Research on the Application of the Full Waveform Inversion Method for Estimating Underground Structures from Seismic Waves in Exploration

李 嘉杭

地震波動から地下構造を推定する探査における全波形インバージョン手法適用の研究

博士論文

著者

李 嘉杭

京都大学大学院工学研究科社会基盤工学専攻に提出

京都大学

Research on the Application of the Full Waveform Inversion Method for Estimating Underground Structures from Seismic Waves in Exploration

Doctoral Dissertation

by

Li Jiahang

Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University

This dissertation is submitted for the degree of

Doctor of Philosophy

KYOTO UNIVERSITY

JAPAN

Dedicated to all splendid and vile of the world.

PUBLICATIONS

during doctoral studies

As at 20-March-2024

The content of this thesis has been published or accepted in some journal papers:

国際学術誌（査読付）：

Chapter II

1. **Li, J***., Mikada, H., & Takekawa, J., 2024. Improving Full-waveform Inversion Based on Sparse Regularisation for Geophysical Data. *Journal of Geophysics and Engineering*, **21**, 810-832. DOI: 10.1093/jge/gxae036 **(13-March-2024)**

Chapter III

2. **Li, J***., Mikada, H., & Takekawa, J., 2024. Improved full-waveform inversion for seismic data in the presence of noise based on the K-support norm. *Pure and Applied Geophysics*, 1-28. DOI: 10.1007/s00024-024-03449-5 **(19-February-2024)**

Chapter IV

3. **Li, J***., Mikada, H., & Takekawa, J., 2024. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomised Singular Value Decomposition. *Journal of Geophysics and Engineering*, **21**, 572–597. DOI: 10.1093/jge/gxae015 **(21-January-2024)**

The content of this thesis has been published in some international conference abstracts:

国際学術講演会（査読付）：

Chapter II

1. **Li, J***., Mikada, H., & Takekawa, J., 2023. Improved full-waveform inversion based on sparse relaxed regularized regression processing of seismic data. In *Seventh International Conference on Engineering Geophysics, Al Ain, UAE, 16–19 October 2023,* pp. 262-266. Society of Exploration Geophysicists. DOI: 10.1190/iceg2023-059.1 **(19-October-2023)**

Chapter III

2. **Li, J***., & Xu, S., 2022. The modified full waveform inversion with the K-Support norm for the noise existing data. In *83rd EAGE Annual Conference & Exhibition*, *Madrid, Spain*, Vol. 2022, No. 1, pp. 1-5. European Association of Geoscientists & Engineers. DOI: 10.3997/2214-4609.202210322 **(06-June-2022)**

Chapter IV

3. **Li, J***., Mikada, H., & Takekawa, J., 2023. Accelerated augmented Lagrangian full-waveform inversion based on truncated randomized singular value decomposition. In *Third International Meeting for Applied Geoscience & Energy*, *Houston, United States of America*, pp. 645-649. Society of Exploration Geophysicists and American Association of Petroleum Geologists.

DOI: 10.1190/image2023-3910542.1 **(31-August-2023)**

More details about the content related to this thesis, especially in the "Chapter I"; the "List of Figures"; the "Discussions"; the "Conclusions"; the "Appendices" and the "References", can also be found in the published preprints:

プレプリント（査読無し）：

1. **Li, J***., Mikada, H., & Takekawa, J., 2023. Improving full-waveform inversion based on sparse regularisation for geophysical data. *arXiv preprint,* arXiv:2311.01688. DOI: 10.48550/arXiv.2309.13871 **(03-November-2023)**
2. **Li, J***., Mikada, H., & Takekawa, J., 2022. Improved full-waveform inversion for seismic data in the presence of noise based on the k-support norm. *arXiv preprint,* arXiv:2212.07074. DOI: 10.48550/arXiv.2212.07074 **(14-December-2022)**
3. **Li, J***., Mikada, H., & Takekawa, J., 2023. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomized Singular Value Decomposition. *arXiv preprint,* arXiv:2309.13871. DOI: 10.48550/arXiv.2309.13871 **(25-September-2023)**

この論文に関連する詳細な内容は、日本一般社団法人資源・素材学会のオープンアクセス·オンラインジャーナルである「Journal of MMIJ」に掲載されます。ご参照いただけますと幸いです（査読無し）：

1. **Li, J***., Mikada, H., & Takekawa, J. Full-waveform Inversion for the Processing of Seismic Data Based on Regularized Regression, 一般社団法人資源・素材学会 2023年度春季大会, 2023年3月15日（水）

2. **Li, J***., Mikada, H., & Takekawa, J. Data Recovery and Inversion Optimization Based on Truncated Nuclear Norm in Full-Waveform Inversion of Marine Seismic Data, 一般社団法人資源・素材学会 第19回若手研究者・学生のための研究発表会, 2022年12月9日（金）

3. **Li, J***., Mikada, H., & Takekawa, J. Full-waveform inversion with the accelerated augmented Lagrangian method in the frequency domain, 一般社団法人資源・素材学会 2022年度秋季大会, 2022年9月7日（水）

***: Corresponding Author・責任著者**

ABSTRACT

In geophysical exploration, frequency domain full waveform inversion is a high-resolution and high-precision velocity modelling method. Conventional full-waveform inversion matches the theoretical seismic records corresponding to the initial model to the actual acquired data, then adopts an optimisation strategy to solve the nonlinear problem, and then continuously updates the velocity model to obtain a high-precision velocity model that can correctly describe the velocity distribution of the subsurface medium. However, due to the inherent limitations of this technology and the problems in its application, it is still difficult to form a complete and mature technology system in industrial production.

Firstly, achieving high-precision full waveform inversion necessitates the use of low-frequency seismic data to address long-wavelength structural issues. High-quality low-frequency data can provide information on large-scale subsurface structures, which is crucial for recovering low spatial frequencies or broadband backgrounds. Therefore, higher-quality low-frequency data can offer better insights into large-scale structures, providing a stronger foundation for the iterative inversion process of more complex, small-scale structures that follow. However, due to factors such as azimuthal coverage or acquisition system strategies, the observed data collected in practical production, especially the low-frequency components, often suffer from poor quality, which significantly impacts the overall inversion accuracy of full waveform inversion, even when multi-scale strategies are employed. In frequency-domain full waveform inversion, forward modelling is achieved by solving the wave equation, specifically the Helmholtz equation, and then iteratively matching the simulated seismic data with observed data. To address inaccuracies in the propagated and back-propagated wavefields during iteration caused by noise contamination and the effects of spatial aliasing due to low-density acquisition, this thesis first introduces the use of sparse

relaxation regularised regression algorithm to optimize the source-receiver data sets obtained in frequency-domain forward modelling, thereby improving the accuracy of full waveform inversion and achieving high-resolution inversion.

Secondly, the complexity of the environmental conditions encountered during actual exploration, including both natural and anthropogenic factors, as well as the physical limitations of electronic devices and sensors, can lead to interference from random and coherent noise. This interference impacts the accuracy of full waveform inversion results, diminishing the effectiveness of the inversion. The least squares algorithm traditionally employed in full waveform inversion is based on the ℓ_2 norm regularization that minimizes the difference. However, the ℓ_2 norm is inherently sensitive to noise, especially vulnerable under low signal-to-noise ratios, making it susceptible to random noise interference. Moreover, to tackle the strongly nonlinear issues characteristic of FWI, traditional least squares methods may fall into local minima, hindering precise convergence of FWI. A common alternative for enhancing sparsity in solutions is the ℓ_1 norm, which, in contrast to the ℓ_2 norm, emphasizes the sparsity of the model rather than its smoothness, encouraging sparse solutions and thus offering better robustness and noise resistance. In light of the aforementioned challenges, this thesis explores the incorporation of a combined norm form within the minimization process of full waveform inversion, specifically the K-support regularization algorithm. The K-support norm introduced in this thesis incorporates a new regularization term in the form of a quadratic penalty, constraining the range of model parameter solutions and mitigating the nonlinearity of the inversion problem.

Thirdly, in an ideal scenario, full waveform inversion determines the direction and magnitude of model updates through gradient computation, subsequently stacking each model update on top of the previous results as the basis for the next iteration, thus completing an iterative cycle. After countless iterations and updates, the inversion results of full waveform inversion theoretically become progressively closer to the real outcomes, ultimately converging towards the true model infinitely. However, the reality is often more complex than the theory. Generally, due to limitations in the acquisition system and environmental factors, obtaining high-resolution inversion results with full waveform inversion becomes highly challenging, especially in intricate exploration settings. A notably challenging exploration scenario is the high-precision inversion of salt domes and salt marshes in the Gulf of Mexico, a region renowned for its extensive distribution of thick salt layers. However, the salt dome and salt canopy structures in the Gulf of Mexico are complex, with these salt layers exhibiting high-velocity characteristics that contrast with the surrounding rock velocities, creating significant velocity anomalies. To address the aforementioned issues, this thesis innovatively proposes a new concept, integrating randomized singular value decomposition with weighted truncated nuclear norm regularization and an inexact augmented lagrangian method to optimize full waveform inversion.

To validate the effectiveness of the aforementioned methods, this thesis employed three synthetic models for testing the algorithm, including the central part of the 2004 BP, Marmousi II and the two-dimensional SEG/EAGE salt and overthrust models. Several numerical simulation experiments demonstrated that the enhanced full waveform inversion algorithm, based on the methods proposed in this thesis, exhibits remarkable stability and efficacy in addressing the aforementioned challenges in the industry.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Chapter IV

CHAPTER I: INTRODUCTION FOR FULL WAVEFORM INVERSION

1.1 Research Background

Applied geophysical exploration has evolved from its early exploratory endeavours to its growing maturity today, traversing a rich trajectory of development (Virieux., *et al*., 2014). Geophysical researchers primarily employ a myriad of methods, such as seismic, gravimetric, and magnetic surveys, to delve into subsurface structures (Plessix *et al*., 2010). Among them, seismic exploration, in particular, is extensively employed in the search for petroleum and natural gas, providing a robust technological pillar for the global energy and industrial sectors (Aghamiry *et al*., 2021). With technological advancements, particularly in the fields of geophysics data acquisition and processing, the utilization of multi-channel geophysics methods has markedly improved the quality of data, and this advancement allows researchers to delve deeper into the understanding of underground structures (Gholami *et al*., 2022). Further, advancements in technology and equipment, especially the proliferation of computational techniques, catalyzed the progress of geophysical exploration methodologies (Zhou *et al*., 2023). High-performance computer and numerical simulation techniques have made it feasible to present more intricate and authentic representations of the subsurface, which offers a more precise theoretical foundation and technological backing for the exploration of oil, natural gas, and mineral resources (Aghamiry *et al*., 2022).

Within this context, Full Waveform Inversion (FWI) emerged, anchored in seismic wave propagation theories, and harnesses the complete waveform information of seismic waves as they travel through subsurface mediums to extrapolate underground parameters (Virieux and Operto, 2009). Initially constrained by computational capabilities, FWI was primarily relegated to modelling and inversion of small and simplistic subsurface models. However, as technological prowess burgeoned, especially with the advent of supercomputers, FWI has witnessed renewed interest from both academia and industry, especially given its demonstrated potential in interpreting intricate subterranean structures (Warner and Guasch, 2016). Venturing into the 21st century, with the ubiquity and application of high-performance computing, FWI has found its niche in larger, more convoluted subsurface contexts. Through FWI, researchers can not only pinpoint parameters such as velocity and density of subsurface constructs with enhanced precision, but they can also offer more accurate insights into the location and scale of oil and gas reservoirs. However, since the conventional algorithm itself is outdated and the real exploration environment has become more and more harsh and complex, the conditions under which FWI can be applied in practice are still strict. Therefore, how to improve the performance of the algorithm itself, reduce the constraints and improve its robustness and resolution, to enable it to have a better performance in the field, is the main direction of the current study by the researchers.

1.2 Current Challenges for FWI in Practical

When FWI is applied in practice, three potential issues may arise. Firstly, due to the limitations of azimuth and acquisition systems, the quality of forward-modelled low-frequency data is often compromised, consequently diminishing the overall performance of FWI (Wu and McMechan, 2021). Secondly, conventional least-squares algorithms tend to have inadequate noise suppression capabilities against random high background noise. They are also prone to getting trapped in local minima, resulting in slower and less accurate convergence rates (Ovcharenko *et al.*, 2018). Lastly, when it comes to the identification and localization of geological boundaries with high-velocity contrasts like salt bodies in the Gulf of Mexico, the challenges are magnified, the presence of strong reflection interfaces and the shadowing effects of high-velocity bodies make the inversion of deep structures a challenging task, limiting the successful application of FWI (Alkhalifah *et al.*, 2021).

1.2.1 Low-Frequency Data Reconstruction

Over the past decade, seismic signal acquisition systems and processes have seen significant improvements. However, conventional methods still struggle to acquire low-frequency seismic data under complex conditions in a noisy environment, primarily because seismic vibrators have difficulties transmitting adequate low-frequency energy to deeper regions (Liu and Fomel, 2011). Additionally, system and environmental noise, along with azimuthal constraints, further degrade the quality of the low-frequency components (Brossier *et al*., 2009). The absence of a sufficient low-frequency seismic dataset can lead to inaccurate large-scale features in FWI inversions, resulting in convergence to local minima instead of global ones. To circumvent this, traditional methods, such as low-frequency data interpolation or extrapolation, have been proposed. Wu *et al*. (2014) treated seismic recordings as modulated signals, using modulation operators to extract low-frequency envelopes and subsequently recover the low-frequency information. Meanwhile, Wang and Herrmann (2016) framed low-frequency extrapolation as a convex optimization problem, employing total variation (TV) regularization to reconstruct the spatial continuity of various shot records. Additionally, Li and Demanet (2017) utilized bandwidth extrapolation techniques, extending the wavenumber extrapolation of images followed by the extended Born modelling to extrapolate the data's frequency bandwidth. However, achieving synchronized reconstruction of high and low-frequency signals and compensating for the spatial discontinuities introduced by low-density acquisition systems remains a challenging endeavour. To overcome this challenge, a novel regression algorithm is proposed to reconstruct seismic data, especially the low-frequency signals, by leveraging optimization algorithms in conjunction with compound regularization to interpolate and reconstruct the low-frequency signals.

1.2.2 Regularization

Regularization algorithms are crucial steps in FWI. Initially, FWI, being a simulation algorithm, only employed the conventional least-squares method based on the ℓ_2 norm as its misfit function. However, its downsides are evident. Firstly, because the inversion problem is highly nonlinear, an unconstrained form of misfit function can easily lead to convergence to local minima rather than the global minimum, resulting in inaccurate convergence or slow convergence rates. Additionally, the noise-sensitive nature of the ℓ_2 norm means that the conventional algorithm has poor noise resistance (Zhong and Zhang, 2013). Therefore, regularization algorithms were introduced to address these issues.

Tikhonov regularization, a representative regularization algorithm, is widely applied for solving inverse problems, especially when they are ill-posed or underdetermined. In FWI, Tikhonov regularization is implemented by adding a regularization term, which is the ℓ_2 norm of the model parameters (like velocity or density). This is then modulated by a regularization parameter to adjust the penalty strength. Introducing this new regularization term can narrow the solution space range, helping FWI avoid falling into local minima. Meanwhile, the convex nature of the ℓ_2 norm further reduces instability during optimization and provides improved robustness while offering some noise resistance (Choi and Alkhalifah, 2012). However, for low signal-to-noise ratio complex conditions, the capabilities of Tikhonov regularization are limited. A more aggressive approach is to incorporate the TV, which is grounded on the ℓ_1 norm, into the misfit function (Schmidt, 2005). Compared to Tikhonov regularization, TV regularization, with its ℓ_1 norm-based terms can provide a stronger sparsity, leading to superior denoising capabilities. However, excessive regularization can sometimes result in underfitting and resolution difficulties (Boyd and Vandenberghe, 2004).

A better solution employs the Lagrangian method, which ensures denoising capability and provides faster convergence rates. The misfit function based on the Lagrangian method can not only tighten the solution space but also be efficiently computed when combined with the alternating direction method of multipliers (ADMM) by introducing convex terms. As a more robust solution, especially when the noise in the data follows a Gaussian distribution, this method offers a sturdier solution strategy. However, even though the Lagrangian method has significantly reduced time complexity, it still tends toward smoother model solutions. Particularly for geophysical inversion problems, the continuity and abrupt changes in some model structures are challenging to resolve (Aghazade *et al.*, 2022). Therefore, a new regularization algorithm combining ℓ_1 and ℓ_2 norms, accompanied by a misfit function solution algorithm with a lower time complexity, is eagerly awaited.

1.2.3 Deep Salt Pillars Boundary Identification

The inversion of complex underground structures, especially the high-precision inversion of geological boundaries with high velocity contrasts like the salt pillars and salt diapirs in the Gulf of Mexico, presents a formidable challenge. It's significant to note that the Gulf of Mexico basin has a wide distribution of salt formations with considerable thickness. According to predictions by O'Brien and Lerche, the temperature decreases by 10.5°C below a salt bed of 1000m thickness, and by 15.8°C beneath a salt bed of 1500m thickness. Consequently, the presence of these salt layers slows down the maturation process of the hydrocarbon source rocks beneath them (O'Brien and Lerche, 1987). Moreover, the high thermal conductivity of salt acts as a heat dissipater, elevating the temperatures of strata above the salt, thereby accelerating the maturation process of hydrocarbon source rocks atop the salt layers. For the deeper primary hydrocarbon source rocks in the Gulf of Mexico, a later hydrocarbon peak period is beneficial for the preservation of oil and gas reservoirs (Billette and Brandsberg-Dahl, 2005). Hence, the high-precision inversion of salt interfaces and boundaries in this region is vital for determining reservoir locations and pressure deformation characteristics.

However, the structure of salt pillars and diapirs in the Gulf of Mexico is intricate. These salt formations are characterized by high velocities, creating a pronounced velocity contrast with the surrounding rocks. This distinction complicates the propagation of seismic waves, frequently distorting their paths and causing multiple reflections and refractions, making FWI's inversion particularly challenging in deeper sections. The high-velocity nature of the upper salt interface results in a strong reflective layer. This reflective interface has a substantial impedance difference, characterized by high P-wave impedance. Seismically, a strong reflective layer often appears as strong amplitude, highly continuous, and relatively low-frequency features, thereby masking reflections from structures beneath the salt (Rabalais *et al*., 2001). Moreover, the unique crystalline nature of the salt leads to strong refraction of acoustic waves, creating a "lens" effect. This "lens" effect is relatively pronounced, especially on the irregular shapes typically found at the edges or flanks of salt formations, making boundary reflections less clear than those at the salt top interface (Turner and Rabalais, 2019). All these challenges heighten the difficulty for FWI in determining the deep salt base structures.

1.3 Objectives of the Thesis

1.3.1 Implementing Regression Regularization Algorithms for Forward Information Completion and Recovery

The low-frequency wavefield plays a crucial role in FWI, as its quality can directly assist FWI in avoiding being trapped in local minima and provide superior large-scale information. However, in practical scenarios, the quality of low-frequency data obtained from exploration is often poor (Zhang, 2010). Therefore, there's a need to employ a more updated algorithm and theory to enhance the inversion accuracy from the algorithmic perspective, rather than relying on hardware upgrades. Thus, to address the issue of poor quality of high-frequency and low-frequency forward wavefields in FWI, especially when the low-frequency information is disturbed by noise interference and spatial discontinuities due to low-density acquisition, I propose optimizing the three-dimensional wavefield data obtained from frequency-domain forward modelling in FWI using sparse relaxed regularized regression algorithms. This enhanced algorithm introduces regularization terms and sparse constraints, which have interpolation effects, thus addressing the aforementioned issues of high-frequency aliasing effects and low-frequency discontinuities. In the numerical fitting section, I optimized the quality of low-quality three-dimensional seismic wavefields, simulating the issues with poor low-frequency data quality encountered by FWI in practical scenarios. By adopting the sparse relaxed regular regression algorithm, I introduced a novel approach for enhancing wavefield data and conducted a comparison between its inversion results and those from conventional algorithms. The comparison served to validate the feasibility of the sparse relaxed regularized regression (SR3) algorithm in addressing such issues, thereby solving the problem of suboptimal FWI inversion results caused by poor data quality.

1.3.2 Developing a New Regularization Algorithm for Seismic Exploration

FWI as an inversion algorithm based on numerical simulation is such that the construction of the misfit function directly impacts the accuracy and convergence rate of the inversion and minimization process. Presently, conventional FWI algorithms still employ the misfit function of subpar performance. These traditional computational methods lead to a high degree of nonlinearity and multiple solutions. Especially under complex conditions, such as low signal-to-noise ratios, many standard algorithms tend to yield smooth model solutions. This excessive smoothening might result in a loss of model details, particularly for intricate underground structures (Zhong and Zhang, 2013). Therefore, a new regularization algorithm is introduced in this thesis as a substitute for the conventional least squares method. Specifically, by combining the sparsity of the ℓ_1 norm and the convexity of the ℓ_2 norm, a novel regularization algorithm constructed in the form of the quadratic penalty method is proposed. The new algorithm preserves the convexity of the ℓ_2 norm while introducing certain sparse properties, endowing the algorithm with enhanced robustness and improved noise removal capabilities. Furthermore, to enhance the rate of convergence, the novel method is augmented with the ADMM, countering the sluggish convergence traits inherent in traditional approaches. In the experimental section, the effectiveness of the newly proposed algorithm is tested through multiple sets of synthetic data simulations, mirroring real complex scenarios.

1.3.3 Utilizing Singular Value Decomposition for Multi-Scale Optimization in the Inversion Process

Beyond drilling techniques, geophysical exploration, especially FWI, stands as the most effective, economical, and primary means of identifying salt layers. Notably, the Gulf of Mexico contains vast deposits of saline gypsum rock layers, which have undergone four stages of evolutionary construction, forming salt bodies of diverse morphologies. The saline gypsum layers in this region not only serve as excellent source rocks for hydrocarbons but also act as effective reservoirs and cap rocks. Therefore, utilizing FWI to delineate and describe the salt bodies and reservoir layers of this region is of paramount importance (Chi *et al*., 2014). To address the shortcomings of conventional FWI, such as inferior imaging resolution, especially in deep zones, and difficulties in determining salt body boundaries, there is an urgent need to employ a more advanced algorithm to enhance FWI's capability to recognize high-velocity anomalies. By integrating a novel image processing technique, the inversion process of FWI is optimized in this thesis. New technology strengthens FWI's capability to describe the primary features of models and its inversion ability for high-velocity body boundaries, thus enhancing FWI's recognition and determination capacity for deep high-velocity anomalies, akin to those in the Gulf of Mexico region.

1.4 Agenda

This thesis aims to elucidate a high-precision multi-scale robust FWI optimization. It updates the FWI from multiple perspectives and processes to achieve higher computational efficiency and inversion resolution. Specifically, Chapter II will introduce a regression regularization algorithm to optimize both low and high-frequency seismic data. This will be applied to a three-dimensional seismic data cube and subsequently incorporated into multi-scale FWI. The algorithm will then be verified in multiple homogeneous media and artificial syntheses. The preprocessing used in Chapter II acts on the forward phase of FWI, laying the groundwork for subsequent inversion processes.

In Chapter III, a novel sparse regularization algorithm will be introduced. Unlike conventional least-squares or Tikhonov regularization, this chapter uniquely employs a sparse regularization algorithm that integrates both the ℓ_1 and ℓ_2 norms to enhance the algorithm's robustness and noise resistance during the optimization process. This chapter will focus on the theoretical basis and algorithmic flow of the new sparse regularization method, applying it to many different types of synthetic data and various low signal-to-noise ratio conditions for testing.

Lastly, Chapter IV will emphasize the initial application of an image processing algorithm in FWI to improve inversion resolution and conduct principal feature analysis. In this chapter, I will substantiate that by adopting the algorithms proposed in this thesis, it is feasible to effectively identify the high-precision contours and boundaries of deep salt domes in salt rock environments.

1.5 Graphical Illustrations of the Highlights

By enhancing the performance of the FWI algorithm itself and altering its logic, the optimized full FWI algorithm process will also be provided. The full flowchart of the improved FWI is shown in Figure 1.1.

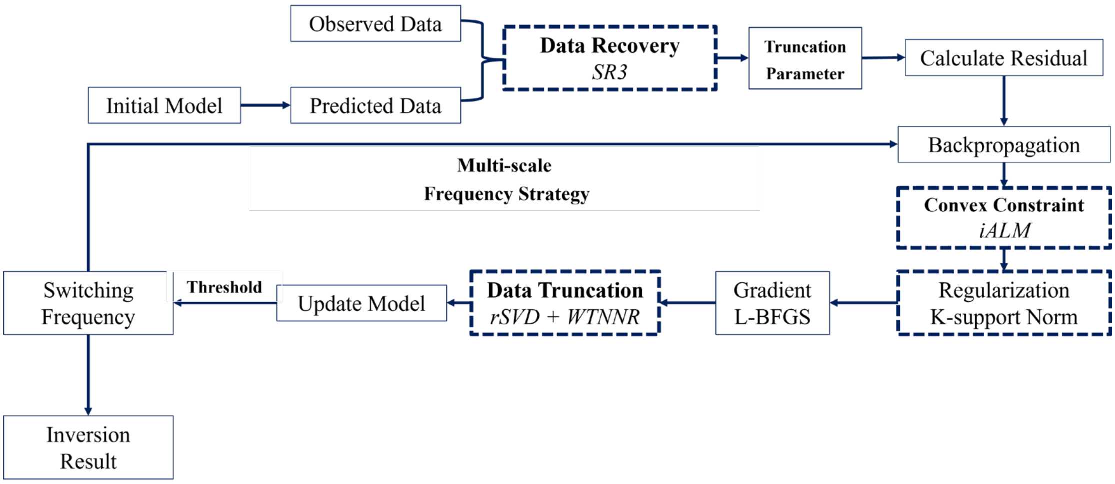

Figure 1.1. Full flowchart of improved FWI proposed in this thesis.

In Figure 1.1, the blue solid-lined box represents the conventional FWI workflow. It begins with observational data and forward modelling, proceeds by calculating the gradient through least squares, coordinates the step length to compute the gradient, and completes one iteration by updating the model. Within the internal loop, single-frequency updates are performed through a finite number of iterations before returning to the external loop to switch frequencies. The process exits the loop when the final threshold is met.

This thesis optimizes and improves the conventional FWI at multiple stages. Firstly, the sparse relaxed regularized regression algorithm is incorporated into the forward modelling process to enhance its effectiveness by denoising and optimizing low-frequency data. Subsequently, a truncation parameter is set in preparation for the subsequent random singular value decomposition. Next, the least squares objective function is optimized based on Lagrangian-type algorithms to achieve convex optimization for the inversion problem. Additionally, a novel K-support regularization algorithm is introduced to achieve denoising and enhance the convexity of the objective function, thereby avoiding local minima and cycle skipping.

During the optimization process, a combined random singular value decomposition and weighted truncated nuclear norm regularization algorithm is used. The combined algorithm truncates the singular value matrix based on its condition number to reduce the rank of the singular value matrix, thereby achieving background separation and enhancing inversion resolution. Finally, after exiting the internal loop, the condition number is switched based on a multi-scale truncation parameter strategy to prevent the truncation parameter from exceeding the rank of the matrix.

Through the above innovations, the proposed improved algorithms are tested on various synthetic datasets. The test results indicate that the proposed improvements can effectively enhance the denoising capability and inversion accuracy of FWI in multiple aspects, providing effective improvement strategies and theoretical foundations for the industrialization of FWI.

CHAPTER II: FULL WAVEFORM INVERSION BASED ON SPARSE RELAXATION REGULARIZED REGRESSION

The content of this chapter has been published in a journal paper and a preprint: *The Journal of Geophysics and Engineering (JGE 2024) and arXiv (2023)*

1. **Li, J***., Mikada, H., & Takekawa, J., 2024. Improving Full-waveform Inversion Based on Sparse Regularisation for Geophysical Data. *Journal of Geophysics and Engineering*, **21**, 810-832. DOI: 10.1093/jge/gxae036 **(13-March-2024)**

2. **Li, J***., Mikada, H., & Takekawa, J., 2023. Improving full-waveform inversion based on sparse regularisation for geophysical data. *arXiv preprint,* arXiv:2311.01688. DOI: 10.48550/arXiv.2309.13871 **(03-November-2023)**

The content of this chapter has been published in a conference abstract: In *Seventh International Conference on Engineering Geophysics.* Society of Exploration Geophysicists. *(ICEG 2023)*

3. **Li, J***., Mikada, H., & Takekawa, J., 2023. Improved full-waveform inversion based on sparse relaxed regularized regression processing of seismic data. In *Seventh International Conference on Engineering Geophysics, Al Ain, UAE, 16–19 October 2023,* pp. 262-266. Society of Exploration Geophysicists. DOI: 10.1190/iceg2023-059.1 **(19-October-2023)**

2.1 Introduction

FWI is recognized for its ability to generate high-definition models of subsurface seismic velocities, making it indispensable in geophysical research, seismic analysis, and the characterization of oil and natural gas deposits. In the frequency-domain FWI, the forward simulation is achieved by calculating the solutions to the wavefield partial differential equation, namely the Helmholtz equation, followed by an iterative match of simulated seismic data with observed data to address the inversion problem and extract the velocity model (Virieux and Operto, 2009). A major advantage of using the Helmholtz equation is that it transforms a second-order partial differential equation into a linear equation, thus simplifying the solution process. The solution matrix of the Helmholtz equation outlines the wavefield relationship between sources and receivers at specific frequencies. Particularly in a three-dimensional view, it provides a multi-scale depiction from low to high frequencies, serving as a useful tool for visualizing wavefield information at various frequencies (Warner and Guasch, 2016). Frequency domain forward algorithms, by calculating the solution to the Helmholtz equation, depict the propagation of seismic waves in the frequency domain. However, FWI faces numerous challenges in practical application, including azimuth, system constraints, and imaging artefacts, especially when forward or collected data is noisy, leading to diminished imaging accuracy (Zhang, 2010). In real-case, noise might contaminate this matrix, impairing the forward wavefield and the consequent imaging and inversion quality. Moreover, in situations where high-density spatial sampling is challenging, the impact of spatial aliasing must be considered. Spatial aliasing arises when the sampling process does not adhere to the Nyquist-Shannon sampling theorem, leading to potential misinterpretation of high-frequency wave components as low-frequency components, which degrades low-frequency information, affecting the accurate recovery of underground structures. Thus this phenomenon most manifestly results in high-frequency artefacts (Wu and McMechan, 2021). Traditional methods of increasing the sampling rate increase data volume and potentially raise computational expenses. Furthermore, FWI encounters numerous problems with low-density sampling, which implies sparse seismic data collection. It's widely acknowledged that low-frequency information is crucial for FWI, assisting in overcoming nonlinearity and facilitating the inversion process to converge to a global optimum rather than local optima (Chen, 2012). Moreover, low-density sampling can cause limited offset distance problems, with insufficient receiver coverage, and a decline in the inversion resolution of subsurface models. These constraints make the FWI updating process heavily reliant on the initial model, diminishing the accuracy of inversion results and affecting FWI's interpretation and recognition of subsurface structures.

Given the aforementioned challenges, this thesis primarily proposes the use of the SR3 algorithm in FWI to optimize the source-receiver data sets obtained in frequency-domain forward modelling, to enhance FWI's precision and achieve high-resolution inversion. The SR3 algorithm, serving as an optimization method, augments the basic regression problem into a new regression problem with additional constraints through an auxiliary matrix (Erichson *et al*., 2020). The main difference between SR3 and standard regression problems is its shrinkage of the loss function's functional space, which can compress the nonlinear wavefield loss space, thereby narrowing the range of singular values (Champion *et al*., 2020). As a result, iterations in regression based on the SR3 algorithm are faster than traditional regression methods and also promote sparse solutions. Specifically, to address spatial aliasing effects which may arise in high-frequency data, a more appropriate image processing technique involves matrix reconstruction and interpolation methods. As a regularization method, SR3 introduces regularization terms and sparse constraints, resulting in inherent interpolation effects. SR3 can mitigate the impact of aliasing effects, while sparsity constraints help to reduce solution space complexity, thus enhancing the accuracy of matrix reconstruction and interpolation. Furthermore, handling low-frequency data and data completion and interpolation of down-sampled data are crucial for FWI. The SR3 algorithm employs multi-regularization methods, regularizing the sparse basis of simulated data through regression and iterative processes and restoring and completing data in the optimization problem's solution via swift iterations. Moreover, by selecting appropriate regularization parameters, this algorithm achieves a balance between sparse constraints and solution smoothness, providing more stable denoising and data completion effects.

This chapter innovatively incorporates the SR3 algorithm into frequency-domain FWI. This incorporation allows FWI to supplement and reconstruct data for low-density spatial sampling, counteract spatial aliasing effects for high-frequency data, and enhance the resolution for noisy datasets. The stability and adaptability of the proposed methods are validated by comparing Tikhonov regularized FWI with SR3-based FWI using two homogeneous medium models and two complex models.

2.2 Frequency-Domain Multiscale Forward Algorithm

In the FWI technology, frequency-domain forward modelling has received widespread attention and research due to its outstanding accuracy, efficient computational speed, and insensitivity to overly idealized time sampling rates, as well as greater flexibility in practical applications. A key part of this algorithm typically involves solving the second-order acoustic Helmholtz equation (Jo *et al.*, 1996). By applying LU decomposition to the impedance matrix, the computational efficiency of the wavefield for compound sources can be enhanced through forward and inverse substitutions:

$$(\Delta + \mathbf{W}^2\mathbf{M})\mathbf{U} = -\mathrm{F}(\mathbf{W})\,\mathrm{G}(\mathbf{X}-\mathbf{X}_\mathrm{S})\,\mathrm{g}(\mathbf{Z}-\mathbf{Z}_\mathrm{S}), \tag{2.1}$$

where $\mathbf{M}$ represents squared slowness, and $\mathbf{W}$ represents angular frequency, $\mathrm{F}(\mathbf{W})$ denotes source signature, $\mathbf{U}$ denotes wavefield, $\mathrm{G}(\mathbf{X}-\mathbf{X}_{\mathrm{S}})$ and $\mathrm{g}(\mathbf{Z}-\mathbf{Z}_{\mathrm{S}})$ denote source term at points $\mathbf{X}_{\mathrm{S}}$ and $\mathbf{Z}_{\mathrm{S}}$, which are in the two directions for two demesional modelling. In this thesis, all of the bold upright letters denote vectors, upright letters denote functions, and italic letters denote variables.

