments like the deep ocean or remote regions. Advanced reconstruction techniques can help reconstruct high-quality data from sub-Nyquist or sparsely sampled data, leading to cost savings in the acquisition process. Meanwhile, more sensors and finer grid spacing result in more data, which subsequently requires more computational power for processing and analysis. Seismic reconstruction aims to regularize field data and increase the folding map for the following processing. Nevertheless, efficient reconstruction algorithms can reduce the computational burden by effectively reconstructing the data from fewer sampling points, thereby allowing for quicker interpretations and decisions. Further, the acquisition process should capture all necessary geological features ideally. However, practical limitations often result in sparse and incomplete data sets. Reconstruction techniques aim to complete this data, estimating what wasn't captured during the acquisition phase.

1.5.1 Seismic data reconstruction: review

Reconstruction is an essential step in seismic data processing, and numerous studies have been done. Two important categories are prediction filtering methods and transform-based methods.

Filter-based methods interpolate by convolution with a filter designed to make the error white noise. The most common of these filters are prediction-error filters (PEFs), which can handle aliased events (Spitz, 1991; Porsani, 1999). Gülünay (2003) introduced the frequency-wavenumber equivalence of prediction filtering methods. Their techniques are capable of interpolating regularly sampled aliased data. Naghizadeh and Sacchi (2007) modified these methods to cope with irregularly sampled data. In the seismic community, difficulties with regularly undersampled data are acknowledged when reconstructing by promoting sparsity in the Fourier domain (Hennenfent and Herrmann, 2008a). For example, Xu et al. (2005) stated that the anti-leakage Fourier transforms for seismic data regularization may fail when the input data has severe aliasing. A wide variety of wavefield-reconstruction techniques exists for regularly undersampled data along one or more spatial coordinates, i.e., data spatially sampled below the Nyquist rate. Methods that employ wavefield operators constitute a distinct interpolation approach that explicitly incorporates the principles of wave propagation into the process. These methods are designed to account for the way waves travel

through a medium, making them particularly useful in applications where understanding and managing the dynamics of wavefields is crucial (Canning and Gardner, 1996; Biondi et al., 1998; Stolt, 2002).

Transform-based methods also provide efficient algorithms for seismic data regularization (Bardan, 1987; Darche, 1990; Sacchi et al., 1998; Trad et al., 2003; Liu and Sacchi, 2004; Sacchi and Liu, 2005; Herrmann and Hennenfent, 2008; Zwartjes and Sacchi, 2007; Trad, 2008). Transform-based methods map the signal to a new domain, and synthesizing the signal from the transformed domain generates data at unrecorded spatial locations. In this case, the seismic signal is represented via a superposition of complex exponentials. Fourier reconstruction methods rely on two fundamental assumptions. Generally, we assume band-limited signals in the wavenumber domain or signals that a parsimonious distribution of Fourier coefficients can represent. Examples of the aforementioned assumptions abound in the signal and imaging processing literature. For instance, Strohmer (1997); Feuer and Goodwin (2005) and Eldar (2006) studied reconstruction of band-limited signals via Fourier methods. In the geophysical literature, Duijndam et al. (1999) and Schonewille et al. (2003) proposed a band-limited signal reconstruction method for seismic data that depends on two and three spatial dimensions. Techniques that used the sparsity assumption to reconstruct data presented by Sacchi and Ulrych (1996) and Sacchi et al. (1998) were expanded to the multidimensional case by Zwartjes and Gisolf (2006) and Zwartjes and Sacchi (2007). However, for irregularly sampled data, e.g., binned data with some of the bins that are empty, or continuous random, undersampled data, the performance of most of these interpolation methods deteriorates (Hennenfent and Herrmann, 2008a).

Besides, machine learning (ML) methods are increasingly becoming a prominent category for reconstructing seismic data. For instance, Jia and Ma (2017) suggested the application of ML in seismic data processing, with a specific emphasis on its use in interpolation. Additionally, Yu and Ma (2021) provided a comprehensive review of the existing and prospective developments in deep learning within the field of geophysics.

1.5.2 Basis functions

CS leverages structure within signals to reduce the required sampling rates. Typically, this structure translates into compressible representations, using an appropriate transform that concentrates the signal's energy into a small percentage of large coefficients. The size of seismic data volumes and the complexity of its high-dimensional and highly directional wave-front-like features make it challenging to find a transform that accomplishes this task (Herrmann et al., 2012). Traditional methodologies for seismic data regularization often depend on analytical bases like Fourier, Wavelet, Seislets, or Curvelet frames for data representation (Xu et al., 2005; Fomel, 2006; Herrmann and Hennenfent, 2008; Naghizadeh

and Sacchi, 2010a; Liu and Fomel, 2010; Ma and Plonka, 2010; Liu et al., 2011; Puryear et al., 2012). For instance, seismic signals admit sparse approximations regarding Curvelets (Candes et al., 2006; Hennenfent and Herrmann, 2006; Demanet and Ying, 2007). Unlike Wavelets (Beylkin et al., 1991), which compose curved wavefronts into a superposition of multiscale "fat dots" with limited directionality, Curvelets (Candes et al., 2006) and wave atoms (Demanet and Ying, 2007) compose wavefields as a superposition of highly anisotropic, localized, and multiscale waveforms, which obey the so-called parabolic-scaling principle. For Curvelets, this principle translates into a support where the length is proportional to the square of the width. At fine scales, this leads to needlelike Curvelets. Still, similarly to any other data-independent transforms, Curvelets do not provide a sparse representation of seismic data in the strict sense (Hennenfent and Herrmann, 2008a).

Therefore, these standard approaches may not offer the most efficient or accurate representation of seismic data specifically. Yu et al. (2015) used a data-driven tight frame method, which a dictionary is estimated from the input training data with small blocks. Further, techniques like proper orthogonal decomposition (POD) or dictionary learning can generate a basis more suitable to the specific characteristics of the seismic data, making the reconstruction more accurate. Thus, in the upcoming section, I introduce a fast and computationally efficient method for data reconstruction anchored in a pre-learned basis library. This approach has been extensively utilized throughout this thesis for data-driven scenarios to demonstrate its efficacy and robustness.

1.6 Pre-learned basis library and data reconstruction

In cases where prior information about the signal type is available, leveraging a specific basis for signal representation becomes feasible (Brunton and Kutz, 2019). Taking inspiration from machine learning methodologies, dominant low-rank features can be extracted from a training dataset to serve this purpose. Various data-driven dimensionality reduction techniques can be employed. Still this thesis gravitates towards proper orthogonal decomposition (POD), which is particularly favored for its widespread application in the reduced-order models (ROM) community and its array of beneficial properties (Pinnau, 2008; Lu et al., 2019; Mendible et al., 2020). Known by alternative names such as principal component analysis (PCA) in the realm of statistical analysis, POD effectively captures the core structures of complex datasets.

In essence, POD expresses a signal in a high-dimensional space as linear combinations of orthonormal modes. Projecting a signal into the POD subspace is like changing coordinates and requires a training dataset to extract the specific basis. The POD basis can be easily obtained from the singular value decomposition (SVD) (Eckart and Young, 1936; Chatterjee,

2000). According to the Eckart-Young theorem, the SVD provides the optimal least-squares approximation to the data of rank r (Golub et al., 1987). In addition, seismic data satisfy the low-rank approximation, and this pre-learned basis can be trained with an available seismic dataset.

Consider again the ideal data matrix $\mathbf{D}$ where each column is a vectorized common source gather (CSG). The SVD can extract an orthogonal basis that permits the data to be modeled. In essence, if ideal data is accessible, one can adopt the SVD to find the rank p approximation of the data

$$\mathbf{D} \approx \mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^T, \tag{1.4}$$

where $\mathbf{U}_p$ and $\mathbf{V}_p$ are the first p columns of the matrix of singular vectors associated with the p largest singular values of the SVD. The diagonal matrix $\mathbf{\Sigma}_p$ contains the singular values sorted in descending order. Note that a column j of the data matrix $\mathbf{D}$ can be approximated $\mathbf{d}_j \approx \mathbf{U}_p \mathbf{a}_j$ where $\mathbf{a}_j$ denotes the j-th column of the matrix $\mathbf{A} = \mathbf{\Sigma}_p \mathbf{V}_p^T$. This clearly shows that $\mathbf{U}_p$ is interpreted as an orthogonal basis capable of representing any matrix column $\mathbf{D}$. In other words, one can write

$$\mathbf{D} = \mathbf{U}_p \mathbf{A}\,, \tag{1.5}$$

and consider $\mathbf{A}$ as the unknown matrix of coefficients required to represent the data $\mathbf{D}$. The observed data can be expressed using the following expression

$$\mathbf{D}_{\boldsymbol{\nu}} = \mathbf{T}\mathbf{U}_p \mathbf{A}\,, \tag{1.6}$$

where $\mathbf{T}$ represents the sampling matrix and is the focal point of the acquisition design process. The subindex $\boldsymbol{\nu}$ indicates that it corresponds to recovered data from observed data with sampling set $\boldsymbol{\nu}$. It should be noted that the primary objective of this thesis, with respect to optimal sampling, is to identify the optimal sampling matrix represented by the sampling set $\boldsymbol{\nu}$.

If letting $\mathbf{\Theta} = \mathbf{T}\mathbf{U}_p$ be the sensing matrix, the reconstructed data can be obtained by solving $\hat{\mathbf{A}}$ via the method of least-squares (Manohar et al., 2018)

$$\hat{\mathbf{A}} = (\mathbf{\Theta}^T \mathbf{\Theta})^{-1} \mathbf{\Theta}^T \mathbf{D}_\nu\,, \tag{1.7}$$

and then using the pre-computed basis functions to estimate the recovered data

$$\mathbf{D}_{\boldsymbol{\nu}}^{rec} = \mathbf{U}_p\, \hat{\mathbf{A}}\,, \tag{1.8}$$

where $\mathbf{D}_{\boldsymbol{\nu}}^{rec}$ is the recovered data after reconstruction. This fast reconstruction method maximizes the utility of the acquired data, which is especially critical for applications requiring real-time or near-real-time interpretation, such as monitoring climate change impacts, earth-

quake early warnings, or adjustments in ongoing exploratory operations.

The SNR as the measure of reconstruction quality is defined. To be more concise, I define $SNR_{\boldsymbol{\nu}}$

$$SNR_{\boldsymbol{\nu}} = 10\log_{10}(\frac{\|\mathbf{D}\|_F^2}{\|\mathbf{D}_{\boldsymbol{\nu}}^{rec} - \mathbf{D}\|_F^2}), \tag{1.9}$$

where the symbol $\|\cdot\|_F^2$ denotes the Frobenius norm.

Some CS-based scheme design algorithms use additional criteria to define an optimal sampling scheme. For example, the nonuniform optimal smapling (NOUS) (Mosher et al., 2012a) has an outer loop, which uses reconstruction error (RE). One distinct disadvantage of RE is that one needs to reconstruct data to get the RE value, which is computationally expensive for large-scale problems. However, the reconstruction adopted in this thesis is fast and computationally efficient, which can serve as the criteria for a data-driven optimal sampling scheme. Aside from its use in data reconstruction, the pre-learned basis library also offers a significant advantage for noise removal, a crucial step in data processing. Detailed information on this application can be found in section 3.5. Notably, the requirement for a training dataset is not a limitation in this context.

1.7 Contribution of this thesis

This research represents a significant departure from existing methods and has both theoretical and practical implications. On the theoretical front, it questions the constraints set forth by CS theory and offers novel approaches for seismic data acquisition. Practically, it has the potential to revolutionize the way seismic surveys are conducted, offering cost savings, operational efficiencies, and reduced environmental impact.

The primary objective of this research is to refine seismic acquisition design with a specific focus on minimizing the number of required sensors. This approach aims to significantly reduce acquisition costs without sacrificing data integrity. The main contributions of this thesis are summarized as follows:

- **Efficient seismic acquisition**: The study introduces novel techniques for optimal seismic acquisition design that utilize fewer sensors than traditional methods. The goal is to cut costs associated with sensor procurement, maintenance, and field operations while maintaining high data quality.
- **Data-driven optimization**: When prior information is available, the study advocates for data-driven seismic acquisition over traditional CS methods. It employs a pre-trained basis library that better characterizes seismic data, enabling a further reduction in the number of sensors needed.

- **Incorporation of reinforcement learning**: The study extends its scope by integrating RL algorithms to solve the sensor placement optimization problem. The goal is to design a more intelligent and efficient data collection methodology.
- **Advancements via deep reinforcement learning**: The research explores the utility of DRL algorithms to conceive a robust and versatile acquisition strategy capable of adapting to complex scenarios.
- **Versatile applications**: The study confirms that the optimized acquisition design is highly adaptable across various seismic acquisition scenarios. These include simultaneous source acquisition, OBN acquisition, time-lapse seismic monitoring in CO_2 geological storage sites, and vibroseis optimal route design.

1.8 Thesis overview

This thesis focuses on optimizing acquisition design and seismic data reconstruction and Figure 1.5 shows the schematic workflow of the thesis with three proposed routes and four possible applications.

In **Chapter 1**, the groundwork is laid by presenting the general landscape of seismic exploration and acquisition design. This chapter elaborates on the limitations of traditional surveys and introduces transformative approaches such as CS seismic acquisition and optimal seismic surveys. The problem is described after an overview of different sampling schemes. Additionally, the interrelationship between acquisition and data reconstruction is explored, highlighting the utilization of a pre-learned basis library.

In **Chapter 2**, a comprehensive review of CS in seismic acquisition and an illustration of how it can be successfully exploited in seismic acquisition are offered. A comparison of different reconstruction methods is made. Moreover, the chapter integrates arbitrary irregular-grid coordinates into the CS framework to better approximate real-world field scenarios.

In **Chapter 3**, a data-driven and cost-efficient approach to seismic acquisition is presented after introducing the optimal sparse sensing theory. Using a basis library pre-trained on existing seismic data, optimal sensor locations are determined through QR column pivoting. Applications to simultaneous source acquisition and time-lapse surveys are presented. A denoising method with a global and local basis is also shown.

In **Chapter 4**, innovative RL techniques are employed for the optimization of seismic acquisition design. A Markov decision process (MDP) is formulated according to the application of OBN acquisition. Q-learning is evaluated and found to produce satisfactory results.

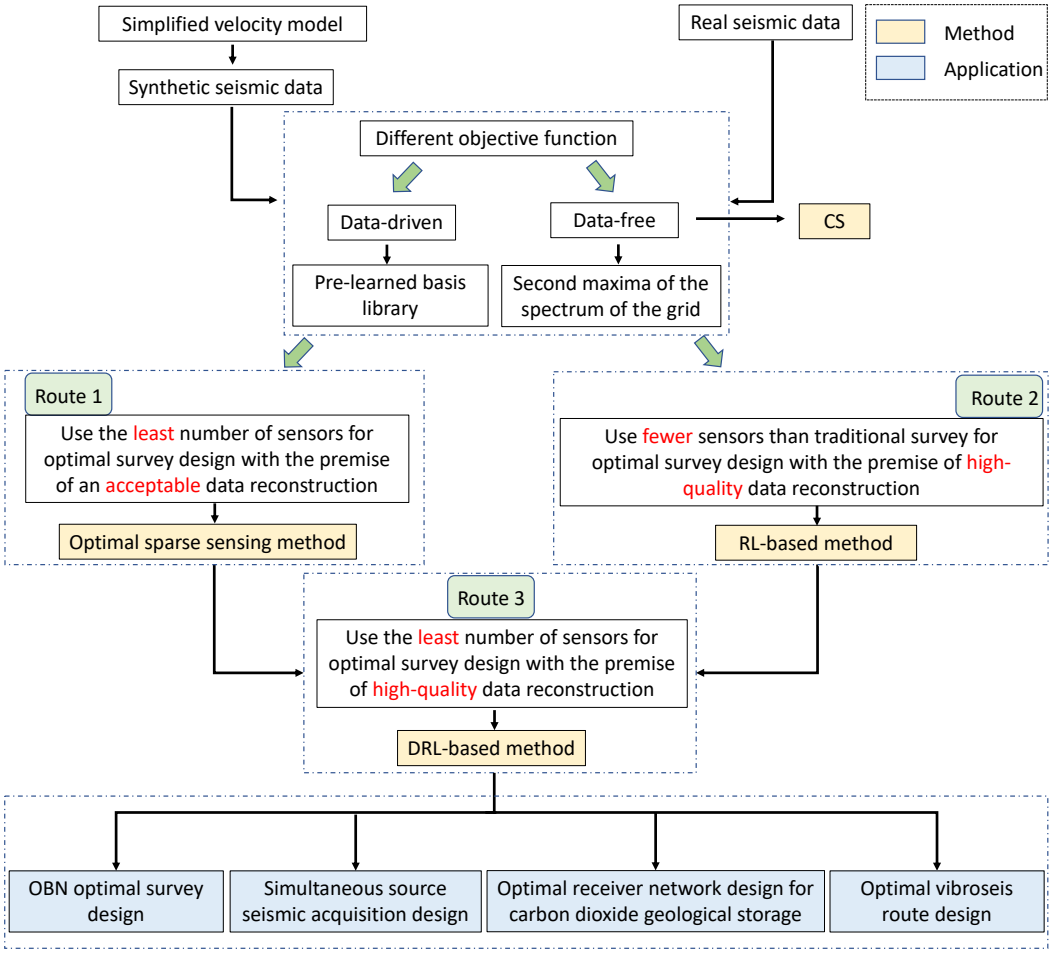

Figure 1.5: Schematic workflow of the thesis.

In **Chapter 5**, the scope of the problem is broadened to include more prominent cases through the implementation of advanced DRL techniques. The MDP formulation is based on the vibroseis optimal route problem. The deep Q-network (DQN) algorithm is adopted to solve this acquisition design problem.

Chapter 6 serves as the conclusion, summarizing the core content, contributions, and limitations of the algorithms developed throughout the thesis. Future research directions and recommendations are also discussed.

CHAPTER 2

Compressive sensing seismic acquisition [1]

This chapter is designed to offer an in-depth review of compressive sensing (CS) in seismic acquisition, focusing on both the method's advantages and limitations. Additionally, I delve into irregular grid scenarios, which mirror real-world conditions more closely. The chapter is organized as follows: First, I present an introduction to the theoretical framework of CS. Subsequently, mutual coherence (MC) is discussed, and after that, I show how CS is applied in seismic surveys using regular and irregular grid setups. To conclude, I showcase some practical examples to illustrate these concepts, with additional discussions of noise-contaminated scenarios.

2.1 Introduction

In traditional signal processing, the Nyquist-Shannon sampling theorem sets a strict criterion for regular spatial sampling. It dictates that a continuous signal can be fully reconstructed from its samples, provided it is sampled at a rate at least twice the highest frequency contained in the signal. However, with the advent of CS techniques, deviating from these strict requirements has become possible.

Using random sampling strategies, CS allows for sub-sampling at rates below what the Nyquist-Shannon theorem prescribes. This has opened the door to more flexible and potentially cost-effective acquisition schemes, as one can sample at significantly lower rates without necessarily losing the ability to accurately reconstruct the original signal.

[1] Part of the version of the work in Chapter 2 of this thesis has been published in a journal paper: Lin R., Y. Guo, F. Carozzi and M. D. Sacchi, 2022, Simultaneous deblending and source reconstruction for compressive 3D simultaneous-source acquisition data via Interpolated MSSA (I-MSSA): *Geophysics*, **87**, no. 6, 1-53.

The seismic reflection method, a primary technique for exploring oil and gas reserves, often accounts for a substantial portion of the exploration budget (Claerbout, 1992). With the advent of 3D seismology (Biondi, 2006) in the late 1980s, a new wave of oil and gas exploration emerged globally, spurring significant efforts to cut acquisition costs. Major oil companies and geophysical contractors worldwide have been at the forefront of these cost-reduction initiatives. In this context, the role of CS seismic acquisition has been pivotal. Initially tested as an experimental approach, CS seismic acquisition has gained considerable traction in the industry. It has emerged as a revolutionary technique in reflection seismology, fundamentally altering the landscape of seismic data acquisition. By allowing for the collection of meaningful data at a fraction of the traditional sampling rates, CS seismic acquisition has proven to be a cost-effective alternative to traditional methods, making it an increasingly popular choice in the oil industry for optimizing acquisition budgets.

2.2 Compressive sensing

Natural signals like seismic data are often highly compressible, meaning they can be efficiently represented using a reduced set of basis functions in a suitable domain. The essence of dimensionality reduction is this: by transforming the signal into a domain where it becomes sparse (mostly zeros except for a few large values), one can store these significant values for later accurate reconstruction. This concept of sparsity allows the recovery of unrecorded or missing data from a limited set of measurements, essentially enabling a form of data compression. This results in substantial savings in storage and computational resources compared to retaining the original, high-dimensional signal.

To demonstrate the CS approach, I begin with a concise overview of CS theory, followed by its relevance and application in exploration seismology. The recovery process from subsamplings involves solving extensive convex optimization challenges, also discussed in the subsequent section.

2.2.1 Compressive sensing framework

CS (Candès et al., 2006; Candes and Romberg, 2006; Donoho, 2006) is one of the groundbreaking innovations of the past decade that inverts the conventional compression paradigm. Instead of compressing and discarding the high dimensional data in the first place, CS proposes to collect the low dimensional data directly and then infer the sparse coefficients in the corresponding transform domain that explain the data. After that, the reconstruction can be done using convex optimization methods (Candes and Tao, 2006; Donoho, 2006; Baraniuk, 2007).

CS relies on a sparsifying transform for the to-be-recovered signal and adopts this sparsity before compensating for undersampling during the recovery process. For reconstructing wavefields in the Fourier (Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007), Radon (Trad et al., 2003), and Curvelet (Hennenfen and Herrmann, 2005; Herrmann and Hennenfent, 2008) domains, sparsity promotion is a well-established technique documented in the geophysics literature. The main contribution of CS is the new light shed on the favorable recovery conditions.

Suppose the data $\mathbf{x}$ has a compact representation in a transform basis Ψ. On a universal basis $\Psi \in \mathbb{R}^{n \times n}$, such as Fourier bases, $\mathbf{x}$ may have a sparse representation

$$\mathbf{x} = \mathbf{\Psi s}, \tag{2.1}$$

where $\mathbf{s} \in \mathbb{R}^r$ is a sparse vector. For the basis Ψ to adequately represent any natural signal, it needs to be complete, encompassing all the necessary basis vectors to capture the inherent characteristics of $\mathbf{x}$.

Consider a set of measurements $\mathbf{y} \in \mathbb{R}^p$, obtained via a measurement matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$ (Baraniuk, 2007)

$$\mathbf{y} = \mathbf{Cx} = \mathbf{C\Psi s} = \mathbf{\Theta s}, \tag{2.2}$$

where $\mathbf{\Theta}$ denotes the sensing matrix. The main challenge is choosing $\mathbf{C}$ for a given $\mathbf{\Psi}$ so that $\mathbf{\Theta}$ is still a good measurement matrix.

For $p < n$, equation 2.2 is under-determined, and there are infinite solutions. The least squares solution is not sparse and typically yields poor reconstruction. So, knowing the natural signals are sparse, one seeks the sparest $\mathbf{s}$ consistent with the measurements $\mathbf{y}$

$$\mathbf{s} = \arg\min_{\mathbf{s}'} \|\mathbf{s}'\|_0 , \text{ such that } \mathbf{y} = \mathbf{C\Psi s}', \tag{2.3}$$

where $\|\mathbf{s}\|_0$ is the ℓ_0 pseudo-norm corresponding to the number of non-zero entries of $\mathbf{s}$. Unfortunately, this requires a combinatorial brute-force search across all sparse vectors $\mathbf{s}$. A major breakthrough in CS is a set of conditions on the measurement matrix $\mathbf{C}$ that allows the non-convex ℓ_0-minimization in equation 2.3 to be relaxed to the convex ℓ_1-minimization, which is computationally tractable.

$$\mathbf{s} = \arg\min_{\mathbf{s}'} \|\mathbf{s}'\|_1 , \text{ such that } \mathbf{y} = \mathbf{C\Psi s}', \tag{2.4}$$

where $\|\mathbf{s}\|_1 = \sum_{k=1}^{n} |s_k|$. This formulation is shown schematically in Figure 2.1.

Within the framework of CS, three key conditions need to be carefully considered so that data can be recovered from severely undersampled data:

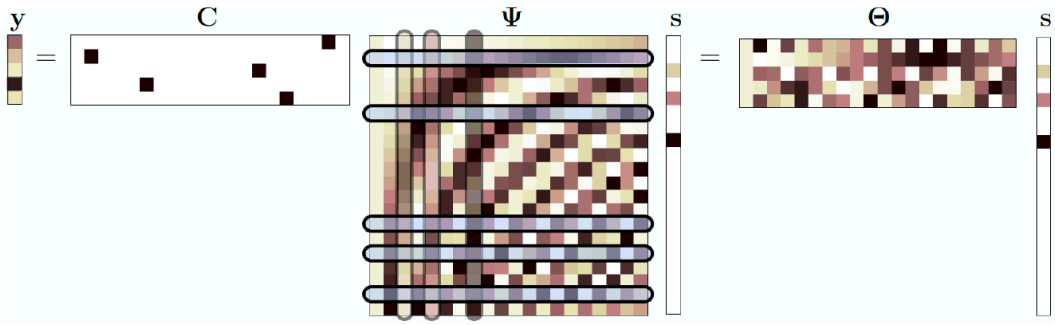

Figure 2.1: Compressed sensing provides the sparsest solution to an underdetermined linear system.

- Selection of sparse transformation: The first condition involves choosing a sparse basis or transformation that can efficiently represent the underlying data that the signal exhibits sparsity in a known transform domain. The choice can vary based on the specific characteristics of the acquired signals.
- Design of the sampling scheme: The design of the sampling matrix is crucial for acquiring enough information from the sparse signals. A poorly designed matrix can result in ineffective or unreliable data reconstruction. Random undersampling is adopted since the artifacts introduced by the undersampling result in incoherent random noise in the sparsifying domain.
- Choice of a reconstruction algorithm: The final condition is to choose a data-consistent, sparsity-promoting procedure used for recovery, that is, an algorithm that can accurately reconstruct the original signal from the compressed measurements. Various algorithms, such as orthogonal matching pursuit (OMP) or basis pursuit (BP), are available for this purpose, and the choice may depend on the computational resources and required accuracy.

Intuitively, the artifacts introduced by undersampling the original signal are not sparse in the transform domain. When this condition on the artifacts is not met, sparsity alone is no longer effective before solving the recovery problem. Albeit qualitative, the second condition provides a fundamental insight into choosing undersampling schemes that favor recovery by sparsity-promoting inversion (Hennenfent and Herrmann, 2008a).

Numerous random matrix ensembles have demonstrated efficacy as compressive measurement matrices. Examples include Gaussian, Bernoulli, and Fourier matrices with randomly chosen rows. A pivotal query arises regarding the relationship between the required number

of measurements for precise recovery and factors like the sparsity level, the number of measurements, and the ambient dimension, sampling frequency, and the size of the sampling grid in traditional sampling theory, respectively. The answer is if the measurement matrix is chosen appropriately, the number of measurements scales only logarithmically with the ambient dimension (Herrmann et al., 2012).

The optimization problem in equation 2.4 finds a sparse or, under certain conditions, the sparsest (Donoho et al., 2001) possible solution that explains the data. For the ℓ_1-minimization in equation 2.4 to yield the sparest solution, the measurements $\mathbf{C}$ must be chosen so that $\mathbf{\Theta} = \mathbf{C\Psi}$ satisfies a restricted isometry (RI) property (Candes, 2008). The RI property condition essentially asserts that a matrix $\mathbf{\Theta} = \mathbf{C\Psi}$ preserves the lengths of sparse signals when they are transformed. More formally, it means that for a K-sparse vector $\mathbf{s}$, the following inequality holds:

$$(1 - \delta)\|\mathbf{s}\|_2^2 \leq \|\mathbf{\Theta s}\|_2^2 \leq (1 + \delta)\|\mathbf{s}\|_2^2, \tag{2.5}$$

where δ is a small positive constant. Various alternative strategies exist to find the sparsest solution to equation 2.3. Among these, greedy algorithms are commonly employed due to their computational efficiency. One such algorithm is the compressed sampling matching pursuit (CoSaMP) (Needell and Tropp, 2008), which iteratively identifies the support of the sparse signal and updates its estimate without needing the measurement matrix to satisfy stringent conditions like the RI property.

The RI property condition is essential for ensuring that ℓ_1-minimization yields the sparsest solution to the problem, effectively reconstructing $\mathbf{s}$ from a set of compressed measurements $\mathbf{y}$. However, verifying that a given matrix $\mathbf{\Theta}$ satisfies the RI property can be computationally challenging or even intractable in practice, especially for large matrices or those derived from complex systems. This remains one of the key challenges in applying CS techniques in real-world applications; that is, determining the RI property may be highly challenging in practice.

2.2.2 Mutual Coherence

A sampling scheme should satisfy necessary and sufficient conditions to obtain error bounds for CS reconstruction and thus quantitatively assess the performance of the chosen sampling scheme. Unfortunately, most of the necessary and sufficient conditions are impossible to calculate. This computational intractability is inherent to some adequate conditions that guarantee an accurate reconstruction, for example, the RI property discussed above. Computationally tractable sufficient conditions include mutual coherence (MC). The MC minimization is another popular strategy for designing a sampling scheme. MC must be

very small to be considered a sufficient condition to guarantee accurate reconstruction. The MC minimization problem is complex since it is a non-convex and non-smooth optimization (Titova et al., 2019). To handle this issue, different kinds of approximations are used (Obermeier and Martinez-Lorenzo, 2017; Lu et al., 2018), and there is ongoing research on developing new computationally verifiable sufficient conditions for sparse recovery.

Mathematically, the concept of MC, μ, plays a significant role in evaluating the performance of CS algorithms. According to Elad (2007), when the following inequality holds

$$\alpha(\mathbf{s}) < 0.5 \left(1 + \frac{1}{\mu(\Theta)}\right), \tag{2.6}$$

one can guarantee an exact solution for equation 2.4, where $\alpha(\mathbf{s})$ is the sparsity of the signal (the number of non-zero elements) and $\mu(\Theta)$ is MC, defined as the maximum absolute correlation along the columns θ of the matrix Θ

$$\mu(\Theta) = \max_{1 \leq i,j \leq Q, i \neq j} \frac{|\theta_i^H \theta_j|}{\|\theta_i\|_2 \|\theta_j\|_2}, \tag{2.7}$$

where θ_i and θ_j are the i^{th} and j^{th} columns of matrix Θ.

For a matrix of size $n_1 \times n_2$, the MC is lower bounded by the Welch bound

$$\mu_W(\Theta) = \sqrt{(n_2 - n_1)/(n_1(n_2 - 1))} \leq \mu(\Theta). \tag{2.8}$$

Interestingly, the Welch bound depends only on the size of a matrix and does not include any information about matrix elements. Also, the Welch bound is not always achieved (Titova et al., 2019).

An important aspect of estimating mutual coherence (MC) is the computation of the Gram matrix linked to the sampling and transform operators. With certain transformations, such as the Fourier transform, this computation is relatively straightforward and enables an analytical calculation of MC. However, when direct matrix representations of these operators are not available, calculating MC becomes a more complex task. To overcome this challenge, Guitton and Loubani (2022) introduced a mathematical method for estimating the Gram matrix that is applicable to any combination of sampling and transform operators. This approach effectively provides a close approximation to the true Hessian for the Fourier dictionary, which aids in the precise calculation of MC. It's worth noting that this technique is computationally intensive, particularly for transformations involving large model vectors, since it involves the explicit formation of the Hessian matrix. Therefore, additional optimization efforts are necessary to improve its efficiency.

Next, I discuss the application and challenges CS faces in the exploration seismology.

2.3 Compressive sensing seismic acquisition

Contemporary seismic data processing, imaging, and inversion methodologies are increasingly dependent on computationally demanding and data-intensive techniques to cater to the persistent demand for hydrocarbons in today's world. Besides, proper seismic wavefield sampling is a crucial element of seismic exploration. But in most cases, the seismic wavefield is undersampled in the spatial domain due to cost limitations and survey acquisition constraints. Such an approach, however, poses challenges as it leads to exponentially rising expenses with the expansion of the targeted survey area.

In an era of CS, a properly sampled wavefield can be achieved from undersampled data by employing a specially tailored signal processing technique. So, drawing inspiration from the innovations in CS and building upon prior research in seismic data regularization (Sacchi et al., 1998) and phase encoding (Romero et al., 2000), CS-based acquisition schemes considerably reduce the costs associated with data acquisition and its subsequent processing, also attain higher spatial bandwidth (Mosher et al., 2012b) in comparison with conventional design. By leveraging the sparse nature of seismic signals, CS enables accurate reconstruction of the subsurface image with fewer measurements, thus leading to substantial cost savings.

This application of CS in seismic exploration has been discussed in various studies (Hennenfent and Herrmann, 2008a; Herrmann et al., 2009, 2012; Li et al., 2012; Ma et al., 2012; Ma and Yu, 2017). Herrmann et al. (2009) identified seismic data regularization and simultaneous acquisition as instances of CS and selected the sparsifying transform and designing randomized sampling schemes realizable in the field (Herrmann et al., 2012). These works underscore the potential of CS not only as a data-efficient methodology but also as a cost-effective alternative to conventional seismic data acquisition methods. CS also sheds light on the concept of simultaneous random time-dithered acquisition. This approach enhances the efficiency of data collection by condensing the average wait time between consecutive shots and may include the deployment of transducers across the seabed in a random fashion. While the application of CS principles can lead to substantial gains in the efficiency of acquisition, these same principles can also enhance the efficiency of wavefield simulations, imaging, and inversion processes. However, the primary focus of this thesis is the application of CS to seismic data acquisition.

According to CS principles, effective dimensionality reduction is rooted in an incoherent sampling approach where coherent aliases transform into white Gaussian noise. When applying this method to the exploration seismology, two primary challenges emerge. Firstly, seismic data collection is bounded by physical limitations related to the placement, type, and number of sources and receivers. Coupled with the enormous scale of seismic data, these limitations necessitate solutions explicitly tailored to seismic challenges. Secondly, while

CS provides promising opportunities for reducing dimensionality, integrating the scientific computing workflow with this new methodology is still challenging (Herrmann et al., 2012).

CS states that a sparse signal can be reconstructed from underdetermined measurements when this signal is acquired according to a specific sampling pattern and reconstructed with an efficient algorithm. In terms of seismic application, a sparse transform is required to get a sparse representation of the seismic wavefield, a suitable arrangement of sources and receivers, and an algorithm for large-scale signal reconstruction. These elements and dealing with noise in seismic data represent the current challenges in developing CS for seismic applications (Baraniuk and Steeghs, 2017).

Let us assume designing a receiver line with N_r receivers. In practice, considering many factors, one wishes to use fewer receivers $K < N_r$. The idea is to use the K measurements to recover the N_r desired measurements. Mathematically, the problem can be described as:

$$\mathbf{d_\nu} = \mathbf{T}\mathbf{d} + \mathbf{n}\,, \tag{2.9}$$

where $\mathbf{d_\nu} \in \mathbb{R}^n$ is the vectorized observed data of length $K \times N_t$, and $\mathbf{d} \in \mathbb{R}^n$ is the vectorized desired data of length $N_r \times N_t$. $\mathbf{T}$ is the sampling operator of size $(N_t \times K) \times (N_t \times N_r)$, which contains information about the seismic acquisition pattern. In addition, N_t represents the number of time samples per trace. Proper temporal sampling following the Nyquist theorem is assumed, and the problem of spatial sampling is the focus. Last, the term $\mathbf{n}$ represents the noise. Since the number of measurements $N_t \times K$ is smaller than the number of unknowns $N_t \times N_r$, the system is underdetermined and consequently has an infinite number of solutions.

If the seismic wavefield is described as a sparse signal $\mathbf{s}$ through some sparse transforms $\mathbf{\Psi}$, i.e., Fourier bases, equation 2.9 becomes

$$\mathbf{d_\nu} = \mathbf{T}\mathbf{\Psi}\mathbf{s} + \mathbf{n} = \mathbf{\Theta}\mathbf{s} + \mathbf{n}. \tag{2.10}$$

The operator $\mathbf{\Theta}$ is the sensing matrix, where $\mathbf{\Theta} = \mathbf{T}\mathbf{\Psi}$. Figure 2.2 is the schematic representation of the CS seismic framework. One possible way of constructing a good $\mathbf{C}$ tailored to a given sparsity basis $\mathbf{\Psi}$ is to choose an appropriate measurement basis $\mathbf{R}$ that is incoherent with $\mathbf{\Psi}$. Once an incoherent $\mathbf{R}$ is chosen, we discard all but p rows from $\mathbf{R}$ and use the resulting $p \times n$ matrix as the measurement matrix. More precisely, we set $\mathbf{C} = \mathbf{H}\mathbf{R}$ where $\mathbf{H}$ is an $p \times n$ restriction matrix (consisting of n rows of the $N \times N$ identity matrix). The universal strategy for choosing $\mathbf{C}$ that does not require prior knowledge of the sparsity basis $\mathbf{\Psi}$ is to choose $\mathbf{C}$ to be an appropriate random measurement matrix (Herrmann et al., 2012).

In equation 2.10, the system is still underdetermined. However, in this case, since we are dealing with a sparse vector, one can take advantage of CS and find the sparse coefficients

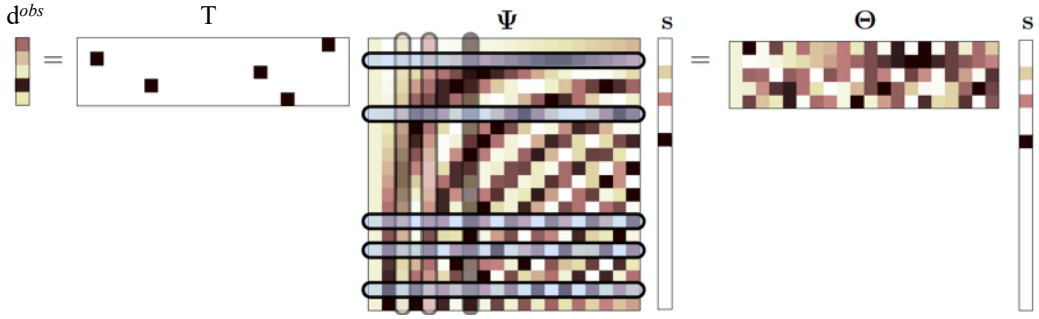

Figure 2.2: Schematic of measurements in the compressed sensing seismic framework. Reproduced from Brunton and Kutz (2019).

s via a convex optimization algorithm. The latter amounts to solving the problem

$$\mathbf{s}' = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_1 \,, \text{subject to} \, \|\mathbf{d}_{\boldsymbol{\nu}} - \mathbf{T}\boldsymbol{\Psi}\mathbf{s}\|_2^2 \leq \sigma, \tag{2.11}$$

where σ is the noise level of $\mathbf{d}_{\boldsymbol{\nu}}$. After computing the solution $\mathbf{s}'$ in equation 2.11, $\mathbf{d}^{rec}$ can be obtained by the synthesis operator $\boldsymbol{\Psi}\mathbf{s}'$. The difference between a properly sampled recorded wavefield $\mathbf{d}$ and a properly sampled reconstructed wavefield $\mathbf{d}^{rec}$ depends on the choice of the sparse transform, the sampling scheme, and the reconstruction algorithm.

2.3.1 Survey design based on mutual coherence

A lower MC value is generally desirable in CS, as it indicates a greater level of incoherence between the measurement matrix and the sparsifying basis. This property is fundamental in seismic acquisition, where a lower MC can potentially lead to more accurate data recovery from a reduced set of measurements, thereby lowering acquisition costs. Using MC as a judging criterion is discussed in the following and belongs to the data-free seismic acquisition design realm.

Significant work has been done to assess the MC minimization for seismic acquisition design. For example, non-uniform optimal sampling (NUOS) proposed by Mosher et al. (2012a) has already been successfully tested in the field. Another example, Bhuiyan and Sacchi (2015), solved the MC minimization problem by considering possible survey requirements. In the papers by Jamali-Rad et al. (2016) and Campman et al. (2017), one may find a deterministically subsampled scheme based on the difference sets and has the smallest MC. Jiang et al. (2018) proposed MC map which is based on the aliased energy (Naghizadeh and Sacchi, 2010b). It is known (Carin et al., 2011) that when the Fourier transform is chosen as a

sparse transform, the aliased energy and MC are the same. In a recent paper by Florez et al. (2018), MC defines the optimal blended distribution of sources. The primary motivation for using MC for sampling scheme design is the computability aspect and the fact that the lower values of MC promote better reconstruction results than the higher values. However, Titova et al. (2019) compared reconstruction results for different sampling schemes: uniform random sampling, jittered undersampling, and sampling schemes with the smallest possible MC and concluded that MC is a very crude way to predict the performance of the sampling scheme. Echoing the views presented in Guitton and Loubani (2022), I maintain that MC, despite its imperfections, remains a valuable indicator of the quality of data reconstruction. In situations where one must choose between two acquisition geometries, assuming all other conditions are identical, the preferable option would be the one that results in the lowest MC. This choice aligns with the principle that lower MC typically indicates better data reconstruction quality.

The concepts of random matrices and the computability of the MC are extensively exploited in the seismic CS-based sampling design. A random matrix is a matrix whose elements could be realizations of a Gaussian distribution or a 0/1 Bernoulli process. However, such matrices are not entirely applicable to seismic acquisition needs (Titova et al., 2019), which is demonstrated with the following 1D and 2D problems. Both cases choose 24 from 70 sensors on a regular grid with a spatial sampling, which satisfies the Nyquist criterion.

1D problem

If only 24 receivers are available and rely on the MC as a design criterion, then in equation 2.11: $\mathbf{s}$ is the Fourier spectrum of the desired signal, $\mathbf{d_\nu}$ is the recorded data, $\mathbf{\Theta} = \mathbf{T\Psi}$ is a matrix of size 24×70, $\mathbf{\Psi}$ is the 1D Fourier transform, and $\mathbf{T}$ is a sampling matrix built with the starting vector formed with elements are either 0 (no receiver) or 1 (an active receiver). As it is a vector, $\mathbf{R}$ is used to represent it to distinguish it from the matrix $\mathbf{T}$. For the matrix $\mathbf{\Theta}$, the Welch bound $\mu_W(\Theta) = \sqrt{(70-24)/(24(70-24))} = 0.17$, which coincides with the value of MC computed by equation 2.7.

2D problem

For the 2D case, a seismic shot gather is considered as an example, and decimate shot gather only in the spatial domain (receivers), while the traces are sampled regularly in time. The equation 2.11 has the following elements: $\mathbf{s}$ is reshaped into a vector of the 2D Fourier spectrum of the properly sampled shot gather, which one aims to find. $\mathbf{d_\nu}$ is the registered shot gather reshaped into a vector. $\mathbf{\Theta} = \mathbf{T\Psi}$ is a matrix of size $24N_t \times 70N_t$. $\mathbf{\Psi}$ is the 2D Fourier transform. The construction of seismic sampling matrix $\mathbf{T}$ follows

certain procedures: first, building a vector of length 70, which elements are either 0 (no receiver) or 1 (an active receiver), same with the 1D case; then we multiply in element-wise fashion an identity matrix of size $N \times N$ with this vector. Then, remove the rows with zeros from this matrix (the amount of non-zero rows equals the amount of $1's$ in the starting vector), thereby constructing a rectangular sampling matrix $\mathbf{T}$. Even if the starting vector is drawn from a 0/1 Bernoulli process, all matrix $\mathbf{T}$ elements are not independent and identically distributed. So, in this sense, the matrix in seismic application possesses a certain structure, and only the starting vector has entries drawn from the Bernoulli or a uniform distribution (Titova et al., 2019), which answers why random matrices are not entirely applicable to seismic acquisition needs.