The standard frequency-domain forward algorithm transforms the model domain into the data domain, promoting a closer alignment between initial and actual models. In the data domain, the least-squares method computes the difference between the observed and actual data to improve their agreement. So, the misfit difference obtained during the inversion process reflects the perturbation of the model, indicating the degree of updating from the current model to a more optimal one. By calculating its derivative value, the direction of model updating can be determined, guiding the next step of updating, and this entire process is collectively referred to as the optimization process. FWI involves numerous updating iterations, involving the exploration of more optimized models based on the current model and gradient direction (Alkhalifah *et al.*, 2021). The goal is for the inversion result to approach the true model more and more closely until the termination criteria are met:

$$\mathbf{M} = \arg\min_{\mathbf{M}\in\mathbf{M}_{\mathrm{A}}} \left\| \mathbf{D}_{\mathrm{ca}}(\mathbf{M}) - \mathbf{D}_{\mathrm{obs}} \right\|, \qquad \textbf{(2.2)}$$

where $\mathbf{D}_{\mathrm{ca}}(\mathbf{M})$ is the calculated data and $\mathbf{D}_{\mathrm{obs}}$ is the observed data.

2.3 Sparse Relaxed Regular Regression: SR3

In practical scenarios, especially during the minimization process of FWI, the precision of the forward-modelled wavefield frequently influences the accuracy of the inversion.

In cases where the quality of the numerical wavefield we adopt is low, or the wavefield exhibits spatial discontinuities due to sparse sampling, data loss and noise can cause the inversion results to deviate from the true values. Even if more advanced inversion algorithms are used, the efficacy and accuracy of FWI can still be significantly impacted. Thus, high-quality forward-simulated wavefields are critical for FWI and play a pivotal role. In light of this, a preprocessing algorithm to optimize the wavefield is employed. The quality of the wavefield matrix and spatial continuity is enhanced by completing and denoising the wavefield data through auxiliary matrices and regression regularization algorithms. This lays a stronger foundation for subsequent inversions.

The goal of seismic wavefield interpolation is to estimate and fill in the wavefield data in unknown areas based on known seismic wave data. This is typically a high-dimensional and sparse data problem, as actual seismic observations are limited and heavily affected by noise. Therefore, an algorithm capable of handling sparse data and robust to noise is required. The SR3 algorithm decomposes the original problem into two subproblems: one for data fitting and the other for regularization. By introducing relaxed variables and relaxation parameters, the original seismic wavefield interpolation problem is relaxed into a new optimization problem, which is solved by alternately optimizing the relaxed variables and the wavefield. For example, seismic wavefield data are often sparse, meaning that there are few observation points. The SR3 algorithm, through regularization, allows the interpolation results to maintain sparsity while effectively filling in the missing data. In regions where seismic wavefield data may not be observed, the SR3 algorithm can infer the wavefield in these areas using the available sparse data. Additionally, seismic data are often accompanied by noise, and direct interpolation may be severely affected by it. By incorporating regularization terms into the optimization process, the SR3 algorithm effectively filters out noise, resulting in smoother and more accurate interpolation outcomes.

Regularization regression, a classic algorithm commonly used in the industry and various other disciplines, plays a pivotal role in areas such as compressed sensing and data completion. As a form of sparse regression, this algorithm helps us select more influential features from a large dataset while discarding redundant ones. In this way, it reduces the complexity of high-dimensional data and computational space complexity (Erichson *et al.*, 2020). Additionally, for datasets of poor quality, the algorithm can assist in predicting missing data, completing spatial discontinuities in the data through iterative processes, and achieving dataset completion. Given these strengths, the algorithm plays an extremely important role in multiple domains. Therefore, it is embedded into FWI as a preprocessing technique, aiming to three-dimensionally optimize the source-receiver data matrix obtained from frequency domain forward simulations (Zheng *et al.*, 2018):

$$\min_{\mathbf{U}} \frac{1}{2}\|\mathbf{AU}-\mathbf{D}\|_{\mathrm{F}}^{2}+\lambda\,\mathrm{R}(\mathbf{SU}), \tag{2.3}$$

where $\mathbf{A}\in\mathbb{R}^{m\times d}$ is the data-generating model for the real data $\mathbf{D}$, $\mathbf{U}\in\mathbb{R}^{d}$ is the recovery data, λ is the regularization parameter, $\mathrm{R}(\bullet)$ is any regularization form and $\mathbf{S}\in\mathbb{R}^{n\times d}$ is a linear map. $\|\ \|_{\mathrm{F}}^{2}$ is the Euclidean norm, m,n,d are the dimensions of the row and column of a matrix space.

I innovatively apply a non-smooth regularizer to FWI intending to enhance the quality of forward simulations cost-effectively affected by sparse sampling, while also achieving denoising of the signal. Compared to other algorithms, the proposed SR3 offers several advantages. Firstly, with the introduction of the auxiliary matrix, this algorithm can effectively identify sparse signals, especially for large matrices with noise. This auxiliary matrix suppresses overfitting and aids in denoising. At the same time, by enhancing the convexity of the regression problem, it accelerates the convergence speed of the recovery and optimization processes.

Furthermore, the proposed algorithm possesses enhanced flexibility and robustness. By adjusting the form of the norm in the auxiliary matrix and the regularization term itself, I can flexibly apply the algorithm to various complex scenarios based on the degree of parameter complexity in different models. This allows us to set different conditions and levels of relaxation for the constraint functions, thereby achieving optimal results under conditions that satisfy fidelity. To this end, I need to construct an auxiliary variable to perform convex relaxation on equation 2.3 (Zheng *et al.*, 2018):

$$\min_{\mathbf{U},\mathbf{W}} \frac{1}{2}\|\mathbf{AU}-\mathbf{D}\|_{\mathrm{F}}^{2} + \lambda\,\mathrm{R}(\mathbf{W}) + \frac{k}{2}\|\mathbf{W}-\mathbf{SU}\|_{\mathrm{P}}, \tag{2.4}$$

where $\mathbf{W}^{k} = \mathrm{prox}_{\lambda k \mathrm{R}}(\mathbf{U}^{k})$ is the auxiliary variable, which gradually approaches with $\mathbf{U}$; $\mathrm{prox}_{\lambda k \mathrm{R}}$ is the proximity operator (prox) for R. λ is the penalty parameter, and k is the relaxation parameter, where k controls the degree of relaxation. The $\|\bullet\|_{\mathrm{P}}$ represents a different form of the regularization used in the optimization, which can be flexibly used as different regularization functions depending on the sparsity of the data matrix, such as ℓ_1 and ℓ_2 norms, even nuclear norms (Champion *et al.*, 2020). In Figure 2.1, I give two different regularizations of unit-ball, where the ℓ_1 norm of Figure 2.1 (a) is emphasized with better denoising ability, while the ℓ_2 norm of Figure 2.1 (b) is convex.

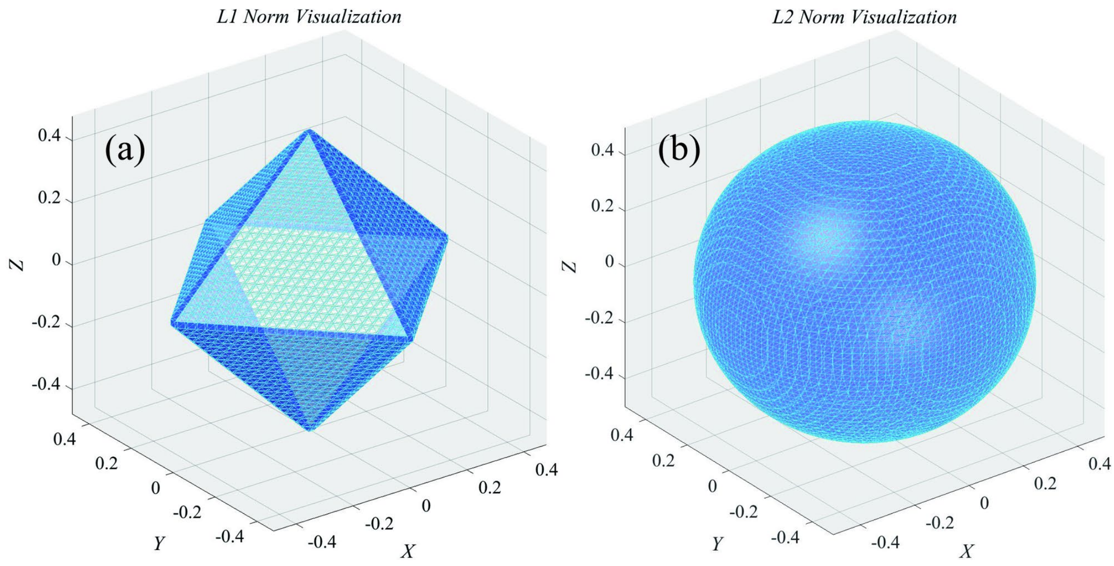

Figure 2.1. Three-dimensional visualization of two common regularizers; (a) ℓ_1 regularizer and (b) ℓ_2 regularizer (Li *et al.*, 2024c).

Equation 2.4 is a form of the value function, which allows us to precisely depict this relaxed framework. A value function offers a quantitative representation of a problem, capable of revealing the nature and characteristics of the problem. It aids in understanding its essence and provides direction for problem-solving (Gholami *et al.*, 2021). Specifically, this value function is obtained by minimizing equation 2.4 over $\mathbf{X}$, a process equivalent to finding a $\mathbf{X}$ that minimizes equation 2.4. This minimum corresponds to the value function I am discussing, which to some extent reflects the properties of the optimal solution to the problem. If I minimize equation 2.4 in $\mathbf{U}$, the minimum value function $\mathbf{v}(\mathbf{W})$, which is the predecessor of misfit, can be obtained (Zheng *et al.*, 2018):

$$\mathbf{v}(\mathbf{W}) = \min_{\mathbf{U}} \frac{1}{2}\|\mathbf{AU}-\mathbf{D}\|_{\mathrm{F}}^2 + \frac{k}{2}\|\mathbf{W}-\mathbf{SU}\|_{\mathrm{p}}. \quad \textbf{(2.5)}$$

I assume that $\mathbf{H}_k = \mathbf{A}^{\mathrm{T}}\mathbf{A} + k\mathbf{S}^{\mathrm{T}}\mathbf{S}$ is invertible, so that $\mathbf{U}(\mathbf{W}) = \mathbf{H}_k^{-1}(\mathbf{A}^{\mathrm{T}}\mathbf{D} + k\mathbf{S}^{\mathrm{T}}\mathbf{W})$ is unique. And:

$$\mathbf{F}_k = \begin{bmatrix} k\mathbf{A}\mathbf{H}_k^{-1}\mathbf{S}^{\top} \\ \sqrt{k}(\mathrm{I} - k\mathbf{S}\mathbf{H}_k^{-1}\mathbf{S}^{\top}) \end{bmatrix}, \tag{2.6}$$

$$\mathbf{G}_k = \begin{bmatrix} \mathrm{I} - k\mathbf{A}\mathbf{H}_k^{-1}\mathbf{A}^{\top} \\ \sqrt{k}\mathbf{S}\mathbf{H}_k^{-1}\mathbf{A}^{\top}) \end{bmatrix}, \tag{2.7}$$

$$\mathbf{g}_k = \mathbf{G}_k\mathbf{D}, \tag{2.8}$$

which provides a closed-form expression for:

$$\mathbf{v}(\mathbf{W}) = \frac{1}{2}\left\|\mathrm{F}_k\,\mathbf{W} - \mathbf{g}_k\right\|_{\mathrm{F}}^2. \tag{2.9}$$

Equation 2.4 then reduces to

$$\min_{\mathrm{W}} \frac{1}{2}\left\|\mathbf{F}_k\mathbf{W} - \mathbf{g}_k\right\|_{\mathrm{F}}^2 + \lambda\,\mathrm{R}(\mathbf{W}). \tag{2.10}$$

To elucidate the superiority of the SR3 algorithm more accurately, I demonstrate partial minimization improves the condition of the problem in Figure 2.2. In Figure 2.2 (a), the coloured ellipses depict the contours of $\left\|\mathbf{AU} - \mathbf{D}\right\|_{\mathrm{F}}^2$, while in Figure 2.2 (b), the contours of $\left\|\mathbf{F}_k\mathbf{W} - \mathbf{g}_k\right\|_{\mathrm{F}}^2$ are vividly portrayed as a circle (Zheng *et al.*, 2018). In Figure 2.2 (a), I exhibit the contour lines of the quadratic part similar to the least absolute shrinkage and selection operator (LASSO) problem (coloured ellipses) and the approximate solution paths (red solid line) in horizontal projection. In Figure 2.2 (b), the contour lines of the quadratic part of the SR3 loss function (coloured circle) and the corresponding approximate paths (red solid line) of the SR3 solution in the relaxed coordinates $\mathbf{W}$ are shown. Additionally, the grey diamonds indicate the contour lines of the ℓ_1 norm of the LASSO problem in each coordinate group.

From Figure 2.2, it is evident that for the widely applied class of LASSO-like problems, the properties of $\mathbf{F}$ are generally superior to $\mathbf{A}$, particularly in terms of the condition number, and $\mathbf{F}$ is typically smaller than $\mathbf{A}$. The ratio of the maximum to minimum singular values of $\mathbf{F}$ is smaller, thereby compressing the contour lines into a shape closer to a sphere, which accelerates convergence and enhances performance. Moreover, executing proximal gradient descent solely in $\mathbf{w}$ naturally resolves these types of problems. The formulas for $\mathbf{F}$ and $\mathbf{G}$ can also be applied to acceleration methods, such as the fast iterative shrinkage-thresholding algorithm (FISTA) algorithm. Overall, the SR3 algorithm reduces the singular values of $\mathbf{F}$ relative to $\mathbf{A}$ and has a weaker impact on small singular values. This effect "squeezes" the contour lines into a near-spherical shape, thereby accelerating convergence and enhancing performance (Zheng *et al*., 2018).

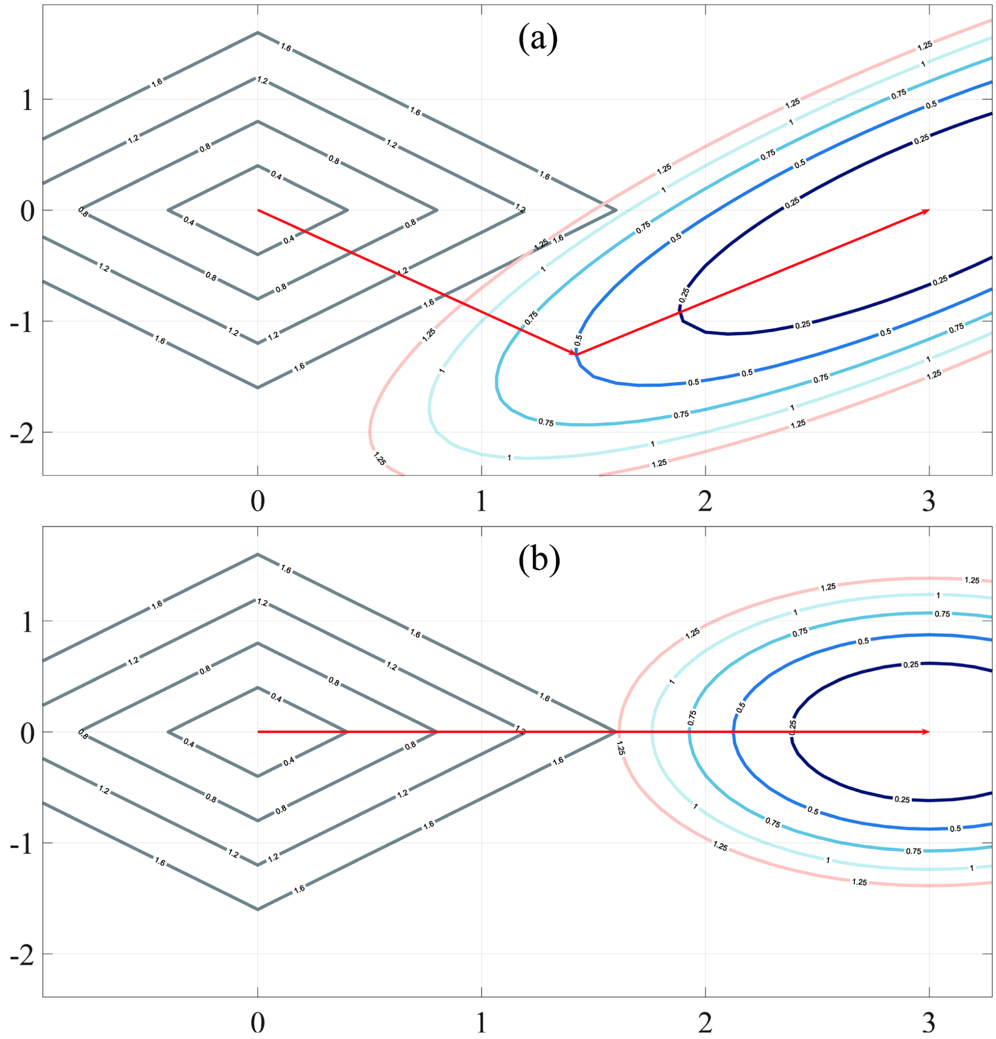

Figure 2.2. Illustrative figures of the gradient iteration process using the LASSO problem as an example, (a) conventional proxy-gradient process, (b) SR3 "tightens" the elliptical contour of the loss function to an approximate circle, thereby accelerating the convergence speed and performance of regression computations. The grey diamond is the contour of the ℓ_1 norm; The solid red line is the direction of the iteration update. (Li *et al.*, 2024c).

Finally, the algorithm requires two parameters to be specified simultaneously; the parameter λ determines the strength of the regularization, while k determines the degree of relaxation. I set the coefficient threshold to $\mathcal{J}$. Equation 2.11 is an empirical formula regarding parameter selection that I derived based on my experience in practical computations. During cross-validation, I used an intermediate parameter to represent the ratio of two unknown parameters, λ and k. I first set a range for one of the parameters and then input it into the algorithm. The algorithm iteratively calculates and optimizes the error between the computed and real wavefield. I then select the value of the $\mathcal{J}$ that corresponds to the minimum error, which gives us the optimal ratio of λ to k. After fixing one of these parameters, then the value of the other parameter can be determined by the above steps. Then:

$$\lambda = \frac{\mathcal{J}^2 k}{2}. \tag{2.11}$$

With equation 2.11, I can change the two-parameter selection problem to a single-parameter selection and significantly improve the algorithm. In addition, I suggest a cross-validation approach to achieve parameter tuning of the automatic strategy.

2.4 Numerical Simulations on Synthetic Data

In this section, to evaluate the efficacy of the algorithm introduced in this chapter, I assess the SR3 algorithm's performance with two homogeneous media and two synthetic data sets, respectively. In numerical experiments, I use the frequency-domain acoustic multi-scale FWI algorithm and introduce a Perfectly Matched Layer (PML) for absorbing boundary conditions (Pratt, 1999). In addition, to quantitatively clarify the intensity of the random noise employed in the numerical tests, I employ the subsequent signal-to-noise ratio equation (de Ridder and Dellinger, 2011):

$$\mathrm{SNR} = 20 * \mathrm{Log}_{10}\left(\frac{\|\mathbf{D}\|_2}{\|\mathbf{E}\|_2}\right), \tag{2.12}$$

where the $\mathbf{E}$ is the noise data. In addition, I adopt the model error measure to quantitatively compare and evaluate the inversion accuracy of various algorithms (Warner and Guasch, 2016):

$$\|\mathbf{M}_{\mathrm{true}} - \mathbf{M}_{\mathrm{inv}}\|_2 / \|\mathbf{M}_{\mathrm{true}}\|_2 , \tag{2.13}$$

where the $\mathbf{M}_{\mathrm{true}}$ and $\mathbf{M}_{\mathrm{inv}}$ represent true and inversion velocity models, respectively.

In addition, in this chapter, I will compare the analytical solution with the numerical solution, for which the distance function first needs to be defined:

$$\mathrm{r}(\mathbf{Z},\mathbf{X}) = \sqrt{(\mathbf{Z}-\mathbf{Z}_s)^2 + (\mathbf{X}-\mathbf{X}_S)^2}, \tag{2.14}$$

where $\mathrm{r}(\mathbf{Z},\mathbf{X})$ denotes the distance between $(\mathbf{Z},\mathbf{X})$ and source $(\mathbf{Z}_s,\mathbf{X}_s)$. In addition, the angular frequency needs to be calculated:

$$\omega = 2\pi f, \tag{2.15}$$

where ω is the angel frequency, f denotes the frequency. subsequently, we compute the analytical wavefield from the 2D Green's function:

$$\mathbf{G}_1(\mathbf{Z},\mathbf{X}) = \frac{i}{4}\mathrm{H}_0^2(\overline{K}\,\mathrm{r}(\mathbf{Z},\mathbf{X})), \tag{2.16}$$

where $\mathbf{G}_1(\mathbf{Z},\mathbf{X})$ is the 2D green function, H_0^2 is the Hankel functions of the second kind, which is the linear combination of the Bessel functions (Bessel function of the first kind) and Neumann functions (Bessel function of the second kind). $\overline{K}$ denotes the conjugate complex of wave numbers. According to the distance function, the distance function of its symmetric source and the 2D Green's function of the symmetric source can be obtained consequently. Finally, the conjugate of the two Green's functions is calculated:

$$\mathbf{G} = \overline{\mathbf{G}_1 - \mathbf{G}_2}, \quad \textbf{(2.17)}$$

where $\mathbf{G}_2$ is the 2D Green's function of the symmetric source, $\mathbf{G}$ denotes the complex conjugate of two Green's functions.

2.4.1 Single-Layer Homogeneous Model

Firstly, I conducted a preliminary test of the SR3 method in the one-layer homogeneous medium. Figure 2.3 (a) provides the basic structure of the single-layer homogeneous medium, while Figure 2.3 (b) depicts the one-dimensional velocity model of this medium with a velocity of 2 km/s. Figures 2.3 (c-e) present the analytical solutions of the wavefield under frequencies of 10 Hz, 13 Hz, and 15 Hz for this model, respectively. The process for calculating the analytical solution in the code is as follows: First, determine the source location and define the distance function. Then, use the Helmholtz function and the distance to define Green's function. Next, calculate the value of the Green's function at a given point. Finally, compute the wave field by taking the complex conjugate of the difference between the Green's functions at different source locations.

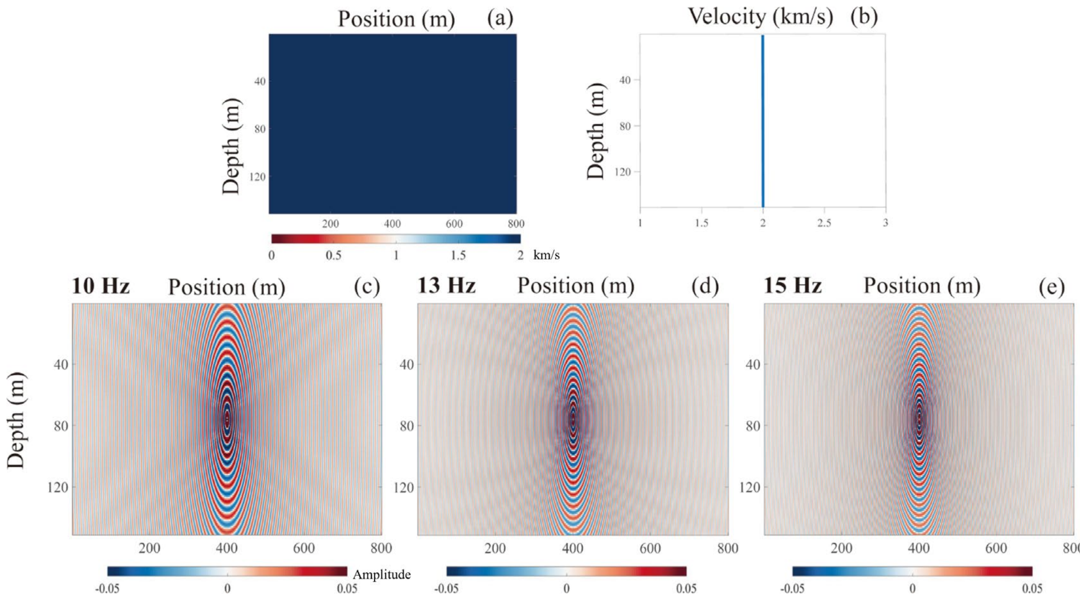

Figure 2.3. (a) The basic structure of a single-layer homogeneous medium, (b) the velocity model of the single-layer homogeneous medium, which achieves a velocity of 2 km/s, (c-e) the wavefield model for the analytical solution of the monolayer homogeneous medium at frequencies of 10 Hz, 13 Hz, and 15 Hz. The source position at (400, 80) of the model's grid. The receivers at the (400, 1) of the model's grid. The horizontal axis is the offset distance, and the vertical axis is the depth (Li *et al*., 2024c).

Initially, I tested the results of the conventional algorithm, as explained in Figure 2.4, where (a) displays the numerical solution at 10 Hz, (b) represents the difference between the real parts of the analytical and simulation solutions, (c) depicts the ratio of the real parts of the analytical and numerical solutions, and (d) describes the angle part of the ratio between the analytical and numerical solutions. Corresponding to Figure 2.4 (a-d), Figures 2.4 (e-h) present the various comparisons between the numerical solutions based on the SR3 algorithm and the analytical solution, with the order of comparison matching that of the conventional algorithm. To enhance visual recognition, I utilized a discrete colour map.

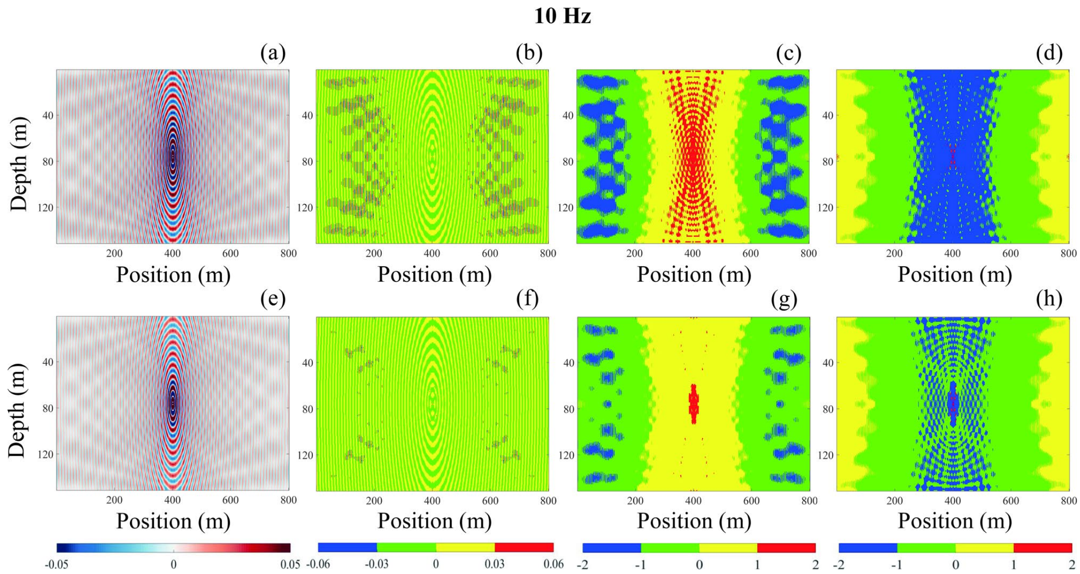

Figure 2.4. (a-d) the results and comparisons of the conventional FWI algorithm at 10 Hz, where (a) shows the numerical solution for the monolayer homogeneous medium at 10 Hz, (b) the real part of the difference between the numerical solution (Figure 2.4a) and the analytical solution (Figure 2.3c), (c) the real part of the ratio of the analytical solution to the numerical solution, (d) the angle part of the ratio of the difference (Figure 2.4b) to the analytic solution. (e-h) show the results after processing by the SR3 algorithm, where (e) shows the numerical solution after SR3 processed for the monolayer homogeneous medium at 10 Hz, (f) the real part of the difference between the numerical solution (Figure 2.4e) and the analytical solution (Figure 2.3c), (g) the real part of the ratio of the analytical solution to the numerical solution, (h) the angle part of the ratio of the difference (Figure 2.4f) to the analytic solution. The discretised colour map is intended to improve recognition performance (Li *et al*., 2024c).

Additionally, I compared results under other higher frequencies. Figure 2.5 (a-h) displays the performance comparison between the conventional algorithm and the SR3 algorithm at 13 Hz, adopting the same comparison method and order as in Figure 2.4. Similarly, Figure 2.6 (a-h) contrasts the conventional wavefield results with the SR3-optimized wavefield results at 15 Hz, maintaining the same comparison method and order as in Figure 2.4.

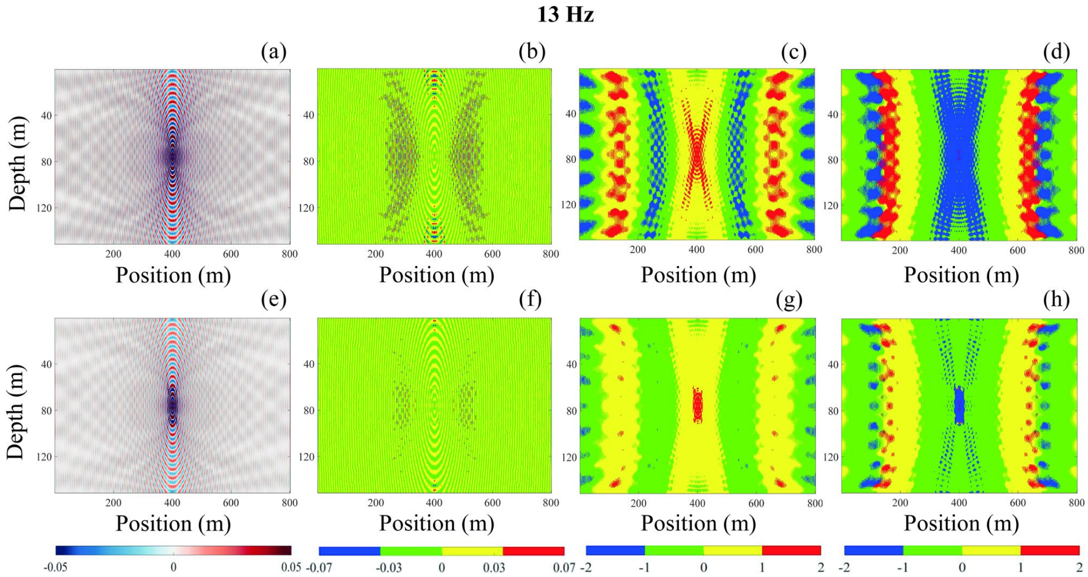

Figure 2.5. (a-d) the results and comparisons of the conventional algorithm at 13 Hz, with the same methods and order of comparisons as in Figures 2.4(a-d). (e-h) After processing by the SR3 algorithm at 13 Hz, the results have the exact ordering as in Figure 2.4(e-h). The discretised colour map is intended to improve recognition performance (Li *et al*., 2024c).

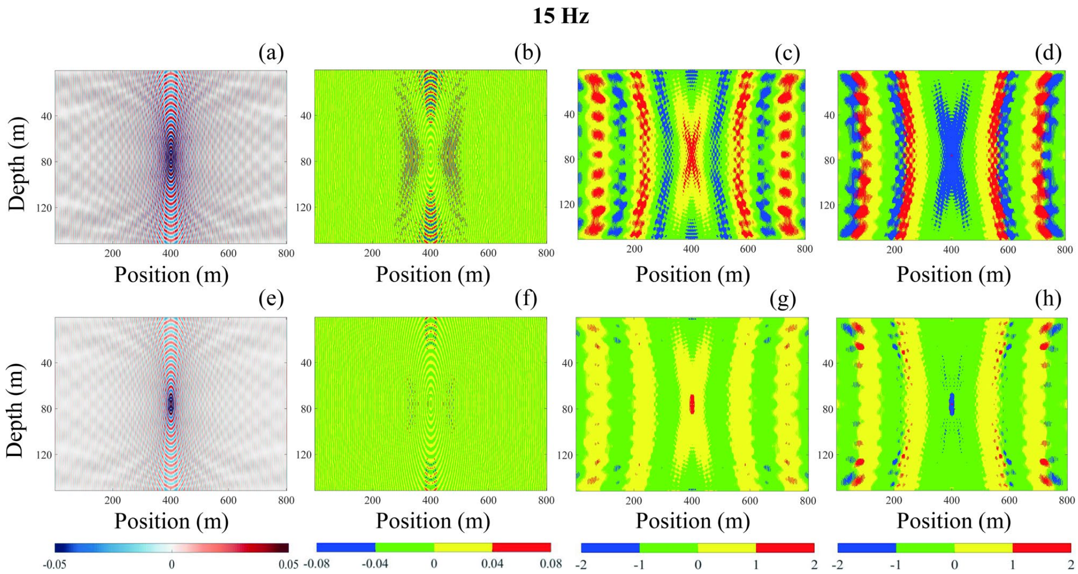

Figure 2.6. (a-d) The results and comparisons of the conventional algorithm at 15 Hz, with the same methods and order of comparisons as in Figures 2.4(a-d). (e-h) After processing by the SR3 algorithm at 15 Hz, the results have the exact ordering as in Figure 2.4(e-h). The discretised colour map is intended to improve recognition performance (Li *et al*., 2024c).

2.4.2 The Central Part of the 2004 BP Model

To better simulate complex exploration conditions, I tested the algorithm proposed in this chapter using synthetic data that is closer to actual conditions.

The 2004 BP model, as a highly representative salt body model, is extensively utilised in FWI. The choice of the salt body model is motivated by the presence of a high-velocity salt anomaly in its central part, which creates a significant velocity contrast with its surroundings. This contrast leads to considerable difficulties in accurately delineating the salt body contours and surrounding velocities under conditions of extremely low SNR. High noise levels exacerbate these challenges. Consequently, the salt body model is an ideal test bed in this experiment for assessing the precision and effectiveness of both conventional and proposed algorithms in high-accuracy inversion under complex conditions.

Figure 2.7 displays the 2004 BP model, representing the core section of the 2004 BP salt body model. This illustrates the geological characteristics of the eastern/central Gulf of Mexico and Angola's offshore regions. One of the complexities in inverting this model is illustrating the boundary of the salt dome intrusion, given the central section of this model consists of rapid salt formations. The presence of two water channels adjacent to the salt structure further complicates the velocity inversion, adding to the inversion challenges, which matches the research challenges described in the research challenge section. Figure 2.7 (b) shows the initial medium used, which increases linearly, and the velocity range is from 1.5 km/s to 5 km/s. I added 10 *dB* of random background noise to the dataset.

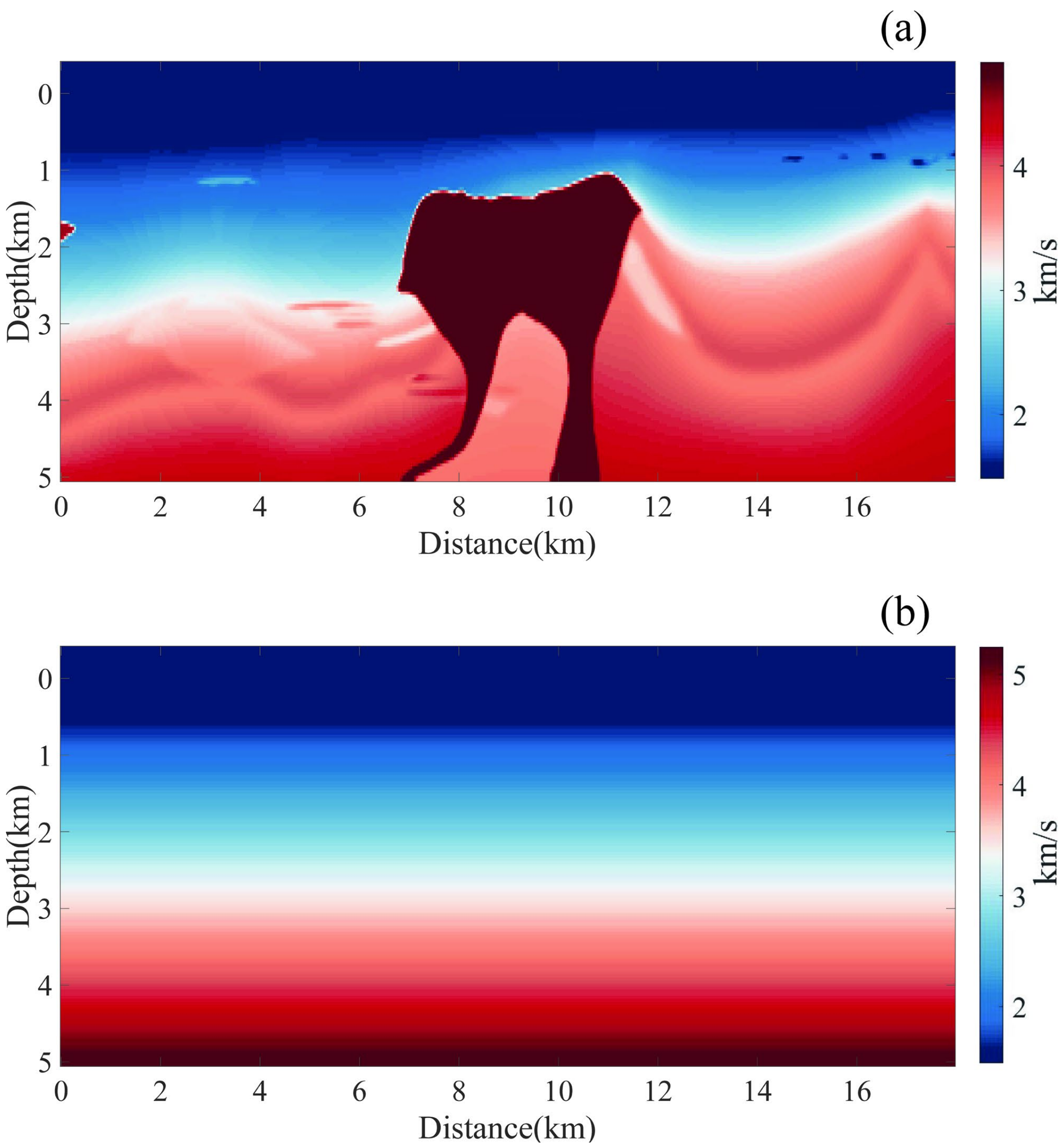

Figure 2.7. The central part of the 2004 BP benchmark; (a) real benchmark; (b) starting medium (Li *et al*., 2024c).

Further, where Figure 2.8 (a) presents the source-receiver data sets data obtained at an inversion frequency of 3 Hz, (b) shows the added 10 dB random background noise, (c) represents the wavefield after linearly adding the noise data, (d) is the downsampled wavefield of the 3 Hz, simulating spatial discontinuities caused by low-density sampling, and (e) shows the test wavefield after linearly stacking the noise-free source-receiver data sets.

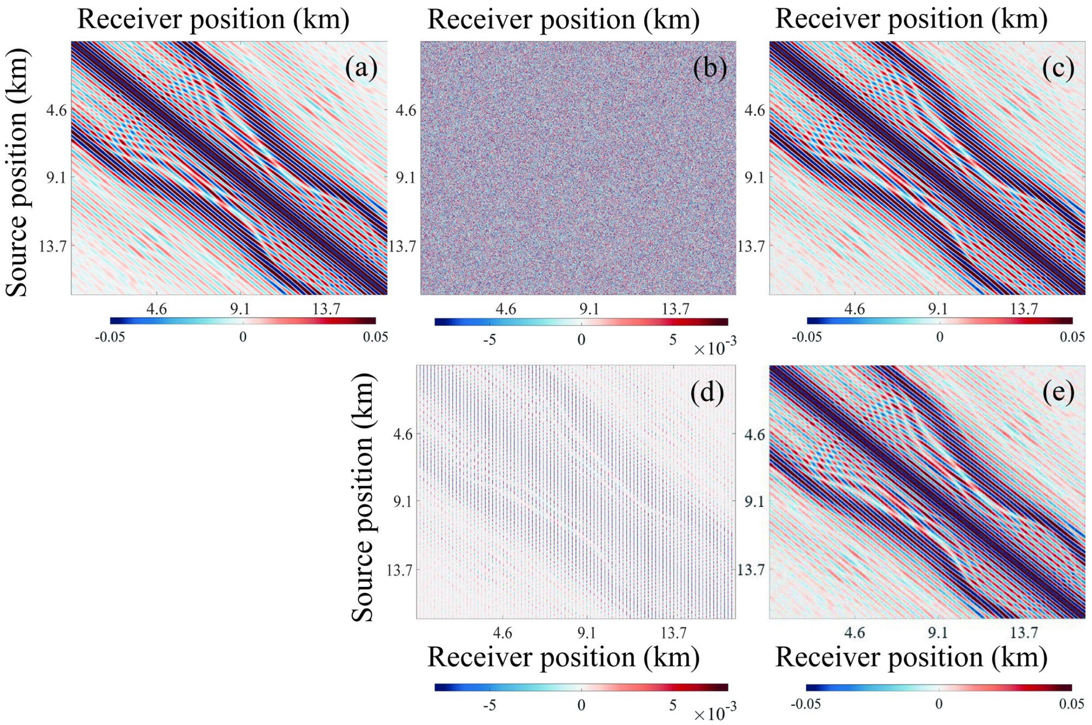

Figure 2.8. Source-receiver domain data set at 3 Hz of the 2004 BP model, the real part of the (a) clean data matrix, (b) 10 *dB* random noise, (c) wavefield matrix with 10 *dB* noise, (d) missing-trace matrix, (e) subsampled wavefield matrix with 10 *dB* noise and missing trace (Li *et al*., 2024c).

Subsequently, I computed the three-dimensional multi-scale wavefield cubes for the above conditions, as shown in Figure 2.9. Figure 2.9 (a) presents the pure wavefield information, viewed from (155,20), (b) is its side view from (105,1), (c) shows the low-quality wavefield stack simulated with linearly added random noise and downsampling from (155,20), (e) its side view from (105,1), (f) the multi-scale wavefield stack optimized by the SR3 algorithm proposed in this chapter, ranging from 0.5 Hz to 5 Hz, viewed from (155,20), and (g) its side view from (105,1). The results indicate that the conventional algorithm, severely disturbed by noise and downsampling, produces very poor wavefield quality with significant artefacts and aliasing. In view(a, b), the 'a' represents the horizontal rotation angle used to rotate the view around the z-axis, while 'b' is the vertical rotation angle used for tilting the view in a direction perpendicular to the horizontal-vertical plane. In contrast, the SR3 algorithm enhances the interpolation and noise removal capabilities, compensates for the spatial discontinuity, and has certain noise resistance, making the wavefield information more coherent, and laying a foundation for subsequent inversion. Particularly noteworthy in Figures 2.9c and 2.9d are the low-quality seismic wavefield data contaminated with noise. In these two images, the continuity of the wavefield information is significantly disrupted, mainly due to the sparsification, which has a severe impact on the lateral continuity of the wavefield data. These low-quality issues undoubtedly present considerable challenges for the seismic data inversion process. In contrast, Figures 2.9e and 2.9f display the wavefield data optimised using the SR3 algorithm. These images demonstrate the remarkable effectiveness of the proposed algorithm in interpolating and denoising low-frequency wavefield data, particularly in compensating for the gaps and discontinuities caused by data sparsification. The results show significant enhancement in the lateral continuity of the data, as particularly evident in Figure 2.9f. In this chapter, subsequent inversion processes will be based on the untreated wavefield data and the wavefield data optimised with the SR3 algorithm. The optimised wavefield data will provide more abundant and precise information for the subsequent processing steps, laying a solid foundation for high-precision imaging inversion.

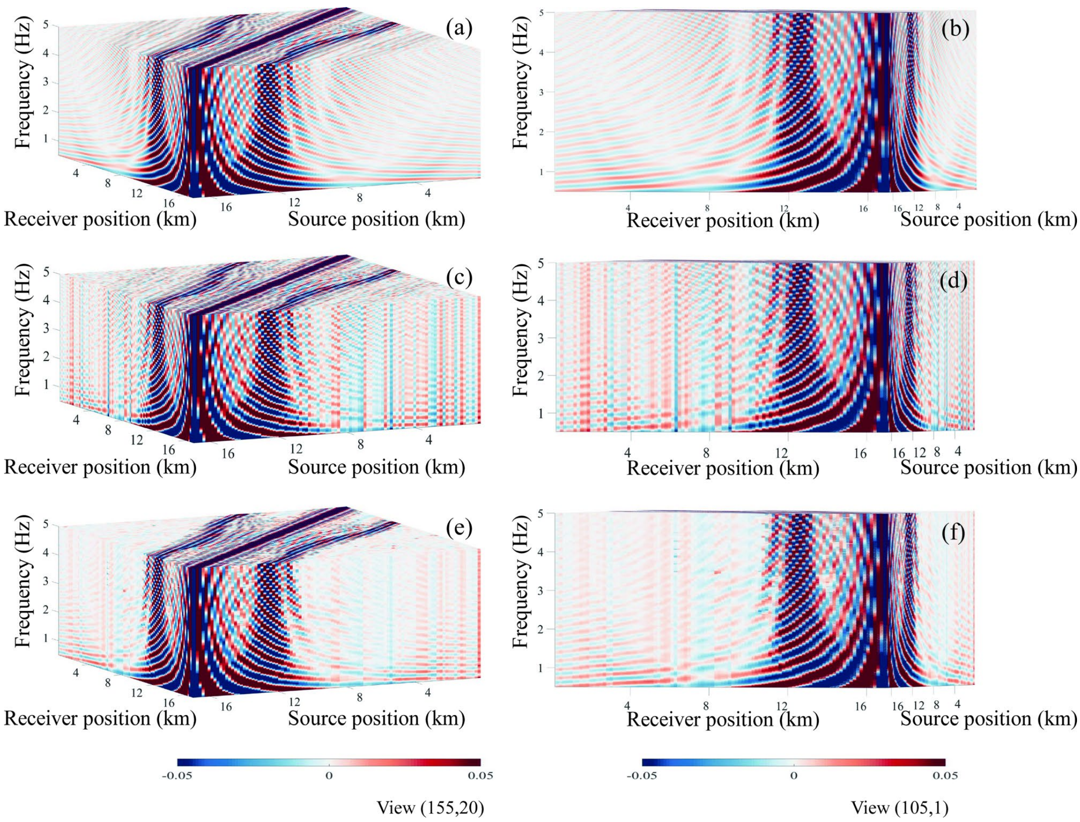

Figure 2.9. Three-dimensional low-frequency source-receiver data sets from 0.5 Hz to 5 Hz of the 2004 BP model, (a) clean data matrix, (b) side view of (a); (c) subsampled wavefield matrix with missing-trace and 10 *dB* noise, (d) side view of (c); (e) wavefield matrix after SR3 optimisation for (c), (f) side view of (e) (Li *et al*., 2024c).

In applying FWI, low and ultra-low frequency data play a crucial role, although there is no clear standard, the industry consensus is that for frequency domain FWI, signals below 3 Hz are considered low-frequency signals (Sun and Demanet, 2020). The significance lies in that FWI relies on accurate and comprehensive frequency content to reconstruct the subsurface velocity structure. High-quality, low-frequency data enhances the precision of the inversion process and is crucial in addressing the challenges of low-frequency paucity and over-dependence on the initial model in FWI.

Figure 2.10 illustrates low-frequency seismic wavefield data impacted by noise and sampling issues, where the characteristics of data sparsification and fragmentation significantly reduce its overall quality. The decline in wavefield quality affects the data's reliability and negatively impacts the subsequent inversion process and interpretation. In this context, the proposed algorithm effectively enhances the quality of low-frequency wavefield data by efficiently interpolating and reconstructing missing or damaged information. Improving the SR3 strengthens the continuity and integrity of the data and compensates for lost information caused by data sparsification and noise introduction. Therefore, the application of the SR3 algorithm has the potential to enhance the quality of inversion results in FWI, leading to more accurate velocity models and providing a more reliable foundation for the detailed imaging and interpretation of complex geological structures.

For a better comparison of the test results, especially the low-frequency effects, I provided side views in Figure 2.10. Figures 2.10 (a) and (d) show the wavefield information at extremely low frequencies of 1 Hz and 2 Hz, respectively, (b) and (e) show the low-quality wavefield stacks with linearly added direct arrivals and random noise, while (c) and (f) show the optimized results for the low-quality wavefields. After preprocessing, the wavefield continuity is improved, especially in the area marked with a box in the figure, artifacts have been significantly suppressed, making it closer to the real clean wavefield information. The experiments prove that the algorithm proposed in this chapter has excellent optimization effects for both high and low-frequency data, capable of resisting certain noise and compensating for a certain range of wavefield omissions.

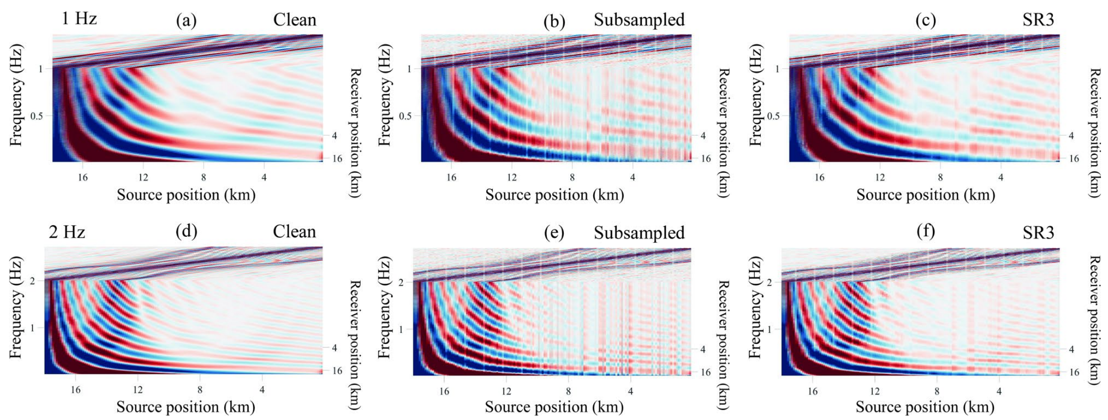

Figure 2.10. Three-dimensional side view low-frequency source-receiver data sets; 1 Hz (a) clean wavefield, (b) subsampled wavefield with missing trace, (c) SR3 processed reconstructing wavefield; 2 Hz (d) clean wavefield, (e) subsampled wavefield with missing-trace, (f) SR3 processed reconstructing wavefield (Li *et al*., 2024c).

Next, I present the results of multi-scale frequency domain FWI for the two algorithms, as shown in Figure 2.11, specifically, (a1-a8) depict the inversion outcomes using the traditional Tikhonov regularization FWI; in contrast, (b1-b8) showcase the inversion results from the enhanced algorithm presented in this chapter; (c1-c8) illustrate the discrepancies between the traditional algorithm and the actual benchmark; meanwhile, (d1-d8) highlight the variances between the outcomes of the enhanced algorithm and the genuine velocity model. Both methods utilize the identical multi-scale inversion frequency and have incorporated 10 *dB* of random noise along with the down-sampling simulation. From the results, it is evident that compared to the conventional algorithm, the improved algorithm offers better contour descriptions and noise reduction capabilities. Specifically, due to the impact of noise and down-sampling, the conventional algorithm's inversion results in the deep part of the model and the water channels on both sides appear blurred, especially the inversion results of the bottom salt dome part are not distinct. This occurrence might be attributed to the subpar quality of the low-frequency data, leading to the larger-scale components of the model being inadequately inverted. However, the inversion outcomes from the enhanced algorithm are markedly better. Specifically, the configurations of the water channels on either side are well-defined, and the salt dome's continuity is commendable.

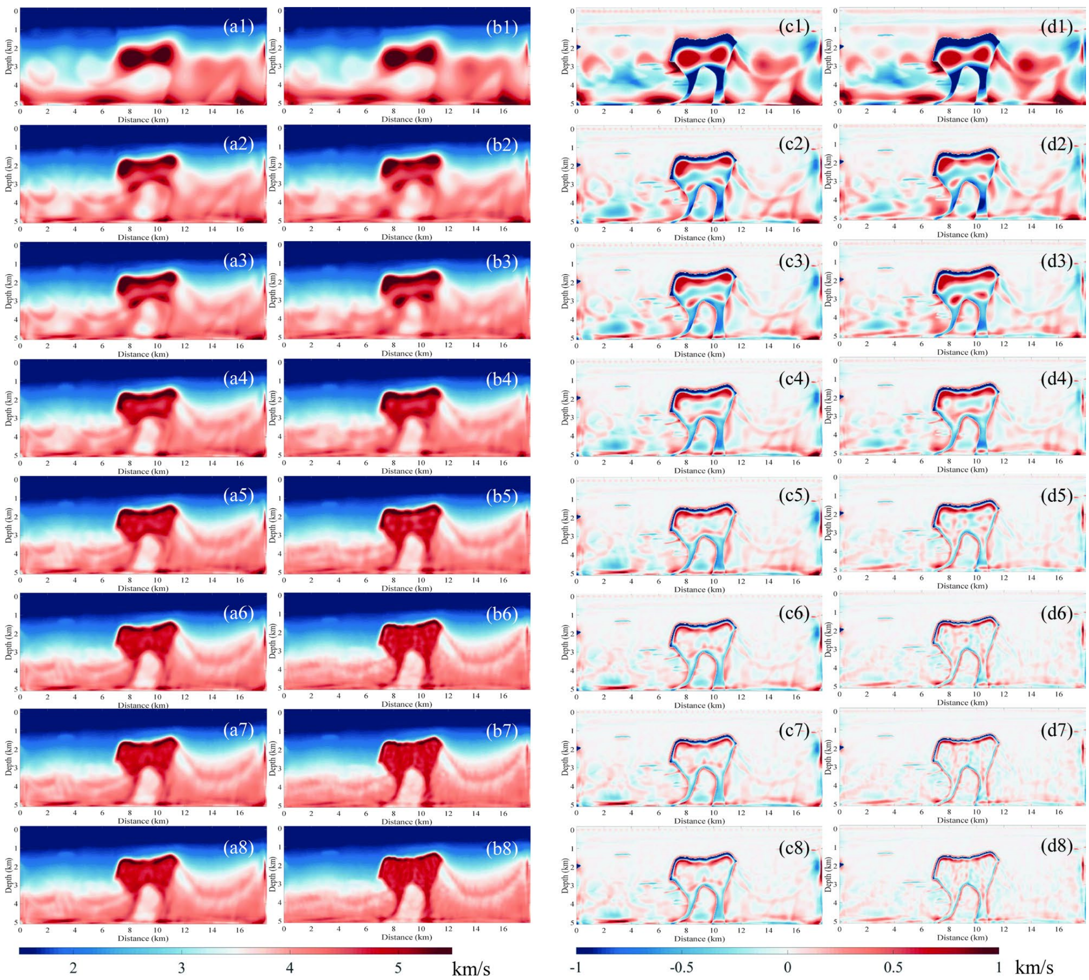

Figure 2.11. 2004 BP model inversion results, (a1-a8) Tikhonov regularisation FWI inversion results in 1.20 Hz, 2.99 Hz, 3.58 Hz, 5.16 Hz, 7.43 Hz, 10.70 Hz, 12.84 Hz, and 15.41 Hz, respectively, (b1-b8) FWI based on SR3 algorithm optimisation inversion results in 1.20 Hz, 2.99 Hz, 3.58 Hz, 5.16 Hz, 7.43 Hz, 10.70 Hz, 12.84 Hz, and 15.41 Hz, respectively, (c1-c8) differences between the Tikhonov FWI and the actual velocity model, (d1-d8) differences between the SR3-based FWI and the actual velocity model (Li *et al.*, 2024c).

To better quantitatively compare the two inversion results, I further contrasted their misfit errors and model errors, as illustrated in Figure 2.12. Where (a) represents misfit error, and (b) represents model error. It can be observed that the convergence rate of the modified algorithm is significantly better, and its model error is also smaller. This indicates that the inversion result of the modified method is closer to the real model.

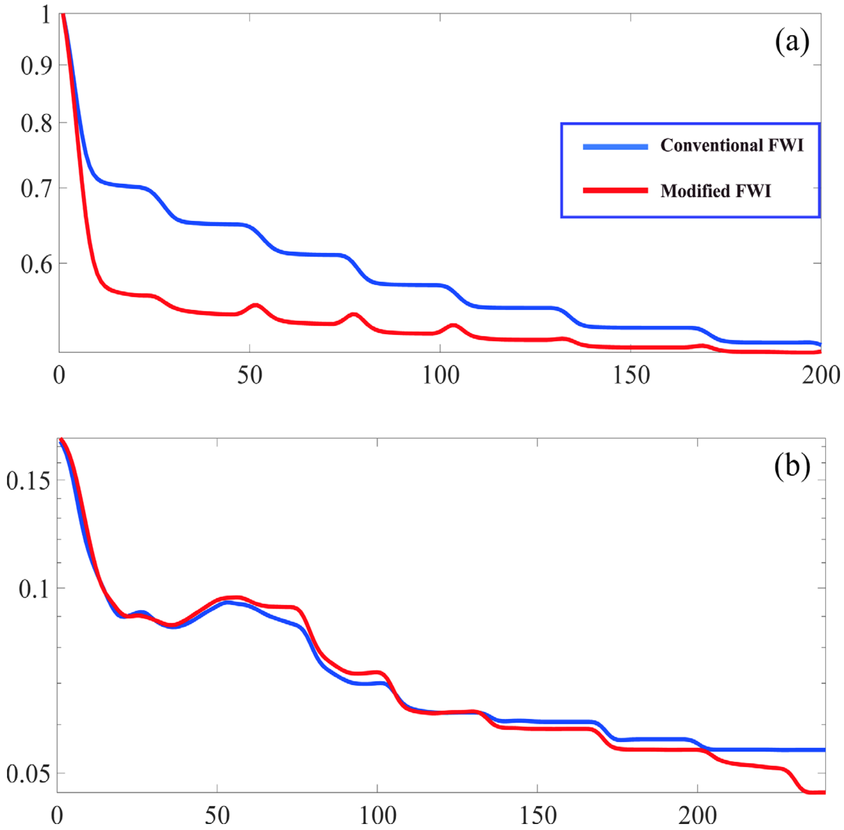

Figure 2.12. Comparison of SR3 algorithm-based FWI with conventional Tikhonov regularisation-based FWI for quantification, (a) misfit error, (b) model error. The horizontal axis is the number of iterations, and the vertical axis is the error value (Li *et al.*, 2024c).

Lastly, the one-dimensional velocity analysis results of the inversion results of the two algorithms are compared, as shown in Figure 2.13, which illustrates a longitudinal velocity comparison across different X coordinates. The red solid line denotes the outcome from the enhanced algorithm, the blue solid line indicates the results from the traditional algorithm, the black solid line stands for the actual velocity, and the grey dashed line signifies the starting velocity.

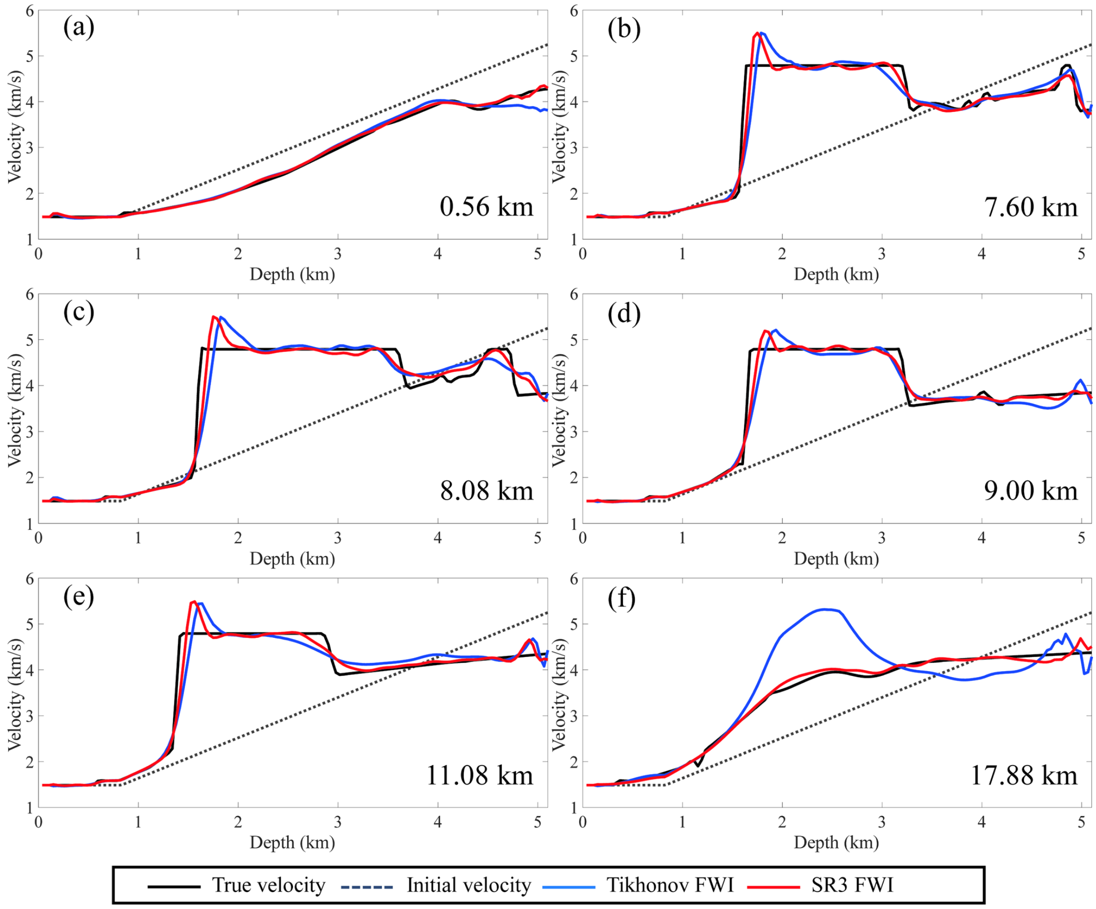

Figure 2.13. 2004 BP model, one-dimensional velocity models at different X-positions, (a) $X = 0.56$ km; (b) $X = 7.60$ km; (c) $X = 8.08$ km; (d) $X = 9.00$ km; (e) $X = 11.08$ km; (f) $X = 17.88$ km, the vertical comparison of the actual velocity model (solid black line), initial velocity model (grey dotted line), the Tikhonov regularisation FWI velocity model (solid blue line), and the SR3-based FWI velocity model (solid red line) (Li *et al.*, 2024c).

Similarly, Figure 2.14 is the lateral velocity model comparison. Compared to the longitudinal comparison, the lateral comparison is more challenging. From the velocity comparison results in both directions, it can be seen that the velocity of the improved algorithm fits the true velocity better, especially in the deep area, the improved algorithm provides better velocity compensation. In contrast, influenced by noise and down-sampling, the conventional algorithm displays noticeable non-fitting. Specifically, in the deep areas and salt dome parts, the velocity fitting results differ significantly from the real results, and in some areas, there are even noticeable velocity artefacts.

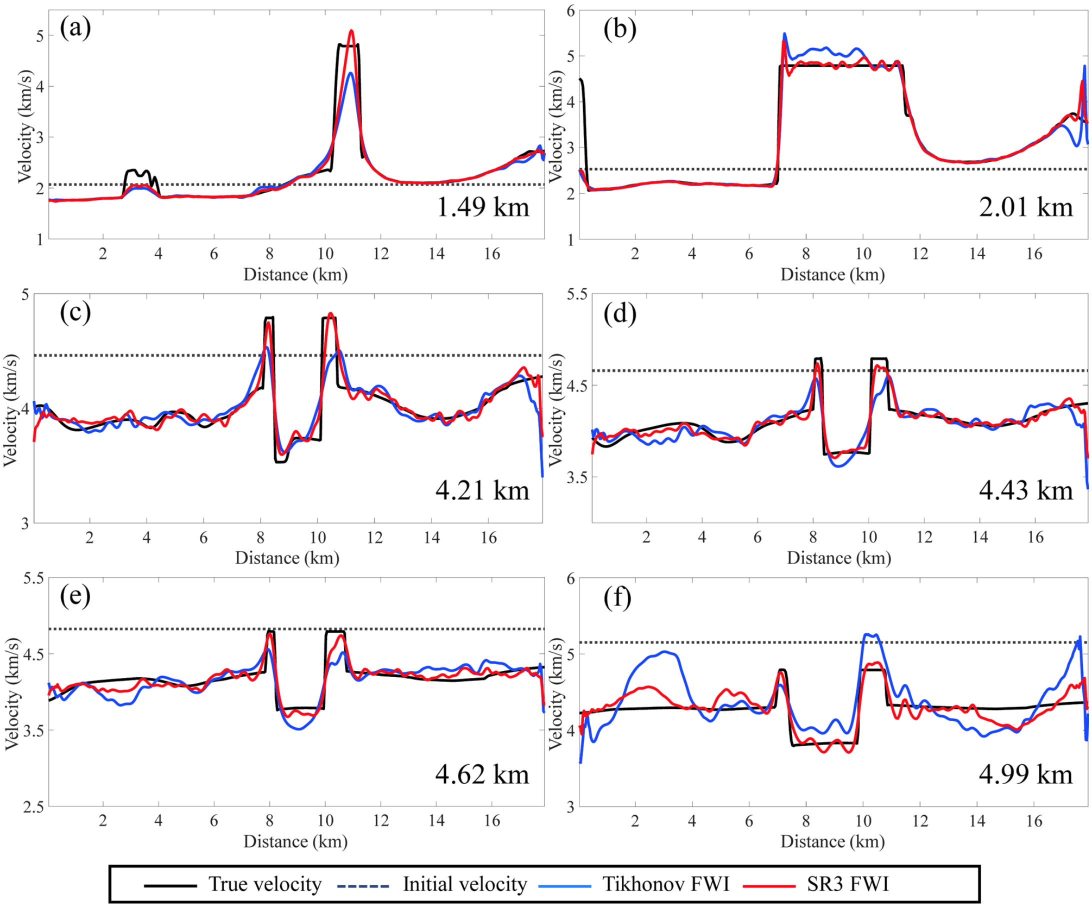

Figure 2.14. 2004 BP model, one-dimensional velocity models at different *Y*-positions, (a) $Y = 1.49$ km; (b) $Y = 2.01$ km; (c) $Y = 4.21$ km; (d) $Y = 4.43$ km; (e) $Y = 4.62$ km; (f) $Y = 4.99$ km, the horizontal comparison of the actual velocity model (solid black line), initial velocity model (grey dotted line), the Tikhonov regularisation FWI velocity model (solid blue line), and the SR3-based FWI velocity model (solid red line) (Li *et al*., 2024c).

To evaluate the performance of my proposed algorithm under low signal-to-noise ratio conditions, I conducted a set of comparative tests with 5 *dB* of random noise present. Figure 2.15 presents these test results. Figure 2.15a shows the clean wavefield information at 2 Hz; Figure 2.15b displays the information with 5 *dB* of random noise; Figure 2.15c illustrates the wavefield after adding 5 *dB* of random noise; Figure 2.15d depicts the simulated missing trace matrix, and Figure 2.15e shows the subsampled wavefield matrix with 5 *dB* noise and missing traces. The 5 *dB* noise significantly disrupts the wavefield, making high-resolution inversion particularly challenging under these conditions.

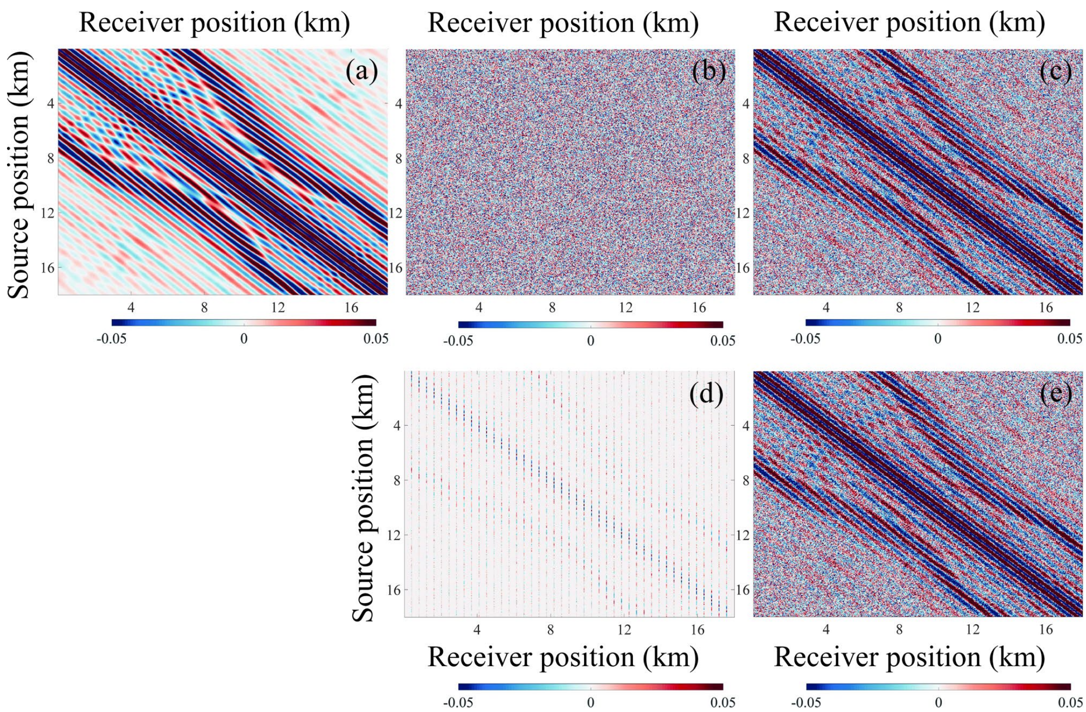

Figure 2.15. Source-receiver domain data set at 2 Hz of the 2004 BP model, the real part of the (a) clean data matrix, (b) 5 *dB* random noise, (c) wavefield matrix with 5 *dB* noise, (d) missing-trace matrix, (e) subsampled wavefield matrix with 5 *dB* noise and missing trace (Li *et al*., 2024c).