Further, a variety of strategies may be used to build the starting vector: a jittered undersampling (Hennenfent and Herrmann, 2008a), a piecewise random subsampling (Wang et al., 2011), a Bernoulli-based random undersampling (Cai et al., 2014), and a local random sampling (Liu et al., 2015). These strategies aim to avoid large gaps in a sampling scheme and use a partitioning approach. However, these strategies are not intended to restrict the reconstruction error and integrate acquisition design logistic requirements. The Welch bound for the matrix $\mathbf{\Theta}$ in this case $\mu_W(\Theta) = \sqrt{(70N_t - 24N_t)/(24N_t(70N_t - 1))} = \sqrt{46/24(70N_t - 1)} \leq 0.17$. As proved by Titova et al. (2019), the MC value 0.17 is the smallest possible, although the Welch bound is not reached. So, to compute MC, first, we obtain the Gram matrix $\mathbf{G} = \mathbf{\Theta^H \Theta}$, then search for the maximum value within $\mathbf{G}$ off-diagonal elements. Following Titova et al. (2019), this expression leads to the conclusion that in the case of regular sampling in time and irregular in space (receiver domain), MC is determined only by the receiver sampling matrix $\mathbf{R}$:

$$\mu(\Theta) = \mu(\mathbf{R\Psi^H}), \tag{2.12}$$

that is if the matrix $\mu(\mathbf{R\Psi^H})$ reaches the smallest MC, then the matrix Θ reaches the smallest MC as well.

2.3.2 Compressive sensing acquisition based on irregular grid

Until now, undersampling strategies based on an underlying fine interpolation grid have been considered. This situation typically occurs when binning continuous randomly sampled seismic data into small bins that define the fine grid used for interpolation (Hennenfent and Herrmann, 2008a). Despite the error introduced in the data, binning presents some computational advantages because it allows for fast implementations of Fourier or Fourier-related transforms. However, binning can lead at the same time to an unfavorable undersampling scheme, e.g., regular or poorly jittered. In this case, one should consider working on the

original data with irregular grids. Despite the extra computational cost, continuous random sampling typically improves interpolation results because it does not create coherent undersampling artifacts (Xu et al., 2005).

Spatially irregularly sampled seismic data is inevitable in field acquisition due to natural obstacles or compressive-sensing design. Based on CS, Mosher et al. (2014a) showed how to use fewer sensors for optimal seismic acquisition design. Mosher et al. (2014b) also proposed a new seismic acquisition design named non-uniform optimal sampling (NUOS) for optimal sources and receivers deployment. Mapping seismic data from irregular grid to regular grid is a long-standing problem for seismic processing. In essence, there are two types of irregularity for field acquisition.

- Natural irregularity

  This type of irregularity is caused by natural obstacles, for instances lakes, buildings, and man-made structures, such as highways and buildings, which is inevitable and quite common for both marine and land acquisition. Significant gaps occur in the observed data, leading to discontinuity or even distorted seismic images (Liu and Sacchi, 2001, 2004; Liu et al., 2004; Trad, 2009; Sacchi and Trad, 2010).

- Human-designed irregularity

  This type of irregularity is predefined for designing the acquisition geometry based on CS theory, which can potentially reduce the survey time and increase the data resolution (Mosher et al., 2014a,b).

Most reconstruction algorithms assume the sources and receivers are positioned in a regular-grid coordinate system. However, real-world seismic surveys often deviate from this ideal due to the above-mentioned irregularities. These irregularities, which are often unavoidable, have been traditionally addressed using classical prestack interpolation methods (Liu and Sacchi, 2004; Trad, 2009). Recent advancements in CS have shifted the focus towards intentionally designed irregular grids to facilitate more accurate seismic data reconstruction (Li et al., 2012; Mosher et al., 2012a). In essence, CS-based techniques employ strategically randomized sampling schemes, enabling the transformation of data collected on irregular grids back to a regular, densely sampled grid (Hennenfent and Herrmann, 2008b). The following describes the reconstruction methods used for CS acquisition.

Reconstruction for irregular grid compressive sensing acquisition

Seismic data processing usually needs an interpolation step due to the incompleteness of the data, which is owed to factors like the presence of obstacles, dead traces, or other physical limitations.

For regular grid scenarios, the applicability of CS to the large-scale problems of exploration geophysics relies heavily on the implementation of an efficient ℓ_1 solver. Even though several attempts have been made to overcome this bottleneck (Tibshirani, 1996; Figueiredo et al., 2007; van den Berg and Friedlander, 2007), a wide range of large-scale applications still use approximate ℓ_1 solvers such as iterated reweighted least-squares (IRLS) (Gersztenkorn et al., 1986), stagewise orthogonal matching pursuit (Donoho, 2006), iterative soft thresholding with cooling (Daubechies et al., 2004; Hennenfen and Herrmann, 2005; Herrmann and Hennenfent, 2008), and fast iterative shrinkage thresholding algorithm (FISTA). Other methods like projection onto convex sets (POCS) are also effective seismic reconstruction methods widely used in industry and academia. It was first introduced into the geophysics area by Abma and Kabir (2006) and then improved by Gao et al. (2013b). Stanton et al. (2015) introduced a soft thresholding function controlled by a single parameter to the POCS algorithm. However, strictly speaking, it is not typically used to solve the ℓ_1 problem. Nevertheless, many applications adopting the POCS algorithm proved workable in seismic exploration.

In the context of an irregular network based on CS, the strategy of *Design First, Reconstruct Later* is introduced. This approach entails initially considering the topography of the acquisition area and the distribution of any obstacles. A judicious random distribution sampling operator is then employed to minimize the number of sensors needed.

Seismic reconstruction aims to regularize the field data and map the data from irregular-grid observation to regular-grid. For the two irregularities mentioned before, different reconstruction methods have been studied.

- Reconstruction for natural irregularity scenario

  Many reconstruction methods have been proposed to solve this problem. To name a few, Liu and Sacchi (2004) introduced a minimum weighted norm interpolation (MWNI) method for multidimensional seismic reconstruction. Xu et al. (2005) proposed the antileakage Fourier transform (ALFT) method for irregular sampling and overcoming the spectral leakage problem. Naghizadeh and Sacchi (2007) adopted prediction error filter (PEF) for aliased regularly and irregularly decimated data reconstruction. Similarly, Zwartjes and Sacchi (2007) reconstructed aliased nonuniformly sampled data with sparse Fourier inversion. Chiu (2014) developed a modified version of the MWNI, incorporating a prior model as a constraint to effectively tackle the issue of aliasing. This alteration is crucial for the MWNI's optimal performance, which depends on the replication of data in the FK domain, necessitating random spatial sampling. This specific challenge in the application of MWNI was earlier identified and discussed by Cary (2011).

- Reconstruction for human-designed irregularity scenario

  Seismic records compressed to lower than Nyquist frequency can be perfectly recovered with sparse Fourier domain inversion by utilizing the coherency of seismic signals in auxiliary domains (Herrmann, 2010; Mosher et al., 2012a; Yu et al., 2020). With an application for seismic acquisition, Li et al. (2012) proposed an interpolated CS method for constructing irregular grid sources or receivers' positions to nominal grid positions. Jiang et al. (2017) extended the POCS method (Abma and Kabir, 2006) to Extended Projection Onto Convex Sets (EPOCS) by incorporating an interpolation operator and extending its usage for under-sampled arbitrary irregular acquisition.

Besides the methods listed above, in recent decades, methods that exploit reduced-rank data approximations for irregularly decimated data reconstruction have been gaining popularity and interest. Rank-reduction constrained reconstruction methods can be divided into two categories: matrix-based reconstruction methods (Sacchi et al., 2009; Trickett et al., 2010; Oropeza and Sacchi, 2011; Ma, 2013) and tensor-based reconstruction methods (Kreimer et al., 2013; Kumar et al., 2015; Gao et al., 2015). Oropeza and Sacchi (2011) presented multichannel singular spectrum analysis (MSSA) method for simultaneous reconstruction and random noise attenuation. Carozzi and Sacchi (2019) modified the tensor completion method (Kreimer et al., 2013) to a robust version for simultaneous 5D seismic volumes reconstruction and erratic noise attenuation. In order to honour true trace coordinates, Carozzi and Sacchi (2021) extended the MSSA reconstruction method to the interpolated MSSA (I-MSSA) method for compressive irregular-grid data reconstruction. Despite the accuracy estimation of MSSA, its computational cost of SVD has always been a concern - a huge memory-load burden when building the Hankel matrix for multidimensional data. Efforts have been made by replacing the SVD with more efficient algorithms, such as randomized SVD (Oropeza and Sacchi, 2011) and Lanczos bidiagonalization (Gao et al., 2013a).

Mathematically, let us examine scenarios where the sensors are distributed in an irregular or quasi-irregular manner on the Earth's surface during seismic data acquisition. Here, sources are used as an example to demonstrate the method. Not every grid point receives traces during the allocation process, necessitating an imputation algorithm to reconstruct the unobserved traces (Oropeza and Sacchi, 2011). For definitions of the various grids involved, readers refer to Lin et al. (2022c) for the description of the different grids.

Initially, we look into a reconstruction situation using a straightforward sampling operator grounded in data binning. This can be formulated as:

$$\mathbf{D}_{\boldsymbol{\nu}} = \mathscr{T}\mathbf{D}. \tag{2.13}$$

Here, $\mathbf{D}_{\boldsymbol{\nu}}$ signifies the data that adheres to the actual observed irregular-grid coordinates,

while $\mathbf{D}$ corresponds to data on the intended regular-grid coordinates.

We characterize $\mathscr{T}$ as the extraction operator, encapsulating two distinct operations: *binning* and *sampling*. The *binning* operation involves mapping arbitrary irregular-grid source coordinates to their nearest regular-grid counterparts. The *sampling* operation entails multiplying by 1 with all regular grid points that have data and by 0 with all regular grid points that are empty (Liu and Sacchi, 2004). In other words, within the extraction operator's framework, the binning step's objective is to map the observed irregularly placed data onto a regular grid. This binning approach can introduce errors in both the amplitude and phase of the seismic traces, and these inaccuracies can adversely impact the quality of the data reconstruction.

In this refined approach, the aim is to reconstruct the seismic sources while preserving the actual positions of these sources, thereby avoiding errors commonly associated with binning. To achieve this, we link $\mathbf{D}_{\boldsymbol{\nu}}$ and $\mathbf{D}$ using a specialized local interpolation operator. Mathematically, this relationship is expressed as:

$$\mathbf{D}_{\boldsymbol{\nu}} = \mathscr{W}\mathbf{D}\,, \tag{2.14}$$

where $\mathscr{W}$ serves as an operator that embodies a 2D Kaiser window tapered sinc interpolation technique (Carozzi and Sacchi, 2021; Wang et al., 2022). This operator is designed to map the target data directly from their regular coordinates to their corresponding irregular coordinates, offering a more accurate and error-minimized reconstruction.

To elaborate on the Kaiser window tapered sinc interpolation operator specified in equation 2.14, let's first set the stage by defining the coordinates of the source positions on the observed irregular grid. These coordinates can be described as:

$$\xi_k = \xi\big(x_k, y_k\big), \tag{2.15}$$

where ξ_k designates the spatial coordinates of the kth source with $k = 1, \ldots, N_s$. N_s is the total number of acquired sources, and $\xi_k \in \mathbf{U}$, representing data in irregular coordinates.

In parallel, we define the coordinates of the desired source positions on the regular grid as follows:

$$\eta_{(i,j)} = \eta\big(\hat{x}_i, \hat{y}_j\big), \tag{2.16}$$

where $\hat{x}_i = \hat{x}_o + i\Delta x$, $\hat{y}_j = \hat{y}_o + j\Delta y$. The coordinates $(\hat{x}_o, \hat{y}_o)$ specify the position of the initial point on the regular grid, while Δx and Δy are the respective intervals between grid points along the x and y axes. Indices i and j range from 1 to N_{sx}, N_{sy}, respectively. The total number of desired reconstructed data in this regular grid is $N_{sx} \times N_{sy}$, and $\eta_{(i,j)} \in \mathbf{D}$.

For the application of CS to be feasible, two primary adjustments to the framework are

essential. Firstly, expecting real-world signals to be strictly sparse is optimistic. A more plausible scenario is for the coefficients to diminish quickly when arranged in order of magnitude. This results in a vector dominated by a few substantial entries and numerous minor ones. While not strictly sparse, such signals can be closely approximated by sparse models and are referred to as compressible. Secondly, in realistic settings, measurements are often contaminated by noise. Ensuring that the CS methodology remains robust in the presence of such disturbances is imperative, as highlighted in Candès et al. (2006). Considering this, the following section concerns the method used for noise-contaminated scenarios.

Noise attenuation in compressive sensing acquisition

Although the focus is on a noise-free, underdetermined system of linear equations, the CS theory and, hence, the related work both extend to the recovery from undersampled data contaminated by noise (Candes et al., 2006; Hennenfent and Herrmann, 2008a).

Erratic noise is non-Gaussian noise. Land 5D volumes, in particular, often contain a large number of "erratic traces" that one needs to ignore or de-emphasize automatically. However, it is not a good idea to automatically ignore traces, which means losing many traces in a problem where we have sparsely populated grids. To name a few, Chen and Sacchi (2014) adopted a robust SSA method to reduce erratic noise. Chen and Sacchi (2017) proposed a robust $f-x$ projection filter for simultaneous erratic noise attenuation. Carozzi and Sacchi (2019) developed a new robust tensor completion algorithm for coping with erratic noise.

Projection Onto Convex Sets (POCS) is a method extensively utilized for seismic data reconstruction. However, a limitation of the standard POCS algorithm is its inability to eliminate erratic noise in seismic traces, which can compromise both the integrity of the original signal and the quality of the reconstructed results. POCS operates through repeated use of the fast Fourier transform, making it an appealing choice for multidimensional reconstruction of large seismic datasets, as noted in Abma and Kabir (2006). In my work, I introduce a robust formula of POCS, which is designed for both erratic noise attenuation and sparse sampling purposes. This advanced POCS formulation incorporates the ℓ_1/ℓ_2 norm as the misfit term, following the methodologies discussed in Carozzi and Sacchi (2019, 2020). This adjustment significantly diminishes the impact of erratic noise on the reconstructed data, rendering the modified POCS algorithm more resilient to both Gaussian and non-Gaussian noise types compared to the traditional approach. Thus, the revised POCS algorithm is more robust to Gaussian and non-Gaussian noise than the traditional method. In other words, even with noise-contaminated scenarios, where Gaussian or non-Gaussian noise, seismic data can be reconstructed with the standard or robust POCS method. Thus, the signal-to-noise ratio of the reconstructed data is significantly enhanced. The superiority of this refined POCS method over the conventional one, especially in the presence of erratic noises, is

demonstrated through the use of 5D synthetic data and real-field data examples. Comprehensive formulas and their derivations are detailed in Appendix B, providing a thorough understanding of this robust approach to seismic data reconstruction.

2.4 Examples

In this section, I employ the FISTA (Beck and Teboulle, 2009), EPOCS (Jiang et al., 2017), I-MSSA (Oropeza and Sacchi, 2011; Cheng et al., 2019; Carozzi and Sacchi, 2021; Lin et al., 2022a), and I-FMSSA (Lin et al., 2022b) algorithms to evaluate the performance of CS in seismic data acquisition through both synthetic and field data examples, considering scenarios with regular as well as irregular grid configurations.

2.4.1 Compressive sensing seismic acquisition based on regular grid

In the first example, I examine the efficacy of CS in a seismic survey that uses a regularly sampled grid featuring three linear events. The grid comprises 30×30 points with an interval of $\Delta x = \Delta y = 20$ in the $x-$ and $y-$ directions. A Ricker wavelet with a central frequency of 20 Hz is used in this synthetic example. To emulate the CS scenario, 50% of the traces are randomly decimated.

Figure 2.3(a) displays the regular grid, which also serves as the ideal grid, while Figure 2.3(b) presents the decimated sampling points superimposed on the regular grid. These two grids illustrate the fundamental premise of CS in seismic data acquisition. The desired regular grid (Figure 2.3(a)) represents the ideal case where full sampling would occur. In contrast, the observed grid (Figure 2.3(b)) signifies the practical scenario in which only a subset of the data points from the ideal grid is acquired, typically due to constraints like time, cost, or other logistical factors. The objective of CS-based reconstruction methods is to accurately estimate the complete dataset represented by the desired regular grid using the sparsely sampled data from the observed grid. The data reconstructed using the FISTA method for this regular grid survey is shown in Figure 2.4. The results are highly satisfactory, with an SNR of 19.78 dB as computed by equation 1.9.

2.4.2 Compressive sensing seismic acquisition based on irregular grid

In this example section, I first evaluate the computation efficiency via the I-MSSA and I-FMSSA algorithms for irregular reconstruction. For the synthetic and real data examples,

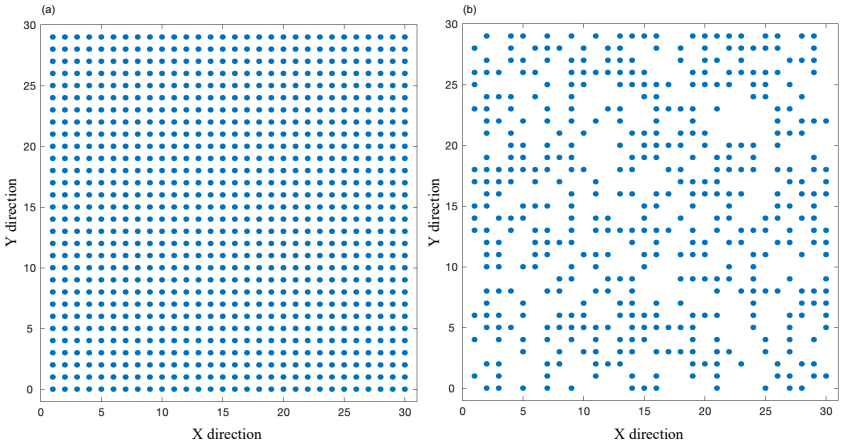

Figure 2.3: (a) The desired regular grid serves as the output grid for the reconstruction process. (b) The observed grid is a randomly decimated subset of the desired regular grid (a).

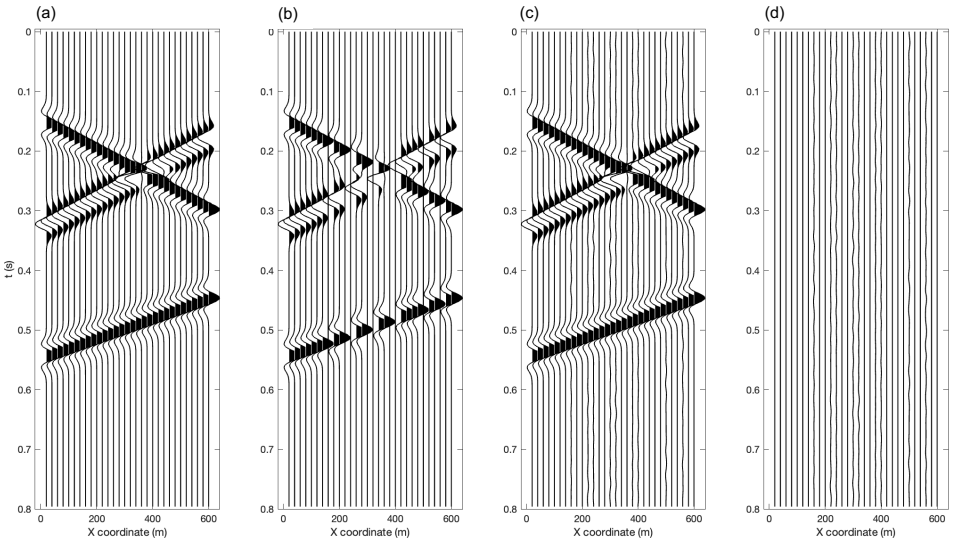

Figure 2.4: Reconstruction results for a clean synthetic data example with a randomly 50% decimated volume along x coordinate direction on a regular grid. (a) Clean data. (b) Observed data with 50% random decimation of (a). (c) Result of the FISTA reconstruction. (d) Residuals between the clean data (a) and (c).

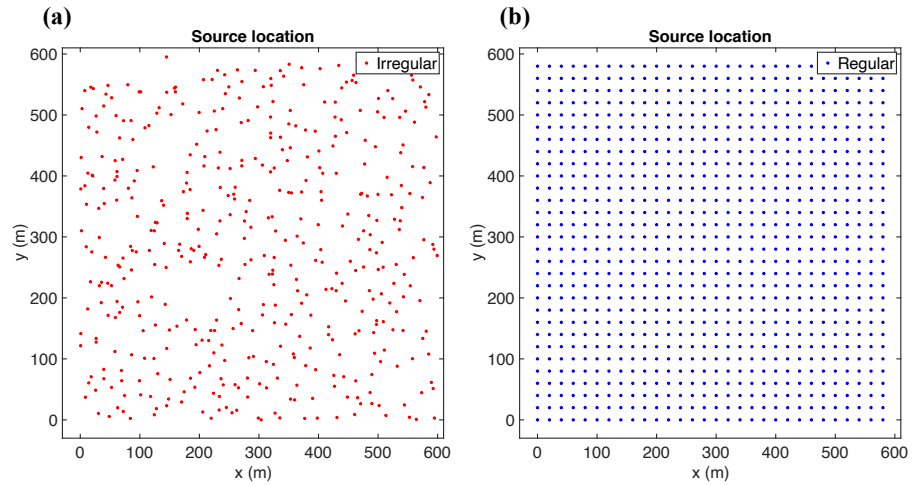

Figure 2.5: Coordinate systems considered in the synthetic examples. (a) Coordinates of the irregular gird after 50% decimation. (b) Desired regular grid of output.

I compare the behavior of the EPOCS method and the I-FMSSA method for compressive irregular-grid data reconstruction with synthetic and real data examples. The expression 1.9 is used to measure the signal-to-noise ratio (SNR) of the recovered data.

Synthetic example

I first consider noise-free 3D synthetic data to compare different algorithms. The synthetic example contains three dipping linear events to mimic a small 3D patch of common receiver gather (CRG). The regular grid consists of 30×30 source points with interval $\delta x = \delta y = 20m$ in the x- and y- directions, and a Ricker wavelet of central frequency 20 Hz was adopted. Then, a perturbation is added to the regular grid to generate the irregular distribution. Specifically, to avoid the generation of significant gaps, source x and y coordinates were perturbed via random deviates drawn from a uniform distribution in the range $[-\Delta x, \Delta x]$ and $[-\Delta y, \Delta y]$. This approach is applicable in scenarios where survey irregularities arise due to physical obstacles or misplacement of sources, as well as in situations where the survey has been intentionally designed with randomness. Utilizing fewer sources aims to minimize the time required for data acquisition, thereby enhancing the efficiency of field operations (Mosher et al., 2017). The geometry of the source coordinate of irregular-grid and desired regular-grid distribution is displayed in Figure 2.5. Figure 2.5(a) shows irregular distribution after 50% decimation, and Figure 2.5(b) represents the desired regular grid output.

Firstly, the computational performance of the I-MSSA method with the I-FMSSA method

is compared. The clean regular grid data and observed irregular grid data can be found in Figure 2.6(a) and 2.6(b). Figure 2.6(c) shows the reconstruction result with the I-MSSA method, and the corresponding error can be found in Figure 2.6(d). A high-quality reconstruction result ($SNR = 46.2$ dB) with no signal leakage for this synthetic data is observed. Similarly, the reconstruction result with the I-FMSSA method can be found in Figure 2.6(e), in which incredibly the same reconstruction quality as Figure 2.6(c) with $SNR = 46.7$ dB is obtained. However, the computing time for this synthetic data shows a considerable difference. Adopting the I-FMSSA method (time = 15.5 s) needs much less computational time than the I-MSSA method (time = 45.8 s). Thus, the I-FMSSA method is an efficient and accurate alternative to the I-MSSA method when applied to the reconstruction problem. Beyond that, Figure 2.6(g) depicts the reconstruction result with the EPOCS method, and the corresponding error can be found in Figure 2.6(h). Comparing Figure 2.6(d) and 2.6(f) with 2.6(h), significant signal leakage of the EPOCS method is observed. Conversely, the reconstruction with I-MSSA and I-FMSSA methods produces satisfactory results without noticeable signal leakage. Also, the I-FMSSA method achieves a much higher SNR value ($SNR = 46.9$ dB) than the EPOCS method ($SNR = 16.2$ dB). The detailed comparison can be found in Table 1. In addition, the EPOCS method requires iterations for each frequency slice to select different threshold values for various frequency slices, which increases the computational complexities when compared with the I-FMSSA method, where the iterations are deployed in the time domain.

Figure 2.7 shows the 2D slices of data in Figure 2.6. Figure 2.7(b) shows the random decimation of the traces, and Figure 2.7(h) shows significant leakage of the EPOCS method lies in the boundary traces of the seismic profile. On the contrary, there is no boundary signal leakage for I-MSSA (Figure 2.7(d)) and I-FMSSA (Figure 2.7(f)) methods.

For noise-contaminated scenarios, Gaussian noise with S/N=1 dB is added to the clean 3D data in Figure 2.6(a) for the next test. The reconstruction results can be found in Figure 2.8. It is observable that except for reconstructing the signal, the EPOCS method also reconstructs the noise, resulting in $SNR = 3.5$ dB. On the contrary, the I-MSSA and I-FMSSA methods generate cleaner and stable reconstruction results with higher SNR values ($SNR = 20.6$ dB and 20.8 dB, respectively).

Again, Figure 2.9 shows the 2D data slices in Figure 2.8. It is observed that the reconstructed random noise for the EPOCS method is shown in Figure 2.9(h). In addition, the signal leakage of the boundary traces also exists. In contrast, the I-MSSA and I-FMSSA methods are effective for simultaneous empty trace reconstruction and random noise elimination. This test's computational time and SNR comparison can be found in Table 2.1.

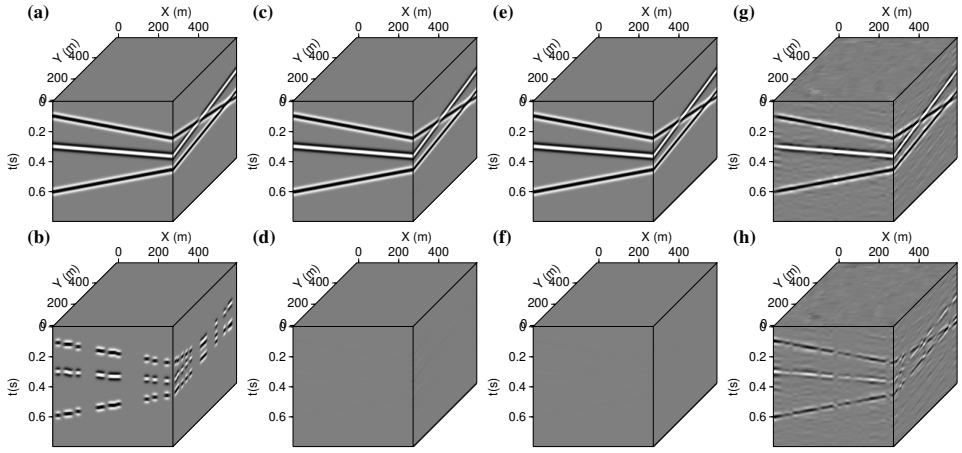

Figure 2.6: 3D irregular reconstruction results comparison for clean synthetic data. (a) Clean regular grid data. (b) Observed decimated irregular grid data after binning. (c) Reconstruction with the I-MSSA method with $SNR = 46.2$ dB. (d) Difference between (a) and (c). (e) Reconstruction with the I-FMSSA method with $SNR = 46.9$ dB. (f) Difference between (a) and (e). (g) Reconstruction with EPOCS method with $SNR = 16.2$ dB. (h) Difference between (a) and (g).

Table 2.1: Computational time and SNR comparison for irregular grid reconstruction.

Cases	**Methods**	**Time(sec)**	**S/N(dB)**
Clean data	I-MSSA	45.8	46.2
	I-FMSSA	15.5	46.9
	EPOCS	1825.7	16.2
Noisy data	I-MSSA	57.2	20.6
	I-FMSSA	27.7	20.8
	EPOCS	1686.6	3.5

Real data example

I also compare the methods with field data from the West Canadian Basin. The coordinate distribution can be found in Figure 2.10. A small survey area with serious irregularity is selected, which contains 113 source points, and the mean intervals between sources and source lines are 100 m and 300 m, respectively. The desired regular grid intervals between sources and source lines are 100 m and 150 m, respectively, which includes $10 \times 29 = 290$ source points.

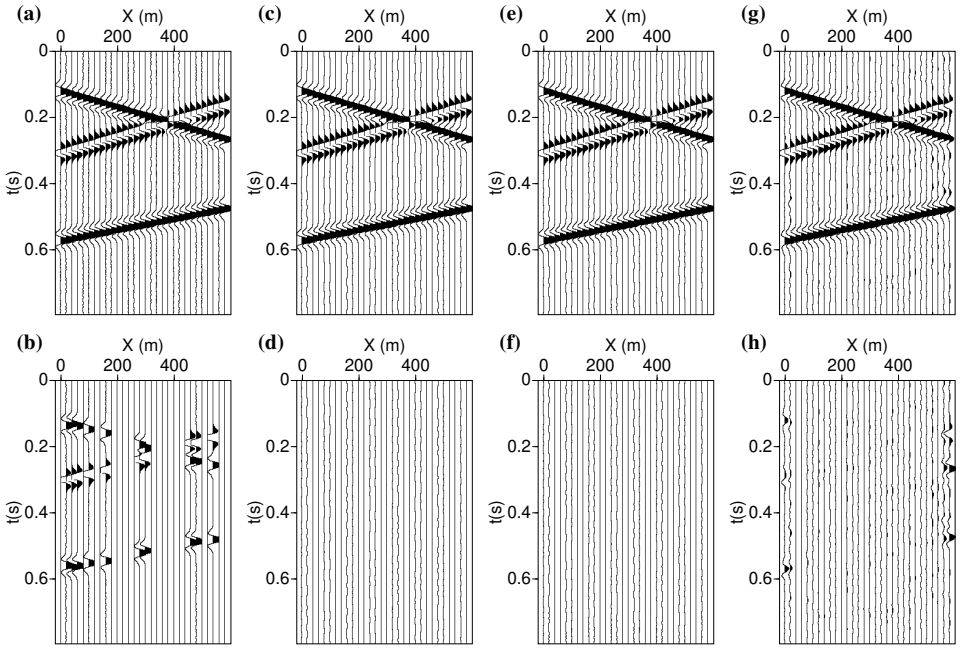

Figure 2.7: One slice display of reconstruction results in Figure 2.15. (a) Clean regular-grid data. (b) Observed decimated irregular grid data after binning. (c) Reconstruction with I-MSSA method. (d) Difference between (a) and (c). (e) Reconstruction with the I-FMSSA method. (f) Difference between (a) and (e). (g) Reconstruction with EPOCS method. (h) Difference between (a) and (g).

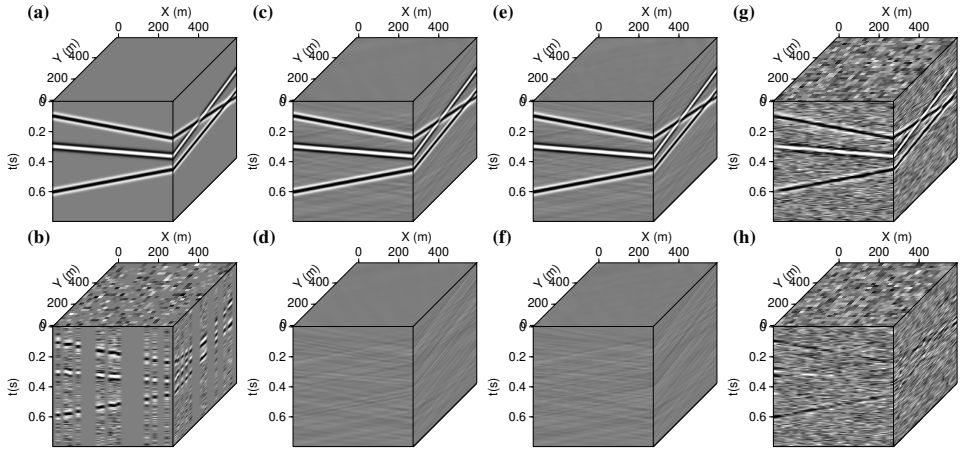

Figure 2.8: 3D irregular reconstruction results comparison for noisy synthetic data. (a) Clean regular-grid data. (b) Observed decimated noisy irregular grid data after binning. (c) Reconstruction with I-MSSA method with $SNR = 20.6$ dB. (d) Difference between (a) and (c). (e) Reconstruction with I-FMSSA method with $SNR = 20.8$ dB. (f) Difference between (a) and (e). (g) Reconstruction with EPOCS method with $SNR = 3.5$ dB. (h) Difference between (a) and (g).

The reconstruction results of the EPOCS method and I-FMSSA method for an inline slice can be found in Figure 2.11. Figure 2.11(a) shows the observed irregular data after binning. Figure 2.11(b) and 2.11(c) display the reconstruction results with the EPOCS method and I-FMSSA method, respectively. Both approaches are effective for irregular grid data reconstruction, and the empty traces have been fully reconstructed. However, the EPOCS method also reconstructs the random noise. On the contrary, the I-FMSSA method generates a slightly cleaner reconstruction result than the EPOCS method. On the other hand, due to lacking the ground truth as a reference, it is difficult to evaluate the signal leakage of the boundary traces with the EPOCS method.

Similarly, the reconstruction results for one cross-line slice can be found in Figure 2.12. Like the observation in Figure 2.11, the I-FMSSA method generates a less noisy result than the EPOCS method. However, it is also practicable to follow a denoising algorithm with the EPOCS method to attenuate the random noise; compared with random noise attenuation, geophysicists are more concerned about signal leakage.

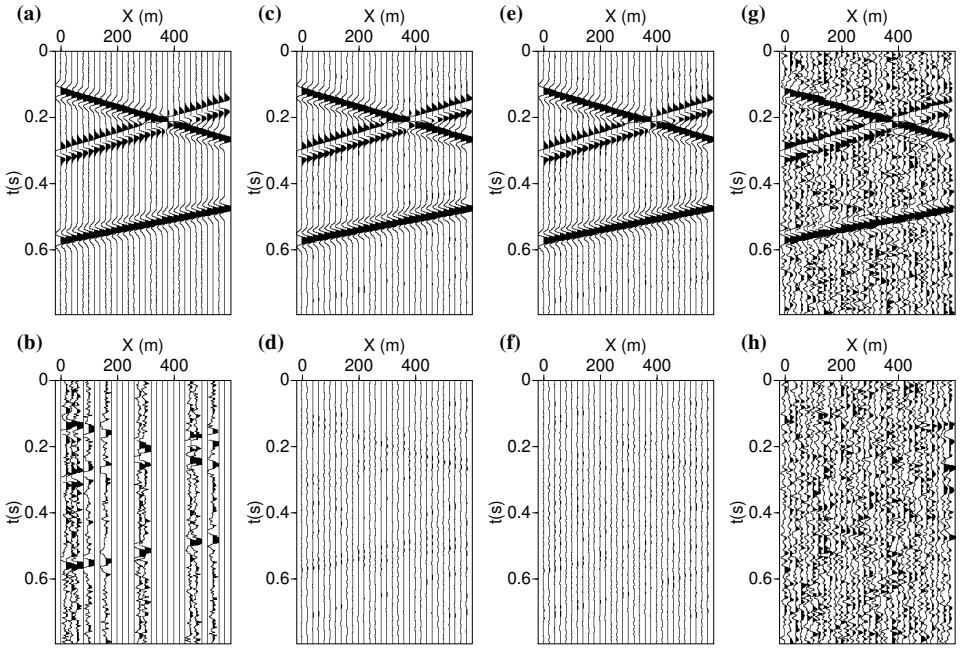

Figure 2.9: One slice display of reconstruction results in Figure 2.8. (a) Clean regular grid data. (b) Observed decimated irregular grid data after binning. (c) Reconstruction with I-MSSA method. (d) Difference between (a) and (c). (e) Reconstruction with the I-FMSSA method. (f) Difference between (a) and (e). (g) Reconstruction with EPOCS method. (h) Difference between (a) and (g).

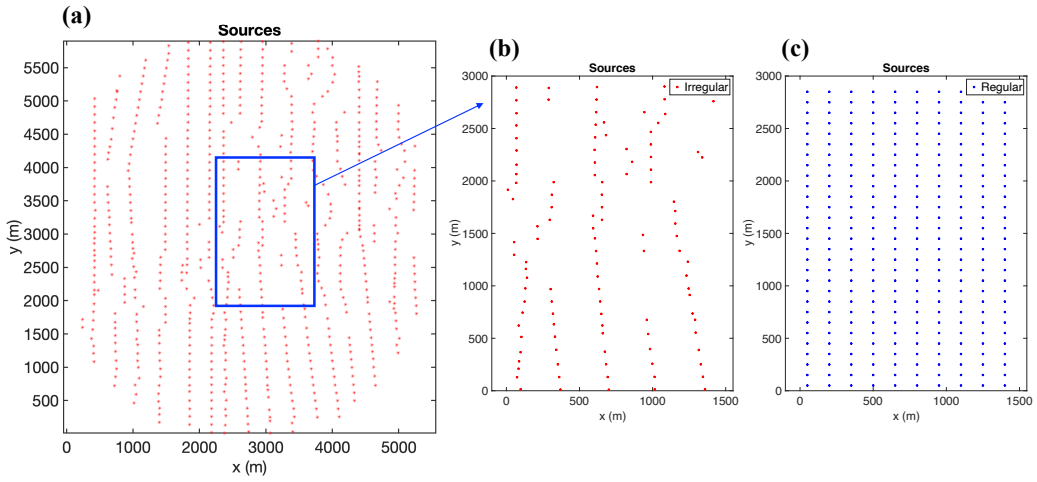

Figure 2.10: (a) Coordinates of the observed source location. (b) Observed irregular coordinates of the selected area, including 113 source points. (c) The desired regular grid of output, containing $10 \times 29 = 290$ source points.

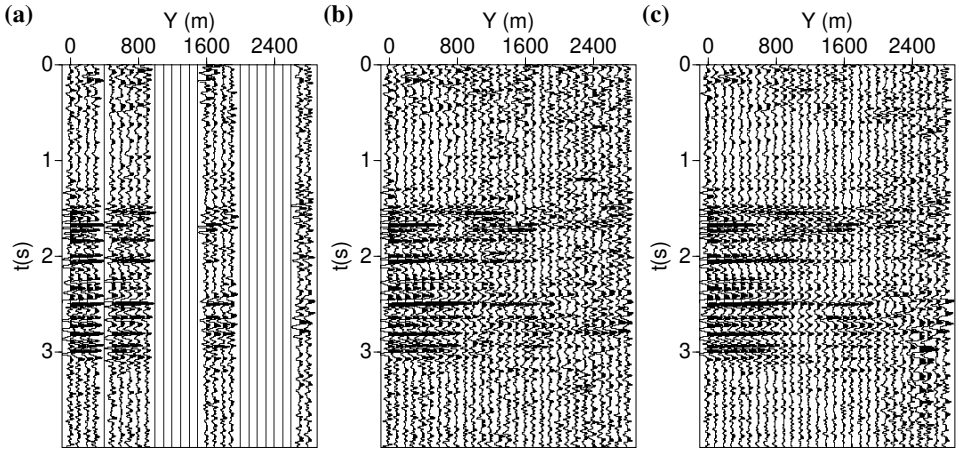

Figure 2.11: Reconstruction result for one inline slice. (a) Observed irregular grid data after binning. (b) Reconstruction result with EPOCS method. (C) Reconstruction result with I-FMSSA method.

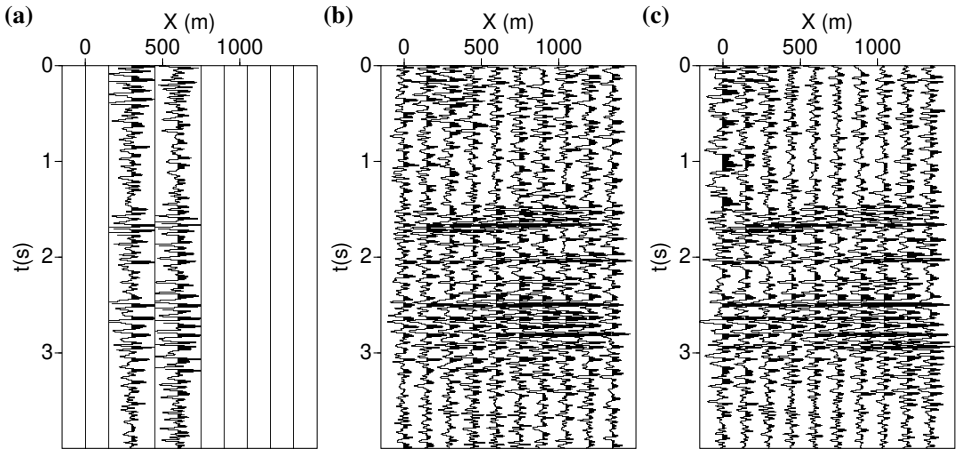

Figure 2.12: Reconstruction result for one cross-line slice. (a) Observed irregular-grid data after binning. (b) Reconstruction result with EPOCS method. (C) Reconstruction result with I-FMSSA method.

2.5 Conclusions

Compressive sensing has emerged as a promising approach for reducing the economic commitment associated with seismic data acquisition. This chapter delves into the subject of CS in seismic data acquisition, discussing its principles and specific application in seismic exploration. Successful wavefield recovery depends on three main factors: the existence of a sparsifying transform, a favourable sampling scheme, and a sparsity-promoting recovery method, which indicates using sparsifying transforms in conjunction with sparsity-promoting solvers that separate signal from sub-sampling artifacts and restore amplitudes; Random sampling break coherent aliases by introducing randomness, i.e., by designing randomly perturbed acquisition grids, or by designing randomized simultaneous sources; and relax complex combinatorial problems into tractable convex optimization problems.

Two scenarios are explored: one involving a regularly gridded survey and the other focusing on irregular or quasi-regular sensor distributions. This chapter also illustrates a comparison study for compressive arbitrary irregular grid acquisition data reconstruction with EPOCS and I-FMSSA methods. An interpolation operator is adopted to map from the observed irregular grid to the desired regular grid. The EPOCS method requires more computational time because iterations are deployed in each frequency slice to select different threshold values suitable for different frequency slices. Both EPOCS and I-FMSSA methods are effective for empty traces reconstruction. For the synthetic examples, the EPOCS method also reconstructs the random noise and shows signal leakage of the boundary traces. On the

contrary, the I-FMSSA method can simultaneously reconstruct and attenuate the random noise. For the real data, a slightly cleaner reconstruction result with the I-FMSSA method is observed. However, a lack of ground truth as a reference makes it hard to precisely evaluate the reconstruction performance of the EPOCS and I-FMSSA methods. Beyond that, it does not mean that the EPOCS method can be replaced with the I-FMSSA method when dealing with irregular grid reconstruction problems, as it is feasible to attenuate the random noise by following a denoising algorithm after EPOCS reconstruction.

CHAPTER 3

Data-driven optimal sparse seismic acquisition design and reconstruction [1]

This chapter presents a data-driven approach for optimal sparse sensing in seismic acquisition. The organization of the chapter is as follows: initially, the theory of optimal sensing is introduced, building upon the pre-learned basis library discussed in Chapter 1. This data-driven method employs QR factorization with column pivoting to select optimal sensor locations. Following this, the chapter explores various applications, particularly emphasizing the selection of optimal sensor locations. The utility of this approach in simultaneous source acquisition and time-lapse surveys is examined. Additionally, a noise removal method involving the learned basis functions and analytical basis functions is provided with synthetic and real data examples.

3.1 Introduction

The seismic acquisition constitutes the initial step of seismic exploration and often accounts for over 80% of the substantial financial outlay, which frequently amounts to tens of millions. Prudent optimization of the acquisition design can lead to direct and significant cost reductions, bearing economic and strategic implications. Additionally, field seismic acquisition has a bearing on Health, Safety, and Environmental (HSE) concerns. Shortening the acquisition

[1] A version of this chapter is published in Y. Guo, R. Lin, and M. D. Sacchi, 2020, Sensor placement optimization for seismic data acquisition and shot reconstruction: *Geoconvention*, Y. Guo and M. D. Sacchi, 2020, Data-driven time-lapse acquisition design via optimal receiver-source placement and reconstruction: *SEG Technical Program Expanded Abstracts*, 66-70, and M. D. Sacchi, 2022, Seismic noise attenuation via learned and analytical basis functions: *Second International Meeting for Applied Geoscience & Energy*, 2867-2871.

time, particularly in challenging environments like the arid deserts of the Middle East or the frigid polar regions of northern Canada, could notably diminish the likelihood of field accidents. This would also mitigate the ecological footprint of field operations. Furthermore, an optimized field survey could enable the completion of comprehensive data acquisition with fewer shots and receivers, which means a relatively smaller dataset could be expeditiously obtained, facilitating faster data transmission during on-site acquisition. Compared to the current practice of amassing voluminous data, the streamlined dataset yielded through optimized acquisition can be more quickly transmitted and analyzed, which, in turn, becomes a critical asset for rapid monitoring applications.

Compressive sensing (CS) provides a compelling alternative to the traditional Nyquist-Shannon theorem for source-receiver sampling. When there's prior knowledge about a signal, customizing a specific basis for that signal becomes viable (Manohar et al., 2018). Instead of resorting to universal bases like Fourier, Wavelets, or Curvelets, a data-driven dimensionality reduction method is leveraged to derive the most appropriate basis for efficient signal recovery. Subsequently, this data-driven methodology can determine the placement of sensors using the optimal sparse sensing technique.

3.2 Data-driven optimal sparse sensing

The CS approach excels in recovering high-dimensional signals of unspecified content by employing random measurements based on a universal basis. But when there's specific knowledge about the signal, one can optimize sensor locations for those particular signals of interest (Manohar et al., 2018). Dominant features are extracted from a training dataset consisting of representative features.

Given a pre-learned basis $\Psi_r \in \mathbb{R}^{n \times r}$ as elaborated in section 1.6, a compressible signal $\mathbf{x}$ can potentially possess a representation with a reduced rank

$$\mathbf{x} = \boldsymbol{\Psi}_{\boldsymbol{r}}\mathbf{a}, \tag{3.1}$$

where $\mathbf{a} \in \mathbb{R}^r$ denote the coefficients.

The primary challenge in this study lies in designing a measurement matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$ that encapsulates a minimal set of optimized measurements

$$\mathbf{y} = \mathbf{C}\mathbf{x} = \mathbf{C}\boldsymbol{\Psi}_{\mathbf{r}}\mathbf{a} = \boldsymbol{\Theta}\mathbf{a}, \tag{3.2}$$

which is the optimal sensing problem, where $\mathbf{y} \in \mathbb{R}^p$ and $\boldsymbol{\Theta}$ represents the sensing matrix.

Effective measurement $\mathbf{C}$ given a basis Ψ_r are chosen so that $\mathbf{\Theta}$ is well conditioned for signal reconstruction. Hence, it is possible to solve for the sparse or the low-rank coefficients $\mathbf{a}$ given the measurements $\mathbf{y}$, by pseudo-inverse of $\mathbf{\Theta}$ as indicated in equation 3.2.

Desired sensor placement results in a measurement matrix $\mathbf{C}$ that is optimized to recover $\mathbf{a}$ from the observed measurements $\mathbf{y}$. The structure of the sampling matrix $\mathbf{C} \in \mathbb{R}^{p \times n}$ is as follows:

$$\mathbf{C} = \left[\mathbf{e}_{\nu_1}\ \mathbf{e}_{\nu_2} \dots \mathbf{e}_{\nu_k}\right]^T, \tag{3.3}$$

where $\mathbf{e}_{\nu_k}$ represents the canonical basis vector for $\mathbb{R}^n$ with a unit entry specifically at index ν_k and zeros elsewhere. Consequently, $\mathbf{y}$ comprises K elements, each corresponding to the elements in $\mathbf{x}$ at the locations specified by each ν_k

$$\mathbf{y} = \mathbf{C}\mathbf{x} = \left[\mathbf{x}_{\nu_1}\ \mathbf{x}_{\nu_2} \dots \mathbf{x}_{\nu_k}\right]^T, \tag{3.4}$$

where $\nu = [\nu_1 \dots \nu_k]$ denotes the set of indices corresponding to the locations of the sensors, which has been defined in section 1.4.

When the signal $\mathbf{x}$ is unknown, its reconstruction can be obtained using

$$\hat{\mathbf{x}} = \mathbf{\Psi}_r \hat{\mathbf{a}},\ \text{where}\ \hat{\mathbf{a}} = \mathbf{\Theta}^{-1}\mathbf{y} = (\mathbf{C}\mathbf{\Psi}_r)^{-1}\mathbf{y}. \tag{3.5}$$

A diagram illustrating sparse sampling using a pre-learned basis library $\mathbf{\Psi}_r$ is presented in Figure 3.1. Optimal sensor placements align with locations that enable the most precise reconstruction. Consequently, the challenge of determining sensor positioning aims to find rows within $\mathbf{\Psi}_r$ that correspond to individual sensor locations, ensuring the matrix $\mathbf{\Theta}$ is optimally conditioned for inversion.

The condition number of a matrix $\mathbf{\Theta}$ measures the sensitivity of matrix multiplication or inversion to errors in its input. It is desirable to explicitly control the condition number of $\mathbf{\Theta} = \mathbf{C}\mathbf{\Psi}_r$ by strategically selecting rows of $\mathbf{C}$. One approach could involve optimizing the product of the magnitudes (determinant) of its eigenvalues or singular value spectrum (Manohar et al., 2018)

$$\nu = \underset{\nu, |\nu|=K}{\arg\max} \left|\det \mathbf{M}_\nu\right|. \tag{3.6}$$

For brevity, the matrix intended for inversion is represented as $\mathbf{M}_\nu = \mathbf{\Theta}^T\mathbf{\Theta}$. Recall that ν determines the structure of $\mathbf{C}$ (which signifies the sensor positions), and hence it affects the condition number of $\mathbf{\Theta}$ and $\mathbf{M}_\nu$. One can indirectly constrain the system's condition number by adjusting the spectral properties of $\mathbf{M}_\nu$ through its determinant, trace, or spectral radius.

The reduced matrix QR factorization with column pivoting decomposes a matrix into three components: a unitary matrix $\mathbf{Q}$, an upper-triangular matrix $\mathbf{R}$, and a column permutation

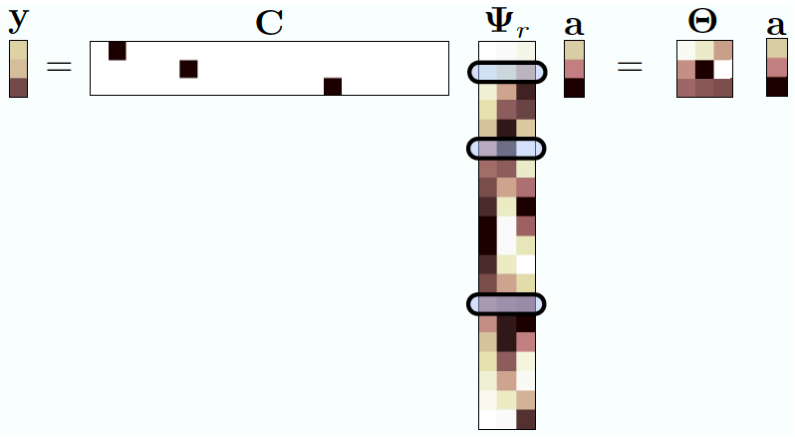

Figure 3.1: A full state reconstruction of $\mathbf{x}$ from point observations is achieved through a least squares estimation of the coefficients $\hat{\mathbf{a}}$ (Manohar et al., 2018).

matrix $\mathbf{C}$. The pivoting procedure provides an approximate greedy solution method for the optimization in equation 3.6, also known as submatrix volume maximization, as the absolute value of the determinant gives matrix volume. In brevity, the main aspect of the pivoted QR approach lies in identifying columns with the largest ℓ_2 norm. Upon identification, these columns are used as a basis to apply the Gram-Schmidt process, orthogonalizing the remaining columns in relation to this direction. Initially, the column with the highest magnitude is selected, followed by the orthogonalization of all other columns relative to it. Subsequently, the column with the next highest magnitude is identified and brought into focus, and a similar orthogonalization is conducted for the remaining columns. This procedure is repeated iteratively, employing Gram-Schmidt orthogonalization to factor out each column in a sequence that resembles a greedy optimization algorithm. A crucial outcome of this process is the identification of pivot locations, which correspond to the rows with the largest magnitude in each iteration. The pivot locations indicate the sensor positions that are optimally aligned with the given basis, and this approach is logically coherent, as it essentially involves selecting sensors that exhibit the highest ℓ_2 norm across various modes in the library. These sensors provide substantial information about the variations in the columns of $\mathbf{\Psi}_r$. Following sensors, in a sequential manner, offer progressively less information about the variances in these columns, adhering to a diminishing order of informational contribution.

During the QR column pivoting process, the submatrix formed by the pivoted columns is incrementally enlarged. This enlargement is achieved by first selecting a new pivot column with the highest two-norm. Following this, every other column has its orthogonal projection onto the pivot column subtracted. This procedure enhances the submatrix volume by

instilling a structure of diagonal dominance. The rationale behind this approach is that the matrix volume also equates to the product of its diagonal elements. Therefore, the QR factorization with column pivoting yields r point sensors (pivots) that best sample the r basis modes $\mathbf{\Psi}_r$ (Manohar et al., 2018)

$$\mathbf{\Psi}_r^T \mathbf{C}^T = \mathbf{QR}. \tag{3.7}$$

Thus, based on the aforementioned discussion, the pivots derived from the QR factorization ensure optimal conditioning of the measurement matrix or the row-selected pre-learned basis.

3.3 Data-driven optimal sparse seismic acquisition

Seismic acquisition design pertains to the selection of locations for sources or receivers. Optimal sparse sensing in seismic acquisition identifies the best positions for a limited number of sources or receivers, ensuring satisfactory reconstruction outcomes in comparison to traditional acquisitions on dense regular grids. Recognizing that this design is synonymous with the sampling scheme is essential.

One distinction between optimal sparse sensing and CS is the sampling scheme. In the realm of CS, the imperative is on random sampling, and the quantity of measurements must adhere to specific conditions to fulfill the restricted isometry (RI) properties (Candes and Tao, 2006; Donoho, 2006; Baraniuk, 2007). Contrarily, in the context of optimal sparse sensing, the sampling is specified to the positions corresponding to the most informative locations. Moreover, the metric for determining the number of measurements is divergent between the two. In CS, it's influenced by the sparsity of the signal within its basis, which mandates the condition $p \approx \mathcal{O}(P \log(n/P))$ as stipulated by Candes and Tao (2006); Donoho (2006) and Baraniuk (2007). However, optimal sparse sensing derives its number of measurements from the intrinsic rank r of the data, often a value considerably lower than the sparsity P. Consequently, this suggests optimal sparse sensing can achieve similar or superior outcomes with fewer measurements than CS, highlighting its efficiency.

The challenge of optimal sensing can be framed as follows:

$$\mathbf{d}_{\boldsymbol{\nu}} = \mathbf{T}\mathbf{U}_r\mathbf{a} + \mathbf{n} = \mathbf{\Theta}\mathbf{a} + \mathbf{n}, \tag{3.8}$$

where $\mathbf{\Theta}$ is the sensing matrix that $\mathbf{\Theta} = \mathbf{T}\mathbf{U}_r$, and $\mathbf{n}$ denote noise. The key is to devise the sampling matrix $\mathbf{T}$ such that it extracts the most pertinent rows from the basis $\mathbf{U}_r$. In essence, an optimal sampling matrix would efficiently capture the most informative rows

from the basis. Regarding the noise attenuation issue, the last section introduces a hybrid method to address it.

To elucidate this concept, let's consider the optimal receiver location problem, using it as a case study to demonstrate how to construct a sampling matrix that efficiently captures pivotal information.

To formulate the sampling matrix in accordance with the problem description presented in section 1.4, let's commence by generating a sparse diagonal matrix $\mathbf{T}$ with dimensions $(N_t \times N_r) \times (N_t \times N_r)$. This matrix, $\mathbf{T}$, is structured with alternating blocks of identity matrices and zero matrices, each having dimensions of $N_t \times N_t$. Notably, the presence of an identity matrix block in the sampling matrix $\mathbf{T}$ signifies that the receiver at that position is active. Conversely, a zero matrix block implies the inactivity of the corresponding sensor. Subsequently, to create the final sampling matrix $\mathbf{T}$, with dimensions $(K \times N_t) \times (N_t \times N_r)$, all zero-filled rows are discarded. K denotes the number of desired sensors. This exclusion is vital as it ensures the sampling matrix only encapsulates the positions of the active sensors, thus streamlining the matrix and facilitating faster computations.

In addition, contrary to CS, equation 3.8 presents an overdetermined system, not an underdetermined one. This distinction arises since there are no zero entities in $\mathbf{a}$, and the volume of the known measurements, represented by $\mathbf{d_\nu}$, is larger than that of the unknown coefficients $\mathbf{a}$. This characteristic inverts the typical CS paradigm where there are more unknowns than known measurements. Figure 3.2 graphically elucidates the application of optimal sensing in seismic acquisition, providing a lucid depiction of the contrasts between the two methodologies.

Considering the three key components discussed in the CS section, I propose the adoption of optimized sparse sampling, and the strategies to address the three questions through optimized sparse sampling are as follows:

- Basis function library: Instead of employing the sparse transformations commonly used in CS, an optimized basis function library tailored for seismic data is developed. This aims at capturing the intrinsic features of seismic data more effectively.
- Optimized sampling matrix: An optimized sampling matrix is designed rather than using a random sampling matrix, as is typically done in CS. This approach aims to provide a more accurate and reliable means of data acquisition.
- Reconstruction algorithm: The least-squares algorithm is employed for data reconstruction, which contrasts to traditional iterative methods and offers the advantage of faster computational speed, thereby making the overall process more efficient.

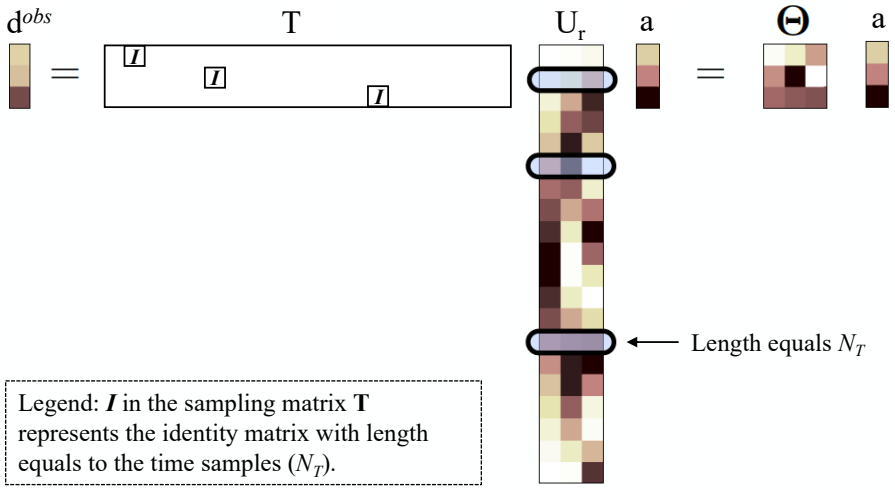

Figure 3.2: Schematic of measurements in the optimal sensing framework. Revised from Manohar et al. (2018) to fit seismic acquisition problem.

These strategies aim to fine-tune the CS framework, adapting it more precisely for seismic data acquisition's unique requirements and challenges. This optimized approach can offer both better performance and greater computational efficiency.

The subsequent key query is: how to determine if a sensor is active? In essence, this means how to optimize the sampling scheme or, equivalently, how to pinpoint the most informative rows of the basis. The foundation for addressing this query is the understanding that the primary objective of designing an effective sampling matrix is to ensure that the sensing matrix $\mathbf{\Theta}$ is well-conditioned for signal reconstruction (Van Loan and Golub, 1996). This implies that the core goal of the data-driven optimal placement design lies in identifying the rows of $\mathbf{U}_r$ that best facilitate the inversion of $\mathbf{\Theta}$.

Multiple strategies exist to bound the condition number of the sensing matrix. While indirect approaches might optimize the spectral content, direct methods often demand a combinatorial search over all feasible configurations, resulting in NP-hard problems. The use of QR decomposition to resolve least-squares dilemmas was introduced by Businger and Golub (1965). Further, Drmac and Gugercin (2016) illustrated the effectiveness of the column-pivoted QR decomposition in judiciously selecting sensor placements. Employing the QR factorization augmented with column pivoting, sensor positions are chosen iteratively, prioritizing those with the most substantial norms. Consequently, QR column pivoting serves to condition the measurements. Following this logic, the sampling matrix $\mathbf{T}$ can be subsequently deduced.

Similarly, I utilized the QR factorization with column pivoting technique as described by Strang et al. (1993) to decompose the transpose matrix $\mathbf{U}_r^T$, specifically addressing the challenge of optimal sensor placement

$$\mathbf{U}_r^T \mathbf{T}^T = \mathbf{Q}\mathbf{R}, \tag{3.9}$$

where $\mathbf{Q}$ denotes a unitary matrix, $\mathbf{R}$ symbolizes an upper-triangular matrix, and $\mathbf{T}$ stands for a column permutation matrix. Subsequently, utilizing the sparse measurements, one can determine the coefficient $\hat{\mathbf{a}}$ through the least-squares method

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{a}\|_2^2 \,, \text{subject to } \|\mathbf{d}_{\boldsymbol{\nu}} - \mathbf{T}\mathbf{U}_r\mathbf{a}\|_2^2 \leq \sigma. \tag{3.10}$$

Equation 3.10 can be recast as minimizing the following cost function:

$$J = \| \mathbf{d}_{\boldsymbol{\nu}} - \mathbf{T}\mathbf{U}_r\mathbf{a} \|_2^2 + \mu\|\mathbf{a}\|_2^2, \tag{3.11}$$

where μ is used to fit the data to give tolerance as expressed in equation 3.10. Then, finding the gradient of the objective function with respect to $\mathbf{a}$ and setting it to zero, the solution is:

$$\hat{\mathbf{a}} = (\mathbf{U}_r^T \mathbf{T}^{\mathbf{T}} \mathbf{T}\mathbf{U}_{\mathbf{r}} + \mu\mathbf{I})^{-\mathbf{1}} (\mathbf{T}\mathbf{U}_{\mathbf{r}})^{\mathbf{T}} \mathbf{d}_{\boldsymbol{\nu}}. \tag{3.12}$$

Ultimately, the reconstruction $\mathbf{d}^{rec}$ can be computed by

$$\mathbf{d}^{rec} = \mathbf{U}_r\hat{\mathbf{a}}. \tag{3.13}$$

3.4 Examples

This section provides potential applications of the proposed optimal sparse sensor design. Initially, I spotlight the selection of optimal sensors to underscore the effectiveness of optimal sensing. Additionally, I navigate through various scenarios, examining the robustness of the design in the face of noise contamination and evaluating the influence of parameter choices. Lastly, I extend the exploration to its application in simultaneous source acquisition and time-lapse surveys.

3.4.1 Optimal sensor location selection

To evaluate the efficacy of the proposed method, I generate a set of common source gathers (CSGs) using a simple model, as depicted in Figure 3.3. Both sources and receivers are

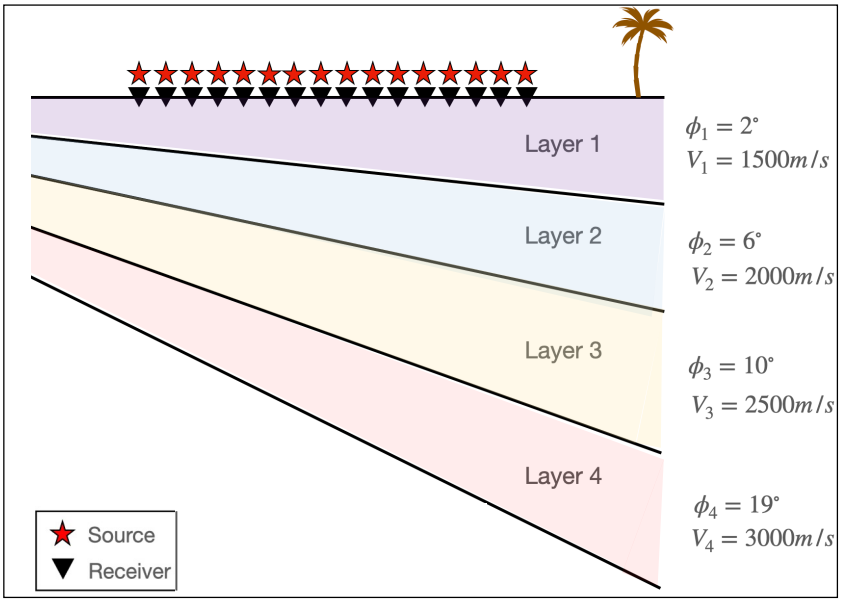

Figure 3.3: A synthetic 2D velocity model of four dipping reflectors.

placed at the model's surface. In total, there are 200 receivers and 400 sources, uniformly spaced. The velocity model is characterized by four distinct layers, each having unique dips and velocities. Synthetic seismic data, representative of these four reflectors, are shown in Figure 3.4.

To appraise the proposed method, I partition the available CSGs into two categories: training and testing data. Collecting the training data, considered a one-time expense, facilitates learning optimal locations. For this experiment, I select shot gather number 175 as the test data, while the remaining CSGs, excluding the test data, comprised the training dataset. In this scenario, the least number of receivers utilized was 86.

Figure 3.5(a) displays the original CSG corresponding to shot number 175, as generated from the model. Meanwhile, Figure 3.5(b) showcases the optimally decimated CSG for shot number 175. The reconstruction outcome is presented in Figure 3.5(c). Upon examination, it's evident that the reconstructed CSG proficiently retrieves all vital events, even with a reduced receiver count compared to the original configuration. Additionally, the achieved SNR value stands at 17.36 dB.

I further explore the method by examining its performance under various training data scenarios and assessing its ability to handle random noise during reconstruction using optimal sensing.

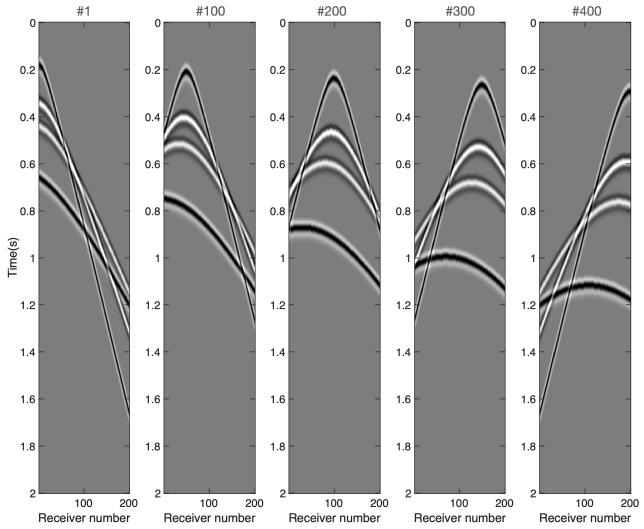

Figure 3.4: CSGs for shots 1, 100, 200, 300, and 400. The four events within each CSG vary due to the inclined layers.

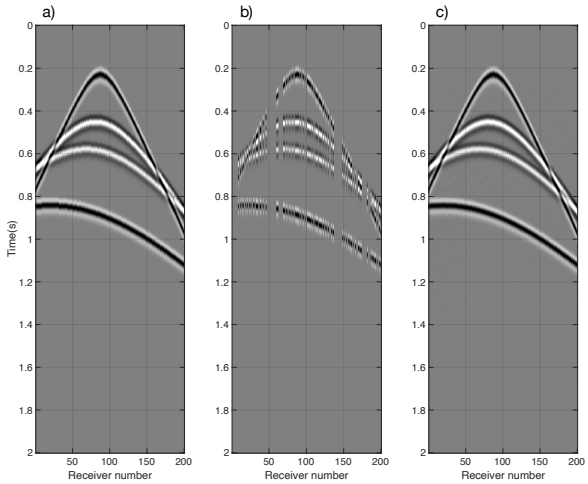

Figure 3.5: A comparison between the original CSG and the reconstructed CSG. (a) The original CSG of shot number 175. (b) Decimated CSG with optimal sampling using 86 receivers. (c) Reconstruction from optimally decimated data of shot number 175.

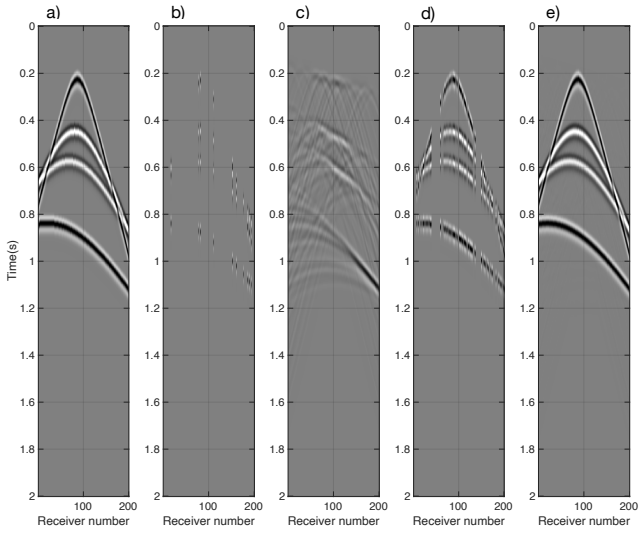

Figure 3.6: Impact of training dataset volume on reconstruction quality. (a) The original CSG for shot number 175. (b) Optimal decimated CSG using a training dataset of 50 CSGs. (c) Reconstruction of shot number 175 based on (b). (d) Optimal decimated CSG when using a more extensive training dataset of 399 CSGs. (e) Reconstruction of shot number 175 based on (d).

Comparison with different numbers of training data

The volume of the training dataset can significantly impact the reconstruction quality. To understand this effect, I conduct experiments using varying amounts of training data while keeping the test CSG consistent at shot number 175. In the first scenario, I randomly select 50 CSGs to serve as the training dataset, ensuring that the test data was excluded. On the other hand, for the second scenario, I employ a more extensive training dataset consisting of 399 CSGs, again excluding the test data. Regarding the number of receivers, only ten are utilized in the first case, while 85 are used in the second.

As visualized in Figure 3.6, the results show a marked difference between the two cases. For the first scenario, where a smaller training set was used, the SNR value was a mere 0.13 dB. In contrast, the second scenario, which benefitted from a more expansive training dataset, recorded an SNR value of 17.50 dB. This substantial increase in the SNR underscores the importance of a comprehensive training dataset. Simply put, the more extensive the training data, the more accurate and high-quality the reconstructed data is likely to be.

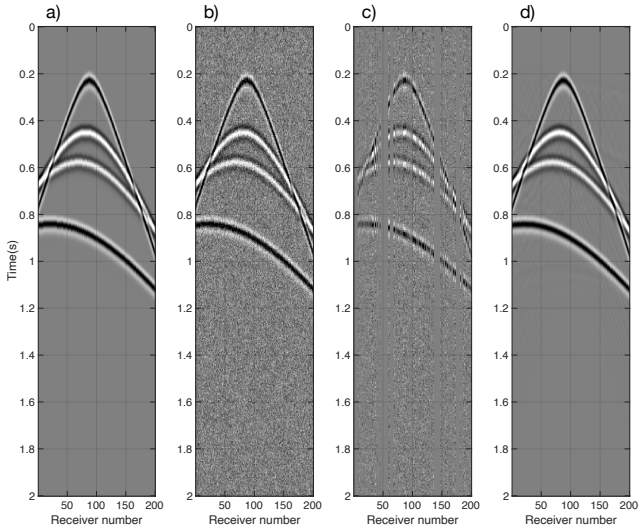

Figure 3.7: Reconstruction from noise-contaminated data using optimally selected traces. (a) The original CSG for shot number 175. (b) CSG for shot number 175 after introducing random noise. (c) Decimated CSG of shot number 175 following optimal selection. (d) Reconstruction result of shot number 175 from the optimally decimated data in (c).

Random noise elimination

Rank reduction techniques are often employed for noise suppression. In my research, leveraging rank reduction methods for constructing the basis inherently eliminates random noise from the data. The noise isn't captured within the basis, so it's naturally excluded during the reconstruction phase.

In this demonstration, I utilize a noise-free dataset for training. The clean test CSG for shot number 175 is depicted in Figure 3.7(a). I subsequently introduce random noise to this test data, as shown in Figure 3.7(b). The decimated CSG with noise contamination is displayed in Figure 3.7(c). Yet, the reconstruction from this noisy decimated data, presented in Figure 3.7(d), is mostly noise-free.

3.4.2 Application in simultaneous source acquisition

In an endeavor to further delve into the efficacy of the method when dealing with optimal sources, I integrate it with a blending acquisition (also called simultaneous source acquisition) design, a cutting-edge technique aimed at economizing on sources. For brevity, the concept of blending acquisition in seismic surveys involves using multiple sources whose

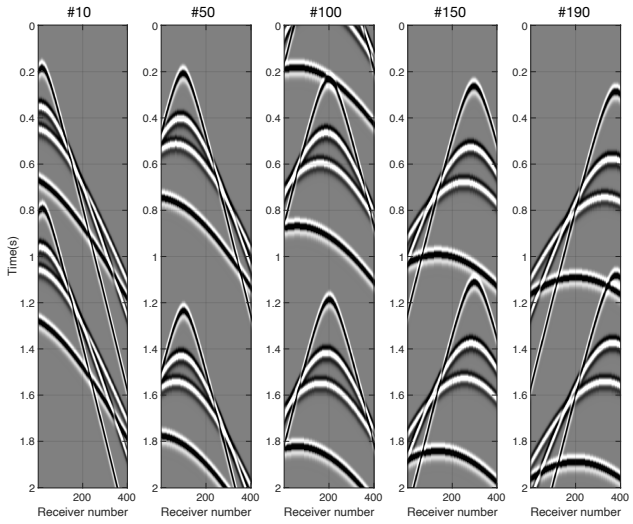

Figure 3.8: Blended CSGs for shots 10, 50, 100, 150, and 190, illustrating the overlapping signals from simultaneous source acquisition.

responses are simultaneously recorded by receivers, as detailed in Hampson et al. (2008); Ibrahim and Sacchi (2014); Cheng and Sacchi (2015); Lin and Sacchi (2020) and Abma and Foster (2020). This approach offers significant benefits, such as cost reduction, shorter survey durations, and diminished Health, Safety, and Environment (HSE) risks. However, a challenge arises during the seismic processing stage, as it necessitates the isolation of individual source responses. This means that there is a need to effectively separate the sources. Put differently, the blending noise, characterized as non-Gaussian, requires efficient removal for accurate data processing. A salient feature of the integrated approach used in this study is its ability to concurrently eliminate deblending noise during the reconstruction process.

I continue to use the four-layer dipping model, identical to the scenario for optimal receiver placement. I evenly distributed both sources and receivers along the model's surface, resulting in a total of 400 receivers and 200 sources. Instead of focusing on CSGs for receiver optimization, the attention in this scenario shifts towards common receiver gathers (CRGs), which naturally align with the optimal source placement problem. Then, I blend the data with a blending factor equals to 2, which means two sources are implemented simultaneously. Figure 3.8 depicts five CSGs of the resultant blended data. Subsequently, when transforming to CRGs (or common channel gathers), the data recorded from the secondary source is treated as blending noise, as showcased in Figure 3.9.

The clean CRG is employed as the training data. Figure 3.10(a) shows the clean test CRG of receiver 200. In contrast, Figure 3.10(b) presents the same CRG but now with

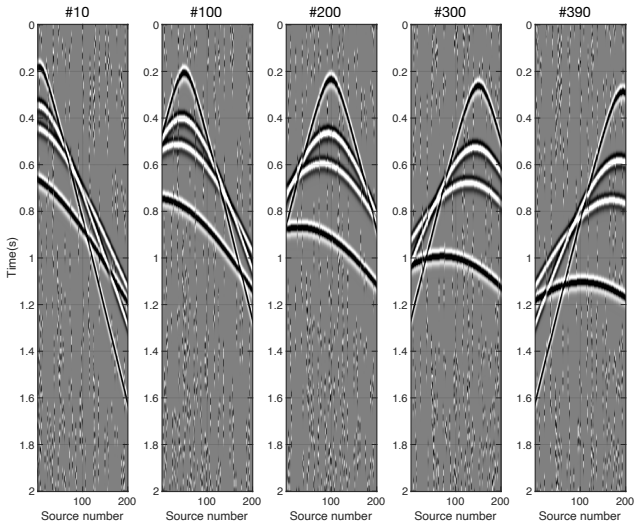

Figure 3.9: Blended CRGs for receivers 10, 100, 200, 300, and 390, showcasing the interference caused by simultaneous sources.

blending noise. When proceed to extract the optimally decimated CRG, which is shown in Figure 3.10(c), the data can be reconstructed, as visualized in Figure 3.10(d). The result underscores that reconstruction can be done simultaneously with blending noise elimination.

3.4.3 Application in time-lapse survey for carbon dioxide storage monitoring

To achieve sustainable development and build a community with a shared future of humanity, the geological storage technology of carbon dioxide (CO_2) is essential to realizing the goals. With the increasing debate focusing on reducing the environmental impact of fossil resources and transitioning away from a hydrocarbon-dependent economy, seismic acquisition methods have the potential to significantly contribute to Carbon Dioxide Capture Utilization and Storage (CCUS) projects, an essential strategy in combating global warming. In particular, geophysical monitoring of CO_2 injected back into the Earth is essential to secure its correct storage and containment in reservoirs (Davis et al., 2003; Fabriol et al., 2011). In terms of technological advancements and monitoring capabilities, seismic techniques are predominantly employed for CO_2 monitoring. These include a variety of methods such as time-lapse cross-well imaging (Spetzler et al., 2006), time-lapse 3D vertical seismic profiling (VSP) (O'Brien et al., 2004), coda-wave interferometry analysis of time-lapse VSP data (Zhou et al., 2010), and 4D seismic monitoring, which have all proven highly effective in monitoring CO_2 sequestration. Specifically, 4D seismic monitoring, following feasibility

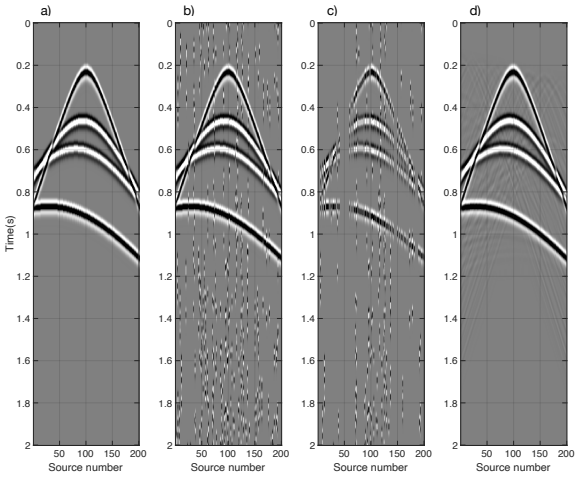

Figure 3.10: Reconstruction result of the deblending noise-contaminated data with optimally selected sources. (a) The original, clean CRG for receiver 200. (b) CRG for receiver 200 is contaminated with blending noise. (c) Optimally decimated CRG of receiver 200 with blending interference. (d) Reconstruction of receiver 200's CRG from (c), demonstrating the removal of blending noise.

studies, proves to be a reliable method for imaging the migration of CO_2 plumes (Fabriol et al., 2011). In essence, 4D seismic method (3D reflection seismology in the time-lapse modality) is necessary for CCUS projects (Arts et al., 2005; Chadwick et al., 2004, 2005, 2009, 2010; Lumley, 2010), and in this case, methods could be leveraged to obtain high-density surveys repetitively and economically, underlining their importance and utility in environmental conservation efforts.

However, how to achieve low-cost, long-term monitoring is a fundamental and cutting-edge issue. At the same time, how to quickly assess the potential risks of leakage and geological disasters and take timely countermeasures against risks is an urgent problem in risk prevention and control. To this end, I propose to apply the optimal acquisition design theory to the sensor network layout design, and the research results will provide theoretical and methodological support for the low-cost monitoring of CO_2 migration changes in timely safety decisions.

Time-lapse seismic monitoring tracks subsurface changes over time by comparing the seismic responses from multiple surveys. This approach can lead to more efficient acquisition strategies when properly leveraged, especially in scenarios where the base and monitor models have significant geological similarities. As shown in diagram 3.11(a), the traditional approach is to deploy sources and receivers in a dense, regular pattern, guaranteeing comprehensive coverage and ensuring that all relevant subsurface information is captured. However, such

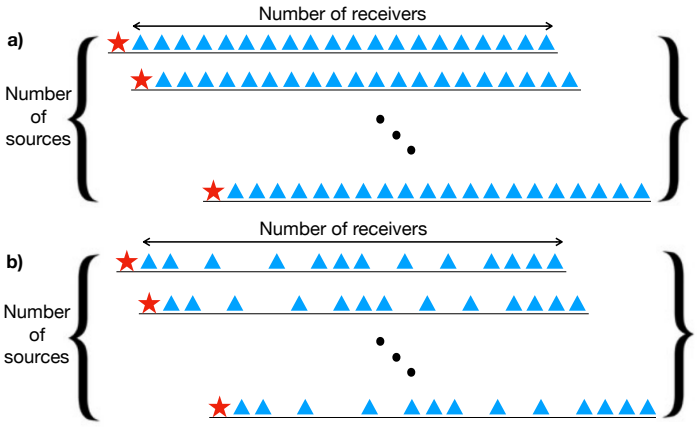

Figure 3.11: Comparison of source and receiver placements between base and monitor seismic surveys. Red stars indicate source positions, while blue triangles denote receiver locations. (a) Conventional acquisition geometry for the base model. (b) Optimized acquisition geometry for the monitor model.

a dense survey can be resource-intensive. The innovative approach proposed in this work leverages the high similarity between the base and monitor models' seismic responses. Using data from the densely acquired base model combined with a subset from the monitor survey, one can train the model to recognize the most informative areas in the subsurface, which enables designing an optimized acquisition geometry for the monitor survey, as shown in diagram 3.11(b).

To demonstrate this idea, I delve into two test cases: the Marmousi and Sleipner models. The Marmousi model is a well-known synthetic model in the geophysics community and provides a complex geological scenario to test the methodology. On the other hand, the Sleipner model simulates a real-world scenario of CO_2 geological storage monitoring, a critical application of time-lapse seismic surveys. By simulating the time-lapse monitoring of CO_2 storage using these models, one can ascertain the efficacy of the proposed optimal sparse acquisition design.

Marmousi model

To evaluate the method's performance, I generate two sets of CSGs by finite-difference modeling for both the base model (Figure 3.12(a)) and the monitor model (Figure 3.12(b)),

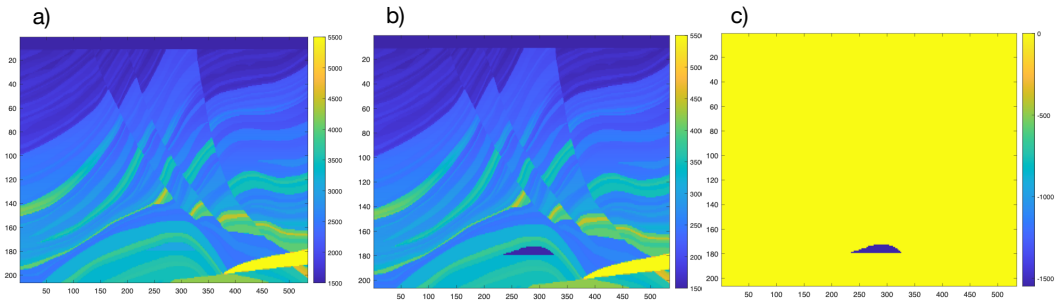

Figure 3.12: 2D Marmousi velocity model. (a) Base model. (b) Monitor model. (c) The model difference corresponds to the reservoir changes over time.

which are based on the Marmousi subsurface model. The difference between the base model and the monitor model is shown in Figure 3.12(c). Both shots and receivers are placed at the surface of the model with equal spacing, and there are 200 receivers and 100 shots in total.

I employ the CSGs derived from the base model and shots numbered from 50 to 60 from the monitor survey as the training data to determine the receivers' optimal positions. The CSGs from the monitor model, not part of the training data, serve as the test dataset. The data exhibits a low-rank characteristic, and the decline of singular values is depicted in Figure 3.13. Within this figure, the blue dots indicate the modes utilized during the reconstruction phase, achieved through the optimal hard thresholding approach for singular values.

Figure 3.14 presents the results utilizing optimal sampling and oversampling. In Figure 3.14(a) and (b) are the original CSGs of the base model and monitor model for shot number 61. The optimally decimated shot for number 61 of the monitor model is depicted in Figure 3.14(c), while its reconstruction is shown in Figure 3.14(d). Figure 3.14(e) shows the oversampled decimated shot for number 61 of the monitor model, with its corresponding reconstruction in Figure 3.14(f). Both reconstructed CSGs demonstrate promising results. Oversampling in this example holds a more incredible promise, suggesting that oversampling is more desirable in areas with complex geology. This aligns with the concept introduced in Chapter 1 as *Idea* 1, which emphasizes minimizing sampling points while allowing flexibility in data quality.

Sleipner model

In a practical example of time-lapse survey applications, I employ the Sleipner models, which are synthetic velocity models representing the Sleipner Field in Norway, a renowned site for

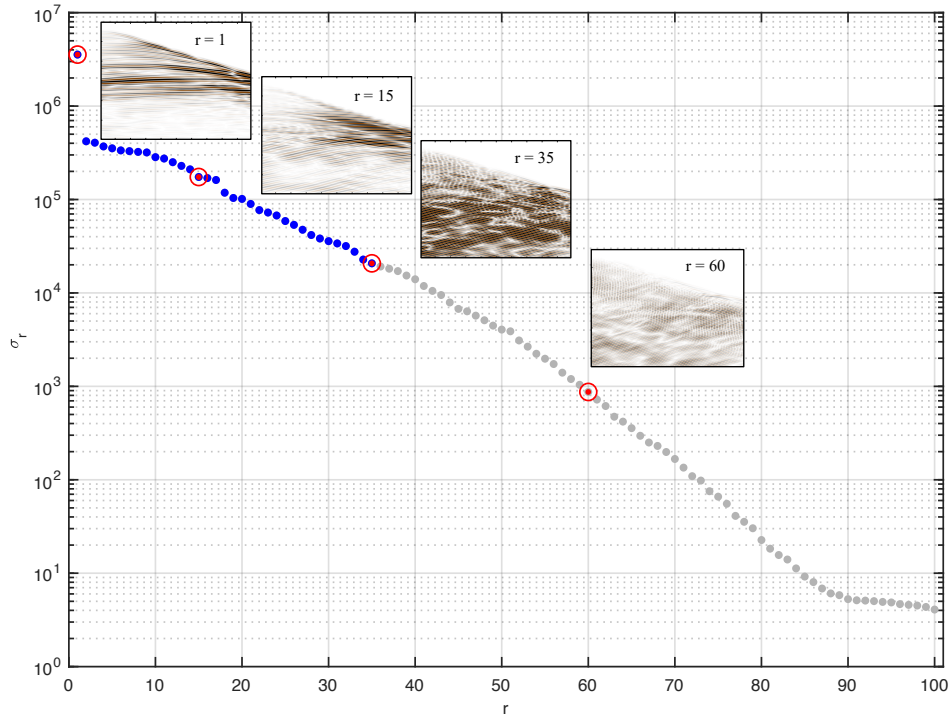

Figure 3.13: Singular values and four selected modes. The optimal singular value truncation threshold occurs at r = 35. The blue dots represent the first 35 singular values.