Subsequently, I demonstrate the test results as shown in Figure 2.16. Figures 2.16a1-a4 illustrate the inversion results using conventional algorithms at frequencies of 2.99 Hz, 3.58 Hz, 5.16 Hz, and 7.43 Hz, respectively, while Figures 2.16b1-b4 display the inversion results using my proposed algorithm at the same frequencies. These outcomes indicate that noise and missing traces have a significant impact on the inversion results, especially the artefacts caused by the very low SNR (5 *dB* of random noise) (although there is no clear standard, the industry consensus is that less than 10 dB is a low SNR). Moreover, in the results of conventional algorithms, the structure of the bottom salt bodies appears very blurred. In contrast, the results obtained by this thesis-proposed algorithm are markedly superior to those of conventional algorithms, particularly in the deeper regions of the model where two salt pillars are located, with higher resolution and more precise structure. These results highlight the efficacy of this algorithm in wavefield interpolation reconstruction and noise reduction, effectively compensating for inversion anomalies due to missing information.

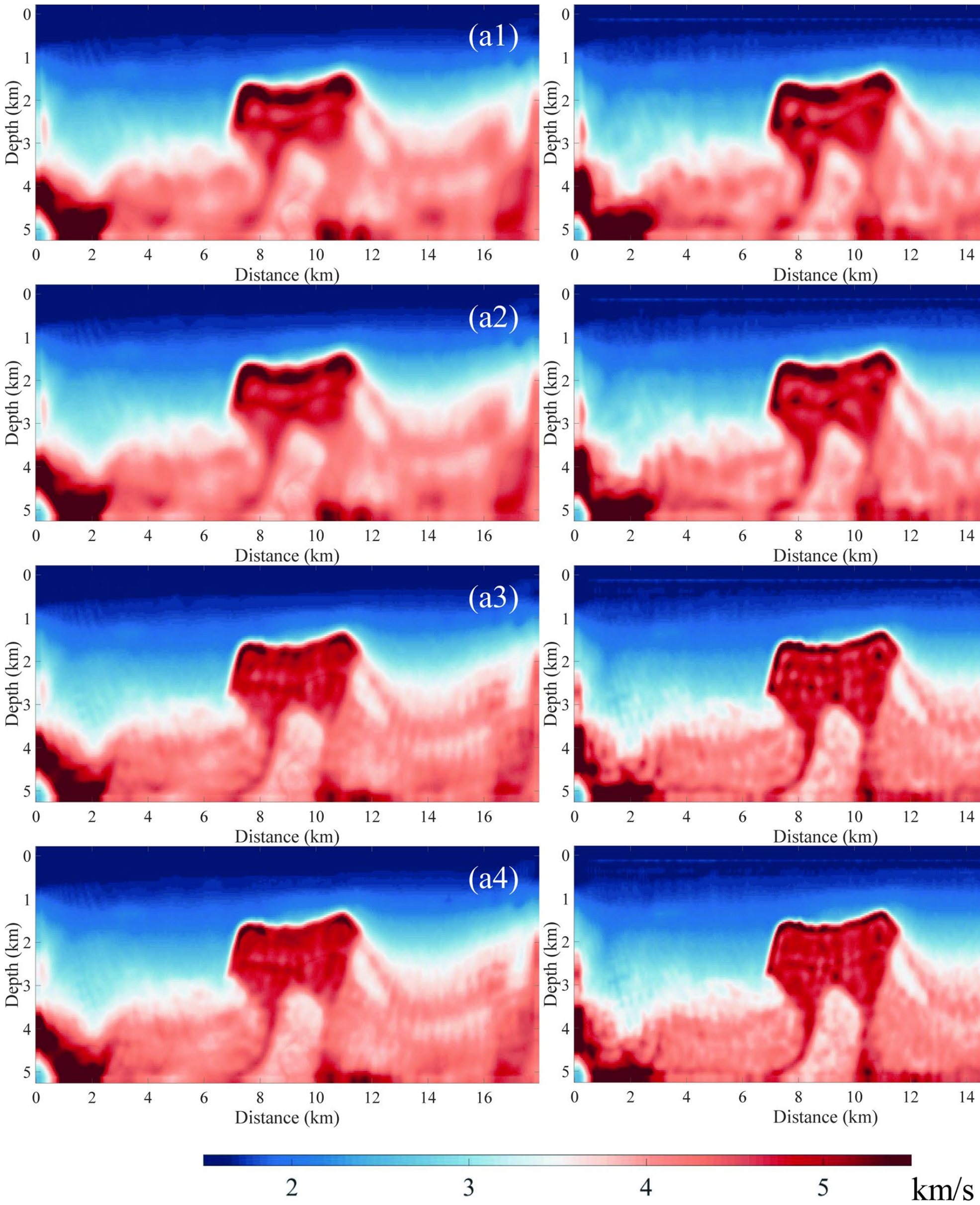

Figure 2.16. 2004 BP model inversion results, (a1-a4) Conventional FWI inversion results in 2.99 Hz, 3.58 Hz, 5.16 Hz, and 7.43 Hz, respectively, (b1-b4) FWI based on SR3 algorithm optimisation inversion results in 2.99 Hz, 3.58 Hz, 5.16 Hz, and 7.43 Hz, respectively (Li *et al*., 2024c).

Finally, I conducted a one-dimensional velocity analysis and comparison of the inversion results, as shown in Figure 2.17. The comparison shows that the red solid line representing this proposed algorithm closely matches the black line indicating the actual velocity, especially in the deep layers where significant velocity differences are observed. The proposed algorithm addresses the noticeable velocity discrepancies present in traditional methods, indicating that the improved algorithm more accurately describes key details of the model.

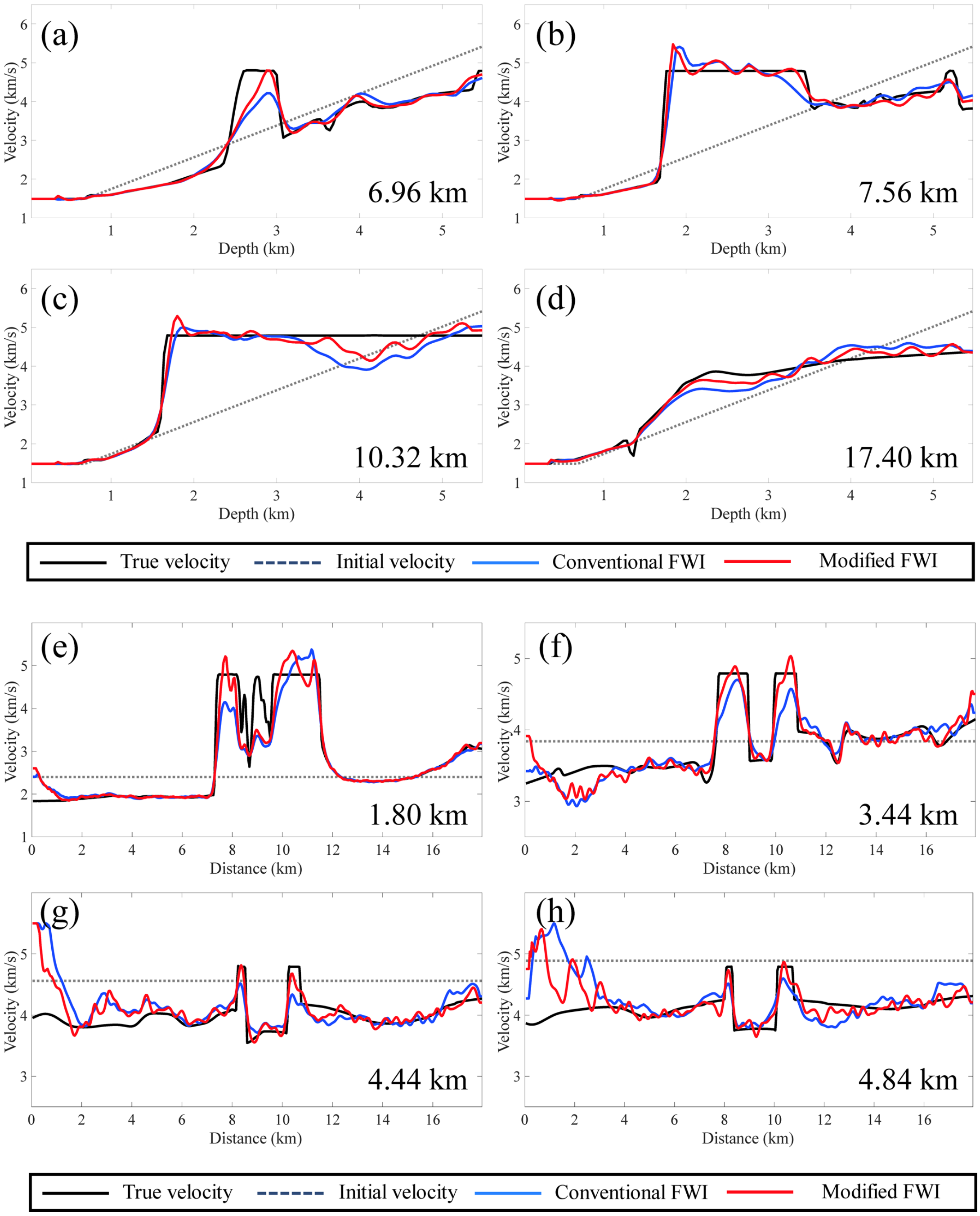

Figure 2.17. 2004 BP model, one-dimensional velocity models at different *X*-positions and *Y*-positions, (a) $X = 6.96$ km; (b) $X = 7.56$ km; (c) $X = 10.32$ km; (d) $X = 17.40$ km; and (e) $Y = 1.80$ km; (f) $Y = 3.44$ km; (g) $Y = 4.44$ km; (h) $Y = 4.84$ km; the comparison of the actual velocity model (solid black line), initial velocity model (grey dotted line), the conventional FWI velocity model (solid blue line), and the SR3-based FWI velocity model (solid red line) (Li *et al*., 2024c).

2.4.3 Two-Dimensional SEG/EAGE Overthrust Model

I also applied the algorithm introduced in this chapter to the two-dimensional overthrust model. This model features several unique geological strata, predominantly marked by overthrust formations. Given its complex structural attributes, the model presents particular challenges for seismic exploration methods, FWI in particular. The primary rationale for selecting this model lies in its distinct representativeness as a geological model. The complexity of the inversion process for the Overthrust model is primarily concentrated on accurately inverting the velocity of high-velocity layers obscured by multiple overlying strata. The inversion difficulty for the Overthrust model involves precisely delineating the velocities of various strata within the covering layers and characterising the features of the high-velocity basement layer.

Figure 2.18 presents the benchmark of the two-dimensional overthrust, with (a) depicting the actual benchmark and (b) representing the starting model.

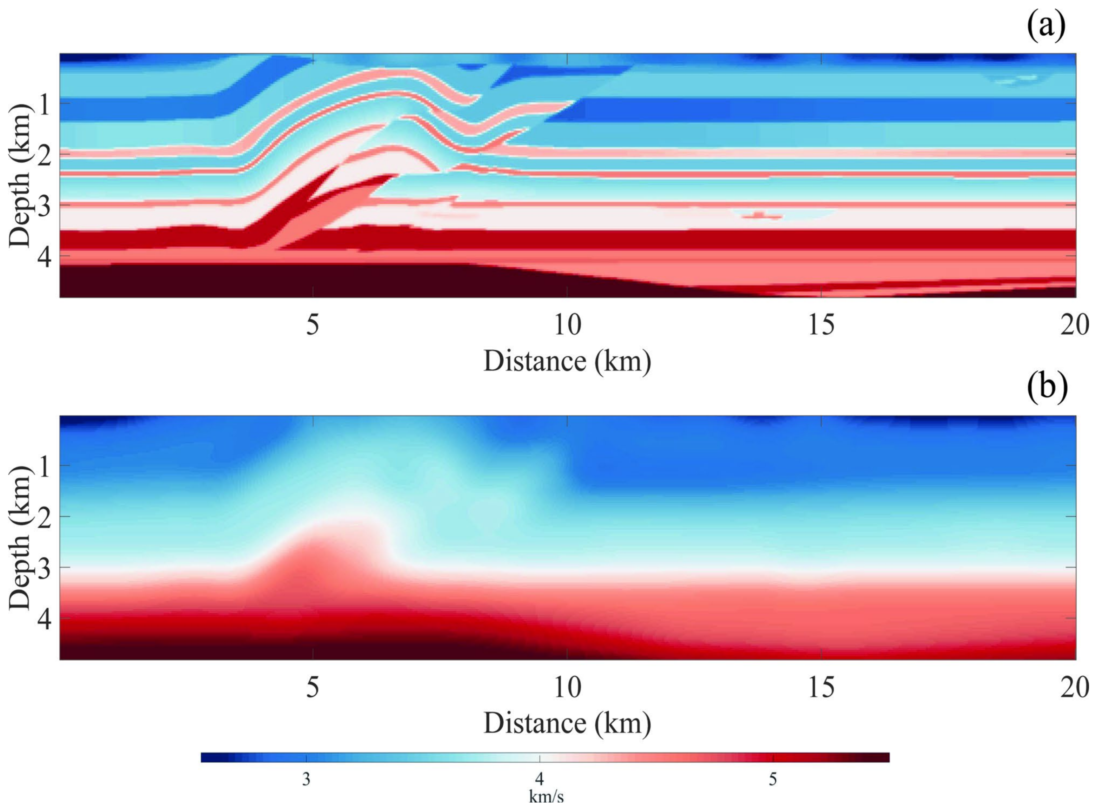

Figure 2.18. Two-dimensional SEG/EAGE overthrust model, (a) actual velocity model. (b) initial velocity model (Li *et al.*, 2024c).

In a manner akin to the 2004 BP Model, I initially showcase the undistorted wavefield of this model at a 3 Hz inversion frequency according to the multi-scale inversion strategy, as exhibited in Figure 2.19 (a). Meanwhile, Figure 2.19 (b) portrays the random noise at 10 *dB*, while (c) is the noisy wavefield obtained by linearly superimposing the random noise onto the clean wavefield, (d) depicts the uniform downsampling simulation, which means that the data corresponding to the set of traces to be zero or very small, and (e) is the result after linearly overlaying both random noise and downsampling simulation onto the clean wavefield.

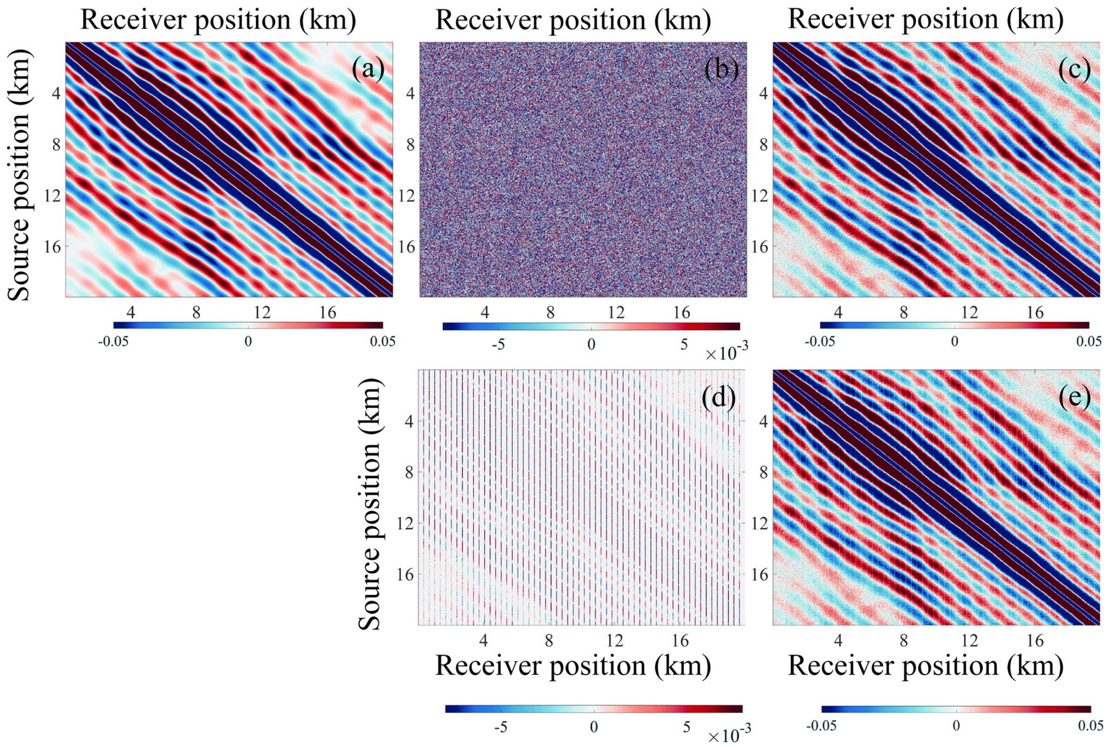

Figure 2.19. Source-receiver domain data set at 3 Hz of the SEG/EAGE overthrust model, the real part of the (a) clean data matrix, (b) 10 *dB* random noise, (c) wavefield matrix with 10 *dB* noise, (d) missing-trace matrix, (e) subsampled wavefield matrix with 10 *dB* noise and missing trace (Li *et al*., 2024c).

Based on the aforementioned models, noise, and downsampling simulations, this experiment compares the results of the conventional algorithm with those preprocessed by the SR3 algorithm for FWI. The test results are shown in Figure 2.20, which displays a three-dimensional multi-scale wavefield stack. Figure 2.20 (a) displays the clean wavefield from frequencies 0.5 Hz to 5 Hz, (b) its side view with an angle of (105,5). Figure 2.20 (c) showcases the noisy, downsampled, low-quality wavefield after following the two aforementioned processing steps, and (d) its side view, again at an angle of (105,5). (e) is the wavefield after SR3 processing, and (f) its side view at

(105,5). As can be seen, the SR3 approach recommended in this chapter has demonstrated its effective and excellent optimization capabilities on various synthetic datasets, especially with robust noise removal and interpolation abilities. The enhanced continuity of the wavefield stack, better noise suppression, and clearer wavefield details are evident.

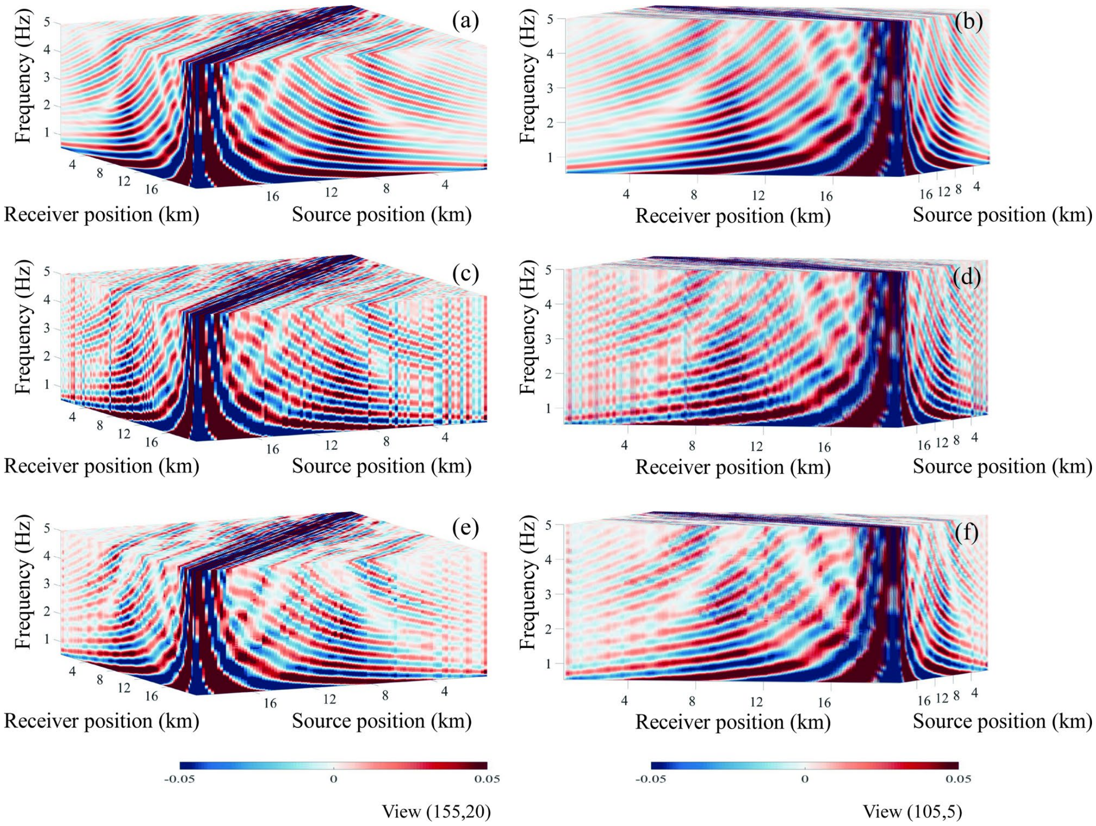

Figure 2.20. Three-dimensional low-frequency source-receiver data sets from 0.5 Hz to 5 Hz of the SEG/EAGE overthrust model, (a) clean data matrix, (b) side view of (a); (c) subsampled wavefield matrix with missing-trace and 10 *dB* noise, (d) side view of (c); (e) wavefield matrix after SR3 optimisation for (c), (f) side view of (e) (Li *et al.*, 2024c).

Subsequently, the final FWI inversion results of the two algorithms are presented in

Figure 2.21. Figures 2.21 (a1-a8) are conventional FWI results, (b1-b8) depict the results based on the SR3 algorithm proposed in this chapter, both methods use a multi-scale strategy. Additionally, Figures 2.21 (c1-c8) depict the error of the outputs obtained by the conventional method and the real benchmark, whereas Figures 2.21 (d1-d8) depict the disparity between the FWI inversion outcomes utilizing the SR3 method introduced in this chapter and the actual velocity model. The comparative results reveal that the improved FWI method has distinct advantages in handling complex structures, not only providing clearer descriptions of shallow parts of the model but also distinctly depicting velocity variations between different rock layers. It enhances the delineation of contour boundaries, avoiding the deep blur often seen in conventional methods.

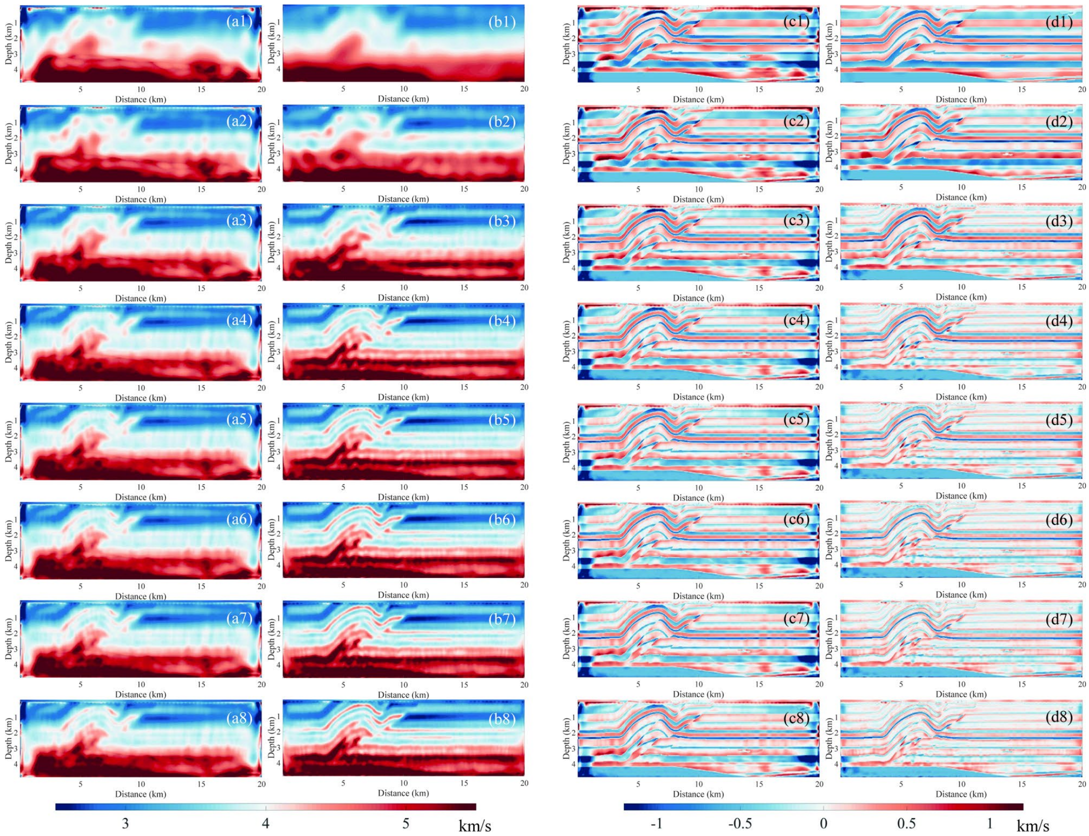

Figure 2.21. The SEG/EAGE overthrust model inversion results. (a1-a8) Tikhonov regularisation FWI inversion results in 2.15 Hz, 3.09 Hz, 5.35 Hz, 7.70 Hz, 9.24 Hz, 11.09 Hz, 13.31 Hz, and 15.97 Hz, respectively; (b1-b8) FWI based on SR3 algorithm inversion results in 2.15 Hz, 3.09 Hz, 5.35 Hz, 7.70 Hz, 9.24 Hz, 11.09 Hz, 13.31 Hz, and 15.97 Hz, respectively; (c1-c8) differences between the Tikhonov FWI and the true velocity model; (d1-d8) differences between the SR3 FWI and the true velocity model (Li *et al*., 2024c).

To further quantitatively compare the two algorithms, Figures 2.22 (a) and (b) respectively present the misfit error and model error comparisons of the two methods. The blue solid line symbolizes the conventional method, whereas the red one indicates the improved algorithm. The improved method's misfit error iterates faster, and its model error is smaller compared to the conventional approach, confirming that the enhanced algorithm provides more accurate inversion results.

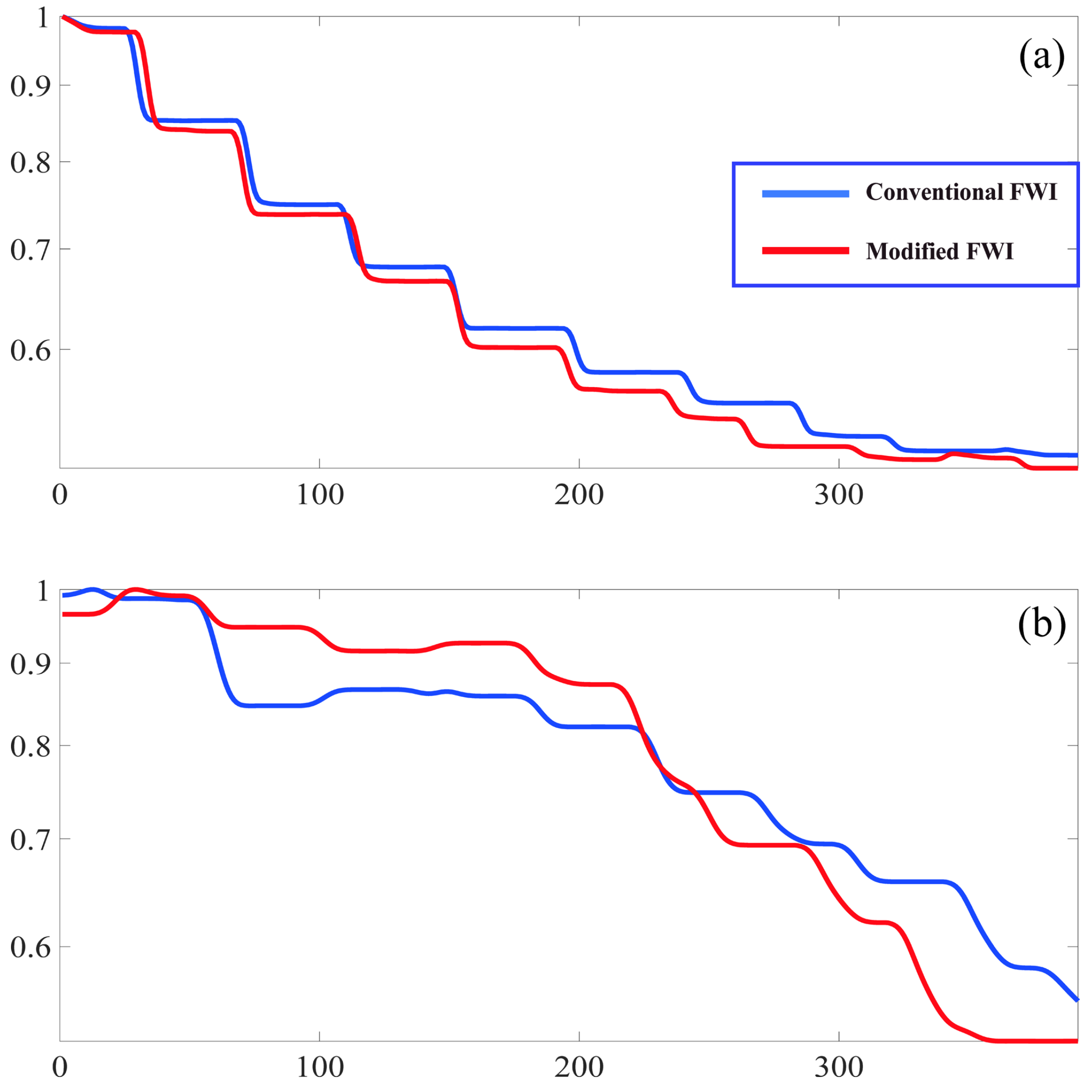

Figure 2.22. Comparison of SR3 algorithm-based FWI with conventional Tikhonov regularisation-based FWI for quantification, (a) normalised misfit error, (b) model error. The horizontal axis is the number of iterations, and the vertical axis is the error value (Li *et al.*, 2024c).

Lastly, several groups of lateral and longitudinal one-dimensional velocity comparisons are provided. Figure 2.23 represents the longitudinal velocity comparison, displaying the differences between results obtained at multiple X positions by the two methods and the true velocity.

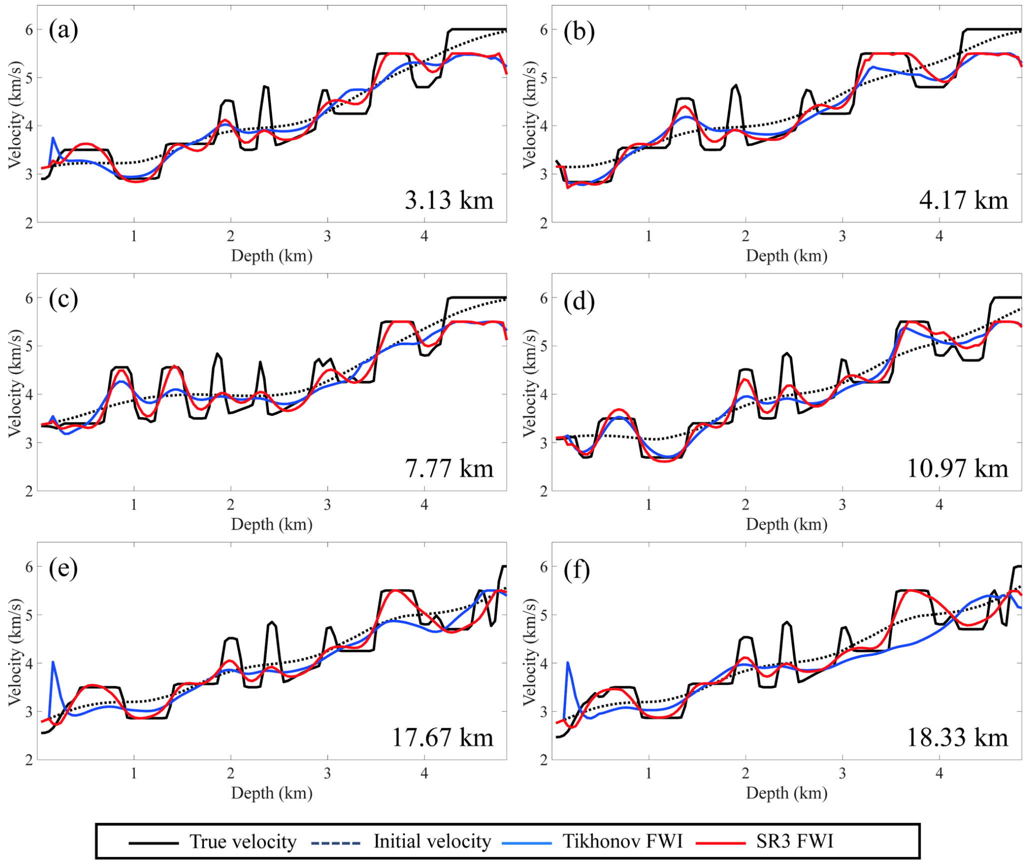

Figure 2.23. The SEG/EAGE overthrust model, 1-D velocity models at different X-positions, (a) $X = 3.13$ km; (b) $X = 4.17$ km; (c) $X = 7.77$ km; (d) $X = 10.97$ km; (e) $X = 17.67$ km; (f) $X = 18.33$ km. The vertical comparison of the actual velocity model (solid black line), initial velocity model (grey dotted line), the Tikhonov regularisation FWI velocity model (solid blue line), and the SR3-based FWI velocity model (solid red line) (Li *et al.*, 2024c).

Figure 2.24, on the other hand, shows the lateral velocity comparison, indicating the differences in results at various *Y* positions between two distinct algorithms with the real benchmark. The comparison highlights that the conventional method, especially in the deep parts of the model, often gives poorer results, sometimes with significant inversion errors that fail to achieve the correct velocity range. However, the more challenging lateral velocity measurement, constrained by seismic wave propagation properties and data collection limitations, becomes evident. Seismic waves radiate spherically from the source, implying a smaller horizontal coverage. Additionally, due to economic and system constraints in actual seismic exploration, receivers cannot be placed at every point on the surface, leading to potentially uneven data sampling that affects the accuracy of lateral velocity measurements. Nevertheless, the test results in Figure 2.24 further underline the significant gap between the inversion capabilities of the two algorithms. The improved algorithm demonstrates excellent continuity and noise resistance, especially near the bottom of the model, even under challenging scenarios, including noise and data loss, the improved method still proves its efficacy and robustness.