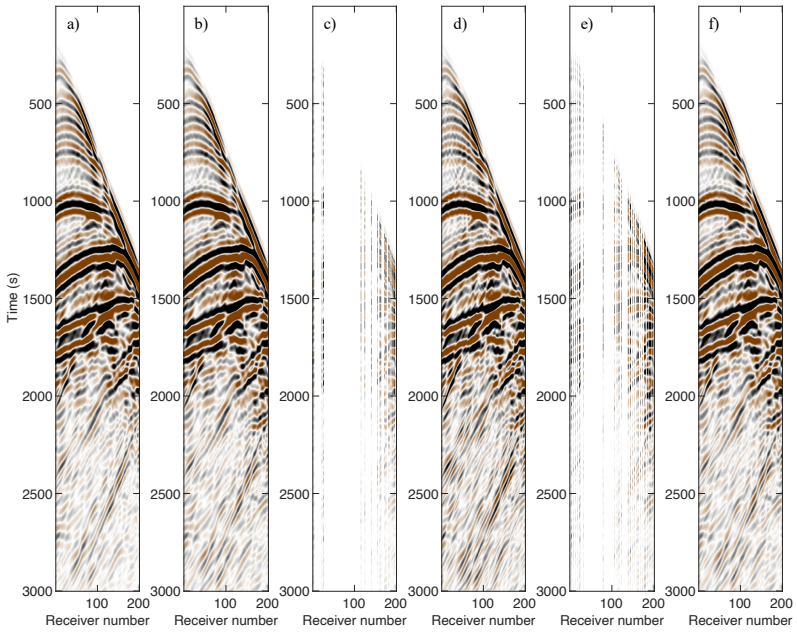

Figure 3.14: Comparison of original and reconstructed CSGs from the base and monitor models. (a) Original CSG from the base model. (b) Original CSG from the monitor model. (c) Optimal decimated CSG using 26 receivers. (d) Reconstruction from the optimally decimated monitor data using 26 receivers. (e) Oversampled decimated CSG using 48 receivers. (f) Reconstruction from the oversampled decimated monitor data using 48 receivers.

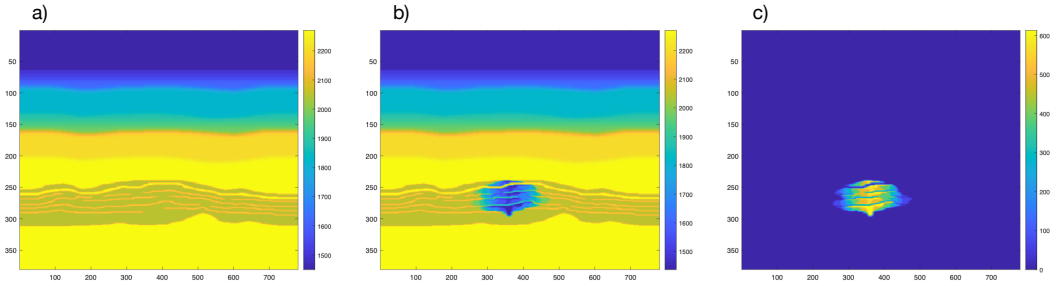

Figure 3.15: Sleipner velocity models. (a) Base model. (b) Monitor model. (c) Difference between the base and monitor models.

CO_2 sequestration. The Sleipner field, situated in the Norwegian sector of the North Sea, has been notable for being the first industrial-scale CO_2 injection project globally, aimed at mitigating greenhouse gas emissions, as described in Chadwick et al. (2005). The CO_2 is injected into the Utsira formation, a major saline aquifer from the late Cenozoic age. The Utsira Sand, characterized as a weakly consolidated sandstone, includes thin intra-reservoir shale layers. Injected CO_2, in its supercritical state near the bottom of Utsira Sand, ascends due to buoyancy until it encounters barriers like the thin shale layers and the top seal shale. Seismic methods are crucial in identifying these layers, appearing as bright subhorizontal reflections largely due to the high compressibility contrast between CO_2 and brine, coupled with constructive tuning effects at the CO_2 accumulations, as discussed in Arts et al. (2004). A combination of time-lapse 2D and 3D imaging has successfully tracked the CO_2 migration within the Utsira brine formation (Arts et al., 2005). Monitoring CO_2 sequestration is essential in any geological setting to trace CO_2 plume migration, manage the injection process, and ensure safety against potential CO_2 leakage. Therefore, a combination of various monitoring methods, including seismic monitoring, is vital throughout the sequestration's lifespan, as suggested in Rubino et al. (2011). The heterogeneous distribution of CO_2 can lead to significant attenuation and velocity dispersion effects, which could profoundly influence surface seismic data (Rubino et al., 2011), underscoring the importance of accurate and comprehensive monitoring techniques.

The Sleipner model comprises both a base model and a monitor model, making it ideal for time-lapse surveys designed to monitor the migration of CO_2 plumes. I produce two sets of data using finite-difference modeling for both the base and monitor models. Both sources and receivers are evenly spaced and positioned on the surface of the model. In total, there are 300 receivers and 300 sources.

Figure 3.16 displays five distinct CSGs from the base and monitor surveys. The differences observed in the CSGs can be attributed to the reflections stemming from differences in the

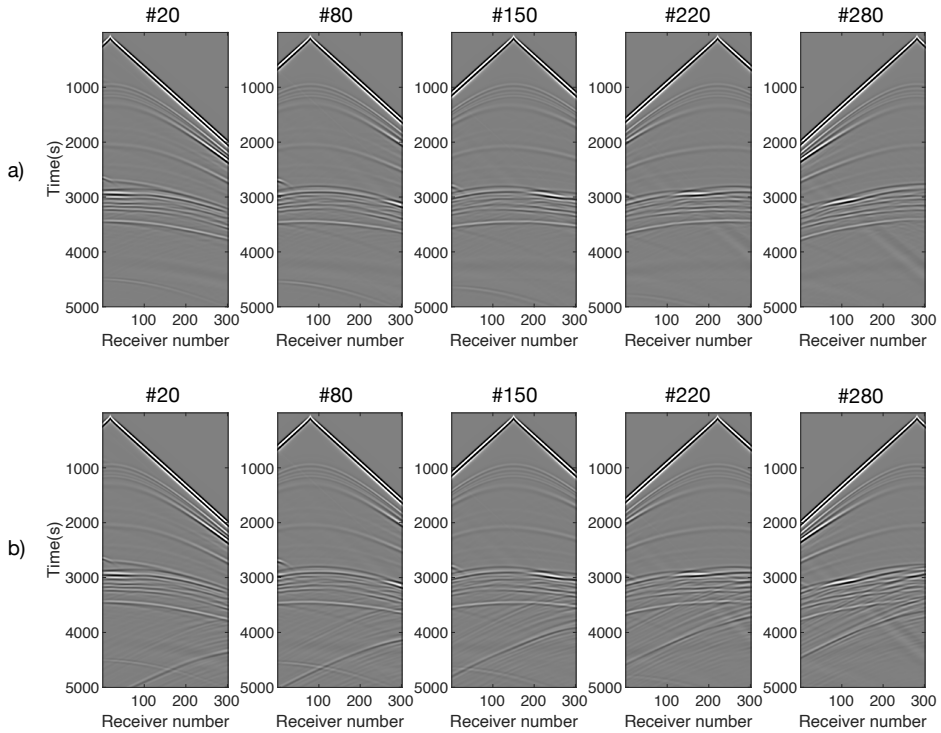

Figure 3.16: Selected CSGs from both base and monitor surveys. Upper figures: CSGs from the base survey numbered 20, 80, 150, 220, and 280. Lower figures: Corresponding CSGs from the monitor survey.

models.

The training data I utilized for an effective reconstruction stems from two distinct parts. Firstly, the CSGs derived from the base model contribute foundational information. Secondly, I incorporate ten randomly chosen CSGs from the monitor model proximate to the test gather to enrich the training dataset and capture variations. The inclusion of data from the monitor survey enhances the reconstruction outcome. This is because the monitor survey introduces information that represent the geological changes, elements that aren't present in the base model survey. Nonetheless, leveraging this approach remains economical for the monitor survey.

Figure 3.17 presents results derived from optimal sampling. Figure 3.17(a) and (b) depict the original CSGs for the base and monitor models of shot number 150, respectively. Figure 3.17(c) showcases the optimally decimated CSG of shot number 150 from the monitor model, with Figure 3.17(d) illustrating its reconstruction. The promising nature of the reconstructed result is underscored by an SNR value of 33.07 dB when employing 89 sensors.

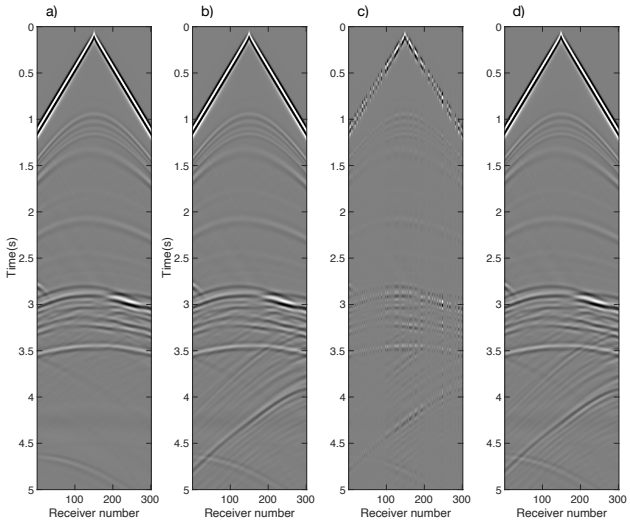

Figure 3.17: Comparison of original and reconstructed CSGs for the base and monitor surveys. (a) Base model's original CSG. (b) Monitor model's original CSG. (c) Optimal decimated CSG of the monitor survey with 89 receivers. (d) Reconstructed CSG derived from the optimally decimated data in (c).

This analysis underscores the potential of the proposed survey design. Through it, one can affirmatively demonstrate that optimally designed time-lapse measurements play an important role in elevating the quality of recovered data and extracting dependable time-lapse signatures. Given the geological similarities between the base and monitor models, their corresponding seismic responses also bear resemblance. Thus, these findings underscore the notion that a densely acquired dataset, based on the base model, can effectively guide the design of an optimal acquisition geometry for the monitor survey. This implies that the dense data acquisition for the base survey provides a robust framework for crafting an ideal sparse acquisition geometry for the subsequent monitor survey, as elaborated by Guo and Sacchi (2020). By training with a reference model, the basis function library, coupled with the corresponding optimal sampling scheme, demonstrates its capability to accurately recover seismic data that aligns with monitoring model data. This research avenue holds the potential to diminish the sensor count required for CO_2 geological storage monitoring. Concurrently, expedited data transmission promises a swift reaction to subsurface carbon dioxide transport shifts.

3.5 Seismic noise attenuation via learned and analytical basis functions

Seismic noise attenuation is a long-term problem in seismic data processing. In equation 3.8, noise in seismic data obscures valuable signals and negatively affects subsequent processing steps, including reconstruction, imaging, and interpretation. Even though many approaches have been proposed for noise attenuation, conventional filtering could harm the signal, and part of the signal energy can leak into the noise estimator. Hence, it is desirable to develop techniques capable of minimizing signal leakage. An example of the latter is a two-step processing flow. In the first step, aggressive filtering is applied, and in the second step, coherent signal (leakage) in the noise panel is added back to the aggressively denoised signal estimator. Such an ad hoc approach often leads to satisfactory results despite not having a sound mathematical framework (Choo et al., 2004).

Many researchers have tackled the signal leakage problem of noise attenuation methods. For instance, Chiu and Howell (2008) proposed to use eigenimages to construct a coherent noise model in a localized time-space window and perform the noise attenuation by adaptively subtracting the noise from the input data. Despite being an exciting workflow, the method above can potentially remove subtle signals depending on the noise level (Chiu et al., 2017). Chiu et al. (2017) used a pattern-based approach to recover the residual signal with the premise that the primaries and residual signals have similar patterns. Similarly, Tian et al. (2018) presented joint sparsity recovery for noise attenuation by exploiting intra- and inter-signal correlation structures, namely commonality, and innovations, a technique inspired by distributed compressive sensing theory. More recently, Li et al. (2019) proposed a learning-based framework when initial models are available. By employing dictionary learning and sparse inversion, Li et al. (2019) showed how one can simultaneously recover the signal and noise.

This section proposes an unexplored approach that entails modeling the signal via two operators for noise attenuation that minimizes signal leakage. Specifically, I combine learned global basis functions in conjunction with an analytical transform to model seismic signals and capture subtle features that often end in the error or leakage panel (Guo et al., 2022). This method is an extended part of section 1.6 regarding the pre-learned basis library. In this case, one doesn't need to worry about the training dataset because the data that needs to be denoised is the training data itself, which conquers the pre-request limitation of the training dataset. The latter is combined with a sparse denoising method to yield a procedure that captures the signal with minimal leakage. While many proposed algorithms attempt to minimize signal leakage via a two-step approach involving a noise-add-back strategy, this method attempts to simultaneously invert for the coefficients that model global features and

subtle signal elements extractable via the popular sparse inversion framework.

3.5.1 Formulation of the problem

Prestack seismic data can be modeled via the superposition of continuous waveforms that slowly change in space. The spatial coordinate can be the source or receiver position or any spatial attribute adopted to represent a particular type of seismic gather.

One wants to extract and model features between adjacent traces in the seismic gather, which permits the construction of a signal model able to synthesize data with spatially variant dips. The global basis captures the most informative signal features of a seismic gather. I name this basis global because it leads to bases the same size as the observed seismic gather. While global basis captures a good portion of the total signal energy, they might not correctly model subtle signal variations with offset, localized features such as diffractions or amplitudes. Therefore, one needs to add an extra transform or basis that cooperatively works with the global estimator to fit complex signals properly.

Consider a noisy 3D seismic data cube $\mathbf{D}_n$, whose size is $n_t \times n_1 \times n_2$. The integer n_t denotes the number of time samples, and n_1 and n_2 are, for instance, the number of sources and receivers or inline and crossline traces, respectively. Then, the noisy data can be represented as follows

$$\mathbf{D}_n = \mathbf{U}\,\mathbf{a} + \mathbf{M}\,\mathbf{b} + \mathbf{n}, \tag{3.14}$$

where $\mathbf{U}$ and $\mathbf{M}$ are the global and analytical bases, respectively. Similarly, $\mathbf{a}$, $\mathbf{b}$ are their corresponding coefficients. The term $\mathbf{n}$ represents random noise. The task is to first learn a plausible $\mathbf{U}$ and then, after selecting a particular forward transform $\mathbf{M}$, one needs to estimate $\mathbf{a}$ and $\mathbf{b}$ in order to synthesize the noise-attenuated signal.

Global bases

I use the pre-learned basis estimated from the data as the global basis (Guo and Sacchi, 2020). For a 3D data cube, the global bases are estimated as follows. First, the noisy data $\mathbf{D}_n$ is filtered via a technique the user chooses according to their preference. I adopt $f-xy$ multichannel singular spectrum analysis (MSSA) described in Oropeza and Sacchi (2011) to obtain preliminary filtered data $\mathbf{D}_f$. Parameter selection at this stage of filtering is not crucial because one can tolerate the presence of signal leakage. In other words, in the case of the $f-xy$ MSSA filter, a small rank parameter can be chosen to apply harsh signal denoising. One is aware that part of the amplitudes is not properly modeled if choosing a harsh MSSA filter.

Then, I use the filtered dataset $\mathbf{D}_f$ as the training data to extract basis functions $\mathbf{U}$. It is worth recalling the pre-learned basis library described previously in Chapter 1.6 in the following sentences. I first reshape the 3D cube $\mathbf{D}_f$ into a $n_t n_1 \times n_2$ matrix $\mathbf{D}$ and compute the singular value decomposition (SVD) of rank p of the matrix $\mathbf{D}$. The user can select the rank, which defines the number of basis functions required to model $\mathbf{D}_f$. Mathematically, the resulting global basis comprises the first p orthonormal left singular vectors $\mathbf{U}$, denoted $\mathbf{U}_p$.

The rationale for calling $\mathbf{U}_p$ the basis is straightforward. Any column of $\mathbf{D}$ can be written as $[\mathbf{D}]_j = \mathbf{U}_p \mathbf{a}_j$ where $\mathbf{a}_j$ is the $j-$column of $\Sigma_p \mathbf{V}_p^T$. In other words, any column of $\mathbf{D}$ can be written down as a linear combination of the columns of $\mathbf{U}_p$. When reshaped back into a data cube, a column of $\mathbf{D}$ corresponds to a data slice representing a common receiver gather, or a common shot gather.

Local bases

The local bases, or analytical bases, can be any conventional transform, like a local Fourier basis, Wavelets, or Curvelets. Because the global basis is trained from the filtered data, it cannot contain the features removed by the filtering procedure. Thus, adding the extra local basis provides the signal that is eliminated back into the estimator of the clean data and thereby, in this manner, harnesses the signal leakage problem.

The local bases with the windowing approach are summarized as follows

1. The first step of the windowing is to break the data into small overlapping data patches

$$\mathbf{S}_k = \boldsymbol{\mathcal{W}}_k[\mathbf{D}], \quad k = 1 \dots N_w, \tag{3.15}$$

   where N_w is the number of windows, and $\boldsymbol{\mathcal{W}}$ represents the windowing operator with the action of extracting a spatio-temporal window with proper tapering. $\boldsymbol{\mathcal{W}}_k$ is the k-th analysis window. The extracted k-th data $\mathbf{S}_k$ is a matrix of size $L_t \times L_x$ where L_t is the number of time samples and L_x the number of seismic traces in the window.

2. Then, apply Fourier transform, soft threshold and inverse Fourier transform to each window

$$\widehat{\mathbf{S}}_k = \boldsymbol{\mathcal{F}}^{-1} \mathcal{T}_\alpha \boldsymbol{\mathcal{F}} [\mathbf{S}_k] \quad k = 1 \dots N_w \,, \tag{3.16}$$

   where $\boldsymbol{\mathcal{F}}^{-1}$ and $\boldsymbol{\mathcal{F}}$ symbolize Fourier transform and inverse Fourier transform, respectively. $\mathcal{T}_\alpha$ denotes soft threshold.

3. Then, the data patches are properly moved back to their position and summed up

$$\widehat{\mathbf{D}} = \sum_{k=1}^{N_w} \boldsymbol{\mathcal{W}}_k^* \, [\widehat{\mathbf{S}}_k], \tag{3.17}$$

where $\boldsymbol{\mathcal{W}}^*$ is the synthesis window operator for data patch k. $\sum_k \boldsymbol{\mathcal{W}}_k^* \boldsymbol{\mathcal{W}}_k = 1$. The latter guarantees the windowing process does not introduce any distortion.

3.5.2 Numerical algorithm

The objective is to attenuate the noise, equivalent to modeling the signal. The cost function to minimize is given by

$$J = \| \, \mathbf{U}_p \mathbf{a} + \mathbf{M}\mathbf{b} - \mathbf{D}_n \, \|_2^2 + \mu \|\mathbf{a}\|_2^2 + \lambda \|\mathbf{b}\|_1. \tag{3.18}$$

The first term of the cost function (equation 3.18) is the data misfit term, which represents the difference between the modeled signal and the observed noisy data. The second and third terms are regularization terms for the coefficients $\mathbf{a}$ and $\mathbf{b}$. Analytical basis functions $\mathbf{M}$ are given by a 3D transform, and $\mathbf{b}$ is the vector containing the corresponding coefficient. Here, I use the complex Fourier transform, but a different basis can also be used. In other words, $\mathbf{b}$ represents frequency-wavenumber coefficients, which one invert uses sparse inversion via classical Fourier reconstruction method that adopts sparsity (Zwartjes and Gisolf, 2007).

The cost function 3.18 must be optimized with respect to unknowns $\mathbf{a}$ and $\mathbf{b}$. I use an alternating minimization algorithm, which first minimizes in terms of the variable $\mathbf{a}$ and then $\mathbf{b}$, which is equivalent to minimizing the cost functions

$$J_1 = \| \, \mathbf{U}_p \mathbf{a} + \mathbf{M}\mathbf{b} - \mathbf{D}_n \, \|_2^2 + \mu \|\mathbf{a}\|_2^2, \tag{3.19}$$

and

$$J_2 = \| \, \mathbf{U}_p \mathbf{a} + \mathbf{M}\mathbf{b} - \mathbf{D}_n \, \|_2^2 + \lambda \|\mathbf{b}\|_1. \tag{3.20}$$

Notice that I adopt quadratic regularization for expression 3.19 and use the orthogonality of $\mathbf{U}_p$ to simplify the inversion. Hence, the first sub-problem (equation 3.19) is the regularized least-squares problem with close form solution given by

$$\mathbf{a}^{k+1} = (\mathbf{U}_p^T \mathbf{U}_p + \mu \mathbf{I})^{-1} \mathbf{U}_p^T (\mathbf{D}_n - \mathbf{M}\mathbf{b}^k) \tag{3.21}$$

$$= \frac{1}{1+\mu} \mathbf{U}_p^T (\mathbf{D}_n - \mathbf{M}\mathbf{b}^k), \tag{3.22}$$

where k represents the iteration number. For the second underdetermined sub-problem, I use the ℓ_1 norm to promote sparsity and adopt the FISTA algorithm to solve it (Beck and Teboulle, 2009). The vector of Fourier coefficients $\mathbf{b}$ is given by

$$\mathbf{b}^{k+1} = \mathcal{T}[\mathbf{b}^k - \lambda \mathbf{M}^T(\mathbf{U}_p\mathbf{a}^k + \mathbf{M}\mathbf{b}^k - \mathbf{D}_n)], \tag{3.23}$$

where λ is an appropriate step size and $\mathcal{T}$ is the shrinkage operator. The algorithm's extra steps involve reshaping predicted components to a cube and vice-versa in each iteration when necessary. Once the algorithm reaches convergence, the final vector solutions $\mathbf{a}^*, \mathbf{b}^*$ are used to estimate the denoised data via the synthesis operator: $\mathbf{D}^* = \mathbf{U}_p\mathbf{a}^* + \mathbf{M}\mathbf{b}^*$.

3.5.3 Examples

The proposed method can be described generally in two steps: extracting the global basis and then iteratively updating global and local basis coefficients (Figure 3.18). For step one, I apply the conventional filtering method to do denoising, and here, I use the MSSA method with minimal parameter tuning. At this stage, signal leakage is permitted, which will be addressed in the second step. By allowing signal or noise leakage in the first step, the method also eliminates the need for fine-tuning parameters, which is usually required for conventional methods. Next, for step two, I use the global basis to generate global features, and the difference between the noisy data and the global features is the combination of local features and random noise. The local features are obtained by adopting local transform to promote sparsity and reconstruction. Then, the final denoised data with minimum signal leakage is obtained by iteratively updating the global and local basis coefficients to minimize the cost functions.

Synthetic examples

The synthetic data example is composed of a cube of three parabolic events (Figure 3.19(a)). I display only one slice of the cube. The data comprises $nt = 340$ samples, $n_1 = 60$ receivers, and $n_2 = 60$ sources. The rank parameter of the MSSA filers is set to $rank = 100$. MSSA works optimally for linear events, so signal leakage is expected when processing events with spatially variant dips as those in this example. Random noise is added to the clean data to obtain the noisy data in Figure 3.19(b). Then, by using the $f-xy$ MSSA method (Oropeza and Sacchi, 2011), the filtered data is estimated, which is portrayed in Figure 3.19(c). The MSSA method Figure 3.19(d) shows the difference between the filtered data and the clean data, and it is clear that some signal has been removed. As mentioned above, this is a consequence of having varying dips in the data and the MSSA algorithm not being able to

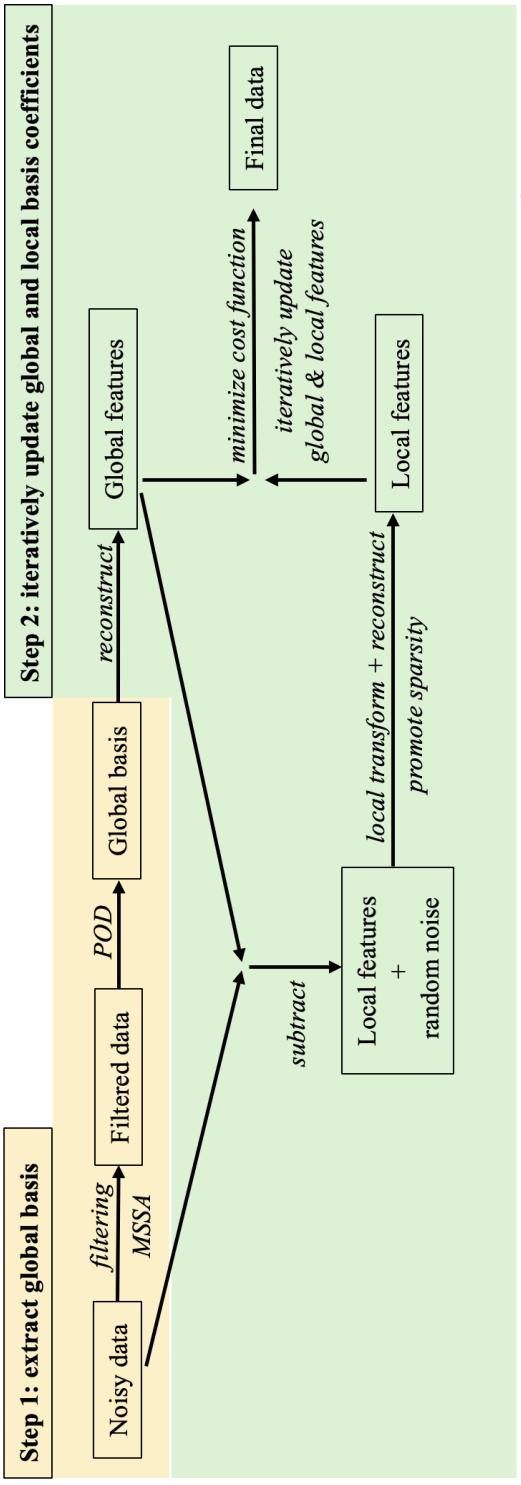

Figure 3.18: Workflow of the two-step noise attenuation method.

adapt to varying dips. The MSSA filter data was used to extract the global bases. The rank parameter is $p = 20$. Figure 3.19(e) shows results from the proposed method and Figure 3.19(f) is the difference between Figure 3.19(a) and 3.19(e). One can appreciate that the signal is not damaged during the noise attenuation. In this example, the trade-off parameters $\mu = 0.01$ and $\lambda = 0.01$ are selected heuristically in a way the variance of the estimated noise approximates the variance of the noise added to the data. Last, it is worth mentioning that the algorithm typically converges in about $10 - 15$ iterations.

The second example is more complex, with 100 receivers and 60 sources, and the slices shown here are CRGs. Figure 3.20(a) and (b) are the clean and noisy data. Figure 3.20(c) is the data after filtering, and Figure 3.20(d) shows the difference between Figures 3.20(b) and (c). A small event is lost after filtering, and the signal leakage can be seen in the difference panel. Figure 3.20(e) is the reconstructed global features using the basis extracted from the filtered data, and Figure 3.20(f) shows the local features extracted from the difference panel. Figure 3.20(g) is the final data with the combination of global features and local features. It is noticeable the leaked signal has been added back. Figure 3.20(h) is the difference between the noisy data and Figure 3.20(g); it has less signal leakage.

A real example from the Gulf of Mexico

The real data example is from the Gulf of Mexico, and I choose a data cube with 100 receivers and 251 sources for the test. As with the synthetic data example, I use the filtered data via $f - xy$ MSSA to learn the global basis function. In this example, the rank of the MSSA filter is $rank = 100$, and the rank parameter of the POD decomposition is $p = 107$.

Figures 3.21 and 3.22 compare the proposed method in the CRG and CSG domain, respectively, with the conventional denoising method. Figure 3.21(a) is the original clean data (one common channel section), and Figure 3.21(b) is the noisy data. Figure 3.21(c) is the result after applying $f - xy$ MSSA filter. It is noticeable from the difference panel (Figure 3.21(d) that a serious signal leakage problem exists. Figure 3.21(e) and (f) show the extracted global and local basis, respectively. Figure 3.21(g) is the data after adopting the proposed hybrid method, where I simultaneously invert the coefficients of the POD expansion and the 3D Fourier transform. Figure 3.21(h) shows the noise difference between Figure 3.21(b) and Figure 3.21(g), and it is clear that most of the leaked signals have been added back to the dataset. In addition, the SNR value of the MSSA filtered data with the clean data is 3.5 dB, while for the new method is 9.4 dB. Similar to the previous example, I choose $\mu = 2 \times 10^{-3}$ and $\lambda = 0.01$ in a way the final variance of the estimated noise approximates the variance of the noise in the input data. In addition, Figure 3.22 shows the corresponding results in CSGs.

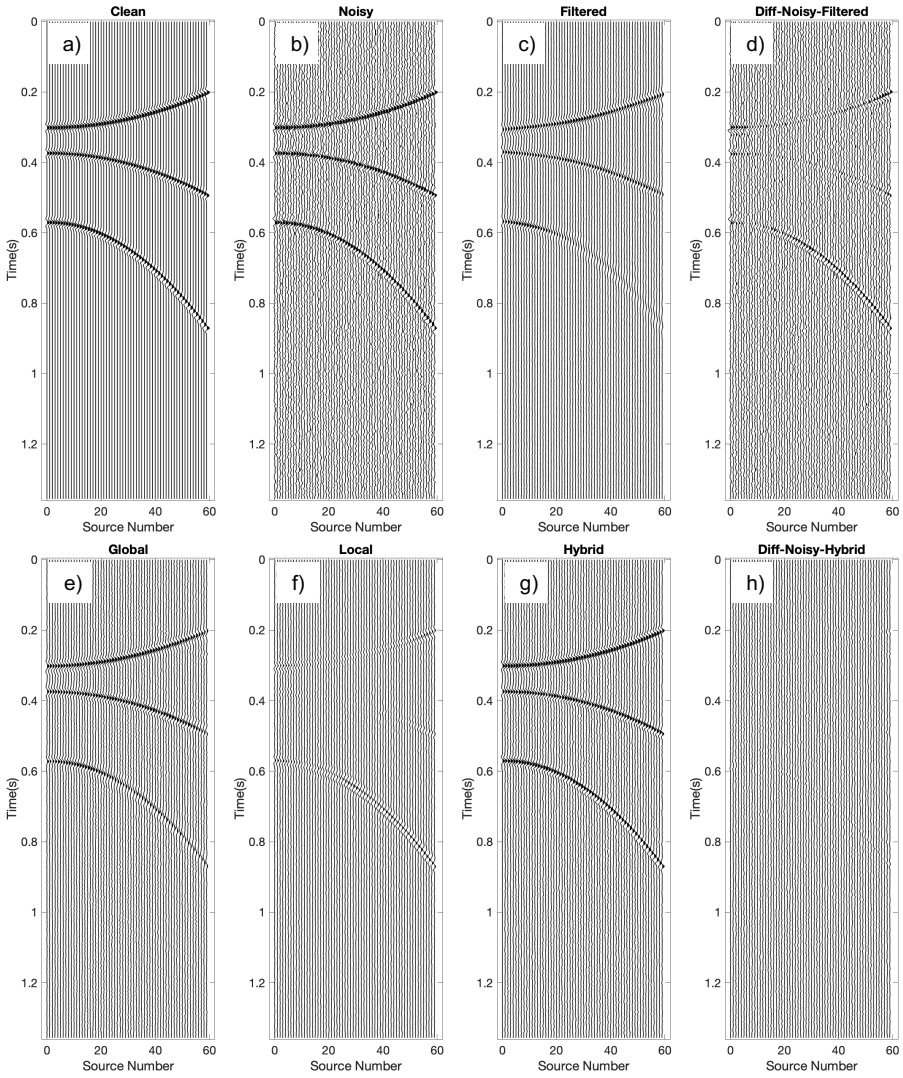

Figure 3.19: Synthetic example. (a) Clean data. (b) Noisy data. (c) Filtered data with a harsh $f - xy$ MSSA filter. (d) Difference between (b) and (c). (e) Global basis. (f) Local basis. (g) Denoised data via the proposed hybrid method. (h) Difference between (b) and (g).

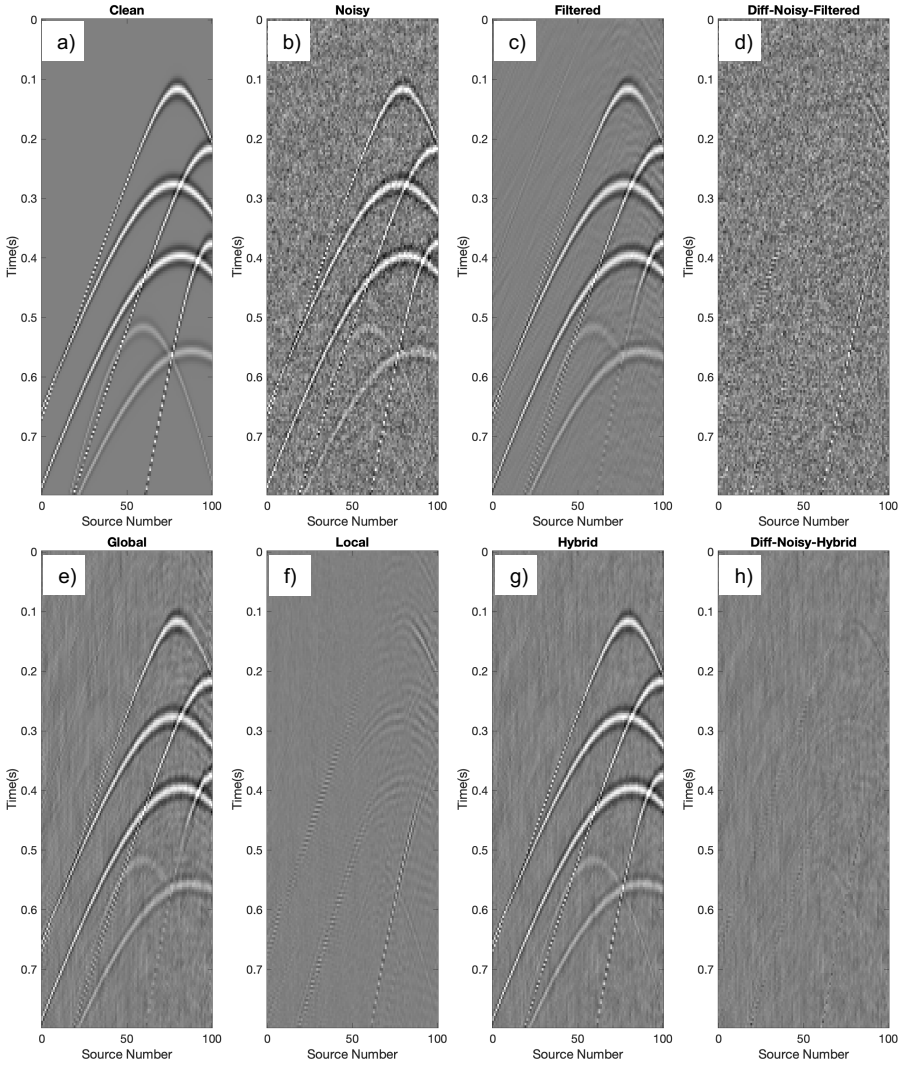

Figure 3.20: Complex synthetic example. (a) Clean data. (b) Noisy data. (c) Filtered data with a harsh $f - xy$ MSSA filter. (d) Difference between (b) and (c). (e) Global basis. (f) Local basis. (g) Denoised data via the proposed hybrid method. (h) Difference between (b) and (g).

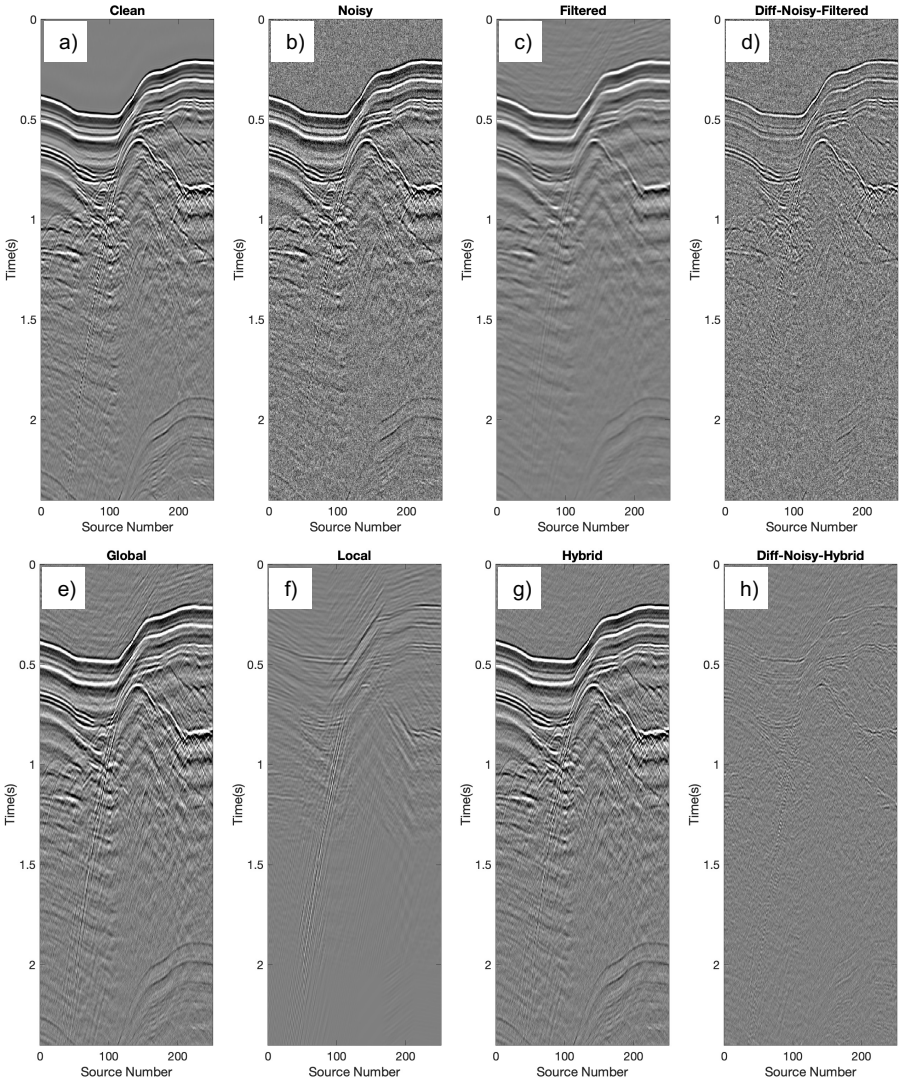

Figure 3.21: Denoising comparison of a common channel section. (a) Clean data. (b) Noisy data. (c) Filtered data with a harsh $f - xy$ MSSA filter. (d) Difference between (b) and (c). (e) Global basis. (f) Local basis. (g) Denoised data via the proposed hybrid method. (h) Difference between (b) and (g).

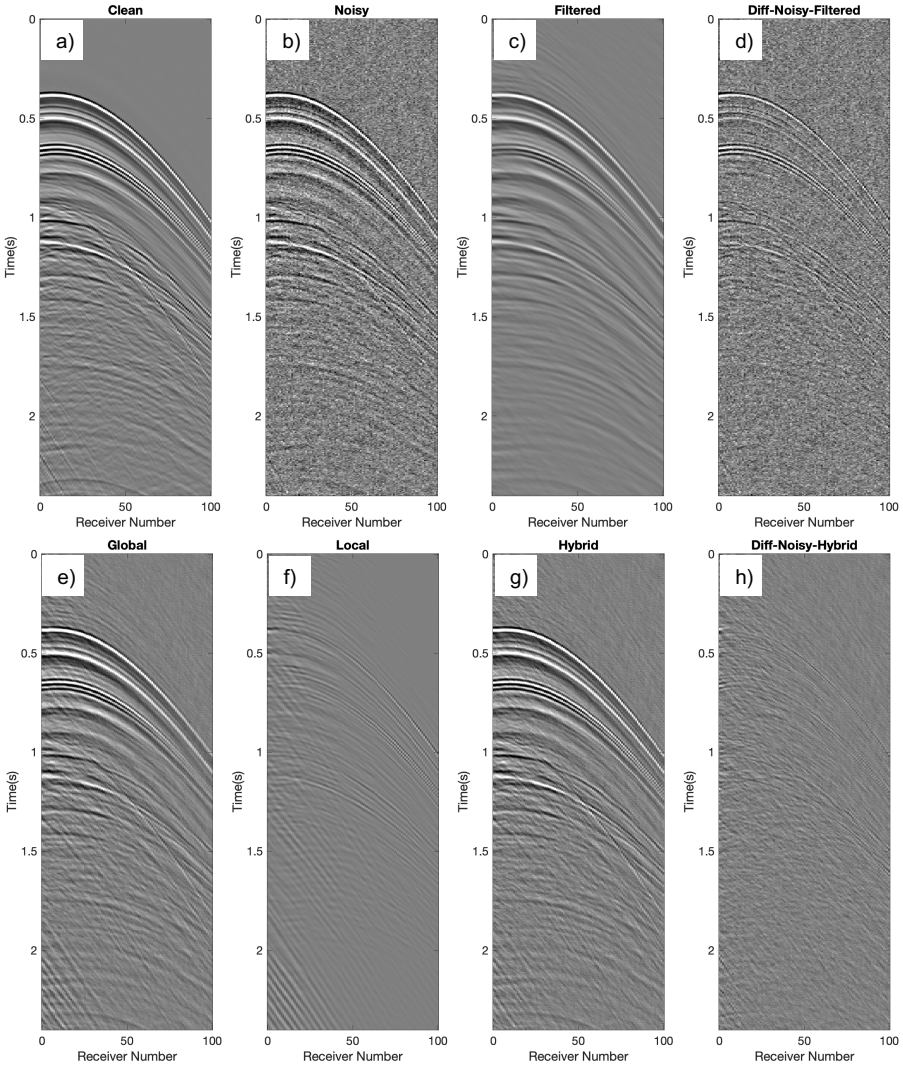

Figure 3.22: Denoising comparison of a common source section. (a) Clean data. (b) Noisy data. (c) Filtered data with a harsh $f - xy$ MSSA filter. (d) Difference between (b) and (c). (e) Global basis. (f) Local basis. (g) Denoised data via the proposed hybrid method. (h) Difference between (b) and (g).

3.6 Conclusions

In this chapter, I adopt optimal sensor selection into the domain of seismic acquisition design. The essence of the method lies in extracting a pre-learned basis library derived from the training dataset. The quest for the optimum acquisition geometry finds its answer in the QR decomposition with column pivoting. Regarding the reconstruction phase, the extracted basis stands as a reliable pillar, allowing the employ of the least-squares method for accurate and efficient results. Through an assortment of case studies, I have embarked on an initial exploration of this technique. Within the framework of seismic acquisition, the concept of optimal sensing has been discussed, and a multitude of scenarios under varying conditions has been deliberated upon.

The results of the data reconstruction prove that savings can be obtained with available training data. Besides, more training data usually yields a notable enhancement of data reconstruction quality, but there is a trade-off between precision and saving. Further, noise, i.e., random or erratic noise, can be removed simultaneously with reconstruction. The time-lapse example results reveal that a previously obtained dense base survey can optimize the monitoring design. This optimal design could lead to strategies that considerably reduce seismic acquisition costs and design intelligent data-driven approaches for seismic data acquisition. In essence, the dense acquisition from the base model serves as a guide, allowing us to infer the most crucial locations for capturing significant subsurface changes. One can deploy a sparser yet highly effective acquisition geometry for the monitor survey by harnessing this knowledge. This methodology promises resource efficiency and paves the way for more sustainable and environmentally friendly seismic monitoring practices.

Further, a hybrid method that involves inverting for coefficients that simultaneously model the signal via a global basis and analytical transform is proposed to solve the signal leakage problem that often arises in seismic processing. The global basis, which contains the general features of the dataset, is extracted from aggressively filtered data. Then, by updating the Fourier coefficients that model the leakage together with the global basis, a balance between the two bases is compromised with the premise that the modeled signal approaches the observed noisy data in the ℓ_2 norm sense. Synthetic and real data examples reveal that the signal component is not damaged compared to traditional noise removal methods.

CHAPTER 4

Optimal seismic sensor placement based on reinforcement learning approach [1]

4.1 Introduction

Seismic data acquisition involves strategically placing sources and receivers to capture seismic waves, subsequently providing insights into the planet's subsurface structure. However, determining the optimal deployment of these sensors presents a challenge due to the vast possibilities of positioning combinations. The challenge is underlined by the NP-hard nature of the optimal sensor deployment problem (Bian et al., 2006), and this complexity means that a straightforward brute-force search method, which would entail checking every potential sensor placement, becomes infeasible due to the exponential time it would require (Wang et al., 2019).

Given this challenge, researchers have sought to develop alternative methods to determine optimal sensor placements without evaluating all possibilities. Nakayama et al. (2019) offered a novel approach that amalgamates the genetic algorithm (GA) with convolutional neural networks (CNN). This combination aims to automatically determine the most suitable acquisition parameters, considering factors such as source blending and the spatial sampling of the sensors. Taking a different route, Guo and Sacchi (2020) leveraged the QR decomposition combined with column pivoting, illustrated in Chapter 3. This method, relying on a pre-learned basis library, is particularly effective in determining optimal sensor placements, especially in time-lapse seismic applications. In a more simulation-based

[1] A version of this chapter is published in Y. Guo, R. Lin, and M. D. Sacchi, 2023, Optimal seismic sensor placement based on reinforcement learning approach: An example of OBN acquisition design: *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-12.

strategy, Zhang et al. (2022) treated data reconstruction as a matrix completion problem, and by employing simulated annealing (SA), they effectively determine the best sensor positions using only the attributes of the sampling operator. These approaches provide feasible alternatives for addressing the problem of optimal sensor deployment in seismic data acquisition. Through these techniques, there's a clear potential for achieving better data quality and significant cost savings by minimizing the number of sensors deployed.

Reinforcement Learning (RL) offers a decision-making paradigm rooted in learning by interacting with an environment. It thrives on the concept of an agent making sequential decisions, learning from the feedback (in terms of rewards or penalties), and adjusting its strategies to maximize some notion of cumulative reward over time (Mnih et al., 2015; Silver et al., 2016; Sutton and Barto, 2018). Essentially, it's a continual process of trial and error, iterating through choices, collecting rewards, and refining decisions.

In geophysics, a few test adopting RL methods has been made. For instance, Ma et al. (2019) dived into automating first-arrival picking using RL. This approach shows promise in tackling the challenge of picking accurate onset times of seismic waves, a critical step in many seismic data processing sequences. On a related note, determining the origin of seismic activities is critical for many applications, from studying the Earth's interior to monitoring human-made earthquakes. Addressing this, Wu et al. (2019) introduced a deep RL method to locate seismic sources. Beyond that, Wang et al. (2020) stated using deep RL for inverting magnetotelluric data in the realm of electromagnetic fields, and Sun and Alkhalifah (2020) used RL to determine the proper time to switch between different misfit functions for Full-Waveform Inversion (FWI). More recently, Feng et al. (2022) used a convolutional autoencoder to estimate waveform database and RL to determine microseismic source location. Similarly, Dell'Aversana (2022) combined geophysical inversion with RL using Q-learning, and Zi et al. (2022) used a data-driven offline RL framework to do gamma-ray well-log depths matching and solve the pattern localization problem.

The application of RL in determining optimal seismic sensor placement remains an open frontier with opportunities for exploration and innovation. The deployment of seismic sensors in a manner that optimally captures vital information about the subsurface structures is a significant challenge due to the vast search space, rendering the problem computationally intensive and complex. In particular, RL can navigate through large and complex search spaces, make decisions in the face of uncertainty, and adjust strategies based on feedback from the environment, which are all pertinent to seismic sensor placement.

This chapter examines and tests the potential adoption of RL for optimizing seismic sensor placements. The narrative unfolds as follows: an introduction to seismic sensor placement utilizing RL is initially provided. This is succeeded by an articulation of how to cast optimal sensor placement into a Markov decision process (MDP) framework. Subsequently, the

concept of Q-learning is explored (Watkins and Dayan, 1992; Sutton and Barto, 2018). Finally, numerical experiments are shown to assess the viability of employing RL in seismic sensor deployment scenarios.

4.2 Seismic sensor placement based on reinforcement learning

This section presents the integrated RL-based optimal seismic sensor placement design. The method is rooted in a data-driven approach and harnesses the pre-learned basis library presented in section 1.6, which allows for robust computational reconstruction through the least-squares method. Notably, the metric utilized to gauge the quality of reconstructions, and hence to be maximized, is the SNR value between the ground truth data and the reconstructed data. Essentially, RL is leveraged to discover the receiver configuration that amplifies the quality of the reconstruction. It's worth noting that while the RL-based design is tethered to the SNR metric, its architecture is adaptable. Various objective functions can be seamlessly integrated into this framework (Wang et al., 2019). As such, users have the flexibility to modify the cost function to their particular problems. Progressing, I shape the problem of optimal sensor placement within the mathematical structure of an MDP, executed via the Q-learning optimization process (Watkins and Dayan, 1992; Sutton and Barto, 2018).

The scenario depicted in Figure 4.1 is first considered, which showcases a pre-defined grid encompassing N_r potential receiver positions to elucidate the concept further. Within this grid, the objective is to discern the optimal K grid points that would best sample the data. Let's recall from section 1.4 the sampling set, represented as $\boldsymbol{\nu}$. The primary objective is to identify a configuration for $\boldsymbol{\nu}$ so that the $SNR_{\boldsymbol{\nu}}$, as articulated by equation 1.9, is maximized. In simpler terms, out of the numerous potential locations available for placing the sensors, one aims to find those locations that would provide the highest SNR value of the data. One of the aims here is to give a broad view of the main approaches and to use a data-driven approach with the pre-learned basis library, which I use routinely for seismic data reconstruction problems.

4.2.1 Formulation of optimal receiver deployment as Markov decision process

The seismic sensor selection challenge seeks to discern the optimal set $\boldsymbol{\nu}$, guaranteeing superior data reconstruction. To achieve this, I turn my attention to RL, rooted in the foundational principles of the MDP framework (Sutton and Barto, 2018).

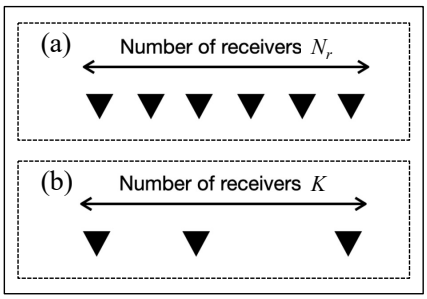

Figure 4.1: Contrasting layouts of seismic survey designs. (a) Conventional design, wherein receivers, depicted by black triangles, are evenly distributed across the survey area, employing a total of N_r receivers. (b) Optimal design, strategically deploying a reduced number of receivers (K, where $K < N_r$) at locations deemed most informative and beneficial for data acquisition. While the conventional approach ensures uniform coverage, the optimal design prioritizes positioning to enhance the quality and relevance of acquired seismic data.

At its essence, an MDP delineates how an agent systematically chooses actions based on its current state within an environment. This decision-making is influenced by the rewards it receives, even in the presence of uncertainties in transitioning between states. Formally expressed, an MDP is represented as the tuple (S, A, R, P):

- S signifies the state space, illustrating the various situations the agent can encounter.
- A represents the action space, capturing the potential actions available to the agent in response to any given state.
- R is the reward function, determining the immediate reward or penalty the agent receives upon taking an action in a specific state.
- P outlines the state transition probability, showcasing the likelihood of transitioning to a new state from the current state after an action.

The strategy or plan the agent follows, known as the policy π, embodies the probability distribution within the state-action domain. It guides the agent in selecting appropriate actions based on the given state, intending to accumulate the highest long-term reward.

The objective is to estimate the optimal policy π^* that satisfies

$$J_{\pi^*} = \max_{\pi} J_\pi = \max_{\pi} E_\pi \left[\sum_{t=0}^{\infty} \gamma^t r_t \right], \tag{4.1}$$

where $\gamma \in [0, 1)$ is the discount factor, r_t is the reward at time-step t, $E_\pi[.]$ stands for the expectation under policy π, and J_π is the expected cumulative reward.

The following illustrates how to mathematically design the optimal sensor placement as a formulation of an MDP concerning state space, action space, and reward, followed by a selection procedure.

State space

The state space in the sensor placement problem is defined as a collection of all the possible combinations of the sampling set $\boldsymbol{\nu} = [\nu_1, \nu_2 \dots \nu_K]$. The state at a specific time-step t is symbolized by s_t, spanning T time-steps for every episode. An inherent challenge arises as the quantity of sensors escalates: the combinatorial explosion of potential sensor placements. For illustration, given an initial setup with $N_r = 20$ sensors and a goal to pick $K = 3$ among them, the potential combination surges to $\frac{20!}{3!(20-3)!} = 1140$. The complexity amplifies when aspiring to select $K = 7$ from the 20 receivers, raising the combination count to $\frac{20!}{7!(20-7)!} = 77520$. The situation becomes even more complicated if the original receiver pool (N_r) substantially surpasses 20 and there's an intent to select an even larger number of sensors. This underscores the inherent NP-hard nature of the sensor placement challenge, where the enormity of potential combinations renders traditional greedy approaches largely inefficient.

When utilizing the classic RL approach rather than deep RL (Mnih et al., 2015; Silver et al., 2016; Sutton and Barto, 2018), the size of the state space constraint surpasses the algorithm can handle. As such, a necessity arises to truncate the expansive set of potential combinations into a more manageable size.

In addressing the optimal sensor selection challenge, a strategic segmentation is performed: the entire set of N_r potential positions is divided into K distinct ranges, with K representing the desired number of sensors. A key stipulation here is that sensors, when selected, must remain confined within these pre-defined segments and are forbidden from trespassing beyond their boundaries. This constraint reflects the real-world presumption that all sensors are interchangeable, so swapping positions doesn't result in any discernible difference. By imposing this constraint, the combinatorial challenge becomes notably less daunting. For instance, if the task is to choose $K = 7$ sensors from $N_r = 20$ candidate positions, the count of possible combinations under these rules would only be $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 2 = 1458$, which is much less than the number of possibilities calculated with the unrestricted scenario and results in a substantial reduction in computational complexity.

This segmentation criterion offers another advantage besides the direct benefit of shrinking the problem size. It inherently ensures that the selected sensors are more evenly distributed,

sidestepping any significant gaps in the design. This strategy echoes the principles of jittered sampling (Hennenfent and Herrmann, 2008a), a methodology known for its efficacy in avoiding aliasing and ensuring a more balanced data acquisition.

Action space

The action space for the sensor placement problem is structured around choosing the location for a particular sensor from the pool of candidate positions. To elucidate, taking an action, denoted as a, entails a two-fold process:

- Activation: This step involves selecting a particular sensor to make an adjustment to its position.
- Directional movement: Once a sensor is activated, the next step involves deciding the direction of its movement. The choices are binary: either move to the left or the right. This movement translates to opting for the adjacent candidate position on the left or right of the currently activated sensor.

It's crucial to note that any action affects only one sensor's position. Thus, when an action a_t is undertaken while in state s_t, the system navigates to a subsequent state, denoted s_{t+1}. The efficacy of the chosen action can then be assessed by comparing the data reconstructed in the newly arrived state, s_{t+1}, with the prior state, s_t. The quality of this comparison is quantified using the pre-established objective function (equation 1.9). Based on this, the algorithm determines the appropriateness of the action, guiding whether or not the move enhances the overall quality of data reconstruction.

Reward

Upon executing a specific action a_t in a state s_t, the agent receives a reward, symbolizing the immediate consequence of that action in the context of the seismic sensor placement problem. The primary objective here is to navigate through possible sensor locations, ultimately identifying a combination that maximizes the cumulative reward. The immediate, one-step reward, denoted as $r(s_t, a_t)$, is calculated by evaluating the reconstruction quality (as per equation 1.9), involving a comparison between states s_t and s_{t+1}. If the data reconstruction quality at the new state s_{t+1} exceeds that of the prior state s_t, the reward is $+10$. If not, a penalty (or a negative reward) of -2 is administered. While the agent amasses rewards at each time-step during its exploration, the algorithm's focal point remains the maximization of cumulative rewards, guided by the discount factor γ, which controls how myopic the agent is in its decision-making.

Selection procedure

Upon defining the state space, action space, and reward, Figure 4.2 shows the sensor selection procedure, delineated into seven distinct stages for clarity. Here, stars symbolize candidate positions, while black triangles indicate the selected position for the sensor.

I begin by aligning candidate positions to a conventional configuration, characterized by densely and uniformly arrayed locations, indicated by stage one shown in Figure 4.2. Then, I set sensor ranges (vertical lines represent the range boundaries) according to the number of sensors one wants to choose in stage two. Once these preliminaries are established, the proposed RL method can be used with the number of episodes to select optimal locations with an initially random given state (shown in stage three). I set the initial state randomly for each episode in the RL algorithm to explore the whole state space fully, which means the local optimum is trying to avoid. For each episode, the state transit from the old state to the new state with action is shown in stage five. Noted, stage five comprises four parts corresponding to the description provided in the action space section. The red triangle represents the activated sensor, and the red arrows indicate moving left or right. Then, suppose in state s_t (stage 5.1 in Figure 4.2), one wants to select the sensor's location from the candidate positions (black stars). The first step of the action is to decide which sensor to move (stage 5.2 in Figure 4.2). After activating the sensor, step 2 of the action is to decide to move left or right (stage 5.3 in Figure 4.2). Afterward, the new state s_{t+1} with a different combination of sensors is reached. Finally, after all the episodes, the optimal state (stage seven in Figure 4.2) corresponds to finding the sensors' combination at optimal positions, which corresponds to the optimal sampling set $\boldsymbol{\nu} = [\nu_1, \nu_2 \dots \nu_K]$.

4.2.2 Q-learning

RL is a vast domain with numerous algorithms tailored for different problems and scenarios. Among these algorithms, Q-learning stands out as a particularly influential method in the RL family. It operates on the principles of value-based learning and is both off-policy and model-free, which means it learns the value of an action without necessarily following it and without requiring a model of the environment.

Central to Q-learning is the Q-function, which attempts to estimate the expected rewards for state-action pairs in an attempt to find the most optimal action for any given state (Watkins and Dayan, 1992). The goal is to ascertain the action that maximizes the expected reward. As the agent interacts with its environment and receives feedback in the form of rewards, it continually updates its Q-values using the Q-learning update rule.

By harnessing the power of Q-learning for the seismic sensor placement problem, the aim

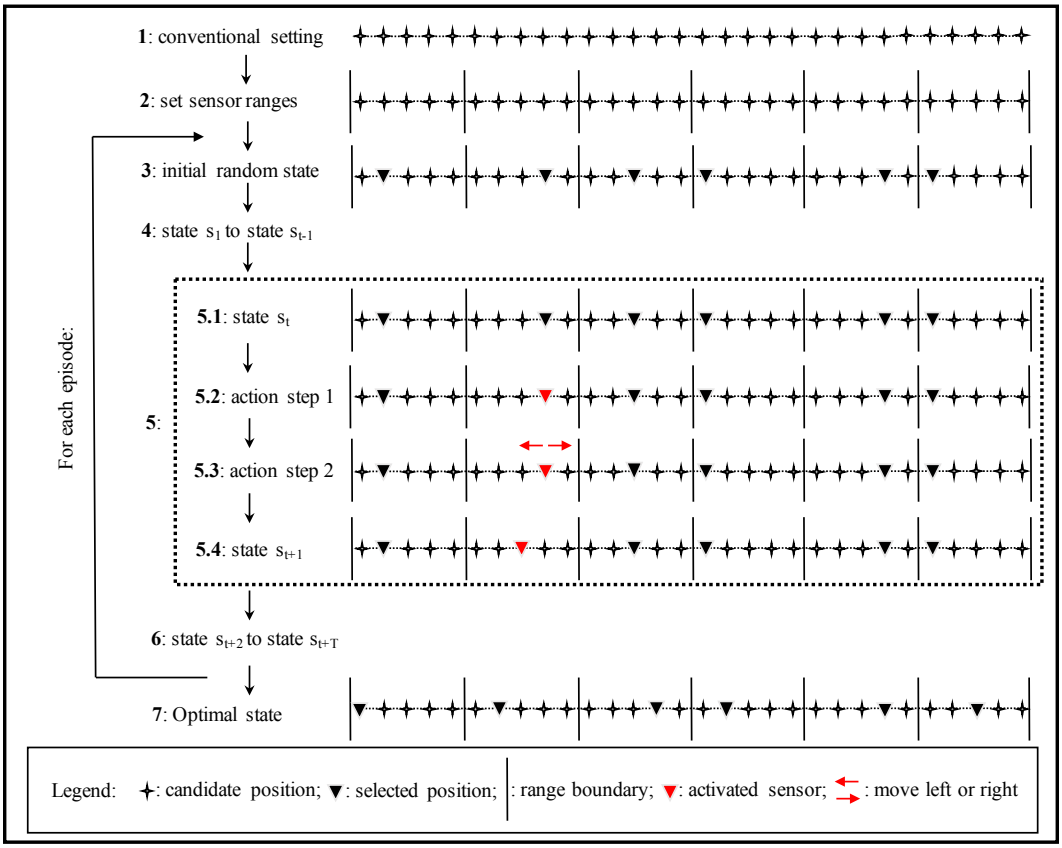

Figure 4.2: Schematic diagram to show selection procedure. The whole selection is summarized as seven stages. Black stars represent the candidate positions to put a sensor, and the black triangles represent the selected positions. The vertical line denotes the ranges defined by the user. Each initial state of an episode is set to be random to explore the whole state. For each action shown in stage five, the red triangle means activating one sensor, and the red arrows indicate moving left or right to a new candidate position. The sub-index t (indicates time-step) runs along the number of time-steps (T).

is to efficiently and effectively navigate the state space to determine optimal sensor positions. Given the characteristics and constraints of the sensor placement challenge, Q-learning emerges as an apt choice due to its ability to handle vast state-action spaces and deliver robust performance.

In general, Q-learning is a subset of temporal difference (TD) learning that can achieve optimal policies from delayed rewards when the agent has no prior knowledge of the unknown environment. The TD methods are grounded in the idea that one can learn by bootstrapping from current estimates to update subsequent estimates. In other words, it uses differences between estimates at different time-steps to update value functions.

In the typical Q-learning process, the agent observes the state s_t at a specific time-step t. Then, it chooses an action a_t according to the ϵ-greedy strategy, which executes an intensified search by exploitation and a diversified search by exploration. The ϵ-greedy strategy's dual nature of exploration and exploitation allows the agent to occasionally try out different strategies, helping ensure it doesn't get stuck in local optima and can uncover potentially better strategies. This is crucial for solving complex problems, especially NP-hard like the sensor placement problem, where the solution space is vast and computationally challenging to navigate exhaustively.

The Bellman equation is one of the foundational mathematical underpinnings of Q-learning and RL in general. The equation is pivotal as it relates the value of a state to the values of its successor states. In the context of Q-learning, the Bellman equation updates Q-values until they converge to the true Q-values iteratively, the expected cumulative rewards for each state-action pair. For those keen on delving deeper into the intricacies of the Bellman equation and its role in RL, the book by Sutton and Barto (2018) serves as an authoritative and comprehensive guide on the subject.

The core algorithm of Q-learning is a Bellman equation as a value iteration update, using the weighted average of the old and new information:

$$Q^*(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)), \tag{4.2}$$

where $Q^*(s_t, a_t)$ is the expected value (cumulative discounted reward) of doing action a in state s at time-step t and then following the optimal policy. $r_{t+1} + \gamma \max_a Q(s_{t+1}, a)$ is the TD target, and the TD is the subtraction between TD target and $Q(s_t, a_t)$. γ is the discount factor, and α is the learning rate.

In the framework of the defined objective function and the formulated MDP, I introduce an integrated RL algorithm for the optimal selection of seismic sensors, detailed in Algorithm 1. The learning procedure unfolds across E episodes, each comprising no more than T time-steps. Actions within this algorithm are chosen to employ the ϵ-greedy strategy, which

permits exploring possible actions and exploits experiences from the reinforcement returns. Notably, the state denoting sensor locations is randomized at the beginning of each episode to enable a thorough exploration of the entire state space. The exploitative aspect of the ϵ-greedy action selection strategy directs the sensors toward states indicative of superior data reconstruction, as delineated by the objective functions. Conversely, the explorative facet entails random selection, ensuring a wide-ranging search across the entirety of the space and mitigating entrapment in local optima.

This procedure operates autonomously, necessitating no user intervention post-engagement of the selection mode. Resultantly, the optimal sensor placement is obtained to provide the best reconstruction performance using the limited set of sensors. For the implementation of the proposed approach, I employed the Julia programming language, particularly leveraging the POMDPs.jl framework (Egorov et al., 2017).

Algorithm 1 Optimal seismic sensor selection using Q-learning for estimating policy π

Inputs: $\mathcal{S}$, $\mathcal{A}$, $\mathcal{R}$, γ, α, E and T
Build pre-learned basis library $\mathbf{U_p}$
Initialize $Q(s, a)$ for all $s \in \mathcal{S}$, $a \in \mathcal{A}$
Initialize the policy π arbitrarily
for *episode* $= 1$ to E **do**
  for $t = 1$ to T **do**
    Initialize random state s_1
    Select action a_t using the ϵ - greedy strategy
    Take action a_t
    Observe next state s_{t+1}
    Reconstruct data for state s_{t+1} (Equations 1.7 and 1.8)
    Calculate $SNR_{\boldsymbol{\nu}}$ value (Equation 1.9)
    Compare the $SNR_{\boldsymbol{\nu}}$ value of state s_t and s_{t+1}
    Give reward (+10) or penalty (-2) for r_{t+1}
    TD target $= r_t + \gamma \max_a Q(s_{t+1}, a)$
    TD = TD target - $Q(s_t, a_t)$
    $Q^*(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha$TD
    $s_t \leftarrow s_{t+1}$
  end for
end for
Obtain optimal policy π
Following the learned policy to obtain optimal seismic sensor locations ($\boldsymbol{\nu} = [\nu_1, \nu_2 \dots \nu_K]$)

4.3 Application in ocean bottom node acquisition

In seismic acquisition, the number of sources and receivers significantly impacts the overall cost. This cost can be particularly exorbitant in marine seismic explorations where special-

ized equipment like OBN is employed. Each node is costly, and hence, the optimization of their layout can lead to substantial cost savings. Several cost factors need to be taken into account in OBN acquisition, including the purchase cost of each node, deployment costs, which include labor and corresponding field operation time, and operational expenses, including data processing and quality control. Applying these methodologies aims to present a cost-effective OBN acquisition strategy that significantly reduces expenses and operational time. This optimized design could be a game-changer for marine seismic exploration, making it more economically feasible and efficient.

This section presents numerical examples to explore the proposed RL-integrated seismic sensor placement problem. Figure 4.3 establishes the whole workflow for reconstruction and sensor selection following the proposed method. I first build the pre-learned basis library as the preparation for reconstruction. Then, the sensor selection with RL is implemented after that.

I employ the acoustic finite-difference modeling from the SeismicJulia package (Stanton and Sacchi, 2016) to simulate a prestack marine OBN dataset using the SAIG velocity model. The preference for the OBN setup arises from the significant cost associated with each node; thus, finding the optimal node locations can yield notable economic advantages. Spatial sampling is set at five meters for vertical coordinates and ten meters for horizontal ones. The velocity model encompasses a range from 1.5 km/s to 3.0 km/s. The velocity model, complemented by the source-receiver geometry, can be observed in Figure 4.4.

The sources (depicted as red stars) and receivers (illustrated as black triangles) are distributed every 20 and 100 meters, respectively. A total number of 350 sources are simulated. Each source fires into a fixed array of 70 receivers. The receivers are situated at 500-meter depth to simulate ocean bottom nodes. A Ricker wavelet with a central frequency of 10 Hz has been utilized for generating the data. 24 out of the original 70 nodes are selected for the optimal setting to test the proposed method.

Based on the pre-learned basis library, the optimal placement for 24 available sensors using the proposed RL-based method and the corresponding learning performance are presented in Figure 4.5. The parameter settings for the iteration are as follows: learning rate $\alpha = 0.1$ and discount factor $\gamma = 0.7$. The ϵ-greedy exploration strategy is applied to investigate the performance of the proposed algorithm. The exploration rate ϵ is set as 1.0 at the beginning and decreases gradually to 0.1. I run 1000 episodes, with each episode iterating over 100 time-steps. The reward trend plotted every ten episodes reveals the converging trend, and it can be observed that the received rewards rapidly grow in the first learning episodes, indicating an intensified and efficient learning process of the RL-based sensor placement method. Besides, I also point out the possible reasons why small-value accumulated rewards around episode 540 are ascribed to the randomness of the initial state or the random selection

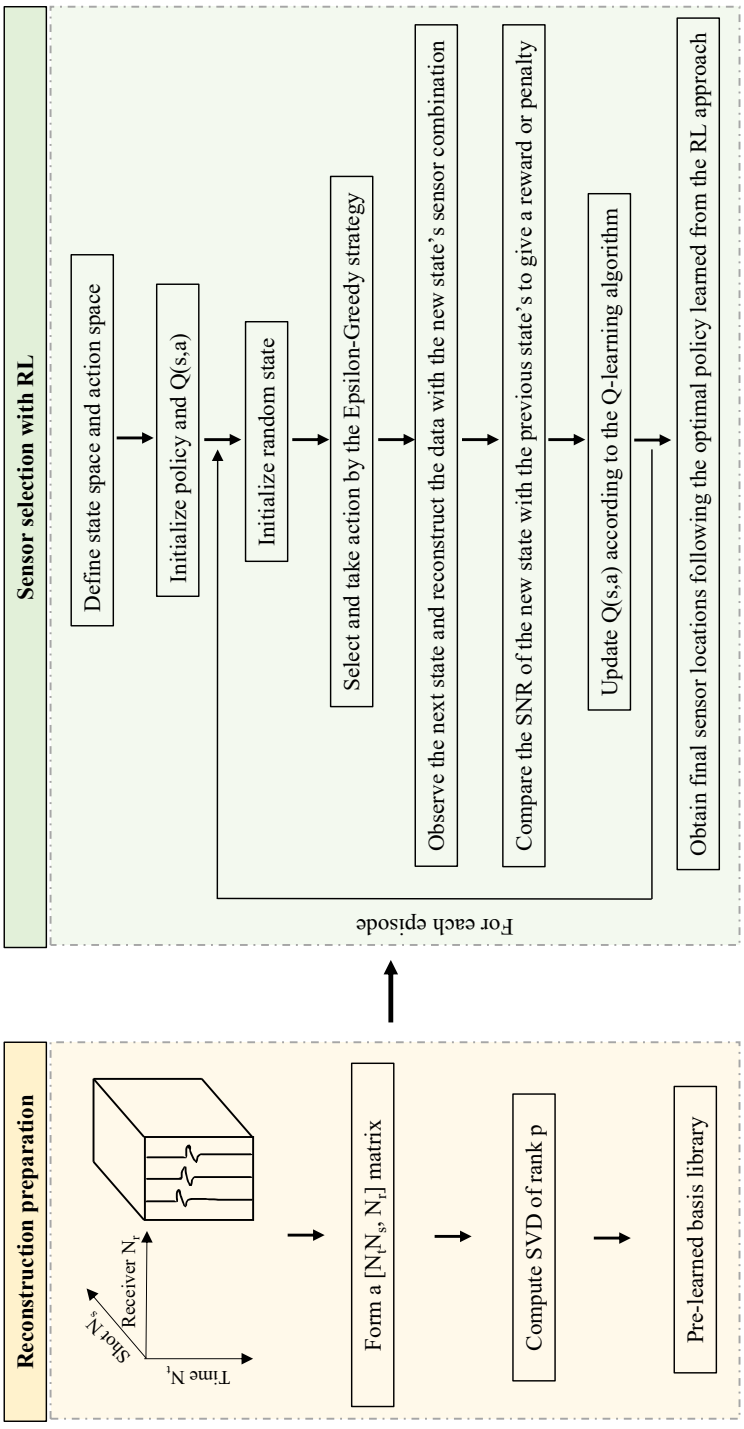

Figure 4.3: The comprehensive workflow comprises two folds: reconstruction preparation and sensor selection. The former delineates the construction of a pre-learned basis library, facilitating fast and efficient reconstruction. Meanwhile, the latter elucidates the Q-learning procedure for choosing optimal sensor locations.

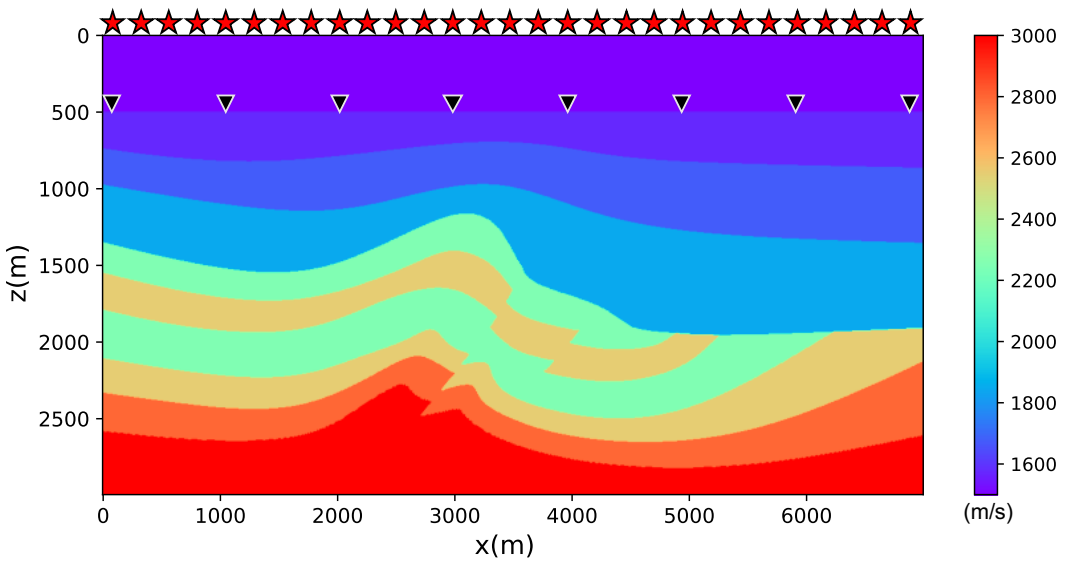

Figure 4.4: The velocity model utilized to simulate ocean bottom node data via finite difference modeling. Red stars represent the sources, and black triangles represent the receivers.

property of the ϵ-greedy strategy. The algorithm converges, guiding us to find the optimal locations of the sensors following the learned policy.

Next, I display the original nine CSGs and CRGs (Figure 4.6(a) and Figure 4.6(b)), respectively, with traditional equal-spaced acquisition using 70 receivers and 350 sources as a reference. Figure 4.7(a) and (b) show the decimated CSGs and CRGs with 24 receivers located at the optimal positions selected by the RL algorithm. Notably, some of the CRGs in Figure 4.7(b) are empty, signifying no receiver was selected by the RL algorithm for those specific indices. Lastly, the reconstructed CSGs and CRGs are depicted in Figure 4.8(a) and (b). It is noticeable that the events, both in CSGs and CRGs, are all fully reconstructed, and the $SNR_{\boldsymbol{\nu}}$ value of the reconstructed data calculated by equation 1.9 is 20.72 dB.

Then I compare the performance of the RL-based optimization method for sensor placement with the state-of-the-art jittered sampling protocol as a benchmark. To clarify, the jittered sampling I mention here only concerns the sampling strategy. Noted, I am not reaching the proposed method with the traditional jittered-CS acquisition scheme but are only comparing the spatial technique sampling regarding where to put the sensors. Both sampling paradigms use the same number of sensors and are all reconstructed based on the pre-learned basis library.

Table 4.1 lists the number of receivers and sources employed in the OBN survey for the

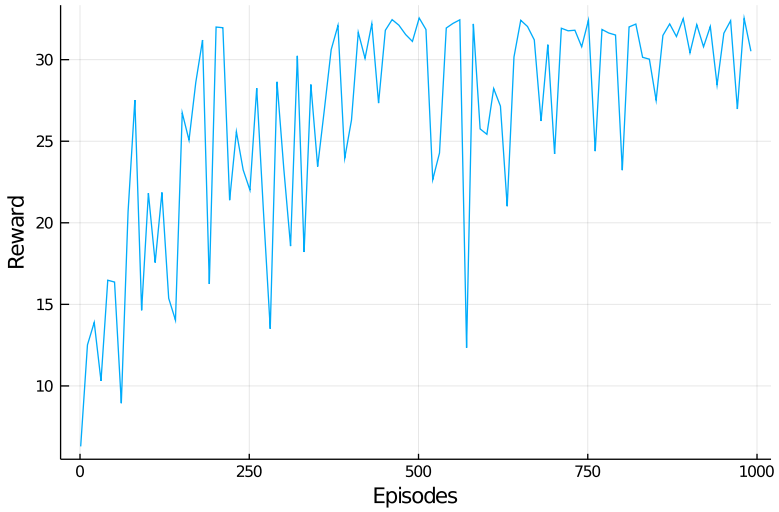

Figure 4.5: Rewards earned during the 1000 episodes. Rewards are plotted every ten episodes to show the general trend.

Table 4.1: Receiver sampling in conventional, jittered, and proposed optimal scenarios in OBN acquisition.

	Standard	**Jittered**	**Proposed**
Receivers numbers	70	24	24
Receiver interval	100 m	irregular	irregular
Sources numbers	350	350	350
Source interval	20 m	20 m	20 m

conventional, jittered, and proposed methods. Given the focus on receiver selection, the number of sources remains consistent across methods. However, the count of receivers in the proposed and jittered methods is reduced compared to the conventional approach.

I randomly initialize ten different jittered sampling acquisition schemes, and Figure 4.9(a) and (b) present the best-reconstructed CSGs and CRGs among the ten jittered sampling trials with the highest $SNR_{\boldsymbol{\nu}}$ value. A closer inspection reveals that slightly more artifacts exist compared with the optimal one (Figure 4.8) and that the optimal sensor locations obtained by the RL-based method provide a closer approximation to the original CRG. It is important to remember that the difference in reconstruction quality is solely due to the spatial sampling difference. Further, the $SNR_{\boldsymbol{\nu}}$ value of the jittered sampling is 17.64 dB, averaged with ten different jittered acquisition schemes.

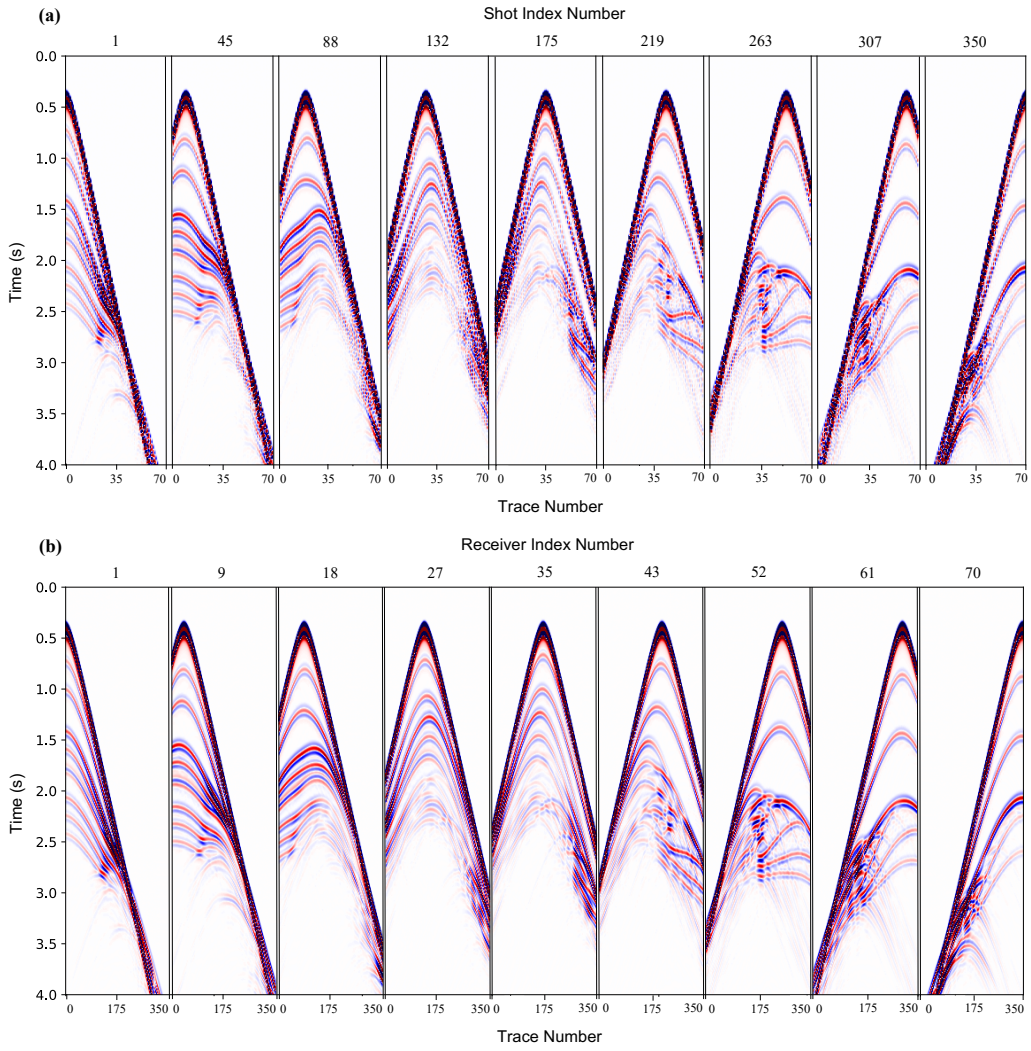

Figure 4.6: Conventional OBN data acquisition using 350 sources and 70 receivers. (a) Nine original CSGs. (b) Nine original CRGs.

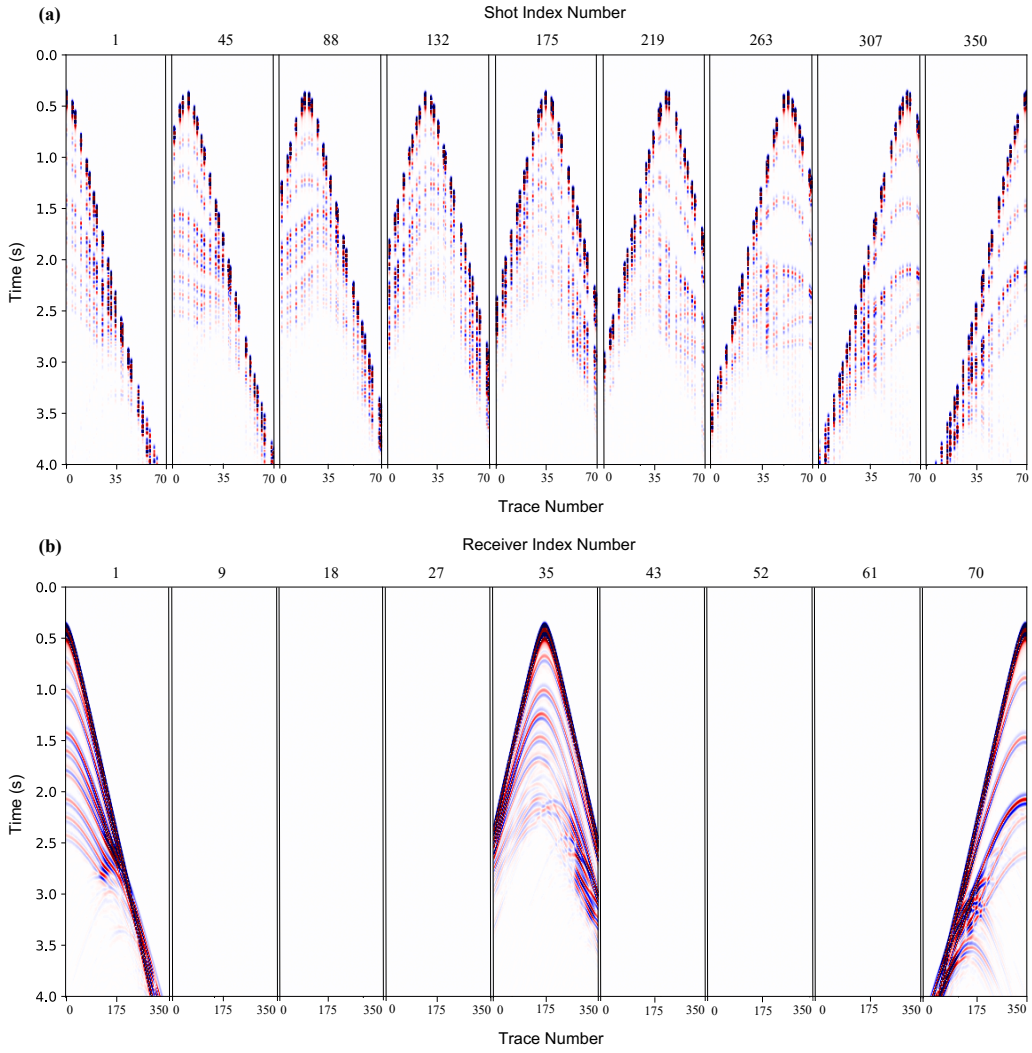

Figure 4.7: Optimized OBN data acquisition using the RL method. (a) CSGs with 24 strategically placed receivers. (b) CRGs with 24 selected receivers.

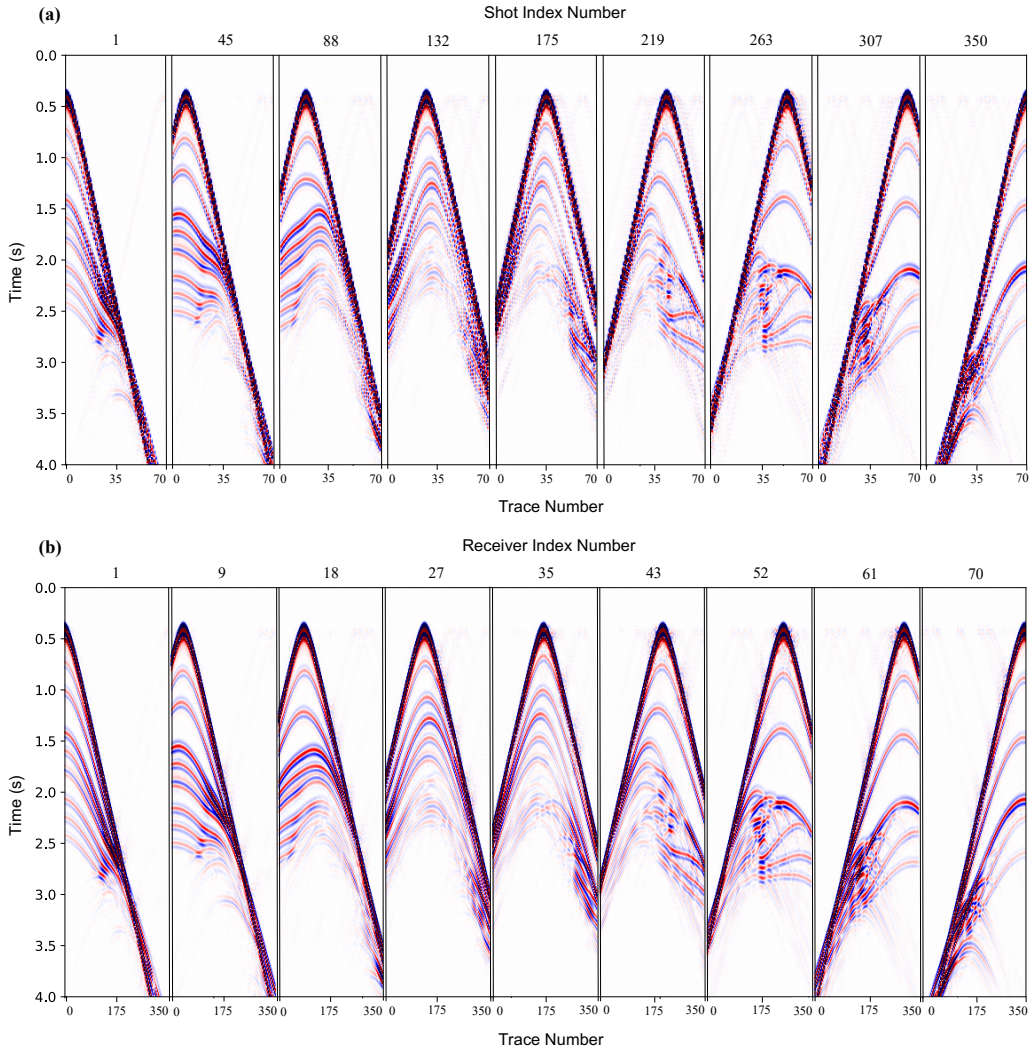

Figure 4.8: Data reconstructed using the optimized OBN acquisition via the proposed RL method. (a) Recovered CSGs. (b) Recovered CRGs.