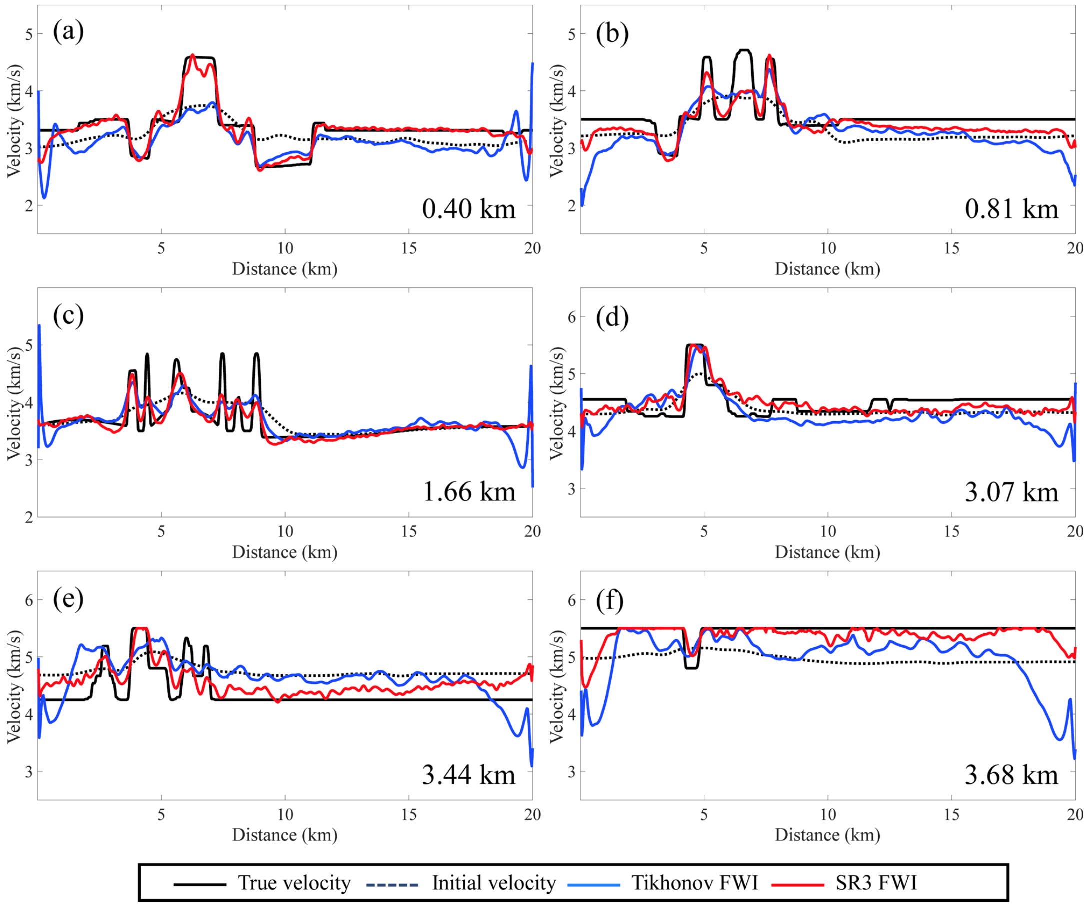

Figure 2.24. The SEG/EAGE overthrust model, 1-D velocity models at different Y-positions, (a) $Y = 0.40$ km; (b) $Y = 0.81$ km; (c) $Y = 1.66$ km; (d) $Y = 3.07$ km; (e) $Y = 3.44$ km; (f) $Y = 3.68$ km. The horizontal comparison of the actual velocity model (solid black line), initial velocity model (grey dotted line), the Tikhonov regularisation FWI velocity model (solid blue line), and the SR3-based FWI velocity model (solid red line) (Li *et al*., 2024c).

2.5 Chapter Discussions

As a simulation algorithm, the quality of the numerical solution of the forward wavefield in FWI has a profound impact on subsequent inversion processes. In other words, a good inversion necessarily relies on an excellent forward modelling result.

Hence, if one wishes to enhance the overall inversion quality of FWI, it's imperative to explore methods for enhancing the quality of the forward wavefield. In actual production, various factors such as natural environments, human factors, and exploration strategies affect and interfere with forward modelling results. Noise interference and azimuthal limitations are the primary influencing factors of the forward results. Therefore, enhancing the forward modelling quality from the economical perspective of merely improving the algorithm is of paramount importance for the further application of FWI.

In this chapter, I propose an improved preprocessing algorithm that numerically optimizes the multi-scale wavefield stacking, which accelerates the convergence rate of the regression algorithm with the aid of auxiliary matrices and the concept of compressing the singular value space. Compared to traditional algorithms, the SR3 algorithm presented in this chapter is a higher-level sparse regression algorithm, manifesting in three main areas. Firstly, regarding the denoising issue of FWI, the SR3 algorithm has already completed it in preprocessing. This eliminates the need for redundancy in the minimization process and avoids the computational burden of introducing model increments twice. Secondly, the SR3 algorithm intrinsically deals with a multi-constraint optimization problem, allowing the use of composite regularizers. This feature grants the algorithm outstanding flexibility and robustness. Lastly, compared to other regression algorithms, the SR3 algorithm tightens the singular value space, greatly accelerating computational efficiency without adding computational pressure to the overall FWI.

In addition, although there is no clear standard, the industry consensus is that less than 10 dB is a low SNR. Multiscale inversion is used to address the challenges posed by the non-flat frequency response of the acquisition system, by gradually incorporating different frequency components to improve inversion stability and accuracy.

Given the above advantages and the significance of forward modelling, I initially declare the feasibility and necessity of employing this algorithm in the second chapter of this thesis. In the subsequent chapters, I will continue to discuss improvements in the optimization stages of the FWI.

2.6 Chapter Conclusions

In this chapter, I have emphasized an innovative improved algorithm that achieves high-resolution velocity modeling by preprocessing the forward wavefield of FWI. As an optimization algorithm that completes, interpolates, and denoises the wavefield information, SR3 can achieve a more comprehensive wideband wavefield optimization by utilizing sparse relaxed regular regression. This overcomes challenges brought about by noise, down-sampling, spatial aliasing effects, etc., realizing the objective of wideband multi-scale frequency domain FWI. I used two homogeneous media and two synthetic datasets for algorithm testing. In the numerical fitting test section, I focused on the improved algorithm's ability to interpolate and denoise the low-frequency wavefield and optimize anti-aliasing for the high-frequency wavefield. Thereby this algorithm achieves the purpose of wideband optimization. Subsequently, the optimized wavefield was integrated back into the FWI process and tested on two synthetic datasets. The test results demonstrate that the algorithm proposed in this chapter displays compelling effectiveness and high performance, including significant noise resistance capability and optimization ability for details. In actual field surveys, the quality of the seismic data we can obtain is inevitably less than ideal, similar to the simulations in this chapter. Therefore, under these extremely challenging conditions, using FWI to image low-quality seismic data becomes exceedingly difficult. However, the preprocessing algorithm proposed in this chapter has proven the feasibility of low-frequency data reconstruction, high-frequency anti-aliasing, and wideband multi-scale FWI, this provides a promising and more rational algorithm and approach for the further industrial application of FWI.

CHAPTER III: FULL WAVEFORM INVERSION BASED ON K-SUPPORT NORM

The content of this chapter has been published in a journal paper and a preprint: *Pure and Applied Geophysics (PAAG 2024) & arXiv (2022)*

1. **Li, J***., Mikada, H., & Takekawa, J., 2024. Improved full-waveform inversion for seismic data in the presence of noise based on the K-support norm. *Pure and Applied Geophysics*, 1-28. DOI: 10.1007/s00024-024-03449-5 **(19-February-2024)**

2. **Li, J***., Mikada, H., & Takekawa, J., 2022. Improved full-waveform inversion for seismic data in the presence of noise based on the k-support norm. *arXiv preprint,* arXiv:2212.07074. DOI: 10.48550/arXiv.2212.07074 **(14-December-2022)**

The content of this chapter has been published in a conference abstract: In the *83rd EAGE Annual Conference & Exhibition.* European Association of Geoscientists & Engineers *(EAGE 2022)*

3. **Li, J***., & Xu, S., 2022. The modified full waveform inversion with the K-Support norm for the noise existing data. In *83rd EAGE Annual Conference & Exhibition*, Madrid, Spain, Vol. 2022, No. 1, pp. 1-5. European Association of Geoscientists & Engineers. DOI: 10.3997/2214-4609.202210322 **(06-June-2022)**

3.1 Introduction

FWI involves aligning the seismic wavefield. During optimization, the traditional algorithm computes the least-squares discrepancy between the observed and the modelled wavefield. It then determines the updated gradient direction through the minimization procedure, setting the stage for the subsequent iteration (Sirgue and Pratt, 2004). Since the optimal fit between the model and the seismic data is achieved at the convergence point of the objective function, quickly finding the global minimum in ill-condition functions is crucial for achieving high-precision FWI.

In geophysics, the least squares method is a classical solution for fitting two types of wavefields. However, when the recorded data contains non-Gaussian white noise or outliers, it can lead to difficulties in the FWI minimization or optimization process. This can result in slow convergence or getting stuck in local minima, which is related to the character of the least squares method itself. The least squares algorithm is based on the ℓ_2 norm, calculating the Euclidean distance between the two. However, this also means that errors caused by outliers in the model are squared and amplified in the ℓ_2 norm, which can be seen as one kind of weight decay. While this prevents overfitting, it also makes the ℓ_2 norm more sensitive to noise and outliers (Lailly *et al.*, 1983; Tarantola 1984). Therefore, under poor observation conditions, the ℓ_2 norm algorithm used by FWI is not sufficient to remove the adverse influence of anomalies and noise in the optimization process. Hence, an improved algorithm urgently needs to adopt a new regularization method to strengthen the constraints of the misfit function and emphasize its noise resistance. Moreover, the nonlinearity of FWI often causes the minimization routine to be trapped in local optima instead of seeking the global optimum, and results in hindering the convergence rate and precision of the inversion process (Virieux and Operto, 2009).

To address the above difficulty, a common sparse norm form of the ℓ_1 norm can be used. Compared with the ℓ_2 norm, it emphasizes the sparsity of the model rather than smoothness and encourages sparse solutions. Therefore it possesses better robustness and noise resistance. A regularization form based on the ℓ_1 norm is TV regularization. This algorithm enforces similarity between adjacent parameters in model space, thus achieving a sparse representation of the model, reducing model complexity while emphasizing the effective information in model parameters (Koh *et al.*, 2007). FWI based on TV regularization utilizes this feature to strengthen the constraints of the misfit function in the optimization phase. This retains effective velocity features, ignores small anomalous perturbations, and thereby achieves better denoising effects. However, the most notable issue with this algorithm is that the ℓ_1 norm is not strictly convex, which can lead to computational difficulties.

Based on the aforementioned issues, in this chapter, I attempt to employ a form combining norms in the minimization process of FWI, specifically, the K-support regularization algorithm. This algorithm was first proposed in the field of mathematics and features a tighter relaxation form while ensuring the characteristics of ℓ_2 norm. This represents that the algorithm maintains the convex characteristics of basic ℓ_2 norm while also ensuring a degree of sparsity and robustness (Bai and Liang, 2020). While providing stricter constraint forms, this algorithm ensures convexity to facilitate the search process in finding the global optimum, thereby aiding the modified FWI in denoising and suppressing outliers.

Additionally, as mentioned above, due to the nonlinearity of the wave equation itself, the calculation process of the least-squares form exhibits strong non-linear characteristics, making it easy to fall into local minima. To address this, the K-support norm introduced in this chapter incorporates a new regularization term in the form of a quadratic penalty to solve the aforementioned problems. The newly added regularization term restricts the range of the solutions to the model parameters, mitigating the nonlinearity of the inversion problem.

This chapter will first introduce the K-support norm for noise resisting, additionally, the ADMM is utilized to accelerate the minimization procedure. In the iterative process, he ADMM algorithm can be used to realize the robust iterative process of K-support under the framework of alternate directions. During the numerical fitting process, the effectiveness of the improved algorithm introduced in this chapter is demonstrated by contrasting it with conventional FWI, and various synthetic examples are employed for testing to showcase the efficacy of the presented algorithm.

3.2 K-Support norm

The wave equation of FWI is as:

$$\min_{\mathbf{M},\mathbf{X}} \left\|\mathbf{AX}-\mathbf{D}\right\|_2^2, \quad \text{s.t. } \mathbf{C}(\mathbf{M})\mathbf{X}=\mathbf{S}, \tag{3.1}$$

where the $\mathbf{A} \in \mathbb{R}^{\mathcal{M}\times\mathcal{N}}$ is linear observation operator, and the $\mathbf{X} \in \mathbb{R}^{\mathcal{N}\times 1}$ represents the model wavefield, meanwhile, $\mathbf{D} \in \mathbb{R}^{\mathcal{M}\times 1}$ stands for the observed data. $\mathbf{C}(\mathbf{M}) \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ designates discretized partial differential equation, $\mathbf{M} \in \mathbb{R}^{\mathcal{N}\times 1}$ represents model parameters, the $\mathbf{S} \in \mathbb{R}^{\mathcal{N}\times 1}$ corresponds to the source term, and $\left\|\bullet\right\|_2$ embodies the Euclidean norm (van Leeuwen and Herrmann, 2013), the detailed derivation can be found in Appendix A.

The research in this chapter is dedicated to a new regularization algorithm, instead of the conventional least-squares method, to help FWI achieve high accuracy in inversion under complex conditions of high background noise. Therefore, it is necessary to combine the two different norms to make the constraint function sparse. I start with the following equation to tighten the relaxation while ensuring convexity:

$$\mathrm{conv}(\mathrm{G}_{\mathcal{K}}) \subseteq \left\{ \mathbf{P} \middle| \|\mathbf{P}\|_1 \le \sqrt{\mathcal{K}}, \|\mathbf{P}\|_2 \le 1 \right\} \subsetneq \left\{ \mathbf{P} \middle| \|\mathbf{P}\|_1 \le \sqrt{\mathcal{K}} \right\}, \tag{3.2}$$

where $\mathcal{K} > 1$ is an adjustable parameter, $\|\mathbf{P}\|_1$ is the ℓ_1 norm for the vector $\mathbf{P}$, and $\mathrm{conv}(\mathrm{G}_{\mathcal{K}})$ is applied as a constraint for the misfit function $\mathrm{G}_{\mathcal{K}}$. I consider add $\mathrm{conv}(\mathrm{G}_{\mathcal{K}}) = \mathrm{conv}\left\{ \mathbf{P} \middle| \|\mathbf{P}\|_0 \le \mathcal{K}, \|\mathbf{P}\|_2 \le 1 \right\}$, as a convex hull of equation 3.2 thus making it a tighter convex constraint. The kernel parameter $\mathcal{K} \in \{1, \ldots, \mathcal{T}\}$ guides the nature of the new constraints to adopt a more adaptable format (Belilovsky *et al.*, 2015).

This approach emphasizes the integration of a convex outer approximation in equation 3.2. By merging the two norms, it establishes a convex hull. Consequently, this can be viewed as a kind of convex constraint that surpasses the conventional sparse constraint. Within FWI, the goal of the optimization process is to identify the model that provides simulated data most closely to the observed data. Regularization acts as a mediator, striking a balance between the complexity of the model and its ability to fit. In complex situations, especially when dealing with noise or when a large number of velocity parameters are required for accurate model description, the model complexity becomes high. If only stricter constraint forms are adopted, it would lead to overfitting. However, relaxing the constraints can result in underfitting. Additionally, the presence of noise and artefacts can lead to poor data continuity, increasing the number of local minima, which elevates the risk of falling into some local minima when searching for the global minimum. Therefore, I propose to employ a flexible regularization method to help us balance the relationship between constraint ability and smoothing ability, and enhance the capability to find the global minimum.

The algorithm introduced in this chapter seeks to strike a balance between model intricacy and inversion precision while providing a certain level of adaptability:

$$\|\mathbf{M}\|_{\mathcal{K}}^{\mathrm{sp}} = \Big(\sum_{i=1}^{\mathcal{K}-\mathcal{J}-1} (|\mathbf{M}|_i^{\downarrow})^2 + \frac{1}{\mathcal{J}+1} \big(\sum_{i=\mathcal{K}-\mathcal{J}}^{\mathcal{T}} |\mathbf{M}|_i^{\downarrow} \big)^2 \Big)^{1/2}, \tag{3.3}$$

where the $|\mathbf{M}|_i^{\downarrow}$ denotes the *i*-th largest element within the model parameter vector $|\mathbf{M}|$, the $\|\bullet\|_{\mathcal{K}}^{\mathrm{sp}}$ signifies the K-support norm. $\mathcal{K}$ regulates the sparsity of the array, while $\mathcal{T}$ standing for the data's dimension. If $\mathcal{J} \in \{0,1,\ldots,\mathcal{K}-1\}$ is considered an integer parameter, then I derive:

$$|\mathbf{M}|_{\mathcal{K}-\mathcal{J}-1}^{\downarrow} > \frac{1}{\mathcal{J}+1} \sum_{i=\mathcal{K}-\mathcal{J}}^{\mathcal{T}} |\mathbf{M}|_i^{\downarrow} \geq |\mathbf{M}|_{\mathcal{K}-\mathcal{J}}^{\downarrow}. \tag{3.4}$$

The K-support norm comprises two parts. One primary part is the ℓ_2 norm, and the other is the ℓ_1 norm. The parameter $\mathcal{K}$ acts as a tunable factor to mediate the balance between the ℓ_1 and ℓ_2 norms. As a result, this newly introduced norm can effectively balance the generalization capacity and the sparsity capability of the algorithm (Lu *et al.*, 2017).

The effectiveness of regularization algorithms in denoising lies in their ability to introduce a penalty term in the objective function to limit the complexity of the model, thereby preventing the model from overfitting to noise and producing smoother and more robust solutions. The ℓ_1 norm promotes sparsity, causing some insignificant parameters to approach zero, thereby removing noise while retaining the main features of the signal. The ℓ_2 norm constrains the sum of the squares of the parameters, preventing them from becoming too large and providing a smoothing effect. Combining the ℓ_1 and ℓ_2 norms leverages the strengths of both: promoting sparsity while also providing smoothing. This combination excels in denoising by effectively balancing signal retention and noise suppression, improving the model's generalization capability and robustness. In practical applications, this combined approach often achieves better denoising performance than methods using a single norm regularization.

A wavefield reconstruction inversion algorithm is a rapidly developing algorithm, based on which quadratic penalty algorithms and augmented Lagrangian-type algorithms have been developed. These algorithms, through distributed iteration, alleviate nonlinearity and enhance the stability of inversion. Although such algorithms have changed the traditional optimization thought and narrowed the solution space under additional constraints, there has not been a discussion regarding the choice of norm forms (Aghamiry *et al*., 2021). Notably, in models with substantial background noise, the quadratic penalty cannot effectively optimize outliers and noise. Based on the quadratic penalty method:

$$\min_{\mathbf{X},\mathbf{M}} \mathrm{R}(\mathbf{M}) \quad \text{s.t.} \quad \mathbf{AX} = \mathbf{D}, \quad \mathbf{C}(\mathbf{M})\mathbf{X} = \mathbf{S}, \tag{3.5}$$

$$\mathrm{R}(\mathbf{M}) = \beta \|\mathbf{M}\|_{\mathcal{K}}^{\mathrm{sp}}, \tag{3.6}$$

where β is the regularization parameter. The values of β and $\mathcal{K}$ can be determined through cross-validation, in line with the principles of regularization-based techniques (Hastie *et al*., 2001). The new algorithm proposed in this chapter takes advantage of the stability of the ℓ_2 norm and uses the characteristics of the ℓ_1 norm to minimize the influence of anomalies on the model, thereby ensuring the algorithm's stability and improving robustness (Lai *et al*., 2014).

Another issue is that the non-differentiable nature brought by the ℓ_1 norm leads to a slow rate of normal solutions. Therefore, to speed up the calculation, this chapter will also adopt an ADMM algorithm, using distributed computing to reduce computational complexity, ensuring a certain level of sparsity while balancing computational efficiency.

3.3 Alternating Direction Method of Multiplier: ADMM

The ADMM introduced in this section operates on a distributed computing framework. During the optimization procedure, this algorithm can be used to optimize equation 3.6 and serve as an iterative form to solve the newly constructed problem (Aghamiry *et al*., 2019; Aghazade *et al*., 2022):

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}}(\left\|\mathbf{AX} - \mathbf{D}^k - \mathbf{D}\right\|_2^2 + \lambda\left\|\mathbf{C}(\mathbf{M}^k)\mathbf{X} - \mathbf{S}^k - \mathbf{S}\right\|_2^2), \tag{3.7}$$

$$\mathbf{M}^{k+1} = \arg\min_{\mathbf{M}}(\mathrm{R}(\mathbf{M}) + \lambda\left\|\mathbf{C}(\mathbf{M})\mathbf{X} - \mathbf{S}^k - \mathbf{S}\right\|_2^2), \tag{3.8}$$

$$\mathbf{S}^{k+1} = \mathbf{S}^k + \mathbf{S} - \mathbf{C}(\mathbf{M}^{k+1})\mathbf{X}^{k+1}, \tag{3.9}$$

$$\mathbf{D}^{k+1} = \mathbf{D}^k + \mathbf{D} - \mathbf{AX}^{k+1}, \tag{3.10}$$

where λ is the penalty parameter. $\mathbf{C}(\mathbf{M}) \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ is discretized partial differential equation (PDE) in constrain function, $\mathbf{S} \in \mathbb{R}^{\mathcal{N}\times 1}$ is the source term, the $\mathbf{X} \in \mathbb{R}^{\mathcal{N}\times 1}$ is the model wavefield, $\mathbf{D} \in \mathbb{R}^{\mathcal{M}\times 1}$ is observed data, $\mathbf{M} \in \mathbb{R}^{\mathcal{N}\times 1}$ are model parameters. Furthermore, with a finite number of iterations and updates, the mismatch function's error will gradually be corrected in a distributed form, thereby aiding the convergence of the wavefield and model parameters to the global minimum within an acceptable accuracy range (Gambella and Simonetto, 2020).

3.4 Numerical Simulations on Synthetic Data

In this section, the efficacy of the newly proposed algorithm is evaluated using three distinct synthetic datasets: Marmousi II, 2004 BP, and the two-dimensional SEG/EAGE Overthrust models. For forward modelling, a frequency-domain multi-scale strategy is employed. The PML algorithm is chosen for the boundary conditions (Pratt, 1999). To offer a quantitative comparison of the inversion outcomes, this chapter utilizes the RMS as the benchmark. This error provides a metric for the disparity between the adjusted model and the true model; a diminished Root Mean Square error suggests a more precise fit:

$$\mathrm{RMSe} = \sqrt{\frac{1}{\mathrm{N_X} \times \mathrm{N_Z}} \sum_{i=1}^{\mathrm{N_X} \times \mathrm{N_Z}} \left(\frac{\mathbf{M}_{\mathrm{true}} - \mathbf{M}_{\mathrm{inv}}}{\mathbf{M}_{\mathrm{true}}}\right)_i^2}, \tag{3.11}$$

where the $\mathrm{N_X}$ and $\mathrm{N_Z}$ indicate the sample size of the matrices for the true and simulated velocity models, respectively (Warner & Guasch, 2016).

The refined algorithm showcased in this chapter harnesses a blend of the ℓ_1 and ℓ_2 norms. As illustrated in Figure 3.1, the ℓ_2 norm expands in a quadratic fashion, which accentuates the detrimental impacts of noise. However, the ℓ_1 norm grows linearly at a relatively slower rate, making it less sensitive to noise, thereby suppressing noise. Another form of combining ℓ_1 norm and ℓ_2 norm is Huber regularization (Brossier *et al.,* 2010). This algorithm uses a threshold to regulate the scope of the two different norms, exhibiting ℓ_1 norm when below the threshold and ℓ_2 norm when above, which, although also a combined algorithm, can be decomposed and viewed as a simple combination of two functions. Therefore, especially when exceeding the threshold, it does not induce sparsity. In contrast, the proposed method presents a hybrid algorithm that still displays tighter constraints when exceeding the Huber norm threshold.

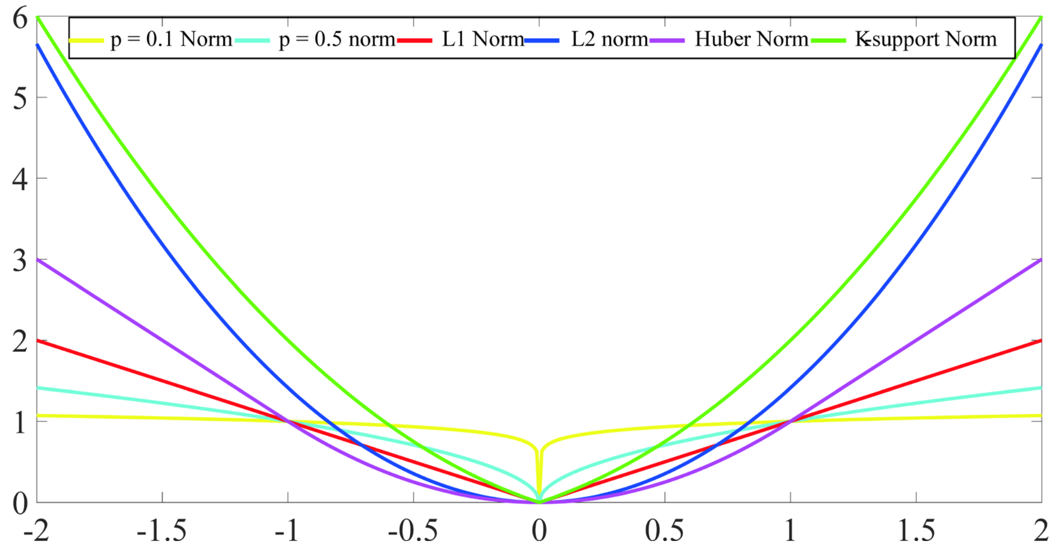

Figure 3.1. The linear growth of norms like p (for values 0.1 and 0.5), ℓ_1, ℓ_2, the Huber norm, and the K-support norm is depicted. The horizontal axis represents the weight and the vertical axis shows the norm's value (Li *et al*., 2024a).

Furthermore, Figure 3.2 provides top-view and contour diagrams of several norms. The ℓ_1 norm exhibits sparsity due to its non-differentiable vertices, while the ℓ_2 norm's contour and top-view are circular, displaying the smooth characteristic of vertices, indicating it might choose more tiny eigenvalues or eigenvectors. In contrast, Huber regularization shows non-differentiable vertices and smooth edges. Compared to them, the K-support norm demonstrates more distinctive features, as depicted in Figures 3.2 (d) and (h), representing the most basic form of this algorithm. As a hybrid algorithm, adjusting the k value allows the regularization term to exhibit either ℓ_2 norm or ℓ_1 norm characteristics solely. Hence, compared to other forms of norms, the new norm is more rational, robust, and flexible.

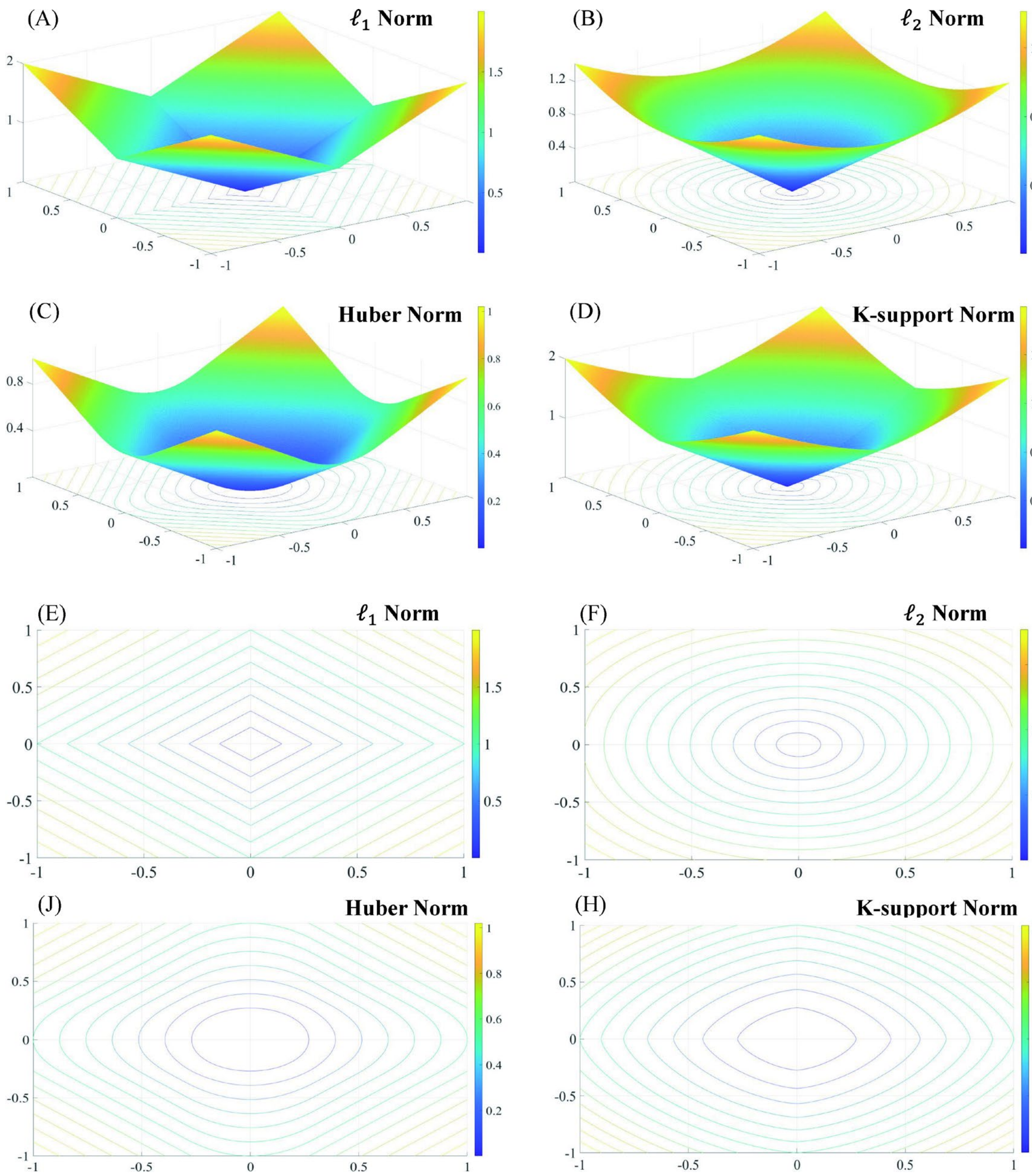

Figure 3.2. (a-d) Bird's eye view of the value surfaces for each norm. (e-h) contour graphs for each norm (Li *et al*., 2024a).

Additionally, the difference between this algorithm and the Huber algorithm should be emphasized. The Huber algorithm is smooth at vertices and almost linear near the theoretical vertices. In contrast, the K-support algorithm is different; its shape is between the diamond shape of ℓ_1 norm and the circular shape of ℓ_2 norm. The shape is mainly adjusted by the *k* parameter, which can be inflated to a circular shape or compressed to a diamond shape, so it is not a simple combination but adjusts the degree of constraint tightening through parameters. Therefore, especially in practical applications, when it is necessary to improve the model's generalization ability, a larger K-support norm value can be used as the regularization term. In contrast, if the aim is to elevate the model's interpretability or pare down its intricacy, a leaner K-support norm value might be more appropriate.

3.4.1 Marmousi II Model

Firstly, the K-support norm algorithm proposed in this chapter is tested on the Marmousi II model. The horizontal scale of this model is 17 km, and the vertical scale is 3.5 km. The inversion frequency range is composed of 1 Hz to 8 Hz. The model is surrounded by equally wide PML boundaries. Additionally, the initial model adopted is a one-dimensional linearly increasing velocity model, with a velocity range from 1 km/s to 4.25 km/s.

Figure 3.3 (a) represents the benchmark, and (b) is the starting model

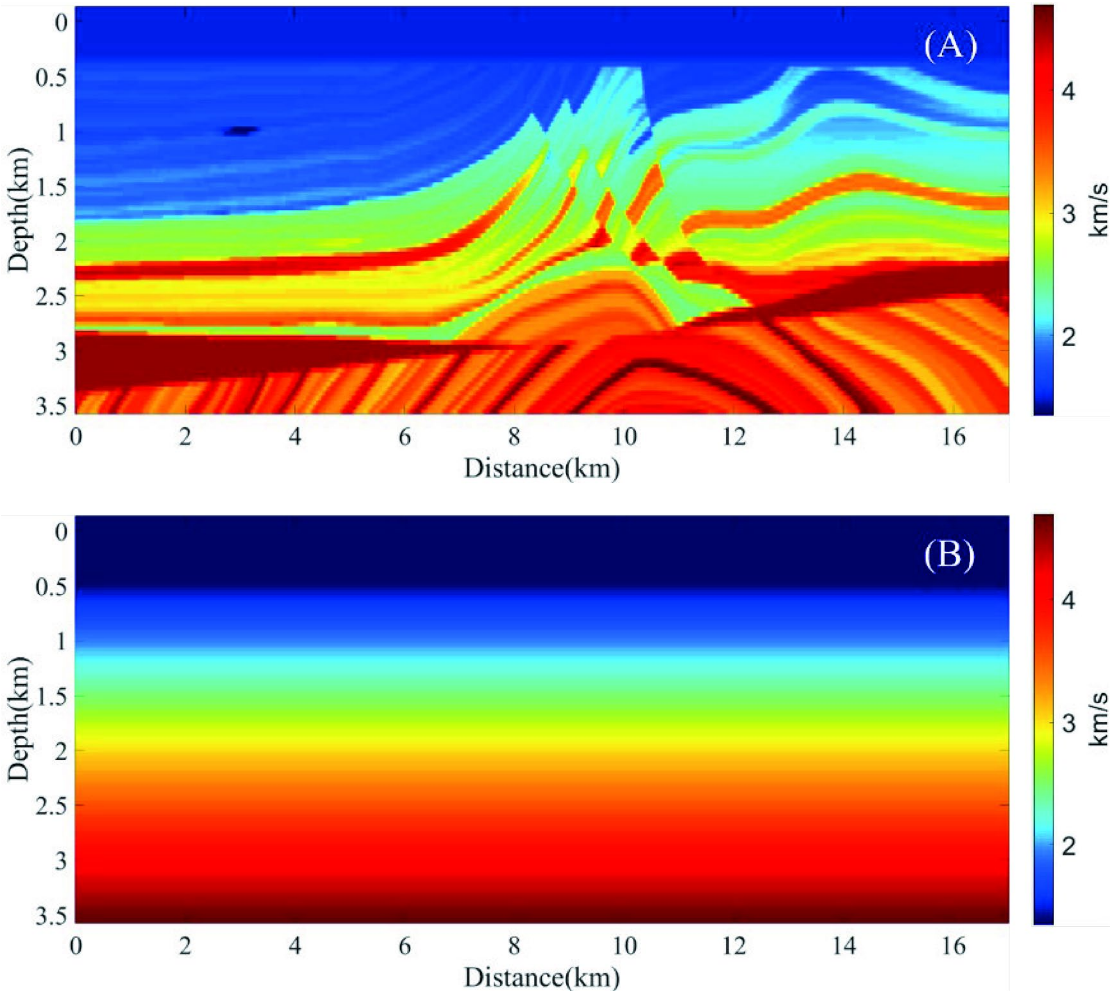

Figure 3.3. Marmousi II benchmark; (a) real benchmark; (b) starting model (Li *et al.*, 2024a).