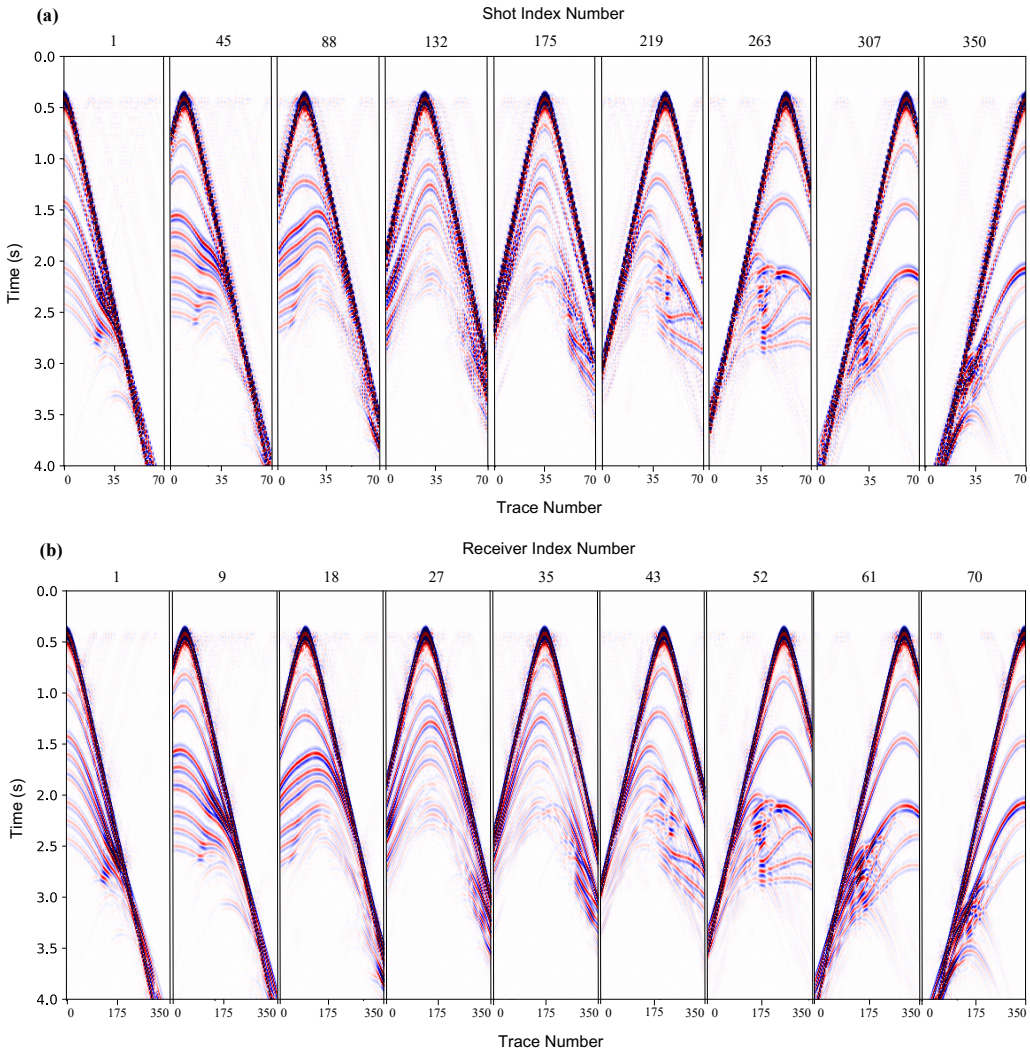

Figure 4.9: Reconstructed data via jittered sampling OBN acquisition. (a) Recovered CSGs. (b) Recovered CRGs.

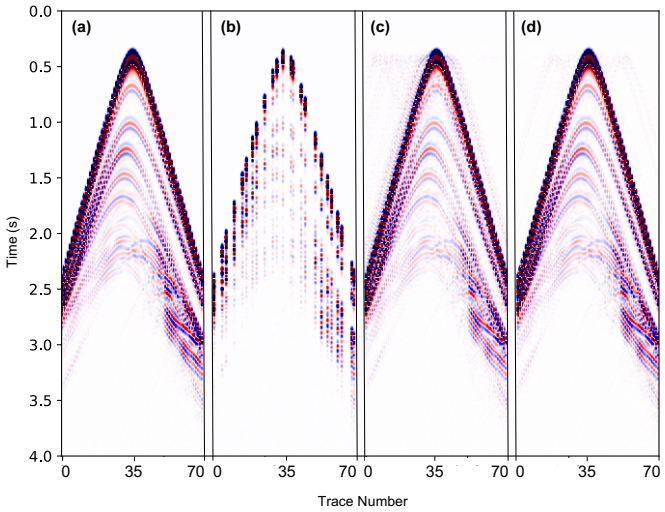

Figure 4.10: Comparison of the original CSG with reconstructions via jittered sampling and the proposed RL-based method at source index 175. (a) Original CSG. (b) Optimal decimated CSG using the RL-based approach. (c) Reconstructed CSG through jittered sampling. (d) Reconstructed CSG via the RL-based method.

Figure 4.10 provides a detailed comparison of one CSG at position x = 3.5 km (corresponds to the 175^{th} shot index in the conventional acquisition setting). The displayed CSGs in Figure 4.10 include (a) the original, (b) optimally decimated, (c) reconstructed using 24 receivers via the best jittered sampling scheme from ten random initializations, and (d) reconstructed with 24 receivers positioned optimally through the RL algorithm. Note that the jittered sampling result shown in Figure 4.10(c) is the best-resulting protocol chosen from ten randomly initialized settings following the jittered sampling scheme. The RL-based method provides a better sensor placement scheme for reconstruction than jittered sampling because RL-based sampling always reaches the same optimal solution. In contrast, jittered sampling is a form of random sampling; hence, one cannot confirm when a jittered sampling realization is the best-stable option.

Similarly, Figure 4.11 highlights a detailed comparison of one CRG at position x = 2.7 km, corresponding to the receiver index 27 in the conventional acquisition setting. Notably, this position was not selected as optimal by the RL method, which means no data was recorded at this particular receiver's location. Nevertheless, the four CRGs displayed in Figure 4.11 are the original CRG, the optimal decimated CRG, the reconstructed CRG with jittered sampling, and the reconstructed CRG with the proposed sampling, respectively. Significantly, the data can still be recovered even without data recorded at position x = 2.7 km in the optimal setup. What's more, the CRG recovered with the proposed sampling

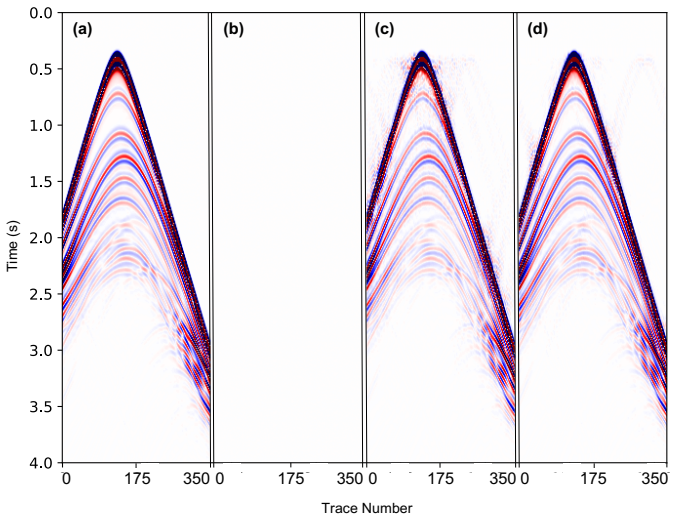

Figure 4.11: Comparison of original and reconstructed CRGs with receiver index 27. (a) Original CRG. (b) RL-optimized decimated CRG. (c) Jittered sampling reconstructed CRG. (d) RL-based method reconstructed CRG.

in Figure 4.11(d) delivers better reconstruction quality compared to the jittered sampling scheme shown in Figure 4.11(c).

Furthermore, I also compare the optimal policy learned by the RL algorithm with the random policy. To clarify, a random policy implies that, irrespective of the state, an agent's action selection is entirely arbitrary and does not stem from a pre-defined or learned policy. Figure 4.12 depicts the total reward gained per episode, and the blue and red dots represent the reward obtained by the Q-learning policy and random policy, respectively. Across a span of 100 episodes, a significant distinction in their performances is discerned, and the mean reward for the Q-learning policy stands at 29.62, while the one for the random policy is 6.23. As anticipated, the Q-learning policy consistently outperforms its random counterpart, emphasizing the superiority and effectiveness of the proposed method. Conversely, given its arbitrary nature, the random policy would necessitate potentially millions of iterations ever to hope to approach the optimal solution achieved by Q-learning.

In the conclusive analysis, Figure 4.13 portrays the images computed by the reverse time migration (RTM) method with the conventional acquired equal-spaced data, shown in Figure 4.13(b), and the reconstructed data using the optimal sensor setting obtained by the proposed method, shown in Figure 4.13(c), in comparison with the original velocity model, which is depicted in Figure 4.13(a). I am inverting for a single stacked image for all the shots for simplicity and memory-saving purposes. Laplacian filtering and gaining with the same parameters are applied to the RTM images. On juxtaposing the original velocity model

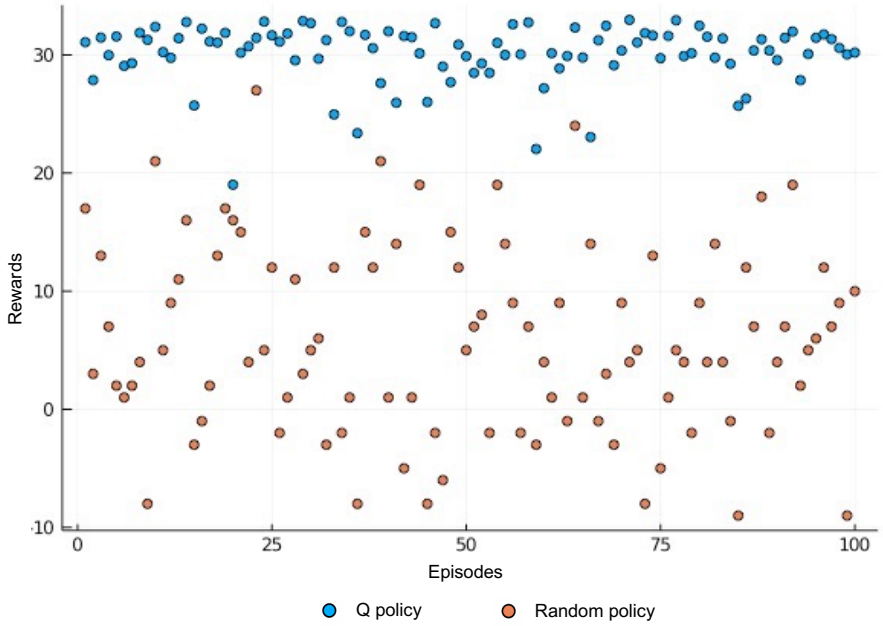

Figure 4.12: Comparison of total reward gained per episode between Q-learning policy vs. random policy.

(Figure 4.13(a)) with the RTM images, prominent geological features become evident — from the seabed located 500 meters below the sea surface to the evident anticlines and fault lines. A striking observation is the remarkable similarity between the RTM images derived from conventional data (Figure 4.13(b)) and the reconstructed data from the RL-based survey (Figure 4.13(c)). This congruence underscores the efficacy and validity of the proposed method, presenting it as a potent alternative to traditional methods.

4.4 Discussions

This section discusses research challenges and issues encountered, casting light on outlooks for future recommendations. First, this chapter introduces an RL algorithm, addressing the seismic sensor placement problem with a fast and computationally efficient reconstruction method. Considering the strategy employed in this chapter is classical RL, the size of the problem is limited compared with deep reinforcement learning (DRL). The next chapter entails adopting more powerful DRL algorithm to tackle more extensive and complex scenarios to remove the size bottleneck. For example, the deep Q-network (DQN) (Mnih et al., 2015; Sutton and Barto, 2018) affords the capacity to manage more convoluted situ-

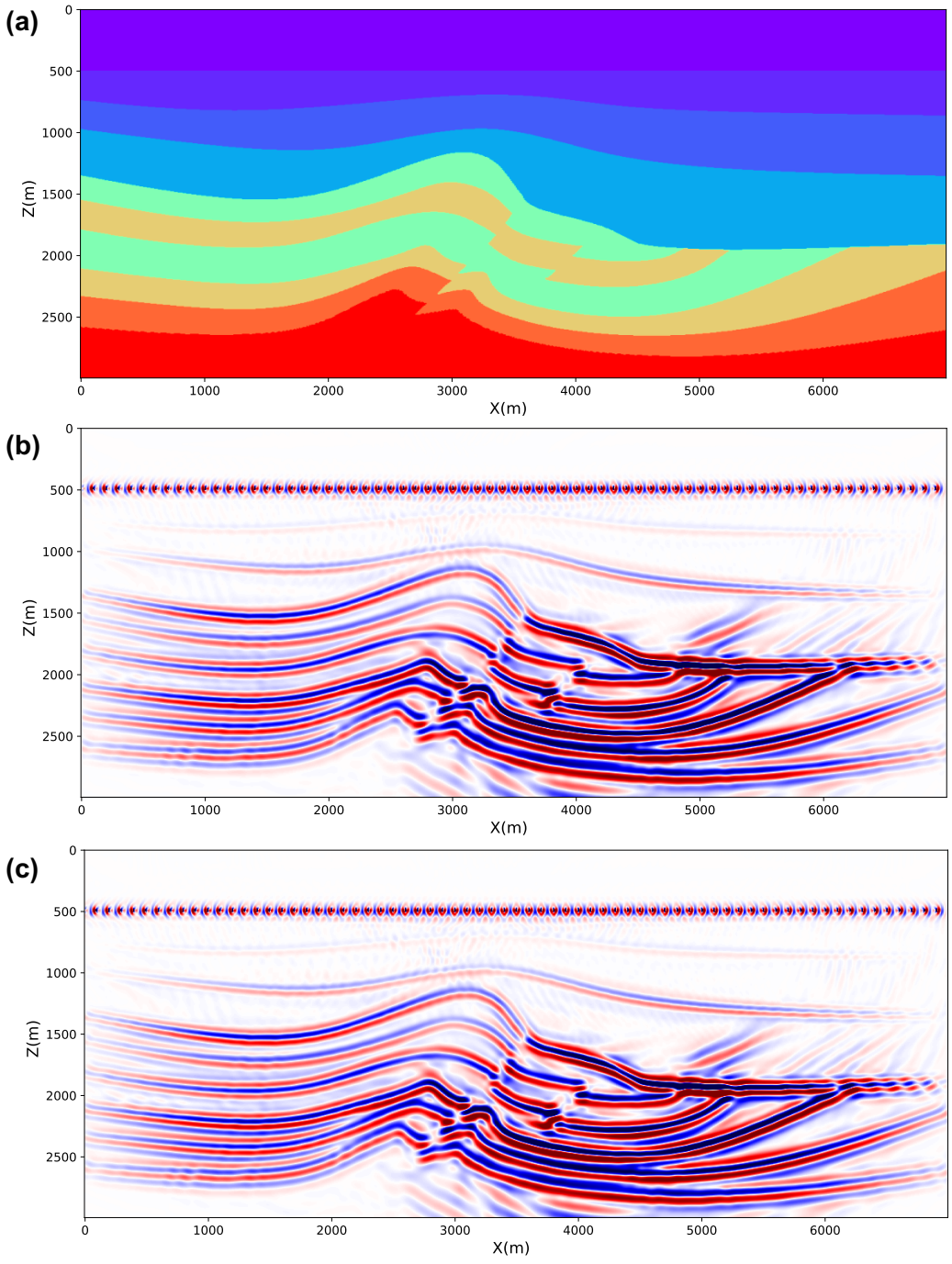

Figure 4.13: Comparison of the velocity model (a) with the RTM images of the conventional acquired data (b) and the reconstructed data via the proposed method (c).

ations involving an increased number of sensors, which forms a pivotal aspect of the future exploration.

In addition, even though the reconstruction method is suitable for the methodology used in this study, this data-driven method also needs training datasets to obtain the basis library. Nonetheless, in real-world applications, regions with prior survey data can serve as training datasets, i.e., time-lapse scenarios. On top of that, despite the objective function I adopted here being explicit and easy to implement, alternative reconstruction methods and objective functions are open for exploration. For example, Zhang et al. (2022) utilized SA to find optimal sensor positions using the properties of the sampling operator and matrix completion for reconstruction. The importance of the theory lies more in introducing RL for optimal seismic sensor location design. Furthermore, the objective function concerns how the agent learns from experience and needs to give a fair amount of thought because defining the rewards to reflect the behaviour the agent wants to know is critical.

Additionally, even though I attempt to simulate the OBN setting with optimal receiver locations, the tentative idea includes but is not limited to, optimizing the nodes within the OBN setting or designing the source firing positions in the land or marine survey scenarios with an aim to reduce acquisition costs. Essentially, the broader potential of the approach can be realized in diverse applications, where either the source or receiver location optimization plays a pivotal role in both land and marine acquisition settings.

After designing and testing the integration of RL with optimal receiver placement, it became evident that with proper reconstruction, fewer sensors could achieve comparable data quality. Alternative designs of the problem could be admitted regarding the different requirements of various issues. All in all, this brand-new approach is advisable and deserves a trial for optimal sensor placement application.

4.5 Conclusions

This chapter introduces a new approach to seismic acquisition design by integrating RL-based optimal sensor placement techniques. The sensor placement problem with the proposed objective function falls under the MDP formulation mathematically, where state space with a practical boundary restriction, action space, and reward is mainly defined for the seismic sensor selection problem. The promising properties of RL enable the sensor selection algorithm to be implemented online, model-free, and fully incremental. The Q-learning result is converged, and optimal sensors can be selected following the learned policy. The resultant synthetic data scenarios are shown in CSGs, CRGs, and RTM images, demonstrating that the proposed method provides better sensor locations for reconstructing data

than the jittered sampling scheme under the same circumstances. Comparison with random policy also verifies the effectiveness of the proposed method. Further, albeit only exemplified with seismic sensor placement application, this general idea of applying RL to solve the NP-hard combinatorial problem could also be used in other fields of science.

CHAPTER 5

Optimal seismic sensor placement based on deep reinforcement learning approach

5.1 Introduction

The seismic acquisition design process is a complex decision-making problem involving a large number of variables and constraints. Traditionally, the acquisition design process is done manually, relying on the expertise of geophysicists and trial-and-error methods. Nevertheless, this approach is time-consuming, expensive, and may not guarantee optimal results.

Artificial intelligence (AI) has made a series of transformative breakthroughs in recent years, consistently reshaping the world and challenging our preconceived notions. Machine learning (ML) techniques, particularly deep reinforcement learning (DRL), have emerged as promising tools for optimizing various complex decision-making problems (Mnih et al., 2013, 2015; Silver et al., 2016; Sutton and Barto, 2018). Notable examples include AlphaGo's victory over the world's top Go champion (Silver et al., 2016, 2018) and the immensely popular ChatGPT (Wu et al., 2023). Both of these milestones owe their success to the advancements in DRL. DRL algorithms can learn optimal policies through trial-and-error interactions with an environment, a simulator, or a real-world system.

Seismic acquisition design has been studied in the geophysical exploration community, and various optimization methods have been proposed. These methods include genetic algorithms (GA) (Nakayama et al., 2019), simulated annealing (SA) (Zhang et al., 2022), and etc. However, these methods have limitations, such as being computationally expensive, requiring prior knowledge of the underlying geology, and not guaranteeing global optimality. DRL has

been applied to various geophysical exploration problems, such as first arrival picking (Ma et al., 2019), seismic source localization (Wu et al., 2019), and inversion (Dell'Aversana, 2022). However, to the best of my knowledge, no study has investigated the use of the DRL algorithm for seismic acquisition design. As a well-established algorithm, the deep Q-network (DQN) algorithm has several advantages, such as handling high-dimensional state and action spaces and being computationally efficient. Therefore, I propose a DQN-based approach for seismic acquisition design in this chapter and demonstrate its effectiveness on a synthetic dataset.

The remainder of this chapter is organized as follows. I first introduce data-free optimal sparse seismic acquisition. Regarding this, the second maxima of the spectrum of the grid are illustrated for optimal design. Then, I describe the proposed DRL-based approach for seismic acquisition design, particularly for the vibroseis optimal route problem. After that is the introduction of DRL, the formulation of the problem as a Markov decision process (MDP), and the DQN algorithm. Then, I present the application in vibroseis optimal route design problem with some experimental results. Finally, the outlines of future research directions and the conclusion of the method are shown.

5.2 Data-free optimal sparse seismic acquisition

Data-free acquisition means no prior data is needed to design optimal sensor placement. Considering this, looking for a proper objective function is essential. Naghizadeh and Sacchi (2010b) stated that 2D signals that are band-limited in one spatial dimension can be recovered by designing a regular acquisition grid that minimizes the mixing between the unknown spectrum of the well-sampled signal and aliasing artifacts, and this concept can be used to define potential strategies for acquisition guided Fourier reconstruction.

Figure 5.1 shows an example of randomly selected four sensor locations and their wavenumber spectra, demonstrating that the second largest maxima of the wavenumber amplitude varies with different sampling schemes. For comparison, Figure 5.2 shows another random sampling; instead of choosing four sensors, I select 24 out of the 49 candidate positions. It is obvious that the wavenumber response is an impulse in the middle of the 2D spectrum for the multidimensional sampling function. Interestingly, with more selected sensors, the second maxima value also decreases.

For those keen on delving deeper, the paper by Naghizadeh and Sacchi (2010b) provides a comprehensive illustration of the subject. Inspired by this interesting paper, this chapter proposes to use the minimization of the second maxima of the grid as the judging criteria, which is data-free. Other than this criteria, readers are free to use others for data-free optimal sparse seismic acquisition design.

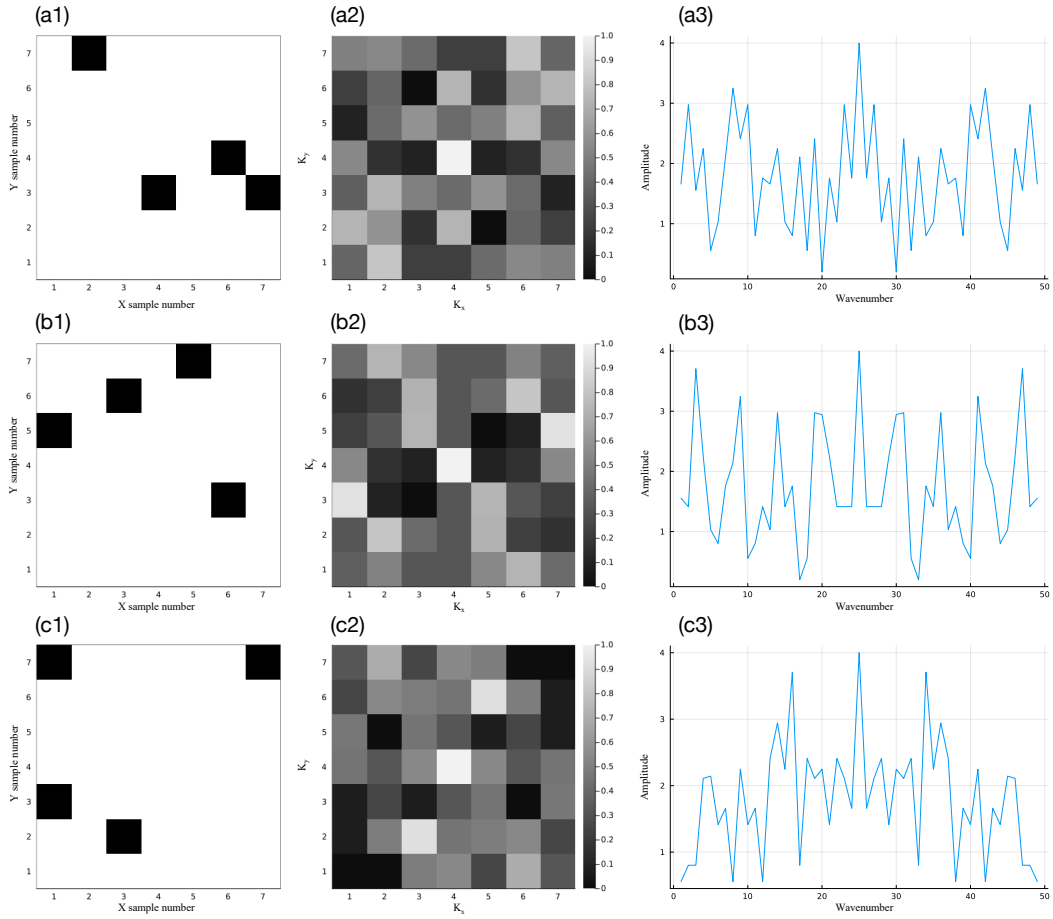

Figure 5.1: Different random sampling schemes and their corresponding wavenumber spectra. Four out of the 7×7 grids are randomly selected. Plots a1, b1, and c1 are three randomly generated samplings. Plots a2, b2, and c2 are the normalized wavenumber spectra. Plots a3, b3, and c3 are wavenumber vs amplitude, showing the second maxima varies with different sampling schemes.

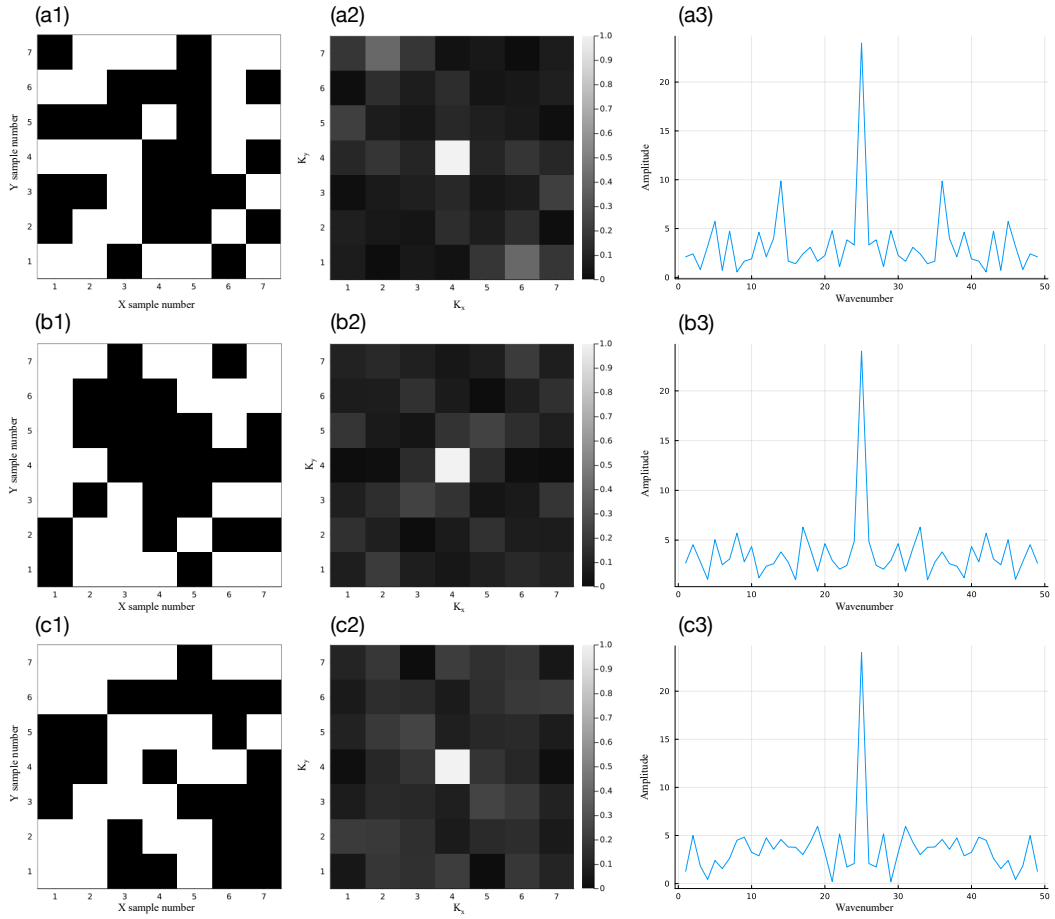

Figure 5.2: Different random sampling schemes and their corresponding wavenumber spectra. 24 out of the 7×7 grids are randomly selected. Plots a1, b1, and c1 are three randomly generated samplings. Plots a2, b2, and c2 are the normalized wavenumber spectra. Plots a3, b3, and c3 are wavenumber vs amplitude, showing the second maxima varies with different sampling schemes.

5.3 Seismic sensor placement based on deep reinforcement learning

The idea of RL can be combined with the concept of optimal seismic acquisition design has been proved and explained in the proposed second route in Chapter 4. The DRL proposed in the third route can be applied to the location of large-scale sensor scenarios to serve more complicated real acquisition. DRL is a subfield of ML that uses RL techniques to train artificial agents to make decisions in dynamic environments. The goal of DRL is to optimize the behavior of the agent to maximize a long-term reward signal, which can be defined in various ways depending on the application. This chapter proposes an optimal seismic survey design method using DRL with the DQN algorithm. In the context of optimal seismic acquisition design, the DQN can process vast amounts of information to determine the best sensor placements for a given scenario. Though RL agents have marked accomplishments in several fields (Tesauro et al., 1995; Diuk et al., 2008; Riedmiller et al., 2009), their scope has predominantly been restricted to areas where valuable features can be manually designed or to situations with transparent, uncomplicated state spaces (Mnih et al., 2015). However, the DQN can make informed decisions to optimize the acquisition process by considering various states. This combination of DRL with seismic design opens many possibilities for more efficient and accurate data collection, ultimately leading to better decision-making in exploration and other related fields.

5.3.1 Formulation of vibroseis optimal route design as Markov decision process

The problem of vibroseis optimal route design is defined as three parts: using the least number of sources, activating them at the optimal positions, and finding the optimal path moving through all the chosen locations. The main idea of designing the optimal route for the vibroseis is threefold: first, when obstacles exist, random sampling is not workable, and designing the path of the vibroseis is required; second, optimal design utilizes minimum sampling point so that the corresponding acquisition cost is lower; last but not least, the field working time is reduced, which is important for HSE concern.

Figure 5.3 depicts the idea of optimal route design. Figure 5.3(a) represents the decision made in the working desert environment. Figure 5.3(b) indicates the vibroseis needs to move from the start point to the end position while placing a source point at every regular dense grid. Figure 5.3(c) indicates while there is an obstacle in the working area, like the mountain in the upper left, how to design the shortest route with the minimum number of shots needs to be studied.

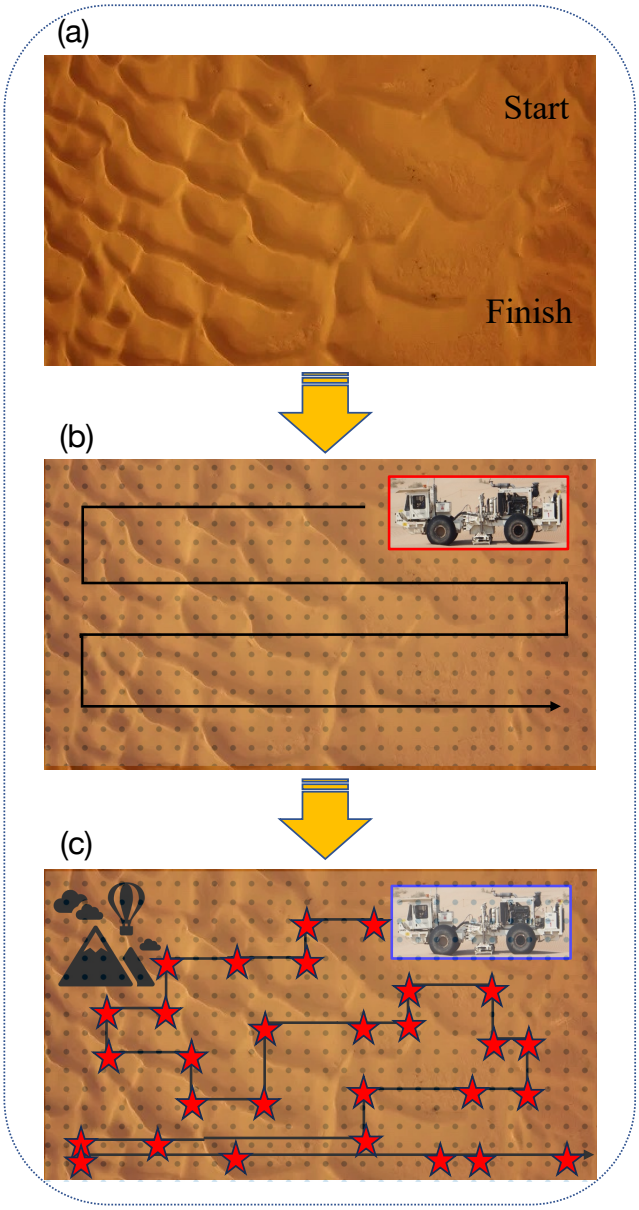

Figure 5.3: Contrasting routes of vibroseis. (a) Define the start and finish position of the vibroseis. (b) Conventional survey with dense and regular grid (black dots), where the vibroseis activates at every sampling point. (c) Optimal route of the vibroseis with the least number of sampling points (red stars). Black lines indicate the vibroseis moving path, and the mountain icon represents an obstacle where data cannot be obtained.

A framework for the MDP consists of an environment and an agent which acts in this environment. In this case, the environment is a regular grid consisting of all the possible candidate positions. The vibroseis is trying to find the shortest route from the start point to the target point and is exploring and exploiting past experiences (episodes) in order to achieve its goal. It may fail repeatedly, but hopefully, after lots of trial and error (rewards and penalties), it will solve the problem. The solution will be reached if the agent finds the optimal sequence of states in which the accumulated sum of rewards is maximal. The following describes the state space, action space, and reward function I defined for the vibroseis optimal route design problem.

State space

Considering the problem is a 2D problem with each grid point possessing an x and y direction coordinates. There are four types of grid points: blocked points represent obstacles like mountains, rivers, or others that prohibit placing a source point; free points indicate the candidate positions that the vibroseis can move and place a source point; labeled points mean the vibroseis placed a source point at this location; and target points denote the start and finish locations. Note that the labeled points previously belonged to the free points, but once the agent moves there and marks it as an optimal position, the free point is marked as the labeled point.

I first define the particular grid status at a specific time-step t, symbolized by x_t, with the process spanning T time-steps for every episode. Note that each x_t is composed of the x and y coordinates of the grid point. Because the agent only observes the current state, it is impossible to fully understand the current situation from the existing grid's status. In other words, the task is partially observed (Kaelbling et al., 1998). Therefore, the state space is defined as the sequences of actions and observations, $s_t = x_1, a_1, x_2, \ldots, a_{t-1}, x_t$, which are the input to the algorithm, and then the strategies are learned depending upon these sequences. Remember, the state definition is the sequences of actions and observations, and these histories of arbitrary length as inputs to a neural network are difficult. So, a function ϕ is introduced as a preprocessing to the m most recent grid status and stacks them to produce the input to the Q-function works on a fixed length representation of histories. All sequences are assumed to terminate in a finite number of time-steps, and this formalism gives rise to a large but finite MDP in which each sequence is a distinct state. As a result, one can apply RL methods for MDPs by using the complete sequence s_t as the state representation at time-step t.

Action space

The action space for the optimal route problem is structured around choosing the location for a particular sensor from the pool of candidate positions. I define eight actions for each position, except the boundaries and the place that are blocked by obstacles: move forward, move backward, turn left, turn right, plus each movement direction being activated, which are eight different choices in total. To elucidate, taking an action, denoted as a, entails a two-fold process:

- Directional movement: This step involves deciding the direction of its movement, and the choices are four directions: move forward, backward, left, or right.
- Activation: This step decides whether the vibroseis activates at this position or not.

When an action a_t is undertaken while in state s_t, the system navigates to a subsequent state, denoted s_{t+1}, that one more position is added to the previous state to form the new state. Further, rather than assess the action at each time-step, I assess the actions at the end of each episode, corresponding to the vibroseis reaching the finish position. The efficacy of the chosen action can then be assessed by the objective function defined, and then the algorithm determines the appropriateness of the action.

Reward function

The task is an agent interacting with the environment in a sequence of actions, observations, and rewards. At each time-step, the agent selects an action a_t; then the action modifies the internal state. The agent observes a vector of values representing the current state and receives a reward r_t representing the change of the whole design. Noted that, in general, the reward may depend on the entire previous sequence of actions, and feedback about the action may only be received after many thousands of time-steps have elapsed (Mnih et al., 2015).

The reward function is designed particularly for vibroseis optimal route problem that should ensure the use of the least number of sensors and have the shortest time spent on the route. Considering these requirements, I define a reward r^- as:

$$r^- = -\beta^- \frac{t_{vibro}}{T_{total} + t_{vibro}}, \tag{5.1}$$

where t_{vibro} denotes the vibration time of the vibroseis and T_{total} means the total time spent from the start point to the finish point without adding a new sensor point, that is, the total time spent for the last episode. If this is the first episode, then T_{total} is set to equal to the

moving time from start to end point without vibration time, which equals to $\frac{L}{V}$, where L is the total distance from start to finish point and V is the velocity of the vibroseis. β^- is a hyperparameter that needs to be tuned.

However, if adding one more source point with valuable information improves the value related to the objective function, then another reward r^+ is evaluated by the end of each episode:

$$r^+ = \beta^+ \frac{SMS_{new}}{SMS_{old}}, \tag{5.2}$$

where SMS_{new} and SMS_{old} are values of the second maxima of the spectrum of the grid computed from the current and previous episodes, respectively. β^+ is the corresponding hyperparameter. Then, with a proper balance between r^- and r^+, the trade-off between the least number of sensors being used as well as the best data quality collected can be achieved.

Besides the r^- and r^+ described, other details about the reward function are as follows: the rewards range from -1.0 to 1.0; each move from one state to the next state is rewarded by a small negative value, to ensure the agent is encouraged to minimize the number of moves, which means the number of optimal sensors. In addition, considering given the same optimal sensor locations, different routes may use the same time from driving from the start to the finish point, so I also give a small negative reward for changing movement direction between consecutive actions to avoid frequent changing directions for the vibroseis in the real world scenario. A reward of 1.0 is given when the agent reaches the finish point, and an attempt to enter the blocked grid point will cost a significant negative reward. Further, the same rule holds for an attempt to move outside the pre-defined grid boundaries. The algorithm ends once the total reward is below the pre-set negative threshold to avoid infinite loops and senseless wandering.

Next, I provide a detailed description of the DQN algorithm and its application to seismic survey design.

5.3.2 Deep Q-network

Mnih et al. (2015) introduced an innovative artificial agent called the deep Q-network (DQN) that connects complex sensory inputs and actions, resulting in the first artificial agent adept at mastering a wide variety of demanding tasks. The novel agent, DQN, is able to combine RL with a class of artificial neural networks (ANN) known as deep neural networks (DNN). Significantly, modern breakthroughs in DNN (Hinton and Salakhutdinov, 2006; Bengio et al., 2009; Krizhevsky et al., 2012), where multiple layers of nodes are used to build up increasingly abstract representations of data, have enabled ANN to grasp concepts

like object categories directly from raw sensory inputs. DQN employs a highly effective design, the deep convolutional network (LeCun et al., 1998), which uses hierarchical layers of tiled convolutional filters to mimic the receptive fields' effect, drawing inspiration from Hubel and Wiesel's foundational research on feedforward processes in the early visual cortex (Hubel and Wiesel, 1963).

The agent aims to interact with the environment by selecting actions to maximize future rewards. In a more technical sense, a deep convolutional neural network is employed to estimate the optimal action-value function

$$Q^*(s,a) = \max_{\pi} \mathbb{E}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots \mid s_t = s, a_t = a, \pi\right]. \tag{5.3}$$

This function in equation 5.3 represents the maximum cumulative rewards r_t, discounted by γ at each time-step t, which can be achieved through a behavior policy π that maps sequences to actions.

If the future discounted return at time-step t is

$$R_t = \sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}, \tag{5.4}$$

where T is the time-step the vibroseis reaches the finish point. Then the optimal action-value function $Q^*(s,a)$ defined as the maximum expected return achievable by following any policy after seeing some sequence s and then taking some action a, same as equation 5.3 can be rewritten as

$$Q^*(s,a) = \max_{\pi} \mathbb{E}\left[R_t \mid s_t = s, a_t = a, \pi\right]. \tag{5.5}$$

The optimal action-value function obeys the Bellman equation, an important identity that if the optimal value of $Q^*(s^{'}, a^{'})$ of the sequence $s^{'}$ at the next time-step is known for all possible actions $a^{'}$, then the optimal strategy is to select the action $a^{'}$ maximizing the expected value of $r + \gamma Q^*(s^{'}, a^{'})$

$$Q^*(s,a) = \mathbb{E}_{s'}\left[r + \gamma \max_{a'} Q^*(s^{'}, a^{'}) \mid s, a\right]. \tag{5.6}$$

The basic idea behind many RL algorithms is to estimate the action-value function by using the Bellman equation as an iterative update, $Q_{i+1}(s,a) = \mathbb{E}_{s'}\left[r + \gamma \max_{a'} Q_i(s^{'}, a^{'}) \mid s, a\right]$. Such value iteration algorithms converge to the optimal action-value function, $Q_i \rightarrow Q^*$ as $i \rightarrow \infty$. However, this basic approach is practically impractical because the action-value function is estimated separately for each sequence (Mnih et al., 2015). Instead, it is expected to use a function approximation to estimate the action-value function, $Q(s,a;\theta) \approx$

$Q^*(s, a)$. In the RL community, this is typically a linear function approximator. Still, a nonlinear function approximator, such as a neural network, which is the deep Q-network, is sometimes used instead. The following describes the neural network function approximator with weights θ as a Q-network proposed by Mnih et al. (2015).

A Q-network can be trained by adjusting the parameters θ_i at iteration i to reduce the MSE (mean squared error) in the Bellman equation, where the optimal target values $r + \gamma \max_{a'} Q^*(s^{'}, a^{'})$ are substituted with approximate target values $y = r+\gamma \max_{a'} Q(s^{'}, a^{'}; \theta_i^-)$ with parameters θ_i^- from some previous iteration. This leads to a sequence of loss functions $L_i(\theta_i)$ that changes at each iteration i

$$L_i(\theta_i) = \mathbb{E}_{s,a,r}\left[(\mathbb{E}_{s'}\left[y \mid s, a\right] - Q(s, a; \theta_i))^2\right] = \mathbb{E}_{s,a,r,s'}\left[(y - Q(s, a; \theta_i))^2\right] + \mathbb{E}_{s,a,r}\left[\mathbb{V}_{s'}\left[y\right]\right]. \tag{5.7}$$

Noted, the targets depend on the network weights, which are different from those used for supervised learning that are fixed before learning begins. At each stage of optimization, the parameters from the previous iteration θ_i^- are fixed when optimizing the ith loss function $L_i(\theta_i)$, ended in a sequence of well-defined optimization problems (Mnih et al., 2015).

DQN represents an approximate value function, $Q(s, a; \theta_i)$, via a deep convolutional neural network as depicted in Figure 5.4. The model architecture has a separate output unit for each possible action, and only the state representation is an input to the neural network. The outputs represent the predicted Q-values for each action based on the provided input state. A primary benefit of this architecture is its capability to determine Q-values for every potential action in a specific state using just one forward pass in the network (Mnih et al., 2015).

When a nonlinear function approximator like the neural network represents the action-value (often termed Q) function, RL has historically been prone to instability or even divergence (Tsitsiklis and Van Roy, 1996). Several factors contribute to this instability: the inherent correlations within the observation sequence, potential significant policy changes resulting from small Q updates that subsequently alter the data distribution, and the interdependencies between Q and their corresponding target values $r+\gamma \max_{a'} Q(s', a')$. Mnih et al. (2015) tackled these instabilities with two primary concepts. Firstly, incorporate a biologically-driven mechanism known as experience replay (Lin, 1992; McClelland et al., 1995; O'Neill et al., 2010), which introduces randomness to the data, effectively eliminating observation sequence correlations and averaging out fluctuations in data distribution. Secondly, applying an iterative update adjustment to the Q aligns them with target values updated at periodic intervals, thus minimizing correlations with the set target.

To implement experience replay, the agent's experiences are recorded and stored at each time-step, $e_t = (s_t, a_t, r_t, s_{t+1})$, in a data set $D_t = \{e_1, \ldots, e_t\}$ after pooled over several

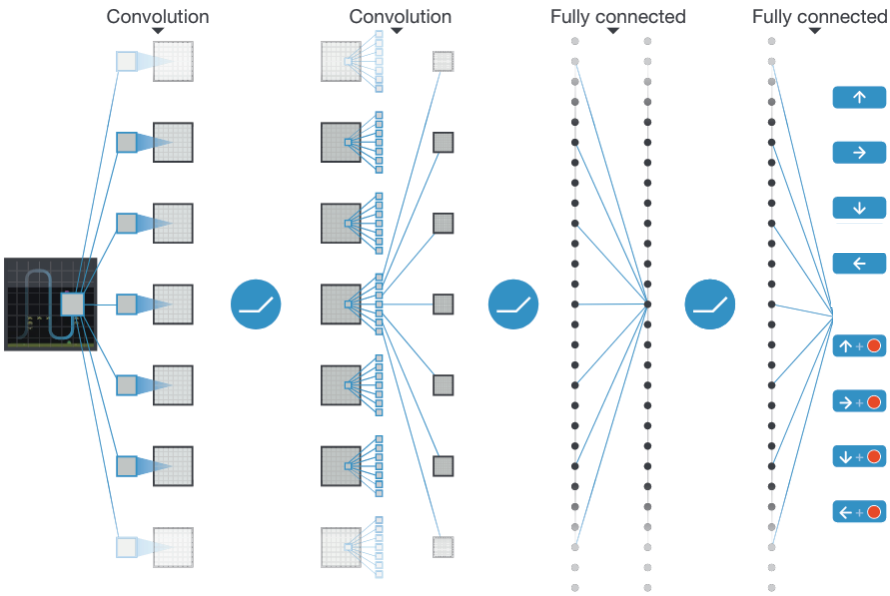

Figure 5.4: Schematic illustration of the convolutional neural network. The input to the neural network is produced by the preprocessing map θ, followed by three convolution layers. The blue line symbolizes the sliding of each filter across inputs and two fully connected layers with a single output for each valid action. A rectifier nonlinearity follows each hidden layer. The red dot represents to put a sensor. Modified from Mnih et al. (2015)

episodes into a replay memory. During the inner loop of the algorithm, Q-learning or minibatch updates are applied to samples of experience, $(s, a, r, s^{'}) \sim U(D)$, drawn at random from the pool of stored samples. This approach has several advantages over standard online Q-learning (Mnih et al., 2015). Using each step of experience multiple times for weight updates enhances data efficiency. Learning from sequential samples is less effective because of the significant correlations between them; randomizing the order of these samples disrupts these correlations, leading to more stable updates. On-policy learning can be problematic because the present parameters influence the upcoming data sample, leading to potential feedback loops where parameters might become trapped in an undesirable local minimum or diverge dramatically (Tsitsiklis and Van Roy, 1996). Through experience replay, the behavior distribution averages over numerous past states, ensuring more consistent learning and preventing parameter oscillations or divergence. It's worth noting that when utilizing experience replay, off-policy learning is essential since the current parameters differ from those that generated the sample, thus favoring the adoption of Q-learning. The Q-learning update for the i-th iteration relies on the subsequent loss function:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim U(D)} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i) \right)^2 \right], \tag{5.8}$$

where γ denotes the discount factor. θ_i denotes the parameters (the weights) of the Q-network, and θ_i^- are the network parameters used to compute the target at the i-th iteration, respectively.

Differentiating the loss function with respect to the weights arrives at the following gradient:

$$\nabla_{\theta_i} L_i(\theta_i) = \mathbb{E}_{s,a,r,s'} \left[(r + \gamma \max_{a'} Q(s^{'}, a^{'}; \theta_i^-) - Q(s, a; \theta_i)) \nabla_{\theta_i} Q(s, a; \theta_i) \right]. \tag{5.9}$$

Instead of computing the full expectations in the above gradient, a computational expedient to optimize the loss function uses stochastic gradient descent. The Q-learning algorithm (Watkins and Dayan, 1992) can be recovered in this framework by updating the weights after every time-step by replacing the expectations using single samples and set $\theta_i^- = \theta_{i-1}$. Since this algorithm solves the RL task without explicitly estimating the reward and transition dynamics $P(r, s^{'} \mid s, a)$, and it learns about the greedy policy $\arg\max_{a'} Q(s, a^{'}; \theta)$ while following a behaviour distribution that ensures adequate exploration of the state space. Thus, this algorithm is model-free and off-policy (Mnih et al., 2015).

The full algorithm for training DQN is presented in Algorithm 2. The agent selects and executes actions according to the ϵ-greedy policy based on Q; that is, the behaviour distribution is selected with probability $1 - \epsilon$ and selects a random action with probability ϵ. The algorithm modifies standard Q-learning to make it suitable for training large neural

networks without diverging (Mnih et al., 2015), employing a distinct network to generate target values y_j in the Q-learning adjustment. Specifically, the network is duplicated for every C iteration to produce a target network, denoted as $\hat{Q}$. This $\hat{Q}$ network generates Q-learning target values y_j for the upcoming C adjustments to Q.

Algorithm 2 Deep Q-learning with experience replay

Initialize replay memory D
Initialize action-value function Q with random weights θ
Initialize target action-value function $\hat{Q}$ with weights $\theta^- = \theta$
for *episode* $= 1$ to M **do**
  Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$
  for $t = 1$ to T **do**
    With probability ϵ select a random action a_t
    Otherwise select $a_t = \arg\max_a Q(\phi(s_t), a; \theta)$
    Execute action a_t and observe reward r_t and x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
    Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D
    Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D

$$\text{Set } y_j = \begin{cases} r_j & \text{if episode terminates at step } j+1 \\ r_j + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases}$$

    Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ with respect to θ
    Every C steps reset $\hat{Q} = Q$
  end for
end for

5.4 Application in vibroseis optimal route design

The vibroseis optimal route design problem is a vibroseis whose mission is to vibrate at multiple positions before arriving at the destination under an objective function's restriction, with the premise using the least number of shot samples and the least amount of time.

The following example is used to test the algorithm for optimal route design. Figure 5.5 shows a small (7×7) grids from a 2D view. For simplicity's sake, I assume the vibroseis starts at the up left, and its destination is at the bottom right of the 2D grids. The experimental setup uses the following minimal prior knowledge: the grid size and the interval between each point are 50 meters, the reward functions, and the number of actions. Suppose the velocity of the vibroseis is 5 m/s, and the vibration lasts up to 30 seconds. So, the time for the vibroseis to move from one grid point to another is 10 seconds. The goal for the design is to use the least number of source points together with the least amount of operational

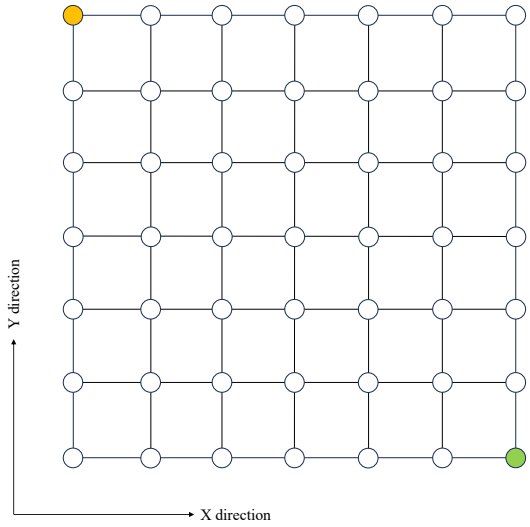

Figure 5.5: Schematic 2D view of the survey area. The area is divided into 7×7 grids, where the blank grid represents the candidate position that the vibroseis can move. The yellow point is the start location, and the green point is the finish point.

time spent in the field. Yellow and green points represent the start and finish positions, respectively. The blank points denote the candidate source positions that the vibroseis can move.

To validate the effectiveness of the proposed method, I compare it with traditional survey design methods and demonstrate its superiority in terms of the working time spent in the field. Figure 5.6 represents the conventional route for the dense and regular grid survey, where the vibroseis visit all the grid points and generate a controlled wavetrain. The red line denotes the movement direction of the vibroseis. In this case, the total time spent for the conventional survey is $48 \times 10 + 49 \times 30 = 1950$ seconds.

Conventionally in the field, in the pre-plot stage, satellite images are used to depict the topography of the survey area; obstacles like mountains, rivers, buildings, or other places that the vibroseis cannot visit, are marked, which are shown in Figure 5.7 by the black points. In this case, conventional and CS random sampling cannot fulfill the requirement, and optimal sampling with the shortest moving time is needed. Under these circumstances, I divide the problem into two steps. I first assume the optimal sampling positions are known and find the shortest path to validate that the shortest time requirement can be met. Then, adding the positions for the vibration points is also unknown, forming the whole optimal route design problem.

Thus, firstly, assume the case if the sensors' locations are known (depicted in Figure 5.8) and

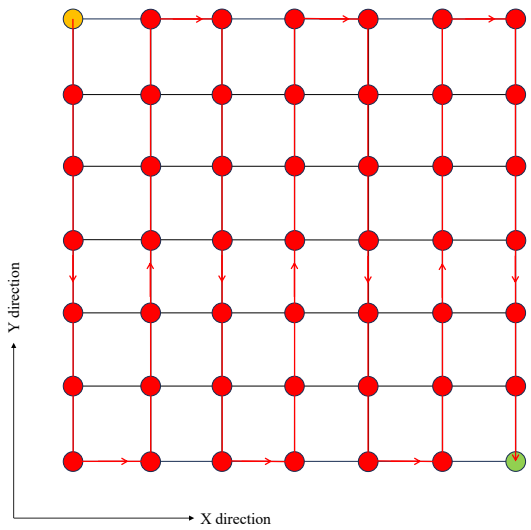

Figure 5.6: Schematic 2D view of the survey area. The red line with the arrow indicating the moving direction represents the vibroseis conventional route, where every grid point is visited, and each point is a source point.

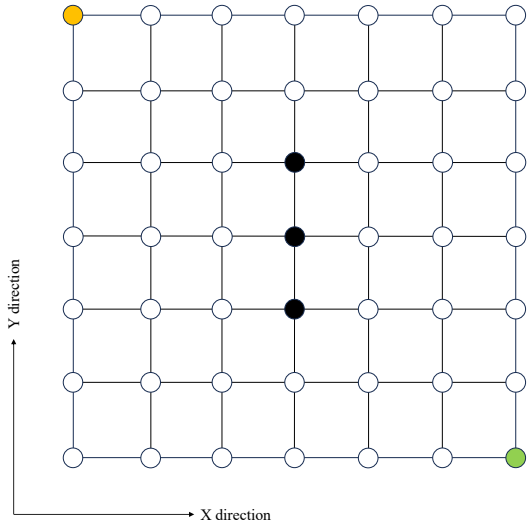

Figure 5.7: Schematic 2D view of the survey area. Black points represent the area where obstacles where the vibroseis cannot move and place a source point.

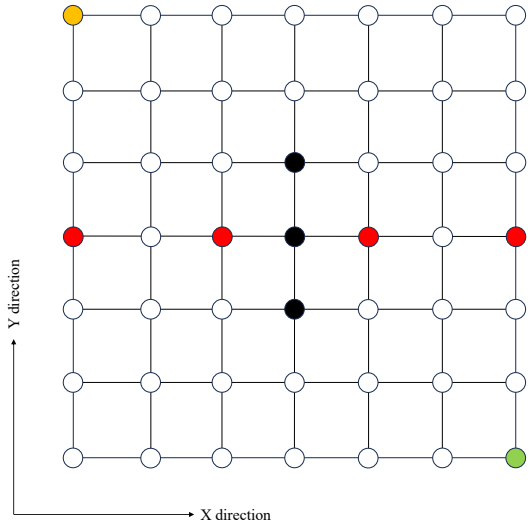

Figure 5.8: Schematic 2D view of the survey area with pre-defined sensor locations.

find the optimal path only. Then, the shortest route calculated with the DQN algorithm with four pre-defined sensors is shown in Figure 5.9, which shows that the shortest path algorithm is working effectively.

Then, I run the whole algorithm considering all the factors: the number of optimal sensors, the location of the optimal sensors and the shortest path problem. Choosing the correct parameters for a suitable neural network model is not easy and requires some experience. The optimizer adopted is Adam, and the loss function is MSE. I use two hidden layers, each equal to the pre-defined grid size. The input layer also has the same size as the grid size since it accepts the grid state as input. The output layer is the same as the number of actions (eight in this case) since it outputs the estimated Q-value for each action. The behaviour policy during training was ϵ-greedy with ϵ annealed linearly from 1.0 to 0.1 and fixed at 0.1 thereafter. I assume that future rewards are discounted by a factor of γ per time-step, and the value is set to 0.99 throughout. The final optimal sensor number is six, and the positions of the selected source points and the optimal route are shown in Figure 5.10. In this case, the total time spent for the optimal survey is $24 \times 10 + 6 \times 30 = 420$ seconds, which is much less than the 1950 seconds that a conventional survey requires. Further, the second largest maxima of the grid in this optimal scenario is 2.99, and the corresponding wavenumber spectra are shown in Figure 5.11. After the model is properly trained, the model is saved and can be loaded at any time to solve the vibroseis optimal route problem as a winning strategy. Further, for larger size problems, the main workflow, functions, and algorithm are the same, only consuming more compute time.

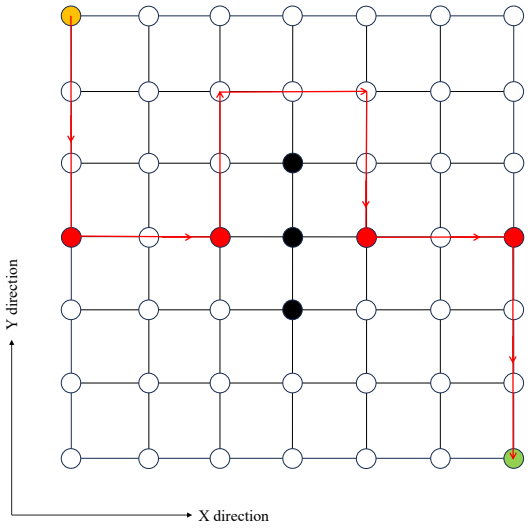

Figure 5.9: Schematic 2D view of optimal route with given sensor locations.

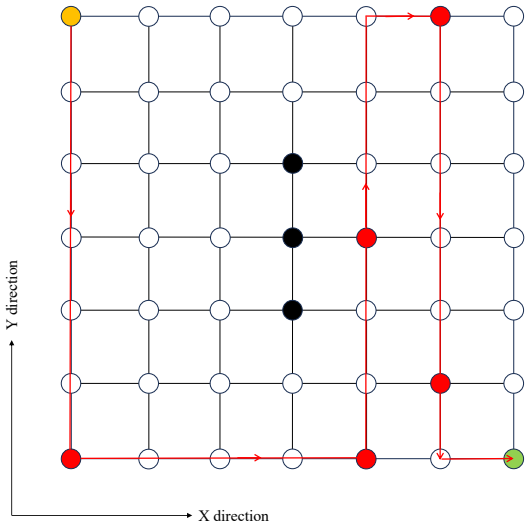

Figure 5.10: Schematic 2D view of optimal route selected following the learned policy.

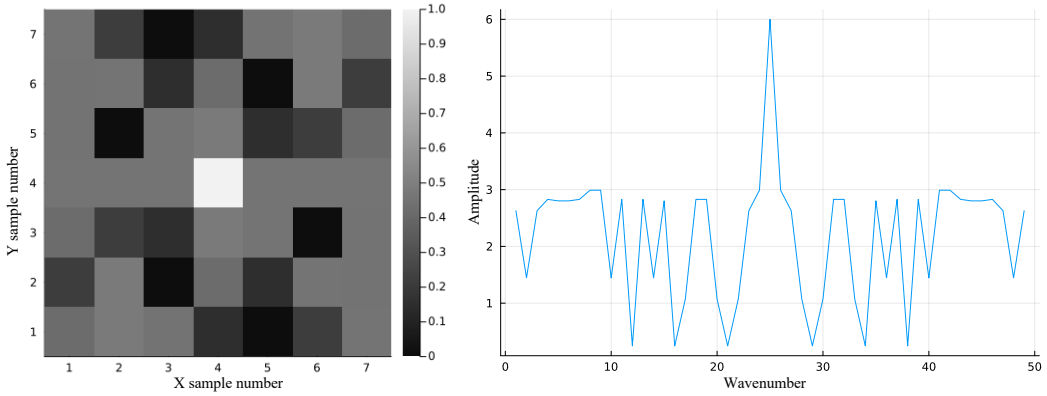

Figure 5.11: Wavenumber spectra of the optimal sampling scheme.

Finally, I also use the random agent as a baseline comparison, and the assessment of the performance of a random agent is when the agent selects an action randomly. Figure 5.12 shows the value of the second maxima of the grid for 1000 episodes of random policy. It is noticeable that after 1000 random simulations, there is only one time the second maxima value reaches the lowest value. Further, there is a possibility that even after more trials, the random policy still cannot reach the optimal setting.

Note that it may happen that in order to reach the goal, the agent will have to endure many penalties on its way. For example, the agent in the environment gets a small penalty for every legal move. The reason for that is that I want it to get to the destination in the shortest possible path. However, the shortest path to the target is sometimes long and winding, and the agent may have to endure many penalties until it gets to the target. After the model is properly trained, the model is saved and can be loaded at any time to solve the vibroseis optimal route problem as a winning strategy. Further, for larger size problems, the main workflow, functions, and algorithm are the same, only consuming more compute time. This chapter serves as a problem proposal and a first trial of DRL application in seismic acquisition design. Future work can cope with the larger and more difficult real data problems.

5.5 Conclusions

This chapter proposes the vibroseis optimal route problem together with the formulation of an MDP framework. DQN algorithm is applied to solve the problem. Minimizing the second maxima of the spectrum of the grid for optimal route design is adopted when some positions

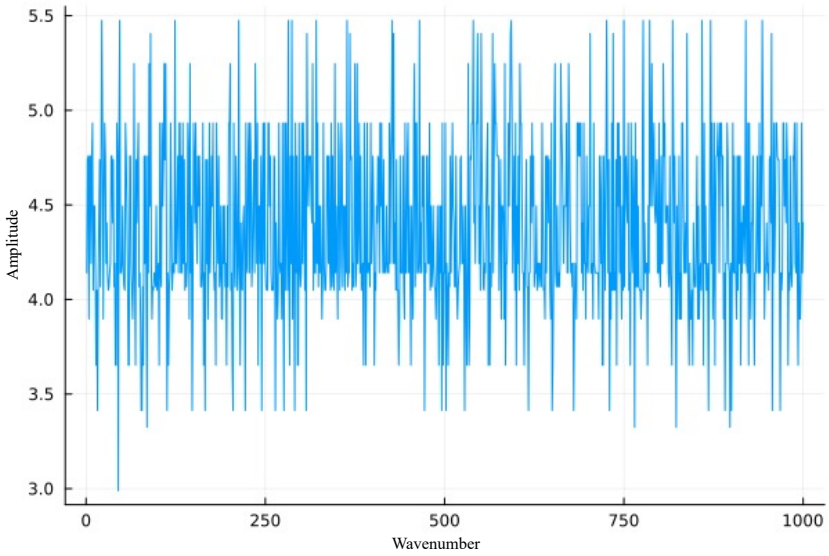

Figure 5.12: The wavenumber's second maxima value using random agent simulates 1000 episodes.

are unavailable. The DQN essentially utilizes a deep neural network to approximate the Q-values, overcoming the traditional limitations of Q-learning in high dimensional state spaces. By feeding the state into the network and obtaining the Q-values for each possible action, the DQN can effectively learn the optimal policy. The inclusion of a replay algorithm is pivotal for the successful integration of RL with deep learning architectures. Results demonstrated that with the proposed method, the least number of sensors is used, and the vibroseis can reach the finish points using the least amount of time. The acquisition cost is lower, and the field operation time is shortened. Preliminary results indicate that DRL holds promise as a robust tool for optimizing the efficiency, cost, and quality of seismic surveys, thereby presenting a transformative approach in the field of exploration geophysics.

CHAPTER 6

Conclusions

6.1 Conclusions and summary

Compared with processing, imaging, and interpretation, seismic acquisition costs most in the seismic exploration stages. Conventional seismic surveys require dense and equally-spaced sensor placement; however, this can be difficult due to expensive initial and maintenance costs. Further, relieving the spatial sampling requirements minimizes HSE exposure in the field and reduces the environmental footprint. While traditional seismic surveys aim ideally at regular and dense sampling, further easing the spatial sampling requirements contributes to the business aspect and reduces the ecological footprint (Jones, 2023). This research seeks to identify the ideal locations for sensors that would necessitate fewer deployments compared to conventional methods. Yet, the prerequisite remains that such configurations should facilitate data reconstructions comparable in quality to those obtained through traditional, denser surveys. Given that the optimal sensor deployment problem is NP-hard, brute-force search methodologies that evaluate all possible sensor positions are infeasible. Thus, this dissertation aims to provide actionable insights into the development of more efficient, cost-effective, and environmentally responsible seismic data acquisition strategies.

This research aims to explore and validate alternative methods that could potentially require fewer sampling points than the current CS theory suggests without sacrificing data integrity. This is accomplished by focusing on three principal avenues for acquisition design: data-driven approaches, reinforcement learning, and deep reinforcement learning. The four principal practical applications of this research focus on critical issues in both the energy and environmental sectors: simultaneous source acquisition, time-lapse monitoring for CO_2 geological storage monitoring, OBN acquisition, and vibroseis optimal route design. Conse-

quently, this research serves as a pivotal contribution to the broader endeavor of optimizing geophysical data acquisition and interpretation.

Chapter 2 provides a concise review of advancements in CS, which offer an appealing alternative for source-receiver sampling by sidestepping the constraints imposed by the Nyquist-Shannon theorem. Specifically, CS leverages random sampling and sparsity-enforcing algorithms to facilitate the recovery of missing seismic data. The primary outcome of this proposed approach is a new technology where overly stringent sampling criteria no longer determine acquisition and processing-related costs. Leveraging the principles of CS theory, this chapter advances the concept of irregular grid seismic acquisition. This design employs randomly distributed sampling positions with fewer shots or receivers and combines them with sparse transformation theory to facilitate irregular off-the-grid data acquisition and reconstruction. In essence, the proposed method can handle acquisition with an irregular-grid geometry based on compressive sensing design. Acquisition time can be reduced further by using fewer sources, thereby improving the efficiency of field data acquisition. A comparative study for reconstructing arbitrary compressive irregular-grid data with commonly used Fourier basis and low-rank constraints is also presented. The I-FMSSA method significantly improves computational efficiency compared with the I-MSSA method. An interpolation operator is adopted for EPOCS and I-FMSSA methods to connect irregular-grid observations and desired regular-grid data without losing accurate spatial coordinates information.

Chapter 3 adopts the optimal sensing theory for optimal source and receiver design for seismic acquisition. The quest for the optimum acquisition geometry is answered in the QR decomposition with column pivoting. The results of the data reconstruction prove that savings can be obtained with available training data. Further, noise can be removed simultaneously with reconstruction. The time-lapse example results reveal that a previously obtained dense base survey can optimize the monitoring design. This optimal design could lead to strategies that considerably reduce seismic acquisition costs and design intelligent data-driven approaches for seismic data acquisition. The dense acquisition from the base model serves as a guide, allowing us to infer the most crucial locations for capturing significant subsurface changes. Harnessing this knowledge, one can deploy a sparser yet highly effective acquisition geometry for the monitor survey. This methodology promises resource efficiency and paves the way for more sustainable and environmentally friendly seismic monitoring practices. Further, a hybrid method that involves inverting for coefficients that simultaneously model the signal via a global basis and analytical transform is proposed to solve the signal leakage problem that often arises in seismic processing. Synthetic and real data examples reveal that the signal component is not damaged compared to traditional noise removal methods.

Chapter 4 introduces a machine-learning-based approach for seismic sensor placement,

specifically leveraging reinforcement learning techniques to address the challenges of acquisition optimization. In the seismic sensor placement problem, the limited number of optimal sensors attempting to be in the optimal locations can acquire nearly the same reconstructed signal quality compared with the conventional dense and regular equal-spacing survey layout. Theoretically, this optimization problem can be incorporated into RL methodology to provide a brand new way to address seismic sensor placement design problems. Therefore, I propose an integrated RL-based optimal sensor location selection and take the OBN acquisition as an example. Defined in terms of optimization of the MDP, RL theory addresses the problem of, with an initially unknown environment, how an autonomous active agent learns the optimal policies while interacting. Thereby, I aim to justify whether the self-learning property can make RL a promising candidate for this optimization problem. Initially, the well-established Q-learning algorithm is adapted to tackle the optimization problem. Then, an MDP is explicitly developed for seismic acquisition design, with focusing the sensor placement issue. Once the optimal policy is trained through this MDP framework, it guides us toward achieving an optimal acquisition design in line with that policy. The resultant synthetic data scenarios are shown in CSGs, CRGs, and RTM images, demonstrating that the proposed method provides better sensor locations for reconstructing data than the jittered sampling scheme under the same circumstances. Comparison with random policy also verifies the effectiveness of the proposed method. Further, albeit only exemplified with seismic sensor placement application, this general idea of applying RL to solve the NP-hard combinatorial problem could also be used in other fields of science.

Chapter 5 proposes the vibroseis optimal route problem together with the formulation of an MDP framework. My intention is to pioneer the application of RL in the domain of seismic acquisition. Specifically, I aim to utilize DRL algorithms to navigate seismic acquisition design's highly complex and multi-variable space. Preliminary results indicate that DRL holds promise as a robust tool for optimizing the efficiency, cost, and quality of seismic surveys, thereby presenting a transformative approach in the field of exploration geophysics. DQN algorithm is applied to solve the problem. Minimize the second maxima of the spectrum of the grid for optimal route design when some positions are unavailable. The DQN essentially utilizes a deep neural network to approximate the Q-values, overcoming the traditional limitations of Q-learning in high dimensional state spaces. By feeding the state into the network and obtaining the Q-values for each possible action, the DQN can effectively learn the optimal policy for even the most complicated scenarios. The inclusion of a replay algorithm is pivotal for the successful integration of RL with deep learning architectures. The network can learn more efficiently and robustly by sampling from this replay buffer and revisiting these transitions. Results demonstrated that with the proposed method, the least number of sensors is used, and the vibroseis can reach the finish points using the least time.

Appendix A gives a toy example of the sampling operator utilizing the Kronecker product.

Appendix B provides a robust POCS version to de-emphasize outliers using the $\ell 1/\ell 2$ norm to reduce the influence of erratic noise. Synthetic and real data are used to examine the performance of the revised POCS method.

6.2 Future recommendations

The content of my dissertation relies on developing new algorithms/methods for seismic acquisition design. During my research, I also found some problems/limitations that can be considered in future research directions.

First, the potential benefits of CS are real and significant. But to realize them, several obstacles need to be surmounted, including overcoming the inertia of entrenched engineering practices and adapting the theoretical framework to practical acquisition schemes and workflows for imaging and inversion. The seismic application of CS and its extensions rely on solving an extensive system of equations that arise from the physical setting of exploration seismology. This puts pressure on developing large-scale solvers that can handle massive data volumes.

Second, for the problem of optimal vibroseis route design, even though I have tested single-agent cases, the field of multi-agent environment is full of research opportunities, which is the case of simultaneous source acquisition. Multiple agents may be demanded to cooperate to collect a mutual maximal reward, or each agent may be allowed to operate on its own egoistically to compete for its maximal reward. About the neural network, several neural networks could be built, and each has to compete against the other neural networks in predicting adversaries' strategies and selecting the best next move. Or only one neural network is needed for predicting a coordinated next move for all the agents.

Third, another potential direction for future work is to explore the use of other DRL algorithms, such as proximal policy optimization (PPO), for seismic acquisition design and compare their performance with the DQN algorithm. Moreover, the proposed approach can be extended to other optimization problems in the geophysical exploration industry, such as well placement design.

I postulate that some of the techniques developed and explored within the confines of this dissertation may be extensible to other applications. These could range from optimal acquisition design in non-destructive testing to advancements in biomedical imaging and telecommunication sectors (Tiwana et al., 2012; Pelivanov et al., 2014; Herrera-May et al., 2016; Manohar et al., 2017).

Last but not least, one promising application is earthquake early warning systems. Quick and accurate seismic data acquisition is crucial for early earthquake detection and warning.

My optimized acquisition techniques aim to contribute to the international academic frontier in earthquake early warning by enabling rapid response and more effective risk mitigation strategies. Although seismology relies on arrays in terms of depth and horizontal resolution, which is different from traditional seismic exploration, the research ideas of array technology (i.e., multi-seismometer acquisition) are first applied in the field of oil and gas exploration, and it has similarities in the principle of seismic wave propagation and the method of data processing. Under the actual conditions in the field, the seismic arrays are limited by the number of available instruments, budgetary funds, geological conditions and research objectives, and other factors. In addition, in recent years, the operation of dense arrays has become a development trend to use and densification of the existing network of arrays. How to design the optimal seismic observation system to improve the signal-to-noise ratio of the seismic signals and the ability of earthquake monitoring is an important research topic. Suppose the optimal acquisition method, by deep reinforcement learning algorithms, can be successfully applied to the design of an earthquake array design. In that case, data can be collected quickly and efficiently to help take rapid response measures to reduce casualties and property damage caused by earthquake disasters. It will be meaningful if this research can make a certain contribution to the timely and effective early warning of earthquake disasters.

Bibliography

Abma, R. and M. S. Foster, 2020, Simultaneous source seismic acquisition: Society of Exploration Geophysicists.

Abma, R. and N. Kabir, 2006, 3d interpolation of irregular data with a pocs algorithm: Geophysics, **71**, E91–E97.

Arts, R., A. Chadwick, and O. Eiken, 2005, Recent time-lapse seismic data show no indication of leakage at the sleipner co2-injection site, *in* Greenhouse Gas Control Technologies 7, 653–660. Elsevier.

Arts, R., O. Eiken, A. Chadwick, P. Zweigel, L. Van der Meer, and B. Zinszner, 2004, Monitoring of co2 injected at sleipner using time-lapse seismic data: Energy, **29**, 1383–1392.

Baraniuk, R., 2007, Compressive sensing [lecture notes]: IEEE Signal Processing Magazine, **24**, 118–121.

Baraniuk, R. G. and P. Steeghs, 2017, Compressive sensing: A new approach to seismic data acquisition: The Leading Edge, **36**, no. 8, 642–645.

Bardan, V., 1987, Trace interpolation in seismic data processing: Geophysical Prospecting, **35**, 343–358.

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.

Bednar, J. B., 1996, Coarse is coarse of course unless. . . : The Leading Edge, **15**, 763–764.

Bengio, Y. et al., 2009, Learning deep architectures for ai: Foundations and trends® in Machine Learning, **2**, 1–127.

Bertero, M. and P. Boccacci, 2003, Super-resolution in computational imaging: Micron, **34**, 265–273.

Beylkin, G., R. Coifman, and V. Rokhlin, 1991, Fast wavelet transforms and numerical algorithms i: Communications on pure and applied mathematics, **44**, 141–183.

Bhuiyan, M. and M. Sacchi, 2015, Optimization for sparse acquisition: SEG International Exposition and Annual Meeting, SEG–2015.

Bian, F., D. Kempe, and R. Govindan, 2006, Utility based sensor selection: Proceedings of

the 5th international conference on Information processing in sensor networks, 11–18.

Biondi, B., S. Fomel, and N. Chemingui, 1998, Azimuth moveout for 3-d prestack imaging: Geophysics, **63**, 574–588.

Biondi, B. L., 2006, 3D seismic imaging: Society of Exploration Geophysicists.

Blacquiere, G. and S. Nakayama, 2019, Optimum seismic acquisition geometry design with the help of artificial intelligence: 89th Annual International Meeting, SEG, Extended Abstracts, 117–121.

Brown, A. R., 2011, Interpretation of three-dimensional seismic data: Society of Exploration Geophysicists and American Association of Petroleum ....

Brown, L., C. C. Mosher, C. Li, R. Olson, J. Doherty, T. C. Carey, L. Williams, J. Chang, and E. Staples, 2017, Application of compressive seismic imaging at lookout field, alaska: The Leading Edge, **36**, 670–676.

Brunton, B. W., S. L. Brunton, J. L. Proctor, and J. N. Kutz, 2016, Sparse sensor placement optimization for classification: SIAM Journal on Applied Mathematics, **76**, 2099–2122.

Brunton, S. L. and J. N. Kutz, 2019, Data-driven science and engineering: Machine learning, dynamical systems, and control: Cambridge University Press.

Businger, P. and G. H. Golub, 1965, Linear least squares solutions by householder transformations: Numerische Mathematik, **7**, 269–276.

Cai, R., Q. Zhao, D.-P. She, L. Yang, H. Cao, and Q.-Y. Yang, 2014, Bernoulli-based random undersampling schemes for 2d seismic data regularization: Applied Geophysics, **11**, 321–330.

Campman, X., Z. Tang, H. Jamali-Rad, B. Kuvshinov, M. Danilouchkine, Y. Ji, W. Walk, and D. Smit, 2017, Sparse seismic wavefield sampling: The Leading Edge, **36**, 654–660.

Candes, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: multiscale modeling & simulation, **5**, 861–899.

Candes, E. J., 2008, The restricted isometry property and its implications for compressed sensing: Comptes rendus. Mathematique, **346**, 589–592.

Candes, E. J. and J. Romberg, 2006, Quantitative robust uncertainty principles and optimally sparse decompositions: Foundations of Computational Mathematics, **6**, 227–254.

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on information theory, **52**, 489–509.

Candes, E. J., J. K. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, **59**, 1207–1223.

Candes, E. J. and T. Tao, 2006, Near-optimal signal recovery from random projections: Universal encoding strategies?: IEEE transactions on information theory, **52**, 5406–5425.

Canning, A. and G. H. Gardner, 1996, Regularizing 3-d data sets with dmo: Geophysics,

61, 1103–1114.

Carin, L., D. Liu, and B. Guo, 2011, Coherence, compressive sensing, and random sensor arrays: IEEE Antennas and Propagation Magazine, **53**, 28–39.

Carozzi, F. and M. D. Sacchi, 2019, Robust tensor-completion algorithm for 5d seismic-data reconstruction: Geophysics, **84**, V97–V109.

——— 2020, Making seismic reconstruction more robust via a generalized loss function: SEG International Exposition and Annual Meeting, D041S077R004.

——— 2021, Interpolated multichannel singular spectrum analysis: A reconstruction method that honors true trace coordinates: Geophysics, **86**, no. 1, V55–V70.

Cary, P. W., 2011, Aliasing and 5d interpolation with the mwni algorithm, *in* SEG Technical Program Expanded Abstracts 2011, 3080–3084. Society of Exploration Geophysicists.

Chadwick, A., G. Williams, N. Delepine, V. Clochard, K. Labat, S. Sturton, M.-L. Buddensiek, M. Dillen, M. Nickel, A. L. Lima, et al., 2010, Quantitative analysis of time-lapse seismic monitoring data at the sleipner co 2 storage operation: The Leading Edge, **29**, 170–177.

Chadwick, R., R. Arts, and O. Eiken, 2005, 4d seismic quantification of a growing co2 plume at sleipner, north sea: Geological Society, London, Petroleum Geology Conference series, 1385–1399.

Chadwick, R., R. Arts, O. Eiken, G. Kirby, E. Lindeberg, and P. Zweigel, 2004, 4d seismic imaging of an injected c02 plume at the sleipner field, central north sea: Geological Society, London, Memoirs, **29**, 311–320.

Chadwick, R., D. Noy, R. Arts, and O. Eiken, 2009, Latest time-lapse seismic data from sleipner yield new insights into co2 plume development: Energy Procedia, **1**, 2103–2110.

Chatterjee, A., 2000, An introduction to the proper orthogonal decomposition: Current science, 808–817.

Chen, K. and M. D. Sacchi, 2014, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, V1–V11.

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

Cheng, J., M. Sacchi, and J. Gao, 2019, Computational efficient multidimensional singular spectrum analysis for prestack seismic data reconstruction: Geophysics, **84**, V111–V119.

Cheng, J. and M. D. Sacchi, 2015, Separation and reconstruction of simultaneous source data via iterative rank reduction: Geophysics, **80**, V57–V66.

Chiu, S., B. Gong, and Z. Meng, 2017, Residual-signal recovery using a pattern-based method: 87th Annual International Meeting, SEG, Expanded Abstracts, 5069–5073.

Chiu, S. K., 2014, Multidimensional interpolation using a model-constrained minimum weighted norm interpolation: Geophysics, **79**, V191–V199.

Chiu, S. K. and J. E. Howell, 2008, Attenuation of coherent noise using localized-adaptive

eigenimage filter: 78th Annual International Meeting, SEG, Expanded Abstracts, 2541–2545.

Choo, J., J. Downton, and J. Dewar, 2004, LIFT: A new and practical approach to noise and multiple attenuation: CSEG National Convention, 1–4.

Claerbout, J., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications, London.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Darche, G., 1990, Spatial interpolation using a fast parabolic transform, *in* SEG Technical Program Expanded Abstracts 1990, 1647–1650. Society of Exploration Geophysicists.

Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, **57**, 1413–1457.

Davis, T. L., M. J. Terrell, R. D. Benson, R. Cardona, R. R. Kendall, and R. Winarsky, 2003, Multicomponent seismic characterization and monitoring of the co 2 flood at weyburn field, saskatchewan: The Leading Edge, **22**, 696–697.

Dell'Aversana, P., 2022, Combining geophysical inversion with reinforcement learning: 83rd EAGE Annual Conference & Exhibition, 1–5.

Demanet, L. and L. Ying, 2007, Wave atoms and sparsity of oscillatory patterns: Applied and Computational Harmonic Analysis, **23**, 368–387.

Dippe, M. A. and E. H. Wold, 1992, Stochastic sampling: Theory and application, *in* Progress in computer graphics (vol. 1), 1–54.

Diuk, C., A. Cohen, and M. L. Littman, 2008, An object-oriented representation for efficient reinforcement learning: Proceedings of the 25th international conference on Machine learning, 240–247.

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

Donoho, D. L., X. Huo, et al., 2001, Uncertainty principles and ideal atomic decomposition: IEEE transactions on information theory, **47**, 2845–2862.

Drmac, Z. and S. Gugercin, 2016, A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions: SIAM Journal on Scientific Computing, **38**, A631–A648.

Duijndam, A., M. Schonewille, and C. Hindriks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, 524–538.

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

Egorov, M., Z. N. Sunberg, E. Balaban, T. A. Wheeler, J. K. Gupta, and M. J. Kochenderfer,

2017, Pomdps. jl: A framework for sequential decision making under uncertainty: The Journal of Machine Learning Research, **18**, 831–835.

Elad, M., 2007, Optimized projections for compressed sensing: IEEE Transactions on Signal Processing, **55**, 5695–5702.

Eldar, Y. C., 2006, Mean-squared error sampling and reconstruction in the presence of noise: IEEE Transactions on Signal Processing, **54**, 4619–4633.

Fabriol, H., A. Bitri, B. Bourgeois, M. Delatre, J. Girard, G. Pajot, and J. Rohmer, 2011, Geophysical methods for co2 plume imaging: Comparison of performances: Energy Procedia, **4**, 3604–3611.

Feng, Q., L. Han, B. Pan, and B. Zhao, 2022, Microseismic source location using deep reinforcement learning: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–9.

Feuer, A. and G. C. Goodwin, 2005, Reconstruction of multidimensional bandlimited signals from nonuniform and generalized samples: IEEE Transactions on Signal Processing, **53**, 4273–4282.

Figueiredo, M. A., R. D. Nowak, and S. J. Wright, 2007, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems: IEEE Journal of selected topics in signal processing, **1**, 586–597.

Florez, K., S. Abreo, and A. Ramirez, 2018, Minimal coherence optimization for coding blended sources: First EAGE Workshop on High Performance Computing for Upstream in Latin America, 1–5.

Fomel, S., 2006, Towards the seislet transform: SEG International Exposition and Annual Meeting, SEG–2006.

Fukushima, K., 1980, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position: Biological cybernetics, **36**, 193–202.

Gao, J., M. D. Sacchi, and X. Chen, 2013a, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: Geophysics, **78**, V21–V30.

Gao, J., A. Stanton, M. Naghizadeh, M. D. Sacchi, and X. Chen, 2013b, Convergence improvement and noise attenuation considerations for beyond alias projection onto convex sets reconstruction: Geophysical prospecting, **61**, 138–151.

Gao, J., A. Stanton, and M. D. Sacchi, 2015, Parallel matrix factorization algorithm and its application to 5d seismic reconstruction and denoising: Geophysics, **80**, V173–V187.

Gersztenkorn, A., J. B. Bednar, and L. R. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: Geophysics, **51**, 357–368.

Golub, G. H., A. Hoffman, and G. W. Stewart, 1987, A generalization of the eckart-young-mirsky matrix approximation theorem: Linear Algebra and its applications, **88**, 317–327.

Gray, S. H., J. Etgen, J. Dellinger, and D. Whitmore, 2001, Seismic migration problems and

solutions: Geophysics, **66**, 1622–1640.

Guitton, A. and J. Loubani, 2022, An approximation of the mutual coherence for compressive sensing problems: 83rd EAGE Annual Conference & Exhibition, 1–5.

Gülünay, N., 2003, Seismic trace interpolation in the fourier transform domain: Geophysics, **68**, 355–369.

Guo, Y., R. Lin, and M. D. Sacchi, 2022, Seismic noise attenuation via learned and analytical basis functions: SEG International Exposition and Annual Meeting, D011S150R003.

——— 2023, Optimal seismic sensor placement based on reinforcement learning approach: An example of obn acquisition design: IEEE Transactions on Geoscience and Remote Sensing, **61**, 1–12.

Guo, Y. and M. D. Sacchi, 2020, Data-driven time-lapse acquisition design via optimal receiver-source placement and reconstruction, *in* SEG Technical Program Expanded Abstracts 2020, 66–70. Society of Exploration Geophysicists.

Hampson, G., J. Stefani, and F. Herkenhoff, 2008, Acquisition using simultaneous sources: The Leading Edge, **27**, 918–923.

Hennenfen, G. and F. Herrmann, 2005, Sparseness-constrained data continuation with frames: Applications to missing traces and aliased signals in 2/3-d: SEG International Exposition and Annual Meeting, SEG–2005.

Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with nonuniformly sampled curvelets: Computing in Science & Engineering, **8**, 16–25.

——— 2008a, Simply denoise: Wavefield reconstruction via jittered undersampling: Geophysics, **73**, V19–V28.