Figure 3.4 (a) and (b) respectively show the inversion results based on Tikhonov regularization and K-support norm after adding 4.5 *dB* random noise, and Figure 3.4 (c) and (d) depict the velocity differences between them and the true velocity model. Traditional FWI algorithms are often sensitive to noise and artefact, exemplified by the pronounced artefact at the 4 km distance in Figure 3.4 (a). Moreover, the deep stratified structures from 10 km to 14 km exhibit significant degradation. The proposed algorithm, however, shows no distinct artefacts at the analogous position near 4 km. Relative to the conventional approach, it maintains a continuous velocity profile, avoiding the disruptions typically induced by noise. The inversion results for this specific layer are also notably more coherent, attesting to the capability of the proposed algorithm to mitigate the effects of artefacts.

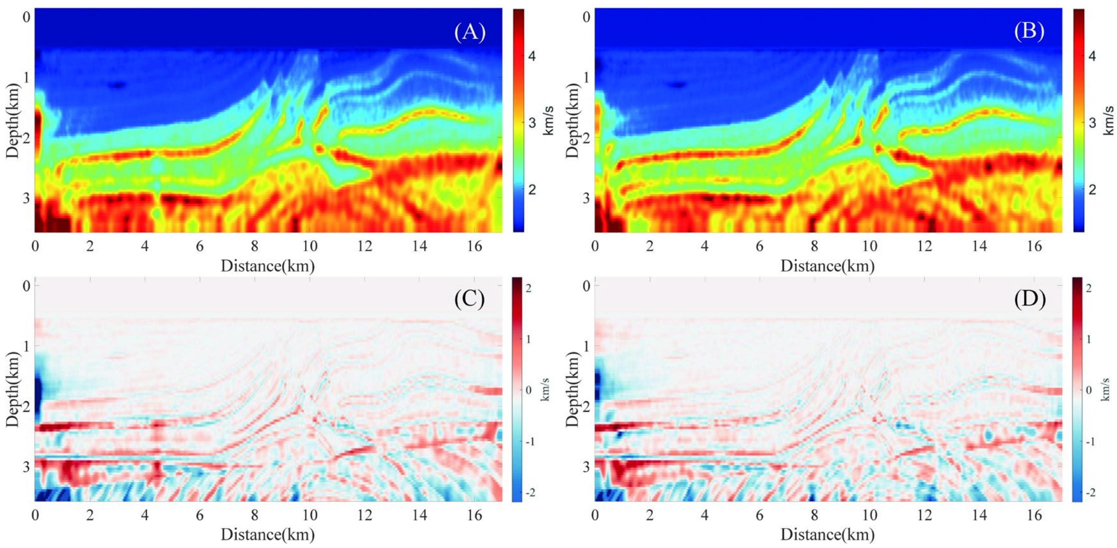

Figure 3.4. Marmousi II benchmark; (a) conventional FWI result; (b) modified method’s output; (c) difference of the conventional result and the real benchmark; (d) difference of the modified method’s output and the real benchmark (Li *et al*., 2024a).

Furthermore, the quantitative comparison of the inversion results is shown in Figure 3.5, where (a) compares the convergence speed of the misfit function of both, and (b) compares the RMSe convergence speed of both. Since ADMM distributed computing is adopted in this chapter, it is evident that the convergence speed of the optimization algorithm is significantly faster.

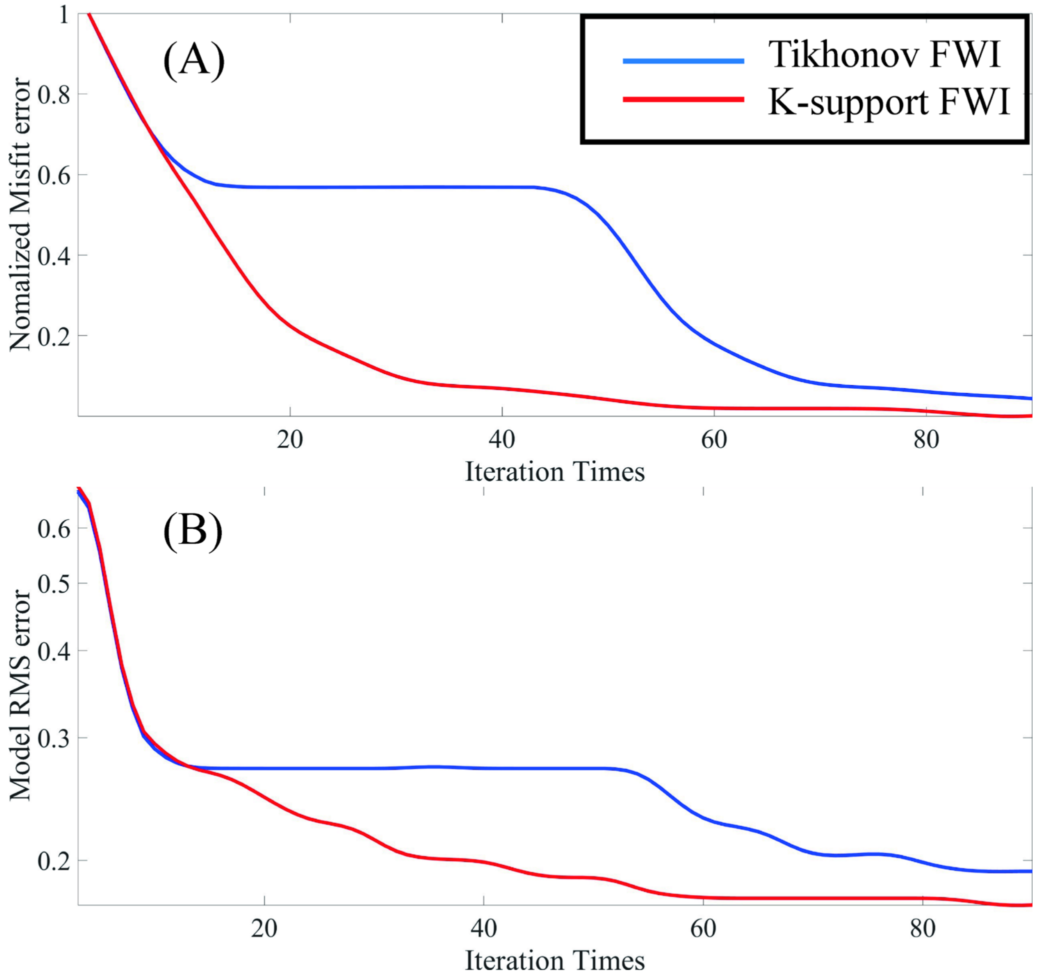

Figure 3.5. Error graphical representations; (a) misfit plot; (b) root mean square plot (Li *et al*., 2024a).

Additionally, to more significantly compare the inversion results of the two, Figure 3.6 presents six sets of one-dimensional vertical velocity comparison results. Among them, the red solid line represents the inversion results based on the K-support norm, the blue solid line is based on Tikhonov regularization, the black solid line is the true velocity, and the grey dashed line is the initial velocity. The velocity results represented by the red solid line based on the K-support norm fit better with the true velocity, indicating that the inversion results of the improved algorithm are also closer to the true velocity model.

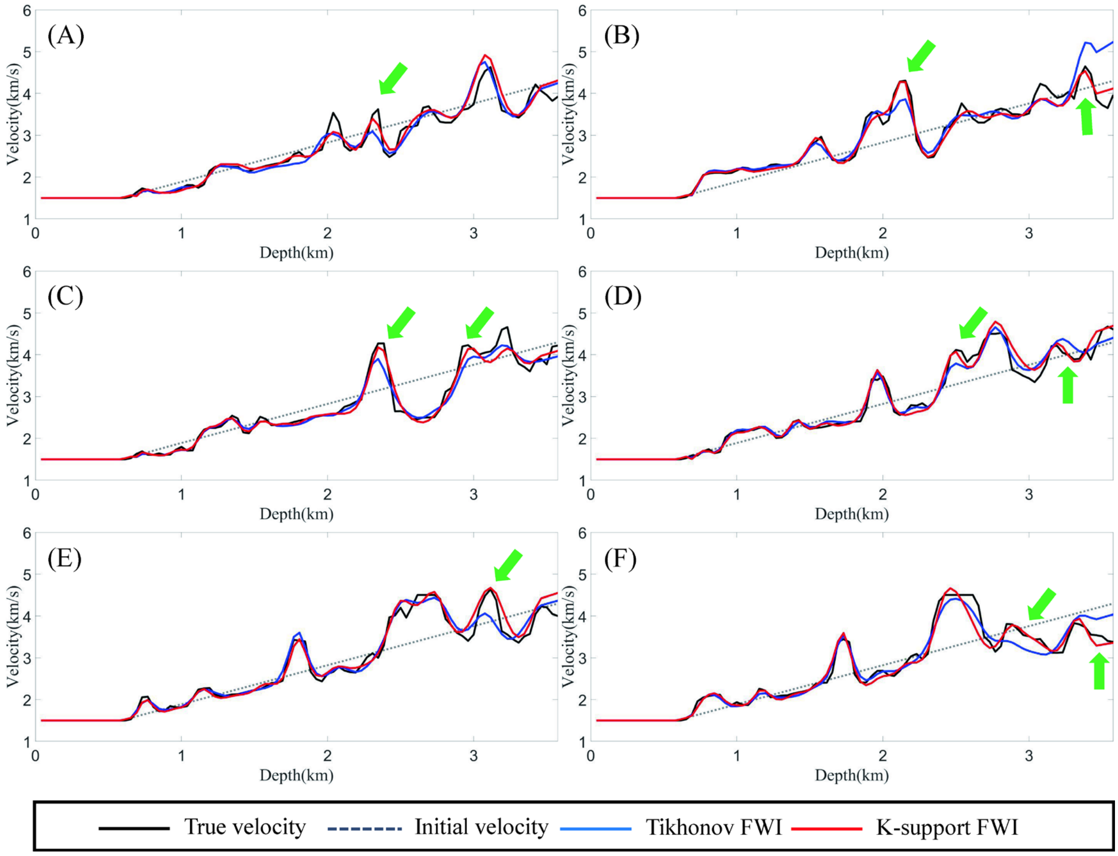

Figure 3.6. Marmousi II benchmark. one-dimensional velocity comparison at six *X*-coordinates separately in 8.72, 9.60, 10.92, 12.56, 13.40 and 15.28 km from (a) to (f) (Li *et al.*, 2024a).

Another more challenging comparison is the one-dimensional horizontal velocity comparison, as shown in Figure 3.7. Due to the propagation characteristics of waves and the greater difficulty of inversion in-depth, the horizontal velocity comparison is more challenging. It can be observed that the fitting effect of the blue solid line based on conventional FWI, especially in the deep part of the model, is relatively poor. In contrast, the improved algorithm is closer to the true velocity in the deep area, indicating that the algorithm proposed in this chapter, even in complex environments and under high background noise, still demonstrates superior performance and behaviour. For complex layered structures like Marmousi, accurately describing the velocities of different layers is crucial. In Figures 3.6 and 3.7, the green arrows highlight areas with significant velocity differences at various test directions and locations. In these areas, the inversion velocities obtained by the proposed algorithm (red solid line) are closer to the actual velocities (black solid line) compared to the results of the conventional algorithm (blue solid line), which show a more significant discrepancy. The areas indicated by green arrows indicate poorer performance of the conventional algorithm in describing the velocities of different layers, especially in deeper regions with a noticeable absence of accurate speed representation. For example, in Figure 3.6 (e), within a range of 3 km, the velocity obtained by the conventional algorithm is less than 4 km/s, which significantly differs from the actual velocity. In contrast, with its superior noise reduction capability, the proposed algorithm yields velocities closer to the actual speeds. For instance, at the exact location, the velocity result from the proposed algorithm is very close to the actual velocity.

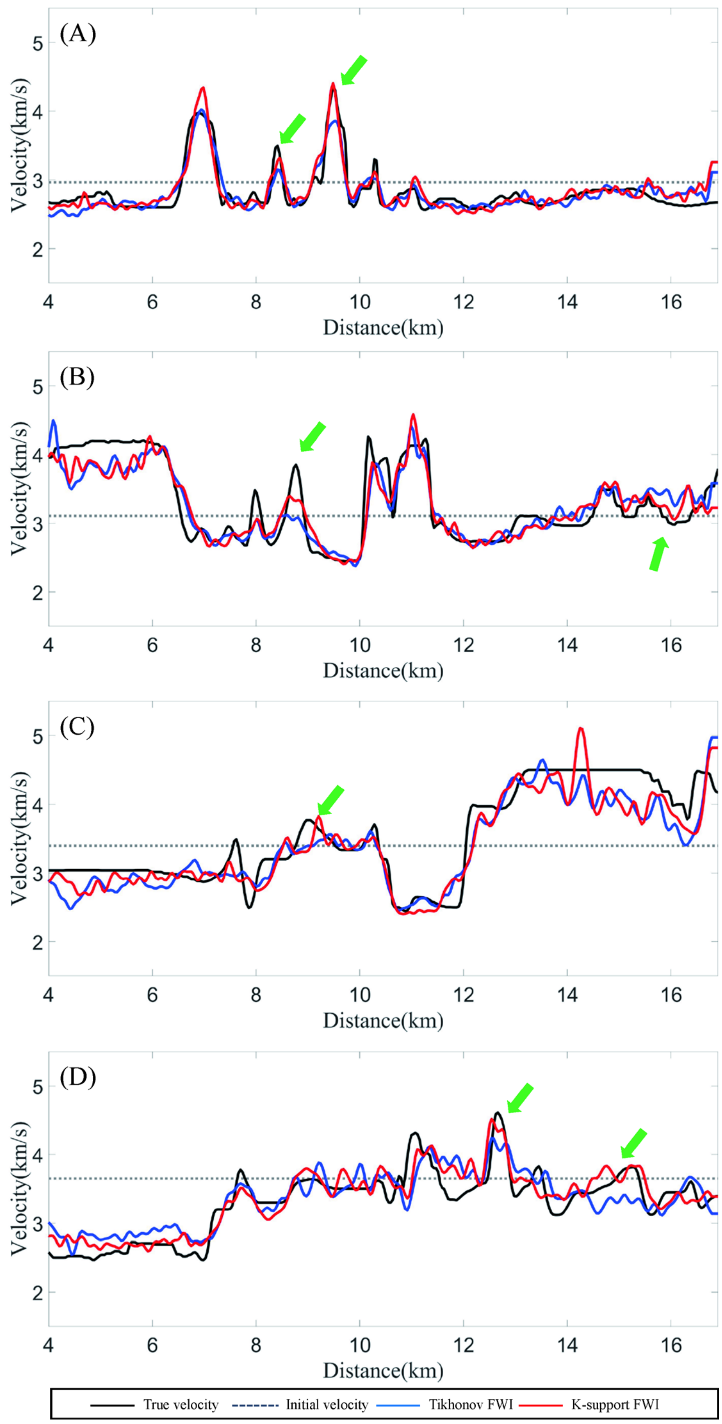

Figure 3.7. Marmousi II benchmark. one-dimensional velocity models at four *Y*-coordinates positions separately in 2.24, 2.40, 2.72, and 3.00 km from (a) to (d) (Li *et al*., 2024a).

3.4.2 The Central Part of the 2004 BP Model

Further, to test the performance of the K-support norm, this section selects another very important model, the 2004 BP model, as a benchmark. The model consists of three parts, with the middle part being used in this chapter. The central portion of this model depicts the underground geological structures of the eastern Gulf of Mexico and offshore Angola. Notably, at the heart of this model is a substantial high-velocity salt formation, outlining this salt body becomes a challenge for inversion. Another challenge lies in accurately inverting, especially for the two salt pillars existing beneath the strong reflective surface, the details of the deep part of this model under complex conditions such as high background noise.

Figure 3.8 (a) showcases the actual benchmark for the central section of the 2004 BP model, while (b) reveals the initial velocity model, which still employs a one-dimensional velocity model with a linear increase, spanning velocities from 1.5 km/s to 5 km/s.

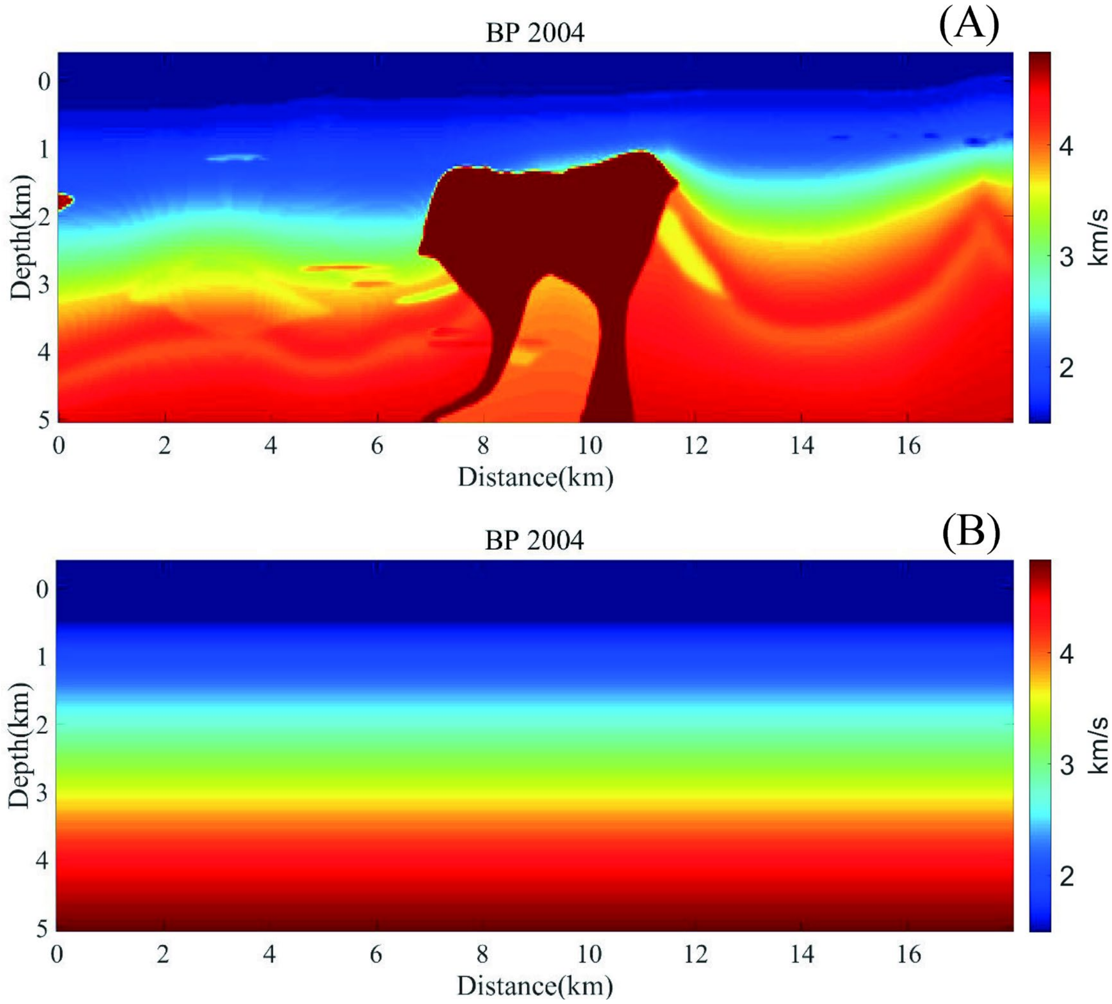

Figure 3.8. The central part of the 2004 BP benchmark; (a) true benchmark; (b) starting model (Li *et al*., 2024a).

Moving on, Figure 3.9 (a) illustrates noise-free data, (b) shows the 4.5 *dB* random noise, and (c) the wavefield post the linear overlay of random noise on the original data. Subsequent experiments in this chapter utilize the high ambient noise scenario displayed in Figure 3.9 (c). The outcomes of the inversion tests are displayed in Figure 3.10, where (a) is the outcome using Tikhonov regularization, while (b) uses the K-support norm FWI. A review of these outcomes reveals that the conventional algorithm’s inversion results are less satisfactory under low signal-to-noise conditions, particularly in the model’s deeper sections where salt pillars become indistinct and there's a lack of velocity. Conversely, the refined algorithm offers notably enhanced inversion results.

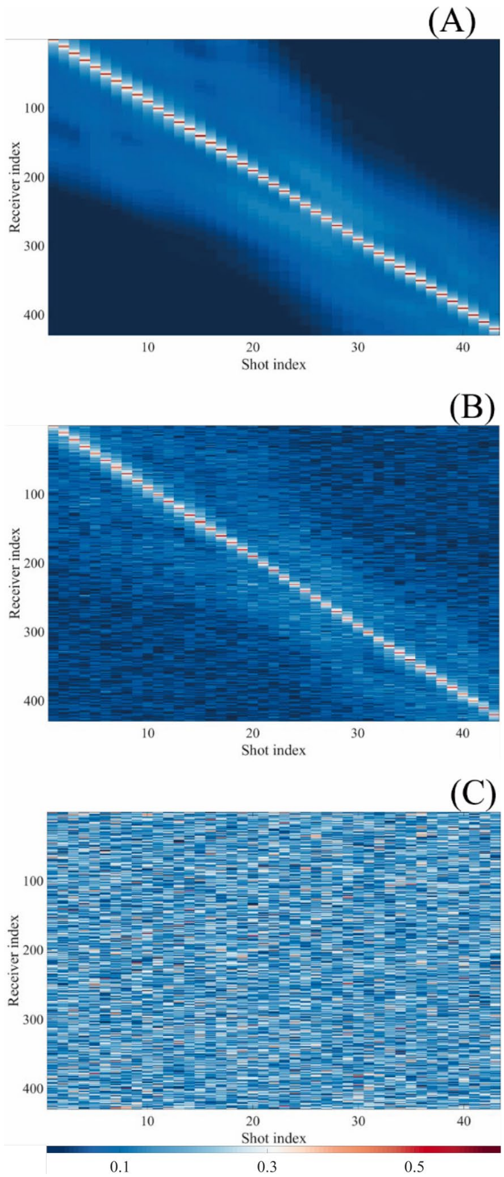

Figure 3.9. Mono-frequency data sets with (a) noise-free; (b) with noise data (c) 4.5 *dB* random noise data (Li *et al*., 2024a).

Figure 3.10 (c) and (d) highlight the variances between the two algorithm outcomes and the true benchmark. Figure 3.10 delineates the error distribution for inversion results using two algorithms on the 2004 BP velocity model. The left column displays the errors from a conventional algorithm, whereas the right column shows those from a proposed algorithm. The magnitude of discrepancies is indicated by colour intensity, with blue reflecting velocities that are underestimated and red representing velocities that are overestimated compared to the actual model. Darker hues of blue or red indicate significant variances between the inversion results and the true model. From the onset at lower frequencies, the proposed algorithm exhibits a swifter rate of convergence and a higher proficiency in denoising, particularly in the depiction of continuous velocities around the salt dome, evidenced by smaller dark regions within the salt bodies. With an increment in iterative processes, the susceptibility of the conventional algorithm to noise becomes more evident, as seen at 1.79 Hz and 5.35 Hz, where artefacts are markedly more distinct. Conversely, the proposed algorithm displays a more robust capacity for noise attenuation at these same frequencies.

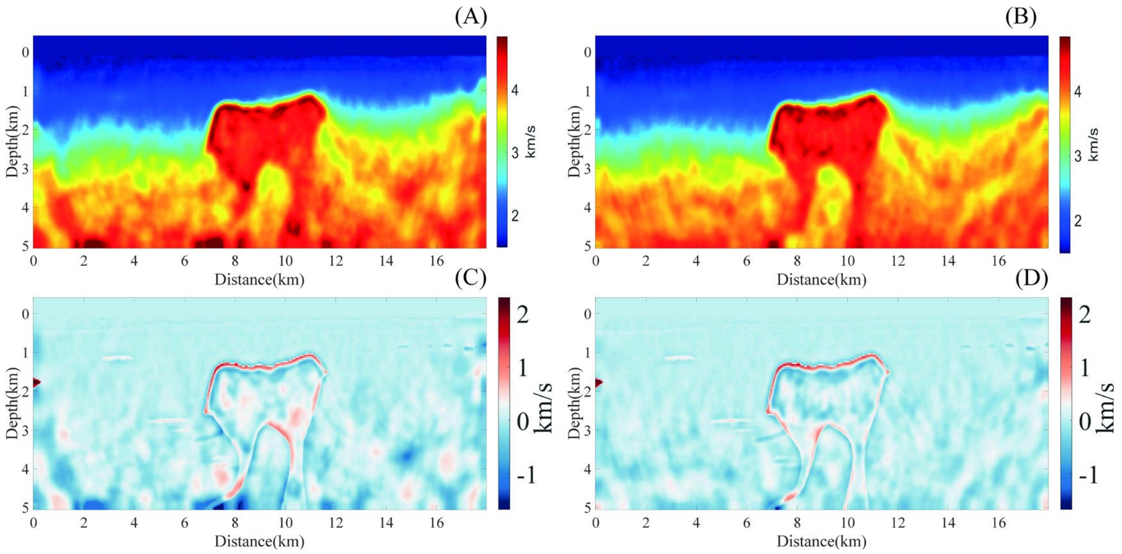

Figure 3.10. The central part of the 2004 BP benchmark; (a) conventional method's results; (b) modified method's results; (c) difference between conventional results and the benchmark; (d) difference between the modified method's results and the benchmark (Li *et al*., 2024a).

Expanding on this, through the multi-scale inversion approach, Figure 3.11 juxtaposes the inversion outcomes of both algorithms across various frequency bands. Here (a-f) represent results from the Tikhonov regularization, (g-l) are derived from the K-support norm, (m-r) depict the differences between the Tikhonov FWI results and the true benchmark, and (s-x) outline the differences between K-support norm FWI outcomes and the authentic velocity model. These findings point out that the refined algorithm boasts a notably quicker convergence rate, evident from the swifter rendering of the salt body framework in the low-frequency domain (below 3 Hz). In this experiment's extremely low signal-to-noise ratio, accurately delineating salt body contours and surrounding velocity becomes difficult amidst high noise levels. Utilizing a multiscale

FWI approach, the low-frequency results indicate that the inversion employing a quadratic penalty in conjunction with the K-support norm surpasses conventional algorithms in convergence speed. The improvement stems from leveraging distributed computing and imposing stricter constraints, facilitating the convergence of the objective function to the global minimum and enhancing the low-frequency fidelity of the proposed algorithm. Results at 5.35 Hz demonstrate that the proposed algorithm yields a more precise representation of the salt body base and avoids the significant velocity anomalies apparent in the results obtained by the conventional algorithm. Additionally, it roughly outlines the shapes of adjacent channels. Consequently, FWI informed by the K-support norm can somewhat insulate against outlier interference, facilitating quicker attainment of correct inversion solutions, with particular efficacy in modelling salt bodies and channels.

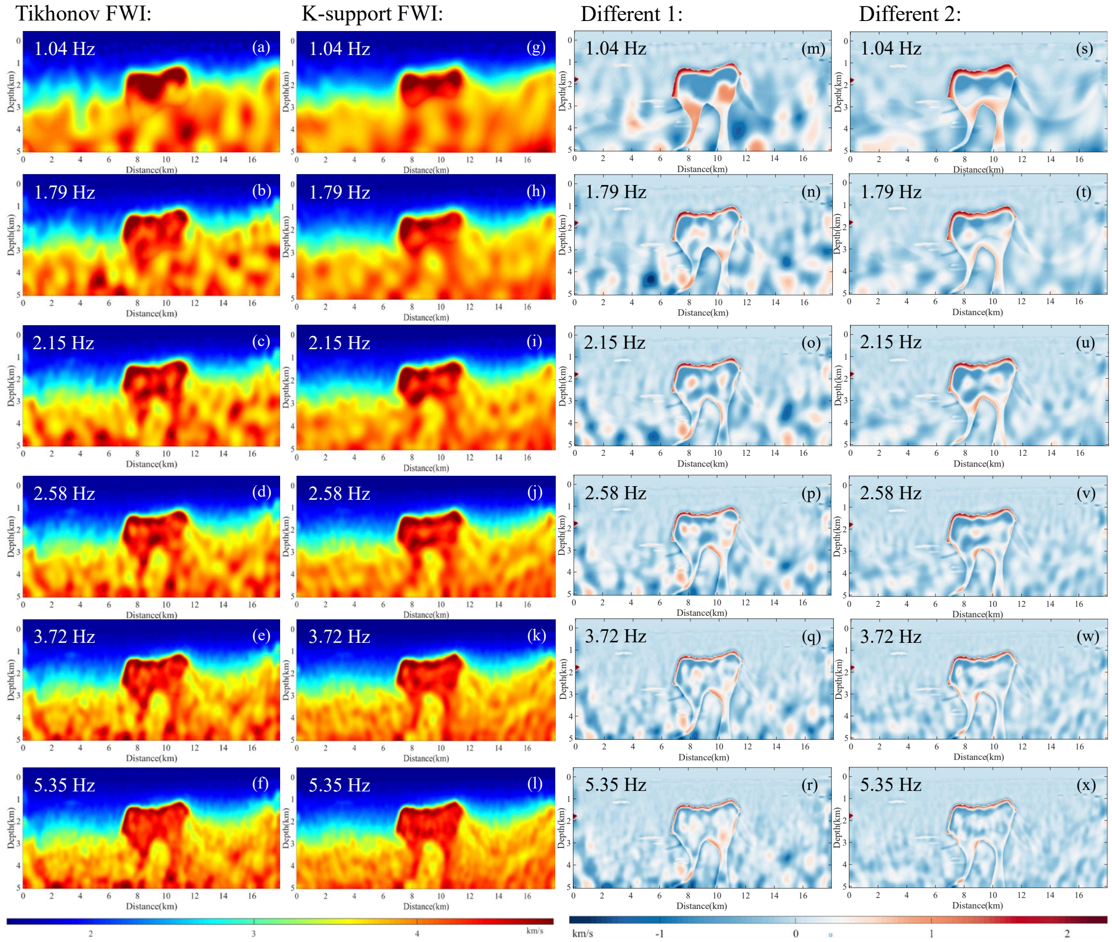

Figure 3.11. The central part of the 2004 BP benchmark; (a-f) conventional method's outputs in 1.04, 1.79, 2.15, 2.58, 3.72, and 5.35 Hz, respectively; (g-l) K-support norm results in the same frequency range; (m-r) differences of the conventional results and the benchmark; (s-x) differences of the modified method's outputs and the benchmark (Li *et al.*, 2024a).

Moreover, Figure 3.12 (a) sets side by side the misfit functions of both algorithms, highlighting that the K-support FWI (depicted by the red line) converges more rapidly than its conventional counterpart (shown by the blue line); (b) contrasts the RMS error between the two, with the enhanced algorithm still showcasing superior efficacy.

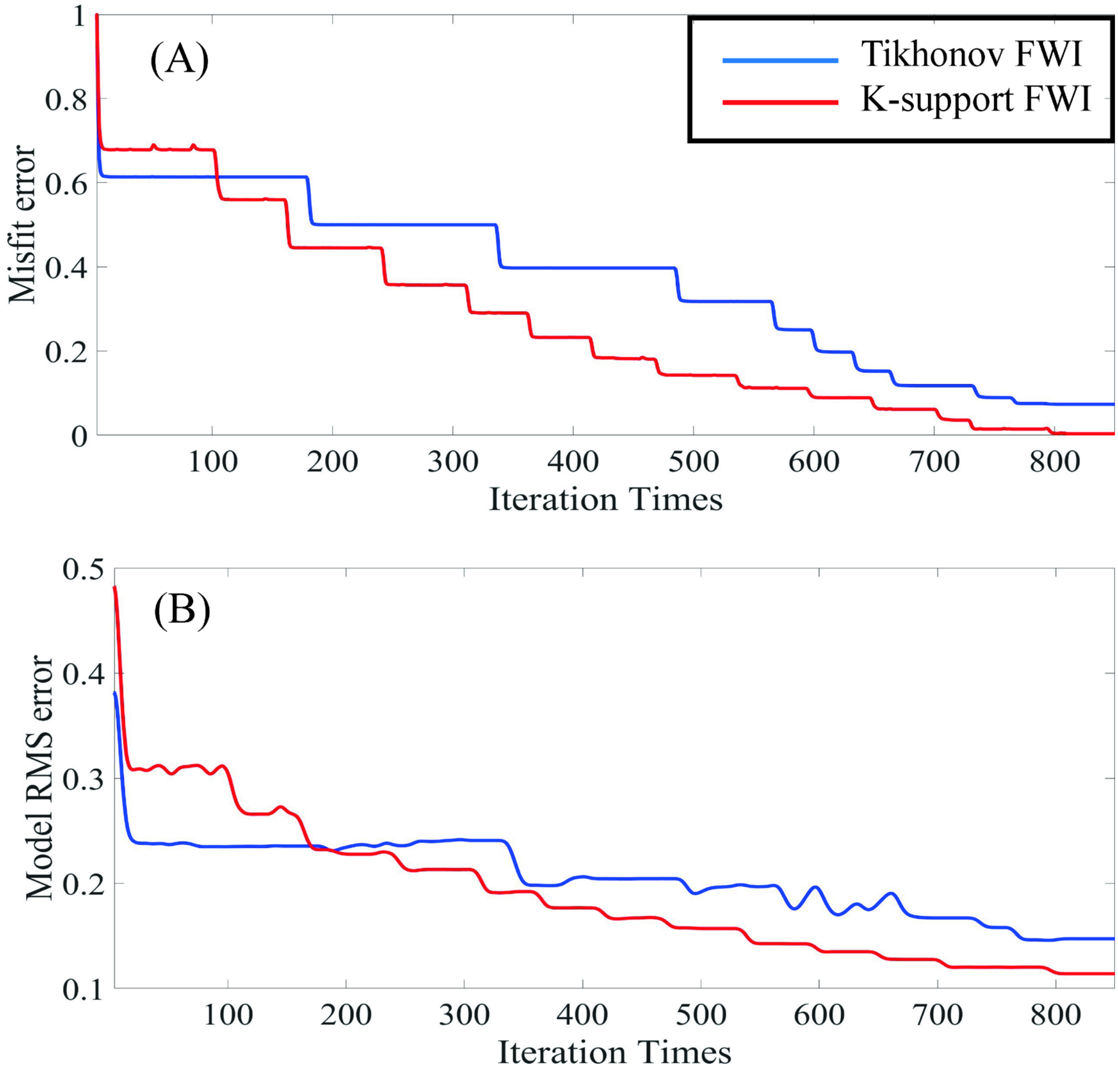

Figure 3.12. Error plot; (a) misfit; (b) RMS (Li *et al*., 2024a).

Additionally, this section still adopts a one-dimensional velocity comparison to perform a more accurate quantitative comparison of the inversion results of the two algorithms. Figure 3.13 shows the vertical velocity comparison at six different locations for the two algorithms, where the green arrows mark the places with significant performance differences between the two algorithms. It can be seen that, under complex conditions with low signal-to-noise ratios, the vertical velocity shows that the improved algorithm has better fitting results.

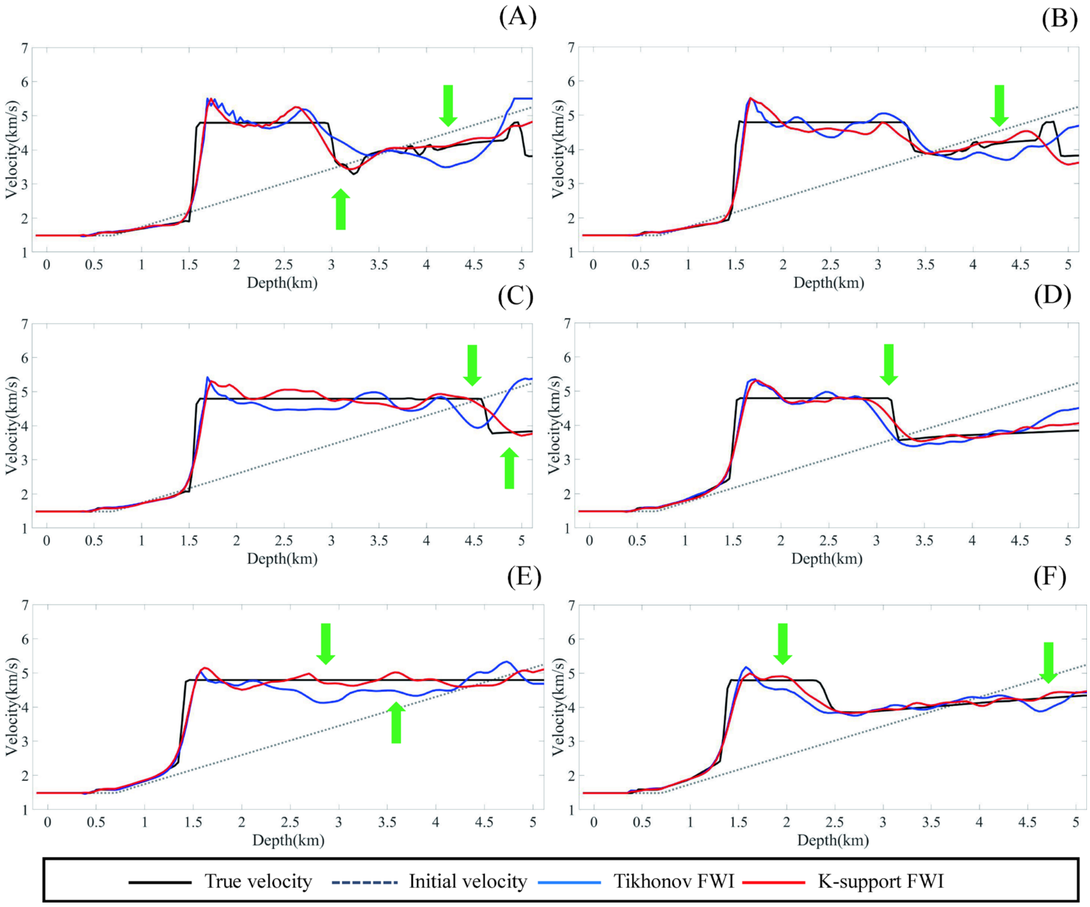

Figure 3.13. One-dimensional velocity at varying X-coordinates; for X = 7.4, 7.8, 8.2, 9.8, 10.4, and 11.3 km separately in (a-f) (Li *et al.*, 2024a).

Moreover, Figure 3.14 compares the horizontal velocities of the two algorithms, compared to the vertical comparison, the horizontal comparison better demonstrates the superior performance of the improved algorithm, especially in the deep areas where the red solid line is closer to the black solid line, while the blue solid line exhibits significant velocity errors. Using both evaluation techniques, it's evident that the K-support norm outperforms in terms of noise reduction and precision of inversion outcomes at six distinct horizontal and vertical placements. Particularly in areas with high velocities, its peaks align more closely with the actual velocity, and notably, no marked velocity discrepancies are observed in zones with low velocities.

The characteristics of the BP model are somewhat distinct from those of the Marmousi model. In conditions with low signal-to-noise ratio noise, accurately inverting the continuity of the surrounding velocities and precisely delineating the contours of the salt dome is of utmost importance. As illustrated by the green arrows in Figure 3.13, due to the pronounced interference of noise on the continuous medium, the conventional algorithm yields inversion results for the suboptimal surrounding velocities, maintaining a consistent deviation from the actual velocities. In contrast, the proposed algorithm, with its enhanced denoising efficacy, accurately inverts velocities in less variable continuous media, exemplified in Figure 3.14 (e) where, within a depth range of 2 km to 4 km, its inversion results (red solid line) closely approximate the actual velocities (black solid line), unlike the conventional algorithm (blue solid line) which consistently displays velocity discrepancies. Moreover, the arrows in Figure 3.14 underscore the comparative accuracy of both algorithms, especially in depicting the contours of columnar salt structures. The inversion results of the proposed algorithm (red solid line) show markedly better fitting to the velocity profiles at the contours of the salt bodies, areas characterized by significant velocity changes. In contrast, the conventional algorithm often exhibits pronounced velocity underestimations or overestimations.

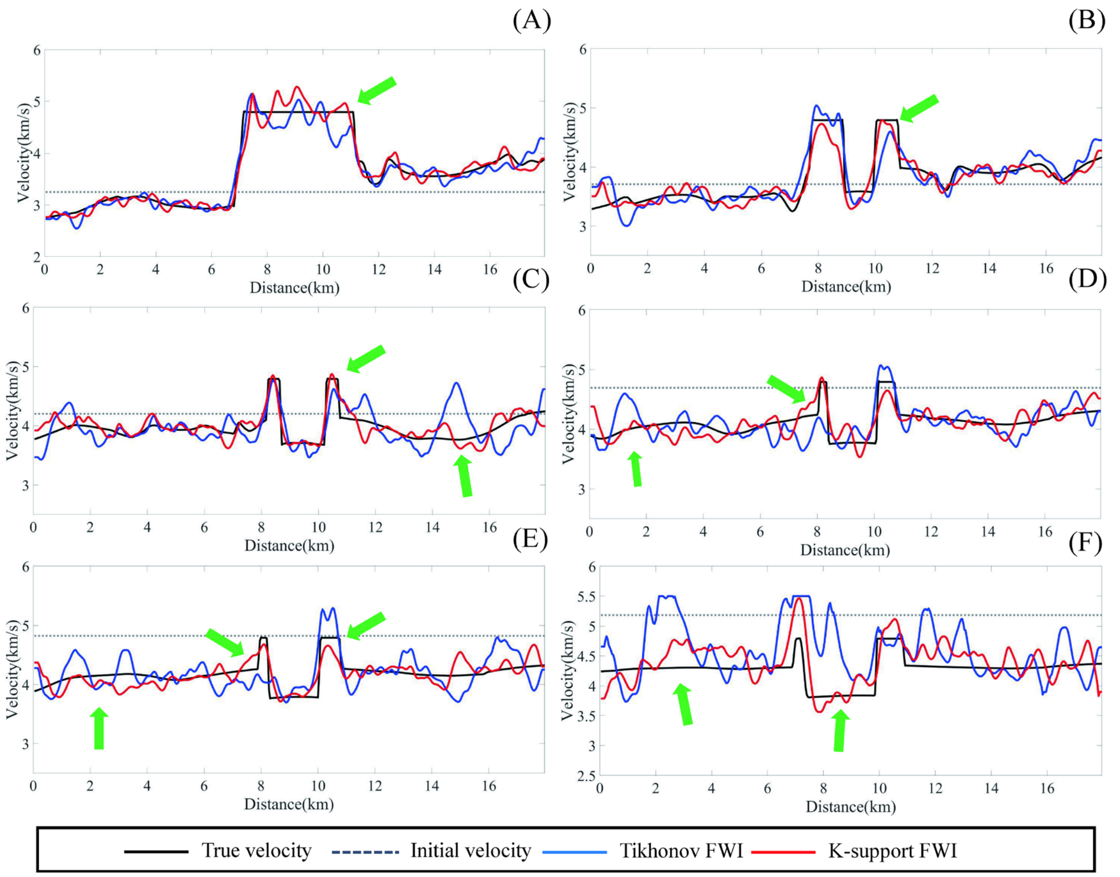

Figure 3.14. One-dimensional velocity at varying *Y*-coordinates; for *Y* = 2.83, 3.35, 3.91, 4.47, 4.61, and 5.03 km separately in (a-f) (Li *et al*., 2024a).

3.4.3 SEG/EAGE Overthrust Model

This section extends its tests to the algorithm introduced in this chapter, applying it to the two-dimensional SEG/EAGE Overthrust model. Recognized as a standard benchmark in FWI evaluations, the SEG/EAGE Overthrust model embodies geological formations characterized by thrust faulting. Figure 3.15 (a) provides a visualization of the actual velocity model, while Figure 3.15 (b) displays the starting velocity model for this section, which is a more generalized or smoothed version.

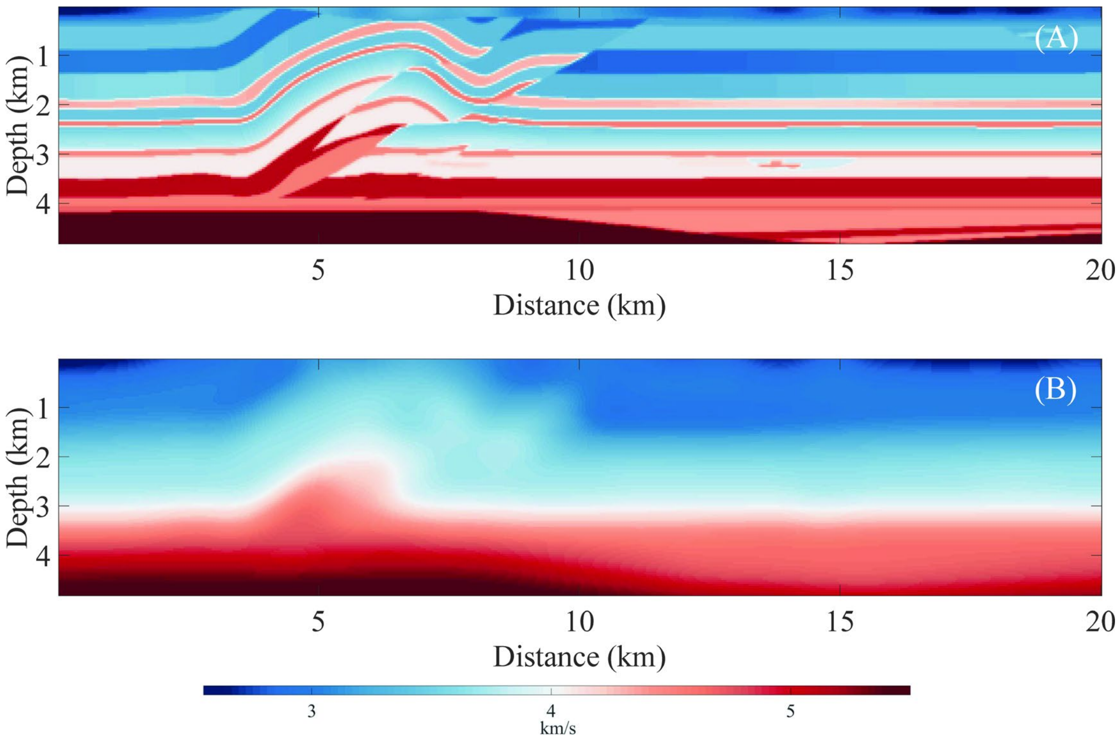

Figure 3.15. Two-dimensional overthrust benchmark; (a) benchmark; (b) starting medium (Li *et al*., 2024a).

To mimic intricate scenarios, Figure 3.16 presents the noise-impacted data for this experiment; (a) highlights the pure data, (b) the data tainted with 4.5 *dB* of random noise, and (d) the data where random noise is linearly overlaid onto the pure data.

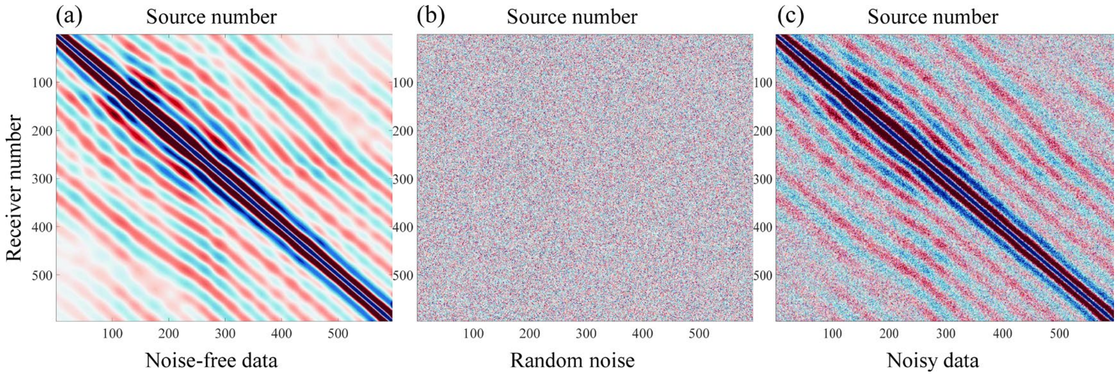

Figure 3.16. Noise-impacted data according to the amplitude for this experiment; (a) pristine data; (b) random noise; (c) noisy data. (Li *et al*., 2024a).

Based on the aforementioned conditions, Figure 3.17 provides a comparison of inversion results between Tikhonov regularization-based FWI and K-support norm-based FWI. From (a-e), I see the inversion results of multi-scale FWI based on Tikhonov regularization at five different inversion frequencies. Meanwhile, (f-j) depicts the inversion results of multi-scale FWI based on the K-support norm at these frequencies. Furthermore, Figure 3.17 (k-o) spotlights the discrepancies between the multi-scale FWI inversion results using Tikhonov regularization and the true benchmark. In contrast, (p-t) underscores the deviations between the results of the K-support norm-based multi-scale FWI and the genuine velocity model. From the results, it's evident that the inversion outcomes using the improved algorithm, especially in the layered structure sections, are much clearer. In contrast, results based on the conventional algorithm are quite blurred, failing to effectively distinguish between different geological layers. The inversion complexity for the overthrust model is centred on the precise inversion of velocities for the high-velocity layers masked by several overlaying strata, the accurate delineation of velocities for each stratigraphic level of the overburden, and the characterization of the high-velocity base layer. The proposed algorithm, particularly at low frequencies, exhibits significant differences from the conventional outcomes due to the implementation of tighter constraints. The algorithm's convergence rate is notably accelerated, and its inversion results at high frequencies are strikingly accurate and precise, evidencing a substantial performance improvement over traditional methods. Additionally, faster convergence and enhanced noise attenuation of the proposed algorithm translate into a more detailed and accurate representation of the base around the depth of 4 km. The conventional algorithm, in contrast, encounters conspicuous artefacts and indistinct stratigraphic inversion at the base and left side, mainly due to noise-induced velocity losses and local minima. By integrating a novel regularization technique and optimized inversion steps, the proposed algorithm has bolstered overall convergence velocity and artefact resistance, showcasing its superior capabilities in the context of commonly used synthetic datasets.

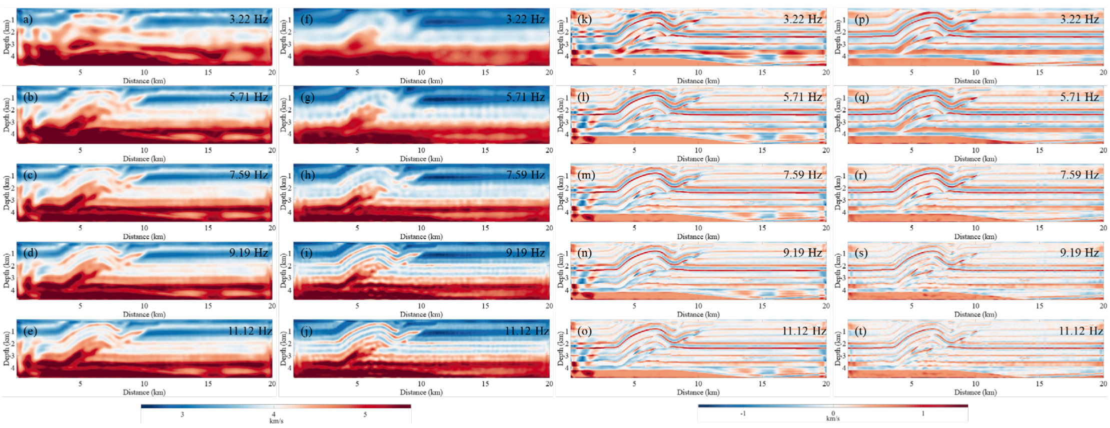

Figure 3.17. Two-dimensional overthrust benchmark; (a-e) conventional method’s outputs in 3.22, 5.71, 7.59, 9.19, and 11.12 Hz, respectively; (f-j) modified method’s outputs in the same frequency range; (k-o) differences of conventional FWI results and the benchmark; (p-t) differences of the modified method’s results and the benchmark (Li *et al.*, 2024a).

Figure 3.17 (k-t) illustrates the discrepancy distribution between the inversion results derived from two algorithms and the actual velocity model of the SEG/EAGE Overthrust model. Figure 3.17 (k-o) delineates the error outcomes from the conventional algorithm, while Figure 3.17 (p-t) conveys those from the proposed algorithm. The intensity of the colour indicates the magnitude of the errors: blue and red indicate lower and higher inversion velocities than the true velocities, respectively. The proposed algorithm demonstrates superior noise suppression and anomaly attenuation capabilities, particularly in the deeper and lateral regions of the model, as indicated by the fewer intense colourations in the right column compared to the left. Notably, in regions beyond a depth of 4 km, inversion results obtained by the proposed algorithm do not exhibit profoundly dark areas, contrary to the conventional algorithm, which shows significant deep-coloured discrepancies on both sides of the model. Figure 3.17 substantiates the assertion that the proposed algorithm offers a markedly improved denoising efficacy, especially at greater depths within the model structure.

Furthermore, a quantitative comparison is presented in Figure 3.18, featuring six sets of vertical one-dimensional velocity comparisons. The conventional algorithm's inversion results are indicated by the blue solid line, in contrast, the red solid line denotes the outcomes from the enhanced algorithm, the black solid line signifies the actual velocity and the grey dashed line represents the preliminary velocity.

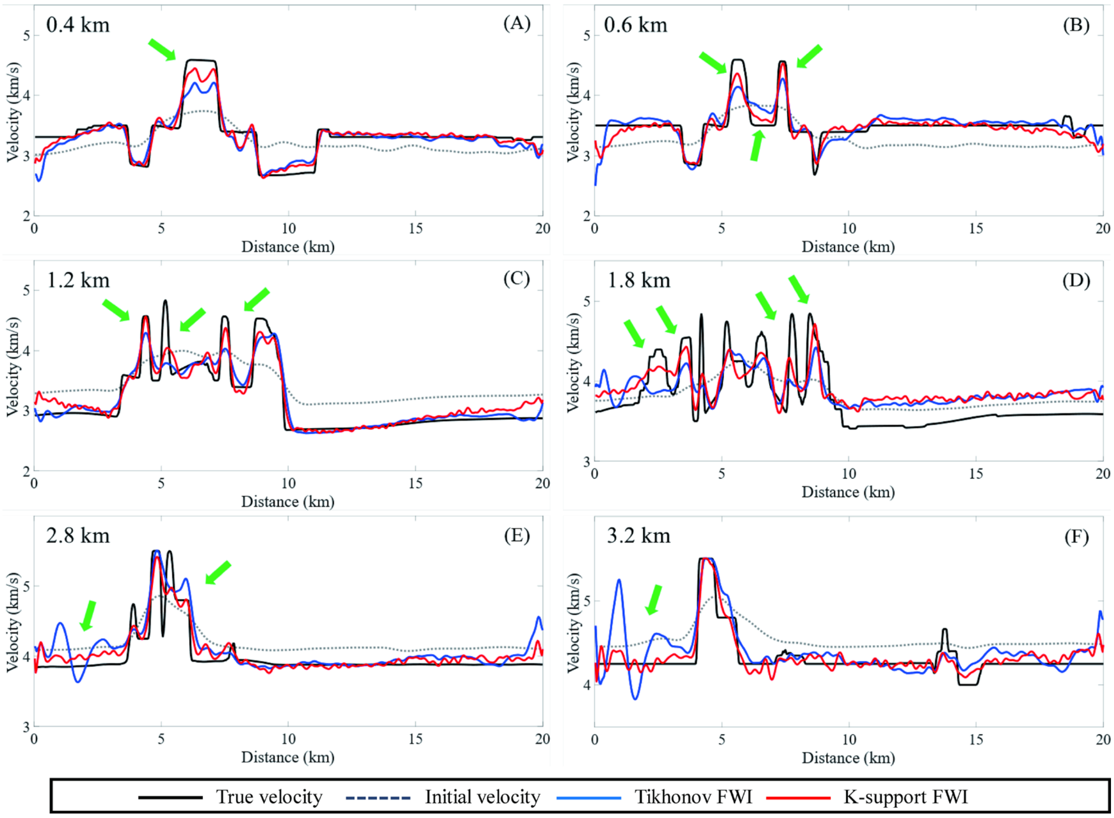

Figure 3.18. One-dimensional velocity at varying *Y*-coordinates; for *Y* = 0.4, 0.6, 1.2, 1.8, 2.8, and 3.2 km separately in (a-f) (Li *et al*., 2024a).

Figure 3.19 provides six sets of horizontal velocity comparisons. In both one-dimensional velocity comparisons, the fit of the improved algorithm to the true results is notably closer, especially for the differentiation of rock layers and imaging in deeper regions. The Overthrust model emulates the structural complexity of thrust faulting and overthrust sequences under significant tectonic stresses. As with the Marmousi archetype, precise delineation of velocity profiles across varied stratigraphic horizons is essential. In Figures 3.18 and 3.19, locations indicated by green arrows demonstrate the comparative performance of two distinct algorithms in rendering stratigraphic velocity profiles and their ability to articulate continuous velocity structures. The green arrows elucidate that although the inversion velocities from the proposed algorithm do not entirely align with the actual velocities within this model, they significantly outperform the traditional algorithm, most notably in the seamless reconstruction of velocities within the 0 km to 4 km depth interval, as depicted in Figures 3.18 (e) and (f). Additionally, the green arrows in Figure 3.19 highlight that despite some residual velocity underestimation by the proposed algorithm across different stratigraphic layers, it nonetheless marks a discernible advancement over conventional velocity estimations.

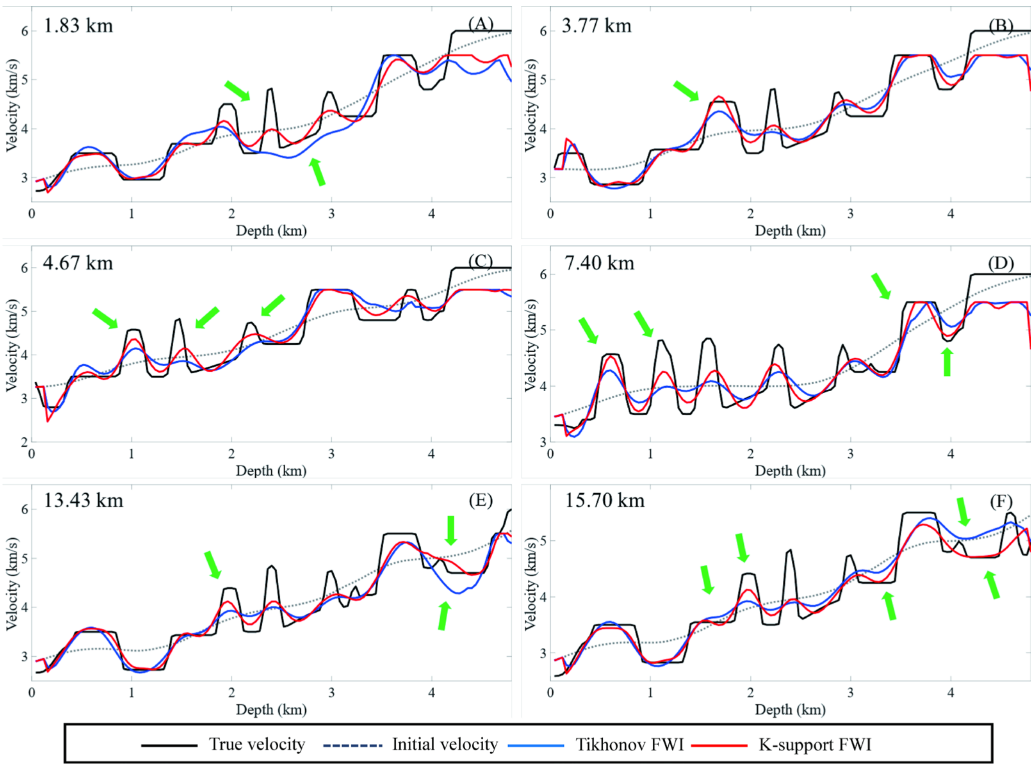

Figure 3.19. One-dimensional velocity at varying X-coordinates; for X = 1.83, 3.77, 4.67, 7.40, 13.43, and 15.70 km separately in (a-f) (Li *et al*., 2024a).

Moreover, this section introduces another type of coherent noise to distinguish from random noise, testing the K-support norm in the presence of coherent noise. As shown in Figure 3.20, (a) represents clean data, (b) a type of coherent noise, and (c) the data containing the coherent noise.

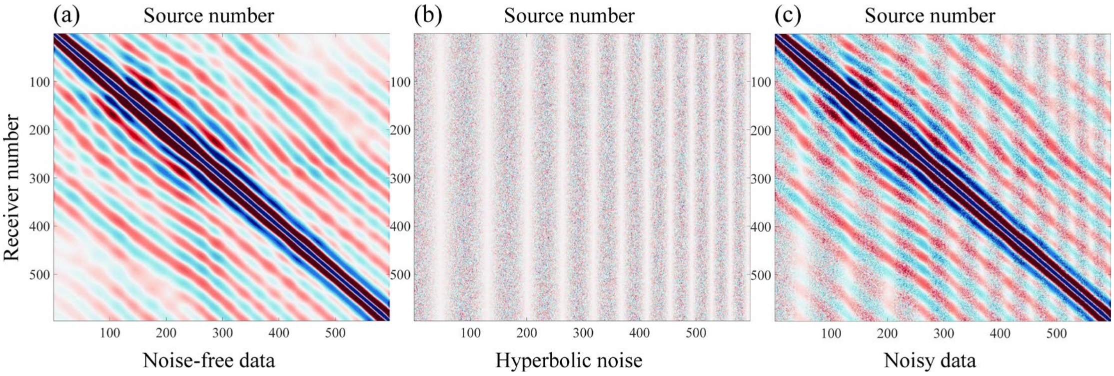

Figure 3.20. Noise-impacted data according to the amplitude for this experiment; (a) pristine wavefield; (b) coherent noise; (c) with noise wavefield (Li *et al*., 2024a).

Finally, Figure 3.21 presents the inversion results of both algorithms under the influence of coherent noise: (a-d) are based on the conventional algorithm and (e-h) K-support norm FWI results. Even with the interference of coherent noise, the improved algorithm still yields superior inversion results, proving its effectiveness not only under random noise conditions but also under coherent noise scenarios.

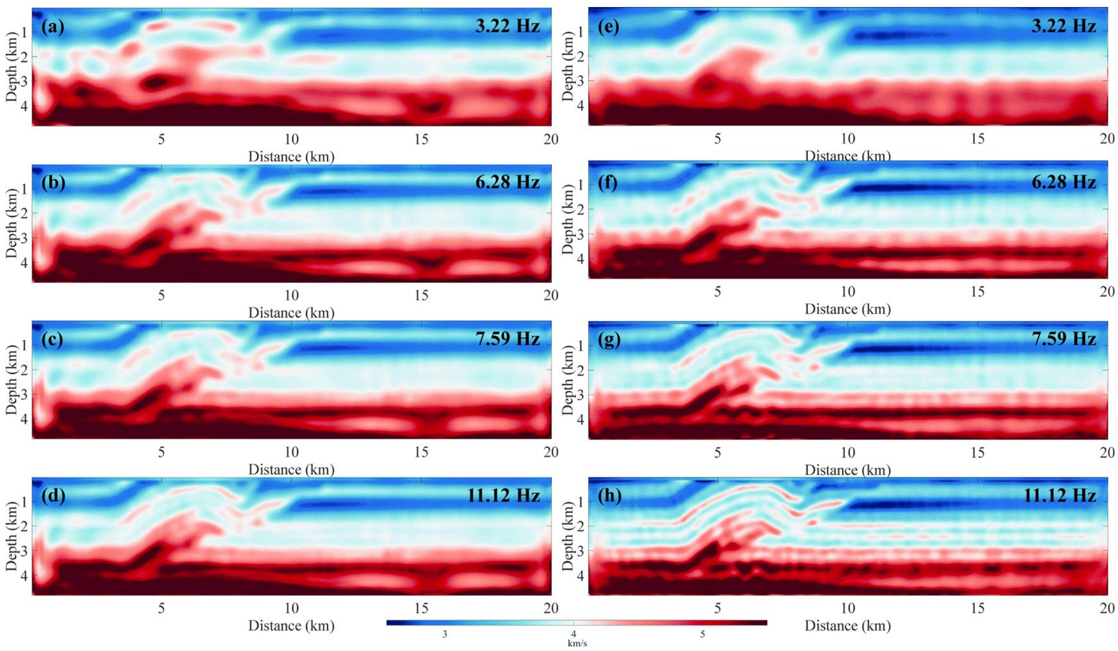

Figure 3.21. Two-dimensional overthrust benchmark; (a-d) conventional method's outputs in 3.22, 6.28, 7.59, and 11.12 Hz, respectively; (e-h) modified method's outputs in the same frequency range (Li *et al*., 2024a).

3.5 Chapter Discussions

The FWI algorithm, as a numerical fitting method, heavily relies on the mathematical optimization process. This optimization can be divided into two parts: 1. Constructing an appropriate misfit function, and 2. Minimizing this constructed misfit function. From these perspectives, building an objective function with more constraints and quickly minimizing this function are the focal points for enhancing FWI performance.

Regularization algorithms add a constraint term to the objective function. By constraining the solution space in a certain way, the likelihood of solutions falling into local minima caused by outliers is reduced. Moreover, due to the introduction of a new constraint term, the range of the solution space is narrowed, leading to increased convergence speed. In practice, the seismic information we collect is often non-prior. Therefore, compared to simply using a single norm form, the algorithm proposed in this chapter has better adaptability. Furthermore, it allows subsequent parameter adjustments to further optimize inversion resolution, eliminating the need for repetitive calculations due to norm selection, thus significantly enhancing computational efficiency (van Leeuwen & Herrmann, 2016).

Compared to conventional algorithms such as Tikhonov regularization and TV regularization, the improved algorithm is a hybrid regularization method. Tikhonov regularization, a ridge regression regularization method, effectively mitigates overfitting but is sensitive to noise. On the other hand, TV regularization is a sparsity-based method that excels in denoising but suffers from convergence difficulties and high computational complexity. The proposed improved algorithm combines sparse regression and ridge regression, adjusting parameters to balance the regularization term between the two approaches. This ensures a certain degree of overfitting control while maintaining robust denoising capabilities. Moreover, by adjusting the parameters, the algorithm can achieve different regularization paths, making it highly robust and flexible to various geological conditions and model characteristics.

Another pivotal issue is the algorithm's adjustable parameters. Parameter choices should not be static but adjusted according to different models. The most straightforward solution is heuristic parameter tuning, conducting multiple inversions, and then balancing between inversion accuracy and computation time to choose the appropriate parameters (Boyd *et al.*, 2011). Another approach is to make the parameters adaptive. Adaptive algorithms link operators with variables in polynomials, obtaining a parameter that can self-correct through the model. While the second approach theoretically aligns with future development trends, there isn't ample research to validate its efficacy. Therefore, for FWI, especially in the optimization segment, a critical direction for the future is to discovery a more suitable parameter selection, which is vital for both traditional FWI and artificial intelligence-based FWI.

3.6 Chapter Conclusions

This chapter introduces the K-support norm algorithm, to address common issues faced by the FWI algorithm under complex conditions. A significant advantage of this algorithm is its ability to suppress the impact of outliers on inversion resolution, especially when seismic data contains noise.

The proposed algorithm in this chapter is built upon the quadratic penalty method. By integrating two distinct norm forms and employing penalty parameters to adjust the constraint forms, a tighter convex relaxation is realized, thereby enhancing the algorithm's robustness and noise resistance. Moreover, the ADMM algorithm decomposes the iterative process into distributed steps, increasing the algorithm's convergence speed by reducing computational complexity.

In the numerical experiments, this section utilizes three different synthetic datasets, a random noise, and a coherent noise, to conduct a comprehensive test of the algorithm proposed in this chapter. The findings indicate that the FWI, when executed using the enhanced algorithm, showcases superior rates of convergence and heightened noise mitigation, most notably in scenarios characterized by extremely low signal-to-noise ratios, it offers commendable inversion results for deeper areas across various models. The algorithm also effectively suppresses outliers, confirming that the proposed method can efficiently achieve robust high-resolution imaging in deeper regions under intricate scenarios.