——— 2008b, Simply denoise: Wavefield reconstruction via jittered undersampling: Geophysics, **73**, no. 3, V19–V28.

Herrera-May, A. L., J. C. Soler-Balcazar, H. Vázquez-Leal, J. Martínez-Castillo, M. O. Vigueras-Zuñiga, and L. A. Aguilera-Cortés, 2016, Recent advances of mems resonators for lorentz force based magnetic field sensors: design, applications and challenges: Sensors, **16**, 1359.

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

Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35–A40.

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

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

Hinton, G. E. and R. R. Salakhutdinov, 2006, Reducing the dimensionality of data with

neural networks: science, **313**, 504–507.

Hubel, D. H. and T. N. Wiesel, 1963, Shape and arrangement of columns in cat's striate cortex: The Journal of physiology, **165**, 559.

Ibrahim, A. and M. D. Sacchi, 2014, Simultaneous source separation using a robust radon transform: Geophysics, **79**, V1–V11.

Jamali-Rad, H., B. Kuvshinov, Z. Tang, and X. Campman, 2016, Deterministically subsampled acquisition geometries for optimal reconstruction: 78th EAGE Conference and Exhibition 2016, 1–5.

Jia, Y. and J. Ma, 2017, What can machine learning do for seismic data processing? an interpolation application: Geophysics, **82**, 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: 87th Annual International Meeting, SEG, Extended Abstracts, 4272–4277.

Jiang, T., Y. Jiang, D. Clark, and R. Gray, 2018, A compressive seismic field trial and reconstruction test using regular indexing: SEG International Exposition and Annual Meeting, SEG–2018.

Jones, A. G., 2023, Mining for net zero: The impossible task: The Leading Edge, **42**, 266–276.

Kaelbling, L. P., M. L. Littman, and A. R. Cassandra, 1998, Planning and acting in partially observable stochastic domains: Artificial intelligence, **101**, 99–134.

Kearey, P., M. Brooks, and I. Hill, 2002, An introduction to geophysical exploration, volume **4**: John Wiley & Sons.

Kotsiantis, S. B., I. Zaharakis, P. Pintelas, et al., 2007, Supervised machine learning: A review of classification techniques: Emerging artificial intelligence applications in computer engineering, **160**, 3–24.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction: Geophysics, **78**, V273–V284.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, Imagenet classification with deep convolutional neural networks: Advances in neural information processing systems, **25**.

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**, V97–V114.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: Proceedings of the IEEE, **86**, 2278–2324.

Leneman, O. A., 1966, Random sampling of random processes: Impulse processes: Information and Control, **9**, 347–363.

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

of Exploration Geophysicists.

Li, C., C. C. Mosher, L. C. Morley, Y. Ji, and J. D. Brewer, 2013, Joint source deblending and reconstruction for seismic data, *in* SEG Technical Program Expanded Abstracts 2013, 82–87. Society of Exploration Geophysicists.

Li, C., Y. Zhang, and C. C. Mosher, 2019, A hybrid learning-based framework for seismic denoising: The Leading Edge, **38**, no. 7, 542–549.

Lin, L.-J., 1992, Reinforcement learning for robots using neural networks: Carnegie Mellon University.

Lin, R., Y. Guo, F. Carozzi, and M. Sacchi, 2022a, Separation and shot interpolation of simultaneous-source data with interpolated mssa (i-mssa): 83rd EAGE Annual Conference & Exhibition, 1–5.

Lin, R., Y. Guo, F. Carozzi, and M. D. Sacchi, 2022b, Interpolated fast and computational-efficient multidimensional singular spectrum analysis (i-fmssa) for compressive simultaneous-source data processing: SEG International Exposition and Annual Meeting, D011S002R001.

——— 2022c, Simultaneous deblending and source reconstruction for compressive 3d simultaneous-source acquisition data via interpolated multichannel singular spectrum analysis: Geophysics, **87**, V559–V570.

Lin, R. and M. D. Sacchi, 2020, Separation of simultaneous sources acquired with a high blending factor via coherence pass robust radon operators: Geophysics, **85**, V269–V282.

Liu, B. and M. Sacchi, 2001, Minimum weighted norm interpolation of seismic data with adaptive weights: 71st annual international meeting, seg, expanded abstracts, 1921–1924: Presented at the .

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: SEG International Exposition and Annual Meeting, SEG–2004.

Liu, G., S. Fomel, and X. Chen, 2011, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.

Liu, W., S. Cao, G. Li, and Y. He, 2015, Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform: Journal of Applied Geophysics, **115**, 129–139.

Liu, Y. and S. Fomel, 2010, Oc-seislet: Seislet transform construction with differential offset continuation: Geophysics, **75**, WB235–WB245.

Lowrie, W. and A. Fichtner, 2020, Fundamentals of geophysics: Cambridge university press.

Lu, C., H. Li, and Z. Lin, 2018, Optimized projections for compressed sensing via direct mutual coherence minimization: Signal Processing, **151**, 45–55.

Lu, K., Y. Jin, Y. Chen, Y. Yang, L. Hou, Z. Zhang, Z. Li, and C. Fu, 2019, Review for

order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems: Mechanical Systems and Signal Processing, **123**, 264–297.

Lumley, D., 2010, 4d seismic monitoring of co 2 sequestration: The Leading Edge, **29**, 150–155.

Ma, J., 2013, Three-dimensional irregular seismic data reconstruction via low-rank matrix completion: Geophysics, **78**, V181–V192.

Ma, J. and G. Plonka, 2010, The curvelet transform: IEEE signal processing magazine, **27**, 118–133.

Ma, J. and S. Yu, 2017, Sparsity in compressive sensing: The Leading Edge, **36**, 646–652.

Ma, J.-W., J. Xu, Y.-Q. Bao, and S.-W. Yu, 2012, Compressive sensing and its application: from sparse to low-rank regularized optimization: Signal Processing(Xinhao Chuli), **28**, 609–623.

Ma, Y., T. Fei, and Y. Luo, 2019, A new insight into automatic first-arrival picking based on reinforcement learning, *in* 81st EAGE Conference and Exhibition, volume **2019**, 1–5.

Manohar, K., B. W. Brunton, J. N. Kutz, and S. L. Brunton, 2018, Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns: IEEE Control Systems Magazine, **38**, 63–86.

Manohar, K., S. L. Brunton, and J. N. Kutz, 2017, Environment identification in flight using sparse approximation of wing strain: Journal of Fluids and Structures, **70**, 162–180.

McClelland, J. L., B. L. McNaughton, and R. C. O'Reilly, 1995, Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory.: Psychological review, **102**, 419.

Mendible, A., S. L. Brunton, A. Y. Aravkin, W. Lowrie, and J. N. Kutz, 2020, Dimensionality reduction and reduced-order modeling for traveling wave physics: Theoretical and Computational Fluid Dynamics, **34**, 385–400.

Mnih, V., K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, 2013, Playing atari with deep reinforcement learning: arXiv preprint arXiv:1312.5602.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., 2015, Human-level control through deep reinforcement learning: nature, **518**, 529–533.

Moldoveanu, N., 2010, Random sampling: A new strategy for marine acquisition, *in* SEG Technical Program Expanded Abstracts 2010, 51–55. Society of Exploration Geophysicists.

Monk, D. J., 2020, Survey design and seismic acquisition for land, marine, and in-between in light of new technology and techniques: Society of Exploration Geophysicists.

Mosher, C., S. Kaplan, and F. Janiszewski, 2012a, Non-uniform optimal sampling for seismic survey design: 74th EAGE Conference and Exhibition incorporating EUROPEC, cp–293.

Mosher, C., C. Li, L. Morley, Y. Ji, F. Janiszewski, R. Olson, and J. Brewer, 2014a, Increasing the efficiency of seismic data acquisition via compressive sensing: The Leading Edge, **33**, no. 4, 386–391.

Mosher, C. C., E. Keskula, S. T. Kaplan, R. G. Keys, C. Li, E. Z. Ata, L. C. Morley, J. D. Brewer, F. D. Janiszewski, P. M. Eick, et al., 2012b, Compressive seismic imaging: SEG International Exposition and Annual Meeting, SEG–2012.

Mosher, C. C., C. Li, F. D. Janiszewski, L. S. Williams, T. C. Carey, and Y. Ji, 2017, Operational deployment of compressive sensing systems for seismic data acquisition: The Leading Edge, **36**, no. 8, 661–669.

Mosher, C. C., C. Li, L. C. Morley, F. D. Janiszewski, Y. Ji, and J. Brewer, 2014b, Non-uniform optimal sampling for simultaneous source survey design: SEG International Exposition and Annual Meeting, SEG–2014.

Naghizadeh, M. and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: Geophysics, **72**, V111–V118.

——— 2010a, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: Geophysics, **75**, WB189–WB202.

——— 2010b, On sampling functions and fourier reconstruction methods: Geophysics, **75**, WB137–WB151.

Naghizadeh, M., P. Vermeulen, A. Crook, A. Birce, S. Ross, A. Stanton, M. Rodriguez, and W. Cookson, 2023, Ecoseis: A novel acquisition method for optimizing seismic resolution while minimizing environmental footprint: The Leading Edge, **42**, 61–68.

Nakayama, S., G. Blacquière, and T. Ishiyama, 2019, Automated survey design for blended acquisition with irregular spatial sampling via the integration of a metaheuristic and deep learning: Geophysics, **84**, P47–P60.

Needell, D. and J. A. Tropp, 2008, Cosamp: Iterative signal recovery from incomplete and inaccurate samples.

Obermeier, R. and J. A. Martinez-Lorenzo, 2017, Sensing matrix design via mutual coherence minimization for electromagnetic compressive imaging applications: IEEE Transactions on Computational Imaging, **3**, 217–229.

O'Brien, J., F. Kilbride, and F. Lim, 2004, Time-lapse vsp reservoir monitoring: The Leading Edge, **23**, 1178–1184.

Oppenheim, A. V., 1999, Discrete-time signal processing: Pearson Education India.

Oropeza, V. and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.

Osisanwo, F., J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, 2017, Supervised machine learning algorithms: classification and comparison: International Journal of Computer Trends and Technology (IJCTT), **48**, 128–138.

O'Neill, J., B. Pleydell-Bouverie, D. Dupret, and J. Csicsvari, 2010, Play it again: reactiva-

tion of waking experience and memory: Trends in neurosciences, **33**, 220–229.

Park, S. C., M. K. Park, and M. G. Kang, 2003, Super-resolution image reconstruction: a technical overview: IEEE signal processing magazine, **20**, 21–36.

Pelivanov, I., T. Buma, J. Xia, C.-W. Wei, and M. O'Donnell, 2014, A new fiber-optic non-contact compact laser-ultrasound scanner for fast non-destructive testing and evaluation of aircraft composites: Journal of Applied Physics, **115**, 113105.

Pinnau, R., 2008, Model reduction via proper orthogonal decomposition, *in* Model order reduction: theory, research aspects and applications, 95–109. Springer.

Porsani, M. J., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, 1461–1467.

Puryear, C. I., O. N. Portniaguine, C. M. Cobos, and J. P. Castagna, 2012, Constrained least-squares spectral analysis: Application to seismic data: Geophysics, **77**, V143–V167.

Riedmiller, M., T. Gabel, R. Hafner, and S. Lange, 2009, Reinforcement learning for robot soccer: Autonomous Robots, **27**, 55–73.

Romanes, G. J., 1883, Animal intelligence, volume **44**: D. Appleton.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Rubino, J. G., D. R. Velis, and M. D. Sacchi, 2011, Numerical analysis of wave-induced fluid flow effects on seismic data: Application to monitoring of co2 storage at the sleipner field: Journal of Geophysical Research: Solid Earth, **116**.

Sacchi, M. D. et al., 2009, Fx singular spectrum analysis: Cspg Cseg Cwls Convention, 392–395.

Sacchi, M. D. and B. Liu, 2005, Minimum weighted norm wavefield reconstruction for ava imaging: Geophysical Prospecting, **53**, 787–801.

Sacchi, M. D. and D. Trad, 2010, Reconstruction of wide azimuth seismic data: assumptions, methodology and field examples: SEG International Exposition and Annual Meeting, SEG–2010.

Sacchi, M. D. and T. J. Ulrych, 1996, Estimation of the discrete fourier transform, a linear inversion approach: Geophysics, **61**, 1128–1136.

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38.

Sava, P. and S. J. Hill, 2009, Overview and classification of wavefield seismic imaging methods: The Leading Edge, **28**, no. 2, 170–183.

Schonewille, M., A. Duijndam, R. Romijn, and L. Ongkiehong, 2003, A general reconstruction scheme for dominant azimuth 3d seismic data: Geophysics, **68**, 2092–2105.

Schultz, W., P. Dayan, and P. R. Montague, 1997, A neural substrate of prediction and reward: Science, **275**, 1593–1599.

Serre, T., L. Wolf, and T. Poggio, 2005, Object recognition with features inspired by visual cortex: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 994–1000.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., 2016, Mastering the game of go with deep neural networks and tree search: nature, **529**, 484–489.

Silver, D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al., 2018, A general reinforcement learning algorithm that masters chess, shogi, and go through self-play: Science, **362**, 1140–1144.

Spetzler, J., Z. Xue, H. Saito, D. Nobuoka, H. Azuma, and O. Nishizawa, 2006, Time-lapse seismic crosswell monitoring of co2 injected in an onshore sandstone aquifer: Proceedings of the 8th SEGJ International Symposium, 1–6.

Spitz, S., 1991, Seismic trace interpolation in the fx domain: Geophysics, **56**, 785–794.

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

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

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, 890–908.

Strang, G., G. Strang, G. Strang, and G. Strang, 1993, Introduction to linear algebra, volume **3**: Wellesley-Cambridge Press Wellesley, MA.

Strohmer, T., 1997, Computationally attractive reconstruction of bandlimited images from irregular samples: IEEE Transactions on image processing, **6**, 540–548.

Sun, B. and T. Alkhalifah, 2020, A data-driven choice of misfit function for fwi using reinforcement learning, *in* 82nd EAGE Annual Conference & Exhibition Online, volume **2020**, 1–5.

Sutton, R. S. and A. G. Barto, 2018, Reinforcement learning: An introduction: MIT press.

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

Tesauro, G. et al., 1995, Temporal difference learning and td-gammon: Communications of the ACM, **38**, 58–68.

Tian, Y., L. Wei, C. Li, S. Oppert, and G. Hennenfent, 2018, Joint sparsity recovery for noise attenuation: 88th Annual International Meeting, SEG, Expanded Abstracts, 4186–4190.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B: Statistical Methodology, **58**, 267–288.

Titova, A., M. B. Wakin, and A. Tura, 2019, Mutual coherence in compressive sensing seismic acquisition: SEG International Exposition and Annual Meeting, D043S094R008.

——— 2021, Two-stage sampling—a novel approach for compressive sensing seismic acqui-

sition: First International Meeting for Applied Geoscience & Energy, 110–114.

Tiwana, M. I., S. J. Redmond, and N. H. Lovell, 2012, A review of tactile sensing technologies with applications in biomedical engineering: Sensors and Actuators A: physical, **179**, 17–31.

Trad, D., 2008, Five dimensional seismic data interpolation: SEG International Exposition and Annual Meeting, SEG–2008.

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

Trad, D., J. Deere, S. Cheadle, et al., 2005, Challenges for land data interpolation: CSEG Annual Convention, Expanded Abstracts, 309–311.

Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse radon transform: Geophysics, **68**, 386–399.

Trad, D. O. and T. J. Ulrych, 1999, Radon transform: beyond aliasing with irregular sampling: 6th International Congress of the Brazilian Geophysical Society, cp–215.

Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: SEG International Exposition and Annual Meeting, SEG–2010.

Tsitsiklis, J. and B. Van Roy, 1996, An analysis of temporal-difference learning with function approximationtechnical: Rep. LIDS-P-2322). Lab. Inf. Decis. Syst. Massachusetts Inst. Technol. Tech. Rep.

Unser, M., 2000, Sampling-50 years after shannon: Proceedings of the IEEE, **88**, 569–587.

van den Berg, E. and M. P. Friedlander, 2007, In pursuit of a root: Technical report, Citeseer.

Van Loan, C. F. and G. Golub, 1996, Matrix computations (johns hopkins studies in mathematical sciences).

Vandewalle, P., L. Sbaiz, J. Vandewalle, and M. Vetterli, 2007, Super-resolution from unregistered and totally aliased signals using subspace methods: IEEE Transactions on Signal Processing, **55**, 3687–3703.

Vermeer, G. J., 1990, Seismic wavefield sampling: Society of Exploration Geophysicists and Shell Research BV.

——— 2012, 3d seismic survey design: Society of Exploration Geophysicists.

Verschuur, D. J., A. Berkhout, and C. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Wang, B., J. Geng, and J. Song, 2022, Off-the-grid blended data deblending via a sinc-based interpolator: SEG 2021 Workshop: 4th International Workshop on Mathematical Geophysics: Traditional & Learning, 5–8.

Wang, H., C. Yin, Y. Liu, X. Ren, J. Li, and B. Xiong, 2020, Deep reinforcement learning for magnetotelluric data inversion, *in* 3rd Asia Pacific Meeting on Near Surface Geoscience & Engineering, volume **2020**, 1–4.

Wang, Y., J. Cao, and C. Yang, 2011, Recovery of seismic wavefields based on compressive sensing by an l 1-norm constrained trust region method and the piecewise random subsampling: Geophysical Journal International, **187**, 199–213.

Wang, Z., H.-X. Li, and C. Chen, 2019, Reinforcement learning-based optimal sensor placement for spatiotemporal modeling: IEEE transactions on cybernetics, **50**, 2861–2871.

Watkins, C. J. and P. Dayan, 1992, Q-learning: Machine learning, **8**, 279–292.

Wu, T., S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, 2023, A brief overview of chatgpt: The history, status quo and potential future development: IEEE/CAA Journal of Automatica Sinica, **10**, 1122–1136.

Wu, Y., J. Wei, J. Pan, and P. Chen, 2019, Research on microseismic source locations based on deep reinforcement learning: IEEE Access, **7**, 39962–39973.

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

Yilmaz, Ö., 2021, Land seismic case studies for near-surface modeling and subsurface imaging: Society of Exploration Geophysicists.

Yu, S. and J. Ma, 2021, Deep learning for geophysics: Current and future trends: Reviews of Geophysics, **59**, e2021RG000742.

Yu, S., J. Ma, X. Zhang, and M. D. Sacchi, 2015, Interpolation and denoising of high-dimensional seismic data by learning a tight frame: Geophysics, **80**, V119–V132.

Yu, S., J. Ma, and B. Zhao, 2020, Off-the-grid vertical seismic profile data regularization by a compressive sensing method: Geophysics, **85**, V157–V168.

Zhang, Y., M. Louboutin, A. Siahkoohi, Z. Yin, R. Kumar, and F. J. Herrmann, 2022, A simulation-free seismic survey design by maximizing the spectral gap: Second International Meeting for Applied Geoscience & Energy, 15–20.

Zhou, R., L. Huang, J. T. Rutledge, M. Fehler, T. M. Daley, and E. L. Majer, 2010, Coda-wave interferometry analysis of time-lapse vsp data for monitoring geological carbon sequestration: International Journal of Greenhouse Gas Control, **4**, 679–686.

Zi, Y., L. Fan, X. Wu, J. Chen, S. Wang, and Z. Han, 2022, Active gamma-ray well logging pattern localization with reinforcement learning, *in* Second International Meeting for Applied Geoscience & Energy, 337–341.

Zwartjes, P. and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: Geophysics, **71**, V171–V186.

——— 2007, Fourier reconstruction with sparse inversion: Geophysical prospecting, **55**, 199–221.

Zwartjes, P. and M. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, V21–V32.

APPENDIX A

Sampling operator using Kronecker product

Consider an ideal seismic data cube $\mathcal{D}(t, r, s)$, where t indicates time, r and s denote the receiver and source position, respectively. Assume the integers N_t, N_r, and N_s are the number of time samples, receivers, and sources, respectively. The cube can be represented as a third-order tensor $\mathcal{D}$ of size $N_t \times N_r \times N_s$.

For illustration purposes, consider an optimal receiver problem, I reshape the ideal data tensor $\mathcal{D}$ into a matrix $\mathbf{D}$ of size $(N_t \times N_r) \times N_s$ where each column of $\mathbf{D}$ corresponds to a common source gather (CSG). Assume the data is acquired with N_s sources and only K receivers ($K < N_r$) are needed for the optimal design problem.

The sampling operator for this problem is $\mathbf{T}_r$. No sampling in time is performed, and the whole seismogram is taken for each sampled receiver. This can be easily accommodated by expressing the data sampling matrix via

$$\mathbf{T} = \mathbf{T}_r \otimes \mathbf{I}_{N_t}, \tag{A.1}$$

where $\mathbf{I}_{N_t}$ is the $N_t \times N_t$ identity matrix and the symbol $\otimes$ stands for Kronecker product. $\mathbf{T}$ is the sampling matrix, a sparse matrix of size $(N_t \times K) \times (N_t \times N_r)$, which extracts K selected seismograms belonging to receivers with sampling positions $\nu_k, k = 1 \dots K$, from the ideal data $\mathbf{D}$. The latter permits the expression of the sampled data, referred to as the observed data for a given sampling set $\boldsymbol{\nu}$, as follows

$$\mathbf{D}_{\boldsymbol{\nu}} = \mathbf{TD}. \tag{A.2}$$

For the toy example, I use $N_s = 10$, $N_r = 20$, and $N_t = 140$. The receivers are deployed at positions given by the sampling set $\boldsymbol{\nu} = [\nu_1, \nu_2 \dots \nu_K]$, and $K = 8$. Figure B.1 shows

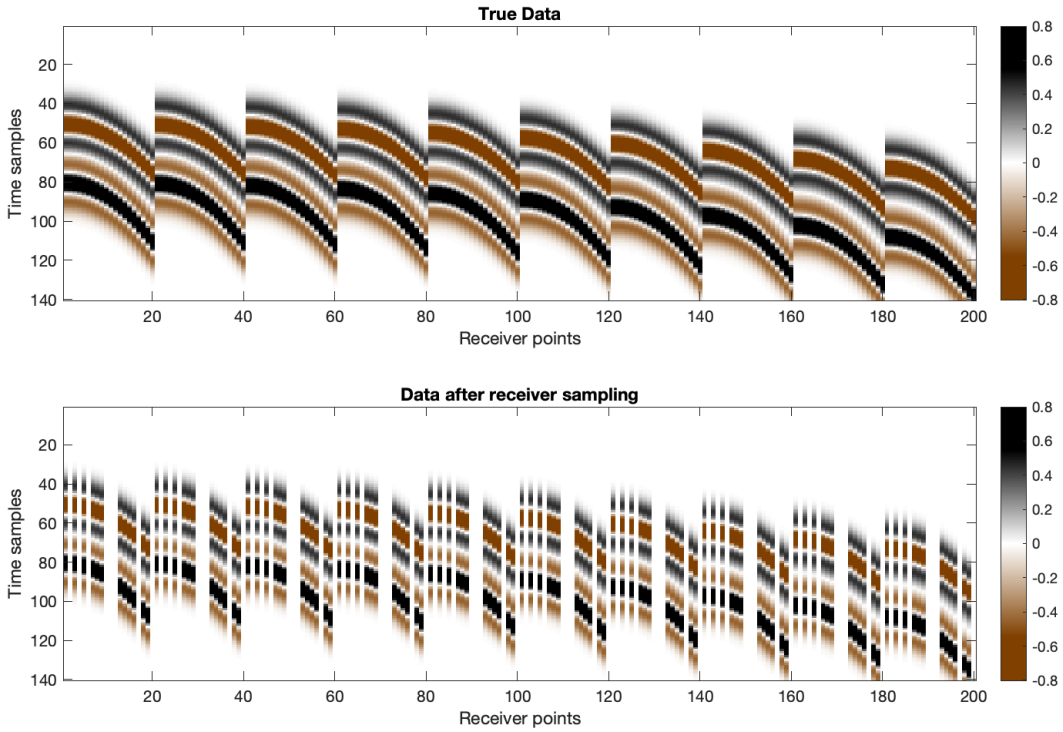

Figure A.1: Toy example showing the sampling operator acting on receivers. The upper figure shows the original data, and the lower one depicts the data after sampling.

the true data ($\mathbf{D}$) with the decimated data ($\mathbf{D}_{\boldsymbol{\nu}}$) after sampling. Clearly, for the optimal receiver problem, the sampling operator acts only on receivers.

APPENDIX B

Robust projection onto convex sets reconstruction for seismic erratic noise attenuation

Even though the POCS method is widely adopted, it cannot remove erratic noise. Appendix B proposes to de-emphasize outliers by using the $\ell 1/\ell 2$ norm to reduce the erratic noise influence. Synthetic and real data are used to examine the performance of the revised POCS method.

B.1 Traditional projection onto convex sets

Traditional POCS algorithm in the $\omega - k$ domain can be expressed as (Gao et al., 2013b):

$$\mathbf{d}^k = \mathbf{d}^{obs} + [\mathbf{1} - \mathbf{P}] \circ \mathcal{F}^T \mathbf{T} \mathcal{F} \mathbf{d}^{k-1}, \tag{B.1}$$

where $\mathbf{d}$ is data in the frequency-space domain, $\mathbf{P}$ is sampling operator. The element of $\mathbf{P} = 1$ means there is a trace and $\mathbf{P} = 0$ indicate there is no trace. $\mathbf{T}$ is a threshold operator that is applied on the spectrum of the data $\mathbf{D}$, and $\circ$ means Hadamard product.

From equation B.1, we can know the reason why erratic noise cannot be attenuated using the traditional method is that the original traces are reinserted in each iteration without any threshold or processing.

B.2 Robust projection onto convex sets

Let us assume the following model is the complete ideal vectorized multi-dimensional seismic data:

$$\mathbf{d} = \mathcal{F}^T \mathbf{x}, \tag{B.2}$$

where $\mathcal{F}$ is the Fourier transform, and $\mathbf{x}$ represents the ideal complete Fourier coefficients. Besides, orthogonality property of Fourier transform guarantee that $\mathcal{F}^T = \mathcal{F}^{-1}$.

Then, noise, including erratic noise, is the difference between the observed data and the sampled ideal data, which can be written as:

$$\mathbf{e} = \mathbf{d}^{obs} - \mathbf{P}\mathbf{d} = \mathbf{d}^{obs} - \mathbf{P}\mathcal{F}^T\mathbf{x}, \tag{B.3}$$

where $\mathbf{d}^{obs}$ is the observed data and $\mathbf{e}$ is the noise we want to attenuate. $\mathbf{P}$ as shown in equation B.4 denotes a diagonal sampling matrix whose elements are also diagonal matrix, which is represented by $\mathbf{P}_i$. Each $\mathbf{P}_i$ corresponds to one trace of the dataset, and n is the total number of traces.

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & & & \\ & \mathbf{P}_2 & & & \\ & & \ddots & & \\ & & & \mathbf{P}_{n-1} & \\ & & & & \mathbf{P}_n \end{bmatrix} \tag{B.4}$$

$\mathbf{P}_i$ is expressed as:

$$\mathbf{P}_i = \begin{cases} \mathbf{I} & \text{if trace is not empty} \\ \mathbf{0} & \text{if trace is empty} \end{cases} \tag{B.5}$$

The symbol $\mathbf{I}$ in equation B.5 means the trace has been sampled contains signal and noise, while $\mathbf{0}$ means the trace is empty. Notice that $\mathbf{P} = \mathbf{P}^T$.

The reconstructed data can be obtained by minimizing the following cost function

$$\boldsymbol{J} = \boldsymbol{J}_M + \boldsymbol{J}_R, \tag{B.6}$$

where $\boldsymbol{J}_M$ is the data misfit term and $\boldsymbol{J}_R$ is the model term.

To deal with the non-gaussian noise, we adopt the $\ell 1/\ell 2$ norm for the misfit term:

$$\boldsymbol{J}_M = \|\mathbf{e}\|_{\ell 1/\ell 2} = \sum_{i=1}^{N} \sqrt{|\mathbf{e}_i|^2 + \varepsilon^2}, \tag{B.7}$$

where N denotes the total number of the complete noise-free seismic data and ε is a small constant.

According to sparse sensing theory, we use $\ell 1$ norm for the model term to let the Fourier coefficients be sparse. Thus, a cost function expressed in equation B.8 is proposed.

$$\boldsymbol{J} = \left\|\mathbf{d}^{obs} - \mathbf{P}\mathcal{F}^T\mathbf{x}\right\|_{\ell 1/\ell 2} + \lambda \left\|\mathbf{x}\right\|_1, \tag{B.8}$$

The derivative for the Fourier coefficients is expressed in equation B.9, which is equal to zero.

$$\frac{\partial \boldsymbol{J}}{\partial \mathbf{x}} = \left(\mathbf{P}\mathcal{F}^T\right)^T \mathbf{A}\left(\mathbf{P}\mathcal{F}^T\mathbf{x} - \mathbf{d}^{obs}\right) + \lambda\mathbf{Q}\mathbf{x} = 0. \tag{B.9}$$

The symbol $\mathbf{A}$ in equation B.9 denotes the weights induced by the $\ell 1/\ell 2$ norm, which is a diagonal matrix with elements given by $\mathbf{A}_{ii} = \frac{1}{\sqrt{|\mathbf{e}_i|^2+\varepsilon^2}}$ Besides, $\mathbf{Q}$ is also a diagonal matrix with elements given by $\mathbf{Q}_{ii} = \frac{1}{|\mathbf{x}_{ii}|+\varepsilon}$.

Grouping,

$$\mathcal{F}\mathbf{P}^T\mathbf{A}\mathbf{P}\mathcal{F}^T\mathbf{x} - \mathcal{F}\mathbf{P}^T\mathbf{A}\mathbf{d}^{obs} + \lambda\mathbf{Q}\mathbf{x} = 0. \tag{B.10}$$

The second term in equation B.10 can be simplified. Because $\mathbf{A}$ and $\mathbf{P}$ are both diagonal matrices, their sequence does not change the final result. Besides, notice that $\mathbf{P}\mathbf{d}^{obs} = \mathbf{d}^{obs}$, which means the observed data stayed the same when the sampling operator was applied to it. Then, the second term in equation B.10 can be rewrite as:

$$\mathcal{F}\mathbf{P}^T\mathbf{A}\mathbf{d}^{obs} = \mathcal{F}\mathbf{A}\mathbf{P}^T\mathbf{d}^{obs} = \mathcal{F}\mathbf{A}\mathbf{P}\mathbf{d}^{obs} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs}. \tag{B.11}$$

Let $\mathcal{F}^T\mathbf{x} = \mathbf{d}$, then equation B.10 can be rewritten as:

$$\mathcal{F}\mathbf{P}^T\mathbf{A}\mathbf{P}\mathbf{d} + \lambda\mathbf{Q}\mathbf{x} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs}. \tag{B.12}$$

Grouping,

$$\lambda\mathbf{Q}\mathbf{x} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs} - \mathcal{F}\mathbf{P}^T\mathbf{A}\mathbf{P}\mathbf{d}. \tag{B.13}$$

The observed dataset can be divided into two circumstances: the traces that are not empty and the empty traces. Let's discuss these two cases:

If, in case 1, the trace is not empty, then the element in the sampling matrix corresponds to an identity matrix, which means that for these traces, the element in matrix $\mathbf{P}_i = \mathbf{I}$, then equation B.13 is as follows,

$$\lambda\mathbf{Q}\mathbf{x} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs} - \mathcal{F}\mathbf{I}^T\mathbf{A}\mathbf{I}\mathbf{d} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs} - \mathcal{F}\mathbf{A}\mathbf{d} = \mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}). \tag{B.14}$$

In this case, the Fourier coefficients are:

$$\mathbf{x} = (\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}). \tag{B.15}$$

If in case 2, the traces are empty, then the element in the sampling matrix corresponds to zeros, which means that for these empty traces, the element in matrix $\mathbf{P}_i = \mathbf{0}$, then equation B.13 is expressed as,

$$\lambda\mathbf{Q}\mathbf{x} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs} - \mathcal{F}\mathbf{0}^T\mathbf{A}\mathbf{0}\mathbf{d} = \mathcal{F}\mathbf{A}\mathbf{d}^{obs}. \tag{B.16}$$

In this case, the Fourier coefficients are

$$\mathbf{x} = (\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}\mathbf{d}^{obs}. \tag{B.17}$$

Because the whole dataset is made of all the traces, whether it is empty or not, that is to say, cases 1 and 2 together form the dataset. Therefore, the final expression for the Fourier coefficients is the sum of the two cases,

$$\mathbf{x} = \mathbf{P}(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}) + (\mathbf{I} - \mathbf{P})(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}\mathbf{d}^{obs}, \tag{B.18}$$

where $\mathbf{I}$ is an identity matrix.

Furthermore, the elements in matrix $\mathbf{A}$ and $\mathbf{Q}$ depend on the unknown $\mathbf{x}$, POCS iteratively can find the Fourier coefficients. So, if we express the equation with iteration, we can rewrite the final formula as follows:

$$\mathbf{x}_{k+1} = \mathbf{P}(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}_k) + (\mathbf{I} - \mathbf{P})(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}\mathbf{d}^{obs}, \tag{B.19}$$

where k represents the iteration number.

Equation B.19 is the iterated equation for the Fourier coefficients. We can see $\mathbf{x}$ in the $k+1$ iteration depends on the difference between the initially observed dataset and the data reconstructed from the k-th iteration and the Fourier coefficients from the previous iteration.

As long as the Fourier coefficients are available, we can easily reconstruct the ideal dataset by applying the inverse Fourier transform.

$$\mathbf{d} = \mathcal{F}^T\mathbf{x} = \mathcal{F}^T\mathbf{P}(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}) + \mathcal{F}^T(\mathbf{I} - \mathbf{P})(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}\mathbf{d}^{obs}, \tag{B.20}$$

Or:

$$\mathbf{d} = \mathcal{F}^T\mathbf{x} = \mathbf{P}\mathcal{F}^T(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}(\mathbf{d}^{obs} - \mathbf{d}) + (\mathbf{I} - \mathbf{P})\mathcal{F}^T(\lambda\mathbf{Q})^{-1}\mathcal{F}\mathbf{A}\mathbf{d}^{obs}, \tag{B.21}$$

B.3 Examples

B.3.1 Synthetic data example

In this section, the proposed method is evaluated with both synthetic data and field data examples. Comparisons between the robust solution and the conventional solution on a 5D seismic data set are made.

Figure B.2 is one slice of the 5D synthetic data contaminated with Gaussian and erratic noise. By using the standard POCS method, the signal has been recovered, but the erratic noise is still there. Figure B.3 shows the [:,:,8,2,2] slice of the synthetic data after the proposed POCS reconstruction with an iteration number of 5. It is noticeable that the reconstructed result is satisfied, and the erratic noise has been successfully attenuated.

B.3.2 Field data example

Figure B.4 shows the [:,4,:,2,2] data slice after standard POCS reconstruction, which means the 2nd, 4th, and 5th dimension is fixed. The iteration number is 100, and it is clear that even though the empty traces are successfully reconstructed, erratic noise is still there. For example, trace 20 is contaminated by erratic noise before and after applying the POCS algorithm. Figure B.5 is the same slice as Figure B.4, but using the proposed algorithm. It is noticeable that the reconstructed result is much better than Figure B.4, and the erratic noise attenuation is successful.

Figure B.6 shows the [:,:,8,2,2] slice of the data after standard POCS reconstruction. It is the same in Figure B.4 that even though the empty traces are successfully reconstructed, erratic noise is still not reduced. Figure B.7 is the same slice as Figure B.6, but using the proposed algorithm. It is clear that the reconstructed result is also much better than Figure B.6, and the erratic noise has been removed.

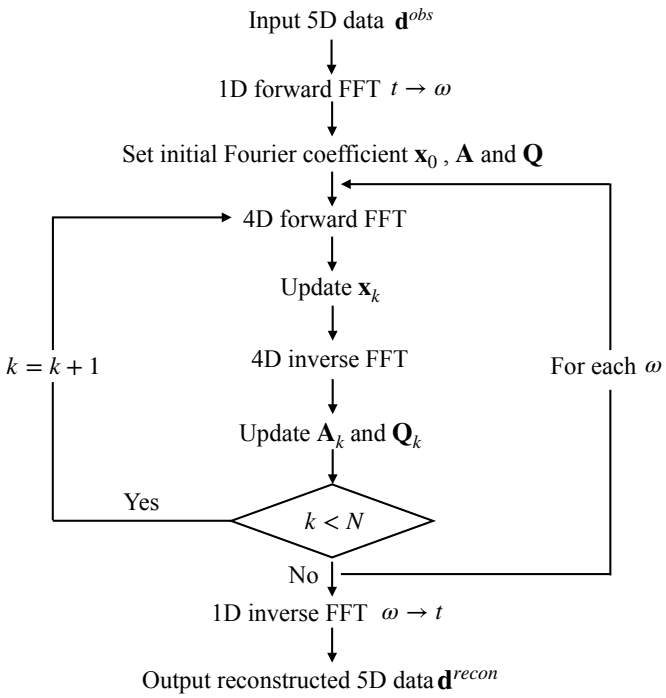

Figure B.1: The flowchart of the 5D POCS reconstruction algorithm with erratic noise sampling operator in the $\omega - nx1 - nx2 - nx3 - nx4$ domain.

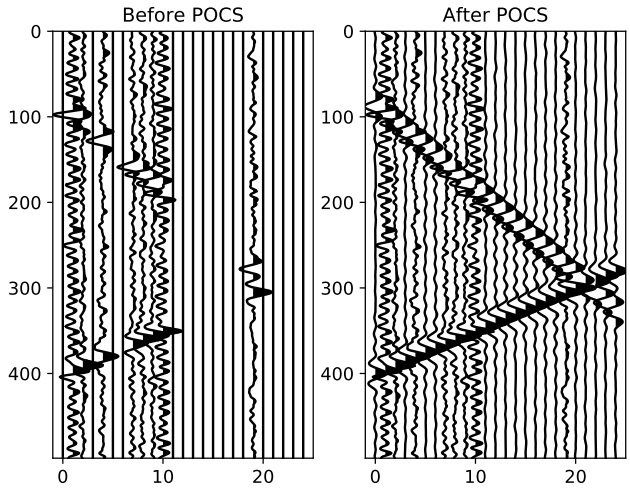

Figure B.2: Left) Original data of slice [:,:,8,2,2] with 50% decimation. Right) Reconstructed data with standard POCS.

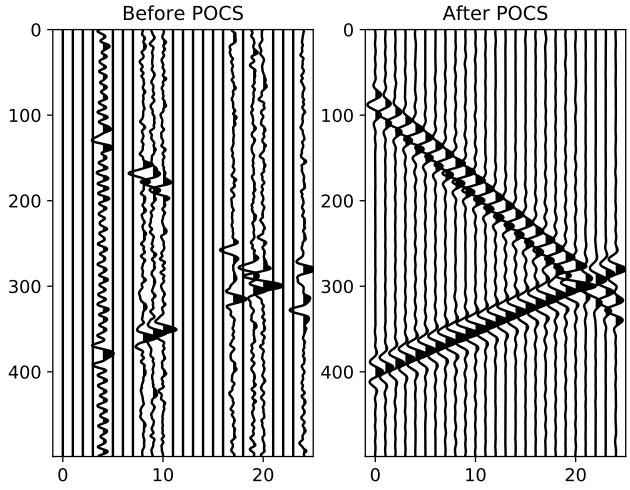

Figure B.3: Left) Original data of slice [:,:,8,2,2] with 70% decimation. Right) Reconstructed data with proposed POCS.

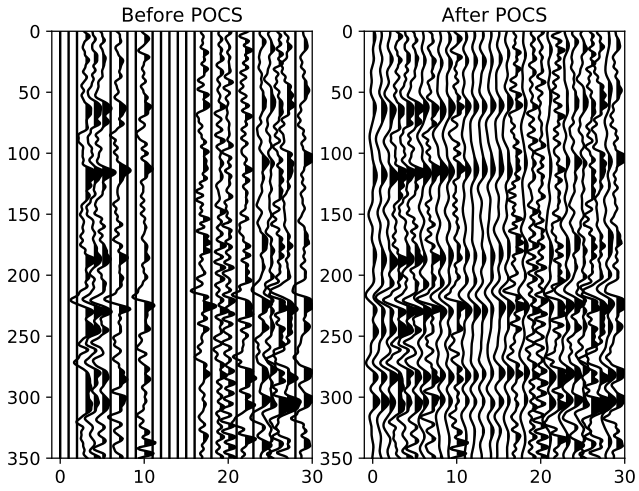

Figure B.4: Left) Original data of slice [:,4,:,2,2]. Right) Reconstructed data with standard POCS.

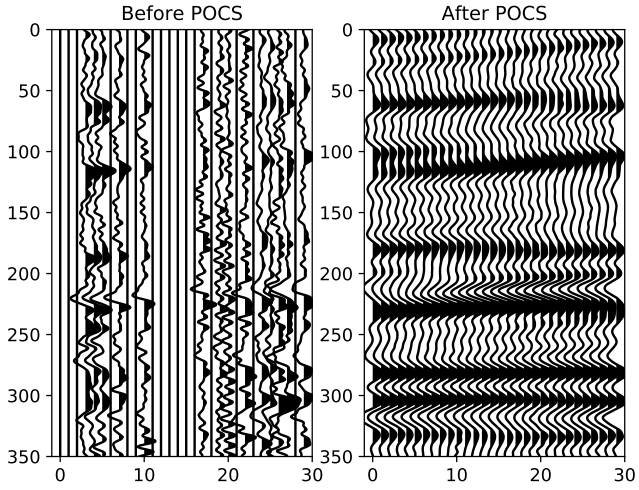

Figure B.5: Left) Original data of slice [:,4,:,2,2]. Right) Reconstructed data with proposed POCS.

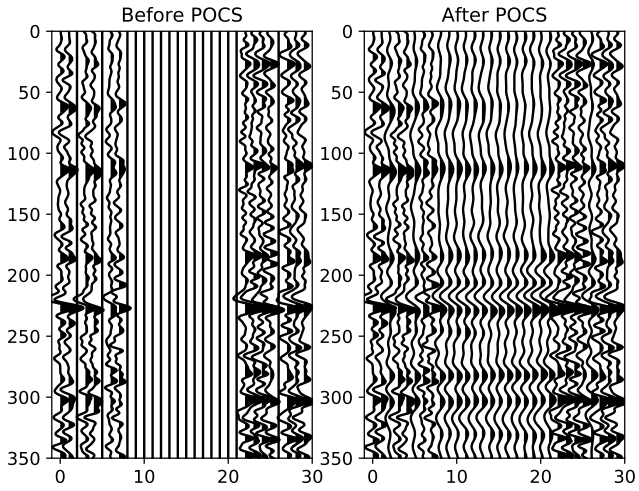

Figure B.6: Left) Original data of slice [:,:,8,2,2]. Right) Reconstructed data with standard POCS.

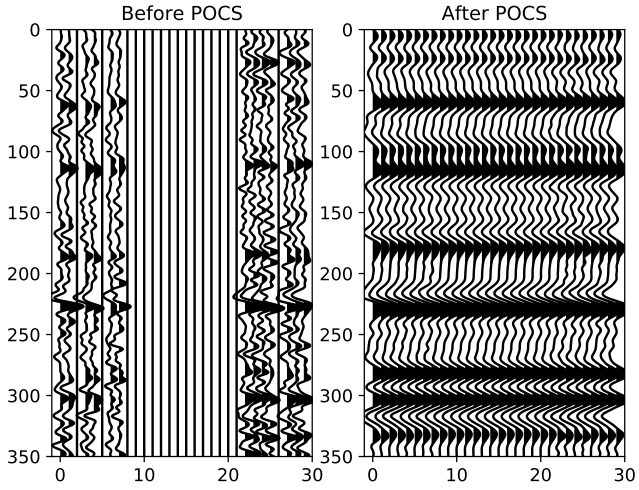

Figure B.7: Left) Original data of slice [:,:,8,2,2]. Right) Reconstructed data with proposed POCS.