CHAPTER IV: FULL WAVEFORM INVERSION BASED ON RANDOMIZED SINGULAR VALUE DECOMPOSITION

The content of this chapter has been published in a journal paper and a preprint: *The Journal of Geophysics and Engineering (JGE 2024) & arXiv (2023)*

1. **Li, J***., Mikada, H., & Takekawa, J., 2024. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomised Singular Value Decomposition. *Journal of Geophysics and Engineering*, **21**, 572-597. DOI: 10.1093/jge/gxae015 **(21-January-2024)**

2. **Li, J***., Mikada, H., & Takekawa, J., 2023. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomized Singular Value Decomposition. *arXiv preprint,* arXiv:2309.13871. DOI: 10.48550/arXiv.2309.13871 **(25-September-2023)**

The content of this chapter has been published in a conference abstract: In *the Third International Meeting for Applied Geoscience & Energy.* Society of Exploration Geophysicists. *(SEG 2023)*

3. **Li, J***., Mikada, H., & Takekawa, J., 2023. Accelerated augmented Lagrangian full-waveform inversion based on truncated randomized singular value decomposition. In *Third International Meeting for Applied Geoscience & Energy*, *Houston, United States of America,* pp. 645-649. Society of Exploration Geophysicists and American Association of Petroleum Geologists. DOI: 10.1190/image2023-3910542.1 **(31-August-2023)**

4.1 Introduction

In an ideal situation, FWI computes the gradient to determine the update direction and magnitude of the model. Then each model update is superimposed onto the previous result to serve as the basis for the next iteration, thus completing an iterative cycle. After a number of iterations with updating the velocity model, theoretically, the inversion result of FWI will gradually approach the real velocity model with increasing iterations, eventually coming infinitely close to the real model. However, practical scenarios are often more complex than theoretical ones. Commonly, due to limitations of the acquisition system and influences from environmental factors, it becomes challenging for FWI to obtain high-resolution inversion results, especially in intricate exploration environments (Zhou *et al*., 2015).

A highly challenging exploration scenario is the high-precision inversion of salt domes and salt diapirs in the Gulf of Mexico. This region is renowned for its widespread distribution of salt rocks with significant thickness. However, the structures of salt domes and diapirs in the Gulf of Mexico are intricate, these salt formations have high-velocity characteristics, which, in contrast with surrounding rock velocities, create pronounced velocity differences. This phenomenon leads to a very complicated seismic wave propagation process, where wave paths are frequently distorted, resulting in multiple reflections and refractions, making the deep inversion of FWI exceptionally challenging (Métivier *et al*., 2013). Moreover, the high-velocity upper interface of the salt rock forms a strong reflective layer. This reflection interface, with a significant impedance difference, has high P-wave impedance characteristics, manifesting as strong amplitude, high continuity, and relatively low frequency in seismic data, thereby masking the reflections of structures beneath the salt. Additionally, high levels of noise can obscure valuable seismic information, hindering the capture of detailed subsurface data, and even slowing the convergence of FWI due to overfitting, leading to erroneous inversion results (Chi *et al*., 2014). All these challenges amplify the difficulty of FWI in determining deep salt bottom structures.

In response to the aforementioned issues, numerous algorithms have been proposed to address noise problems and poor resolution. Common noise-mitigating solutions include forward interpolation and misfit function reconstruction. However, many enhanced algorithms are not perfectly suited to handle models like the high-velocity salt formations of the Gulf of Mexico with high discontinuities or those with vast velocity contrasts. For instance, Tikhonov regularization can result in smoother solutions, blurring the outlines of salt bodies in inversion results, and hindering the effective identification of salt diapirs and domes (Qu *et al*., 2019). Another approach involves using image processing techniques to accentuate primary features while ignoring secondary ones. Sparse dictionaries and sparse coding algorithms are advanced optimization techniques in this regard. Sparse dictionary methods construct a training set via singular value decomposition (SVD) and then learn from this dataset, forming a highly adaptive dictionary (Li and Harris, 2018). Additionally, sparse coding can represent signals as sparse linear combinations, aiding in noise separation from signals. However, traditional sparse dictionary construction and sparse coding solutions are computationally intensive, posing challenges for large-scale seismic exploration, and additionally, if training data are inadequate, the effectiveness of the dictionary can diminish (Guo *et al*., 2020).

Building upon these real-world issues and existing algorithm foundations, this chapter innovatively proposes a novel concept, the new algorithm combines random singular value decomposition (rSVD) with weighted truncated nuclear norm regularization (WTNNR) and the inexact augmented Lagrange method (iALM) to optimize FWI (Liu and Peter, 2020). Firstly, unlike the conventional SVD, the advantage of rSVD lies in its faster processing speed and superior truncation capabilities. It assists FWI in truncating redundant eigenvalues effectively during velocity increment compression, with the compression scale determined by a truncation coefficient. This coefficient is embedded in FWI's internal loop, enabling stratified optimization during the inner loop process, effectively stripping noise and emphasizing eigenvalues. Furthermore, the WTNNR algorithm, a method to shrink matrices, effectively complements rSVD, processing the rectangular diagonal matrices decomposed by rSVD, achieving a more precise eigenvalue truncation (Deng *et al.*, 2020). Finally, iALM, an optimization method for the minimization process, is based on the augmented Lagrange method but employs an adaptive parameter setting for faster convergence.

4.2 Randomized Singular Value Decomposition: rSVD

The minimization process of frequency domain FWI can be expressed as follows:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}-\mathbf{F}[\chi,\mathbf{S}]\|_F^2, \tag{4.1}$$

where $\mathbf{A}$ is the input actual observed seismic data, χ is the model parameters, and $\mathbf{S}$ is the source matrix, where the fitted seismic wave field function **F** can be obtained by forward modelling in the frequency domain. Subsequently, the FWI's optimization procedure encompasses both the objective function and the gradient method. Once the objective function is determined, the next step involves solving the inverse of the Hessian matrix and applying it to the gradient. This, when coupled with a suitable step size, yields the model increment.

$$\tilde{\chi} = \mathbf{H}^{\dagger}\mathbf{g}, \tag{4.2}$$

$$\chi = \chi + \alpha\tilde{\chi}, \tag{4.3}$$

where $\mathbf{H}^{\dagger}$ is the approximation of the Hessian matrix, $\mathbf{g}$ is the gradient, α is the iteration step size, and $\tilde{\chi}$ is the model update. The above equations are the classic quasi-Newton-type algorithm and the classical solution to the FWI inverse problem (Li *et al*., 2012b). Yet, given that every internal cycle of the FWI demands the calculation of the gradient and reduction of the discrepancy between the observed and the modelled data, the computational load is quite substantial. More importantly, especially in the low-frequency initial phase of multiscale inversion, if the quality of the updates in the optimization process is poor, it will significantly affect the subsequent iteration speed and the quality of the fitting model. This causes the fitting results to deviate from the true model and severely affects the overall inversion resolution of FWI (Pan *et al*., 2016).

To address the above issues, this chapter will adopt a technique combining matrix dimensionality reduction and image processing. In the FWI optimization process, by using truncation operations, I can reduce the dimensionality of the matrix, which not only helps speed up the optimization process but also isolates small eigenvalues representing outliers and noise, retaining only large eigenvalues representing the large-scale structures in the model, thus this approach not only effectively denoises but also reduces nonlinearity. Specifically, a truncation parameter k is needed to compress the original velocity increment, followed by truncation and matrix restoration steps to obtain a matrix approximate in size to the original matrix but with a different number of eigenvalues. In the approximate matrix, some eigenvalues are zero, meaning the rank of this matrix is much less than the original matrix. The proportion of valid eigenvalues in the approximate matrix is determined by the truncation parameter.

Specifically, first, I need to compute the product of the velocity increment and the Gaussian random matrix, compressing the size of the original matrix to a new size controlled vertically by the truncation parameter:

$$\chi' = \tilde{\chi}\mathbf{\Omega}, \tag{4.4}$$

where $\chi' \in N_{\mathrm{pr}}^{m\times k}$ is reshaped velocity increments, pr denotes the dataset for the improved algorithm to differentiate it from the conventional algorithm. $\mathbf{\Omega} \in N_{\mathrm{pr}}^{n\times k}$ is the Gaussian random matrix. Subsequently, a special type of QR decomposition is required to decompose the process matrix χ' under the condition of saving computational time. Since the right matrix **R** obtained after the decomposition is redundant, an economic QR decomposition is used in rSVD to minimize the size of the **R** dimension from $\mathbf{R} \in N_{\mathrm{pr}}^{m\times n}$ to $\mathbf{R} \in N_{\mathrm{pr}}^{m\times k}$, which in turn further accelerates the updating speed at each step:

$$\mathbf{Q} = \mathrm{eqr}(\chi'), \tag{4.5}$$

where $\mathrm{eqr}(\ )$ is the economic QR decomposition, where $\mathbf{Q} \in N_{\mathrm{pr}}^{m\times k}$. Regular QR decomposition decomposes a $m\times n$ matrix (where $m \geq n$) into $m\times n$ upper triangular matrix R. In contrast, the economic QR decomposition decomposes the same matrix into an $n\times n$ upper triangular matrix R (Song *et al.* 2017). Next, I calculate the transpose of Q to get $\mathbf{Q}^* \in N_{\mathrm{pr}}^{k\times m}$, and multiply $\mathbf{Q}^*$ with the velocity increments $\tilde{\chi}$ to get a preliminary compression matrix:

$$\chi_c = \mathbf{Q}^* \tilde{\chi}, \tag{4.6}$$

where $\mathbf{Q}^*$ is the transpose of matrix Q, and χ_c is the compression matrix I initially obtained. It is of size $\chi_c \in N_{\mathrm{pq}}^{k\times n}$ compared to the original matrix of size $\tilde{\chi} \in N_{\mathrm{pr}}^{m\times n}$, and pq denotes the dataset after the rSVD step. The scale of the approximation matrix is contracted by the compression action of a Gaussian random matrix of small size.

Similarly, the compression property of a Gaussian random matrix of a prescribed size allows obtaining a small-scale compressed approximation matrix regardless of the original size of the matrix. The number of columns of this approximation matrix is the truncation parameter *k* (Halko *et al.*, 2011).

Subsequently, the complete SVD process should be acted the $\boldsymbol{\chi}_c$ to obtain the singular values matrix:

$$\tilde{\boldsymbol{\psi}}\boldsymbol{\Sigma}_k\mathbf{V}^* = \boldsymbol{\chi}_c, \tag{4.7}$$

where $\tilde{\boldsymbol{\psi}}$ is the left singular vector, $\boldsymbol{\Sigma}_k$ contains the first *k*-th singular values in the diagonal elements, $\mathbf{V}^*$ is the right singular vector.

The above process describes the complete compression process of rSVD, which obtains the first *k*-th singular values of an approximate matrix from an original matrix through the action of the Gaussian random matrix, facilitating subsequent optimization processes. Through the above steps, I can quickly obtain the approximate singular value matrix of the velocity increment and its first *k* singular values. To intuitively compare the computational efficiency of rSVD with that of full SVD, Figure 4.1 (a) provides a comparison of their computation times. In the depicted graph, the blue solid line signifies the computation duration for the full SVD, whereas the red solid line indicates the computation time for the rSVD. As can be seen from the figure, due to the truncation parameter *k*, rSVD has extremely high computational efficiency in the early stages. This is because the value of the truncation parameter is small at the initial stage, meaning the low-rank of the approximate singular value matrix. Consequently, it makes the initial computational efficiency of rSVD significantly advantageous compared to full SVD. The red solid line in Figure 4.1 (b) shows that when the truncation parameter *k* is 20, which means the column entries of the Gaussian random matrix are 20, the rank of the approximate matrix processed by rSVD is significantly smaller than the rank of full SVD, which indicates that in the approximate singular value matrix processed by rSVD, most of the singular values are forced to zero.

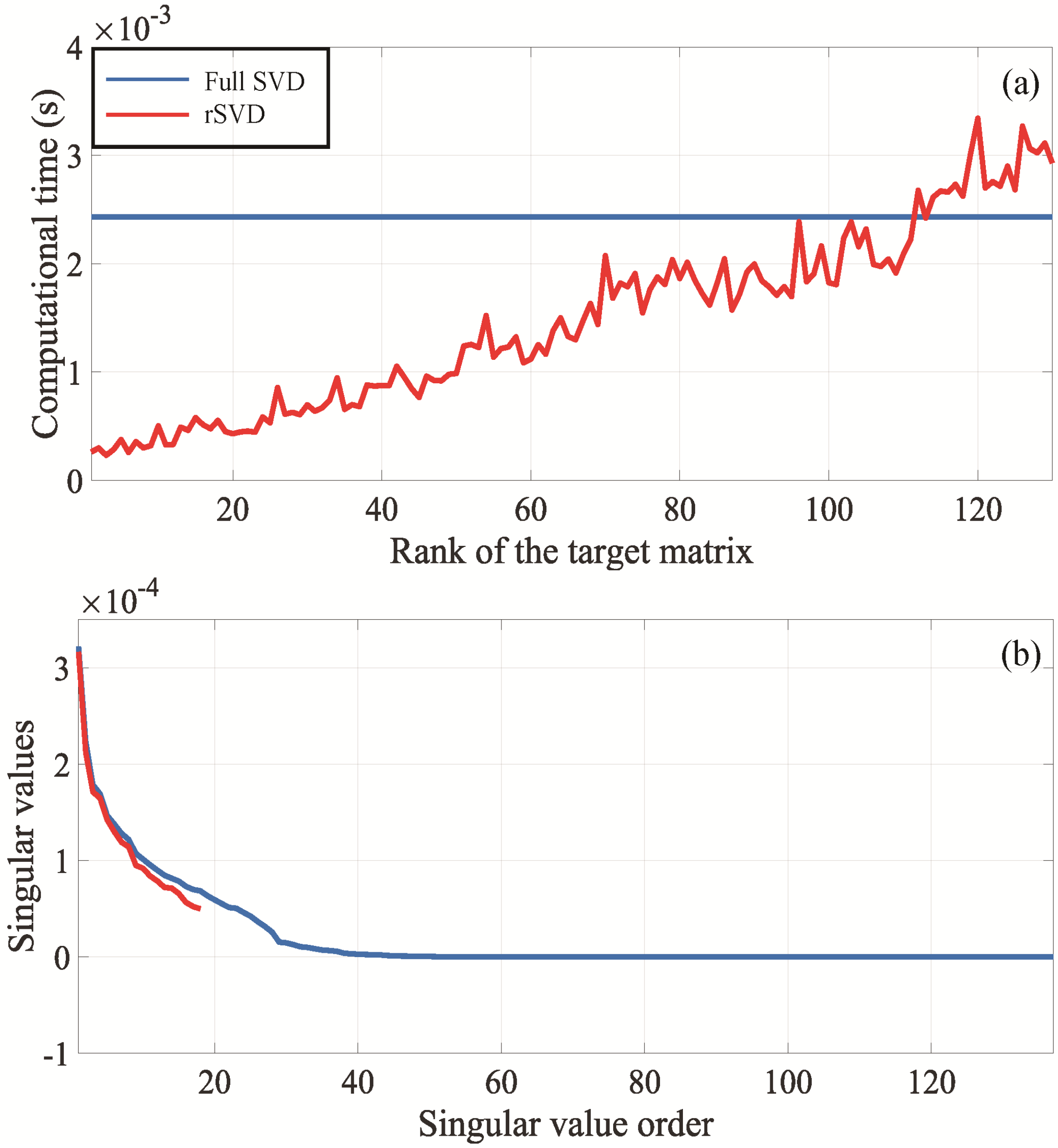

Figure 4.1. (a) The computation time for the full SVD (blue) and the rSVD (red); (b) the full SVD estimated singular value curve (blue) and approximate singular value curve based on rSVD (red) (Li *et al*., 2024b).

Subsequently, I need to recover the left singular vector $\tilde{\boldsymbol{\psi}}$ and combine this vector with the singular value matrix $\tilde{\boldsymbol{\Sigma}}_k$ which will be optimized by weighted truncated nuclear norm regularization (WTNNR) to reconstruct the output matrix $\tilde{\boldsymbol{\chi}} \in N_{\mathrm{pr}}{}^{m\times k}$, which has the same size as the input:

$$\tilde{\boldsymbol{\chi}} = \boldsymbol{\psi}\,\tilde{\boldsymbol{\Sigma}}_k\,\mathbf{V}^*, \qquad \textbf{(4.8)}$$

where $\tilde{\boldsymbol{\psi}}$ is the recovered left singular vector, and $\tilde{\boldsymbol{\Sigma}}_k$ is the singular value matrix.

The last issue to address is the selection of the truncation parameter *k*. Unlike the selection of other parameters which might require empirical judgment or heuristic methods, this parameter is initially set to a very small value before the forward modelling of FWI begins. Before the inversion process within a single-frequency inner loop, a step size of increment of the parameter is set. This allows for a gradual increase in the truncation parameter according to the step size during the optimization of the single frequency. At this point, an upper limit needs to be set for the truncation parameter, which is the larger number between the row count and column count of the singular value matrix. This ensures that the selection of the truncation parameter does not exceed the matrix scale, and in turn, ensures that the column count of the Gaussian random matrix does not surpass the row or column count of the singular value matrix. After completing the iteration of a single frequency, FWI returns to the forward modelling phase. At this time, the value of the truncation parameter is reset to zero, ensuring that it can continue to increase in the next single frequency without exceeding its upper limit due to changes in the inversion frequency. The maximum limit here refers to the largest value between the row and column counts of the matrix. Figure 4.2

provides a schematic diagram of the truncation parameter update process.

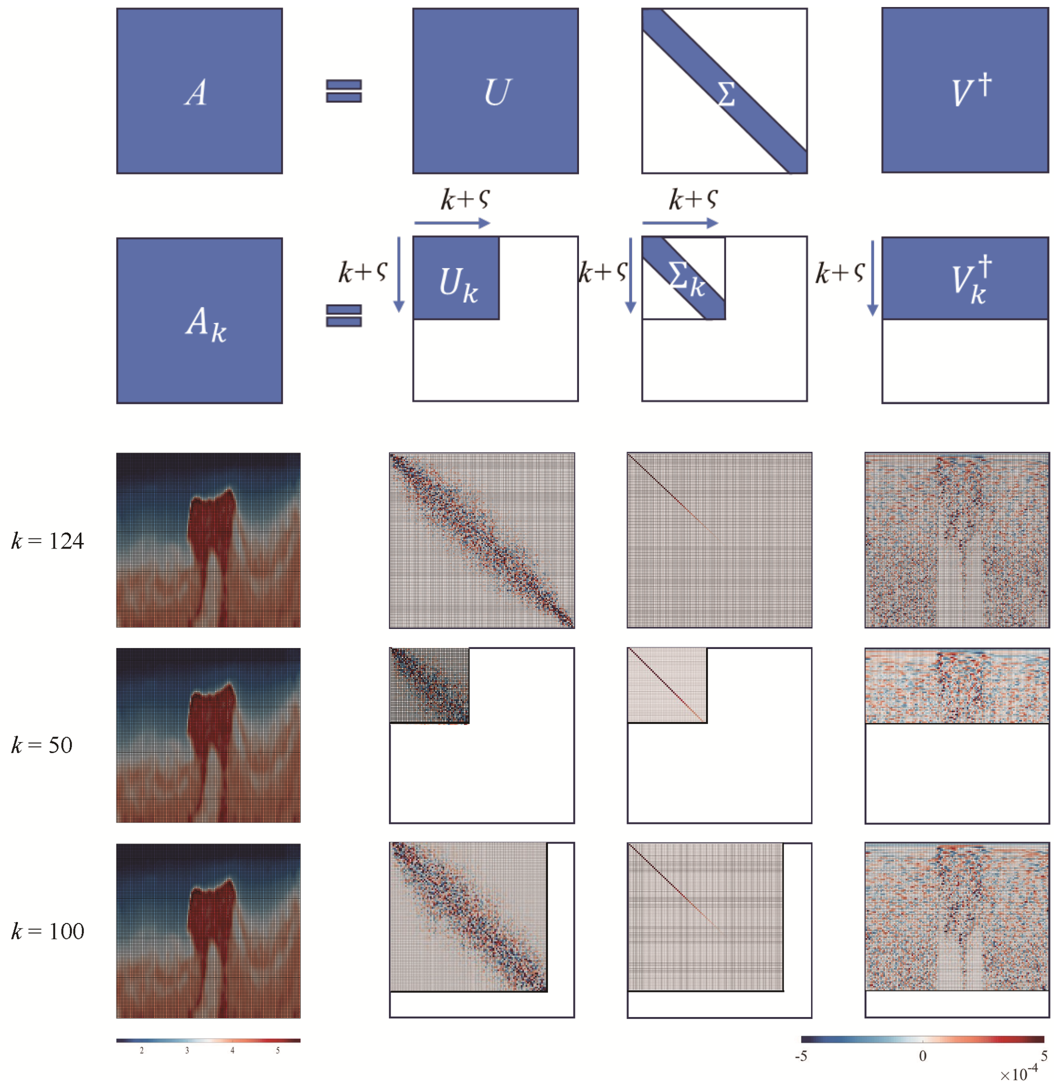

Figure 4.2. Schematic representation of the increase in the truncation parameter with the number of internal iterations, where k is the truncation parameter and ς is the step size (Li *et al.*, 2024b).

Furthermore, for the overall process described, Figure 4.3 provides a visual demonstration using the 2004 BP model mentioned in earlier chapters as an example, of which size is 124×124, it offers an intuitive reference for a single iteration process of rSVD, where the truncation parameter is set to 50.

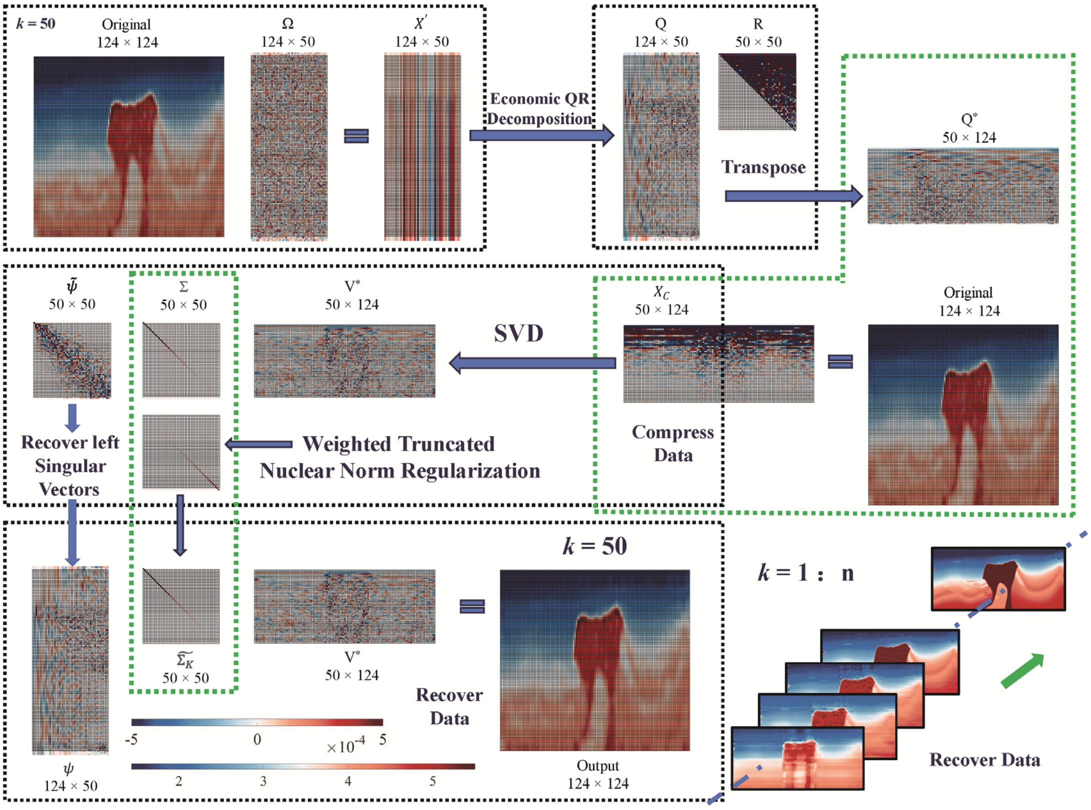

Figure 4.3. Visualization flowchart for rSVD, where $k = 50$ is the truncation parameter and the model size is 124 × 124 (Li *et al.*, 2024b).

4.3 Weighted Truncated Nuclear Norm Regularization: WTNNR

The singular value matrix can be obtained after rSVD processing, and then a matrix shrinkage algorithm will be used in this section to further optimize this singular value matrix. It can effectively reduce the load of FWI by further truncation operation and optimize the iterative updating term to strengthen the main feature elements (Horst *et* *al.,* 2000):

$$\delta\chi = \arg\min_{\delta\chi} \left\|\tilde{\chi}\right\|_{\mathrm{WT}}, \tag{4.9}$$

where $\|\bullet\|_{\mathrm{WT}}$ is the weighted truncated nuclear norm, and $\delta\chi$ is the optimized increment item. The WTNNR problem can be understood as the following:

$$\left\|\tilde{\chi}\right\|_{\mathrm{WT}} = \left\|\tilde{\chi}\right\|_{*} - \max_{\psi\psi^{k}=\mathbf{I},\mathbf{V}\mathbf{V}^{k}=\mathbf{I}} \mathrm{Trace}(\psi_k\, \tilde{\chi}(\tilde{\boldsymbol{\Sigma}}_k)\mathbf{V}_k^T), \tag{4.10}$$

where $\psi_k \in N_{\mathrm{pk}}{}^{k\times k}$ and $\mathbf{V}_k \in N_{\mathrm{pq}}{}^{k\times n}$ satisfying $\psi_k\psi_k^T = \mathbf{I}_{k\times k}$ and $\mathbf{V}_k\mathbf{V}_k^T = \mathbf{I}_{k\times k}$, to prevent the truncation parameter from exceeding the upper bound, it is therefore necessary to satisfy $k \le \min(m,n)$ and $\mathrm{Trace}(\bullet)$ represent the trace (Chang *et al.*, 2000). With the WTNNR algorithm proposed in this section, the singular-value matrix that has been processed by rSVD can be further shrunk to isolate more nonzero singular-value elements. As a result, the proportion of small singular values can be decreased significantly:

$$\mathbf{W}(\tilde{\boldsymbol{\Sigma}}_k) = \max(\tilde{\boldsymbol{\Sigma}}_k - \frac{\mathbf{W}_k}{2}, 0), \tag{4.11}$$

where $\mathbf{W}_k$ is the truncation matrix, $\mathbf{W}(\tilde{\boldsymbol{\Sigma}}_k)$ is the new singular value matrix after the weighting process. If $\frac{\mathbf{W}_k}{2} \ge \tilde{\boldsymbol{\Sigma}}_k$, $\mathbf{W}(\tilde{\boldsymbol{\Sigma}}_k)$ is taken to be zero.

Subsequently, to better illustrate the principle of the WTNNR algorithm, a schematic figure is provided in Figure 4.4. In the displayed chart, the blue solid line illustrates the ordering of the diagonal elements of the singular value matrix from the full SVD. The red solid line shows the ordering of the diagonal elements of the singular value matrix post-optimization exclusively by rSVD. The black solid line portrays the ordering of the diagonal elements of the singular value matrix from the weight shrinkage matrix crafted by the WTNNR algorithm. Meanwhile, the purple solid line indicates the ordering of the diagonal elements of the singular value matrix following optimization via the WTNNR algorithm. From the figure, it can be seen that the singular value arrangement constructed by WTNNR is in the opposite direction from others, playing a shrinkage role, and truncating some of the smaller singular values.

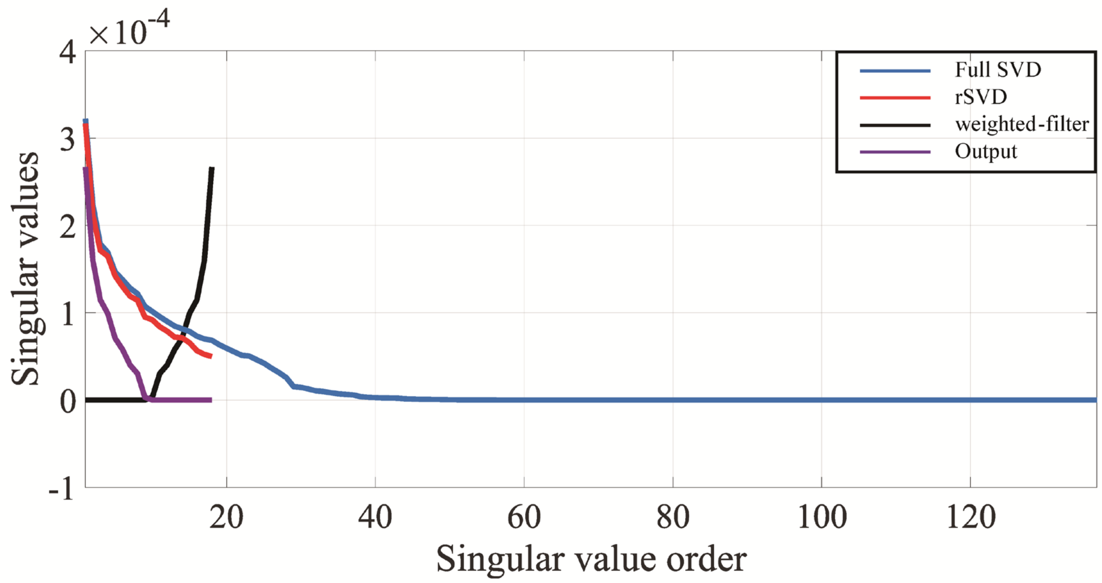

Figure 4.4. The blue solid line is the full SVD estimated singular value curve, the approximate singular value curve of rSVD is represented by the red line, the black solid line is the weight curve based on WTNNR, and the purple solid line depicts the approximate singular value curve optimized by rSVD-WTNNR (Li *et al*., 2024b).

Additionally, Figure 4.5 presents a more intuitive illustration, which depicts two different scales of matrix optimization processes, (a-c) are of size 9 × 9, while (d-f) are of size 19 × 19. Figure 4.5 (a) is the singular value matrix after rSVD processing, (b) is the WTNNR matrix, and (c) is the matrix after WTNNR shrinkage. As inferred from the figure, the WTNNR algorithm can effectively shrink the singular value matrix, further reducing the proportion of non-zero elements and small singular values, thereby achieving the purpose of singular value shrinkage.

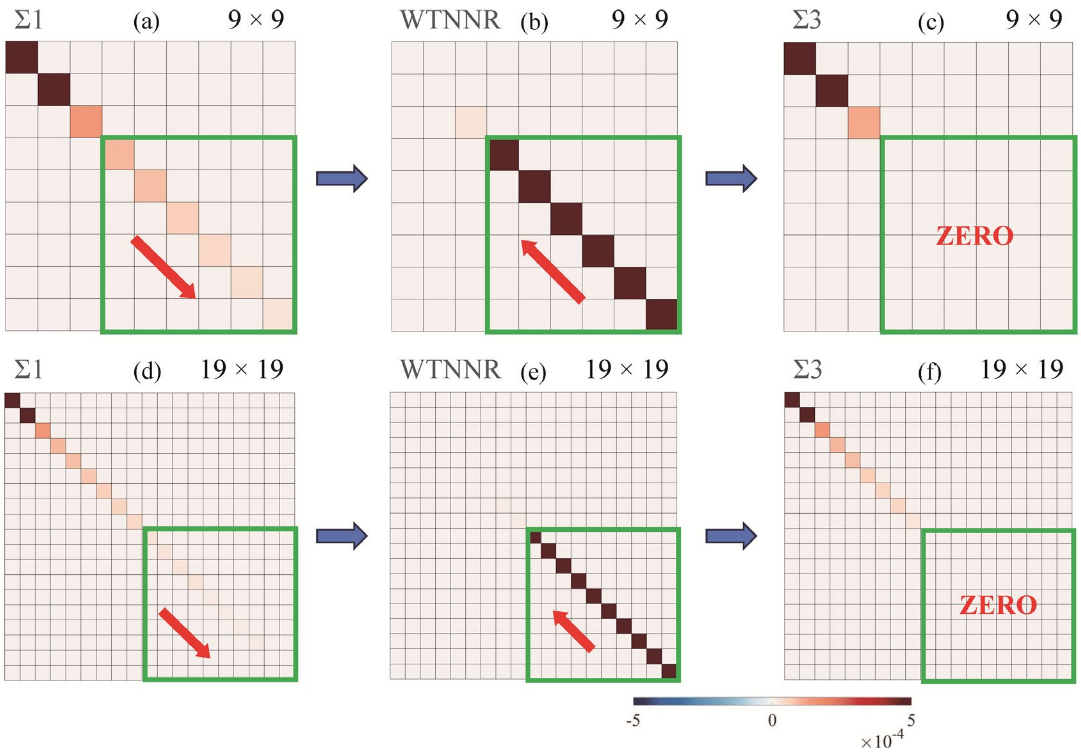

Figure 4.5. (a-c) 9 × 9 test data, (a) singular value matrix after rSVD decomposition; (b) weighted singular value matrix constructed based on WTNNR with singular values growing in the opposite direction of (a); and (c) optimized singular value matrix; (d-f) 19 × 19 test matrices; (d) Singular value matrix after rSVD decomposition; (e) weighted singular value matrix constructed based on WTNNR with singular values growing in the opposite direction to (d); (f) optimized singular value matrix (Li *et al*., 2024b).

The reason for this approach is that the singular value matrix represents the characteristics of the model, large singular values typically represent the large-scale features of the model, whereas smaller features are more susceptible to artefacts and noise. Therefore, as the noise increases, more singular values need to be isolated, and as the noise decreases, fewer singular values need to be isolated. Thus, the amount of shrinkage allocated to the *i*-th singular value is inversely proportional to its amplitude:

$$\mathbf{W}_i = \beta\sqrt{\gamma} / (\boldsymbol{\sigma}_i + \varepsilon), \tag{4.12}$$

where β is a nonnegative constant, $\boldsymbol{\sigma}_i$ is the locally estimated variance at the *i*-th position, γ is the number of similar patches, and $\varepsilon = 2^{-52}$ to avoid the divisor being zero:

$$\hat{\boldsymbol{\sigma}}_i = \sqrt{\max(\boldsymbol{\sigma}_i^2 - \gamma\boldsymbol{\sigma}_k^2, 0)}, \tag{4.13}$$

where $\boldsymbol{\sigma}_i$ is the *i*-th singular value of the matrix. Similarly, in the case of $\gamma\boldsymbol{\sigma}_k^2 \geq \boldsymbol{\sigma}_i^2$, $\hat{\boldsymbol{\sigma}}_i$ is taken to be zero.

Additionally, using the 2004 BP model as a benchmark, the WTNNR algorithm was tested again, as shown in Figure 4.6. In it, (a) represents the input increment, (b) is the singular value matrix after truncation by rSVD, where I can see a noticeable reduction in rank. (c) is the WTNNR matrix, and (d) is the output increment after WTNNR optimization, it is evident that only a few large singular values have been retained.

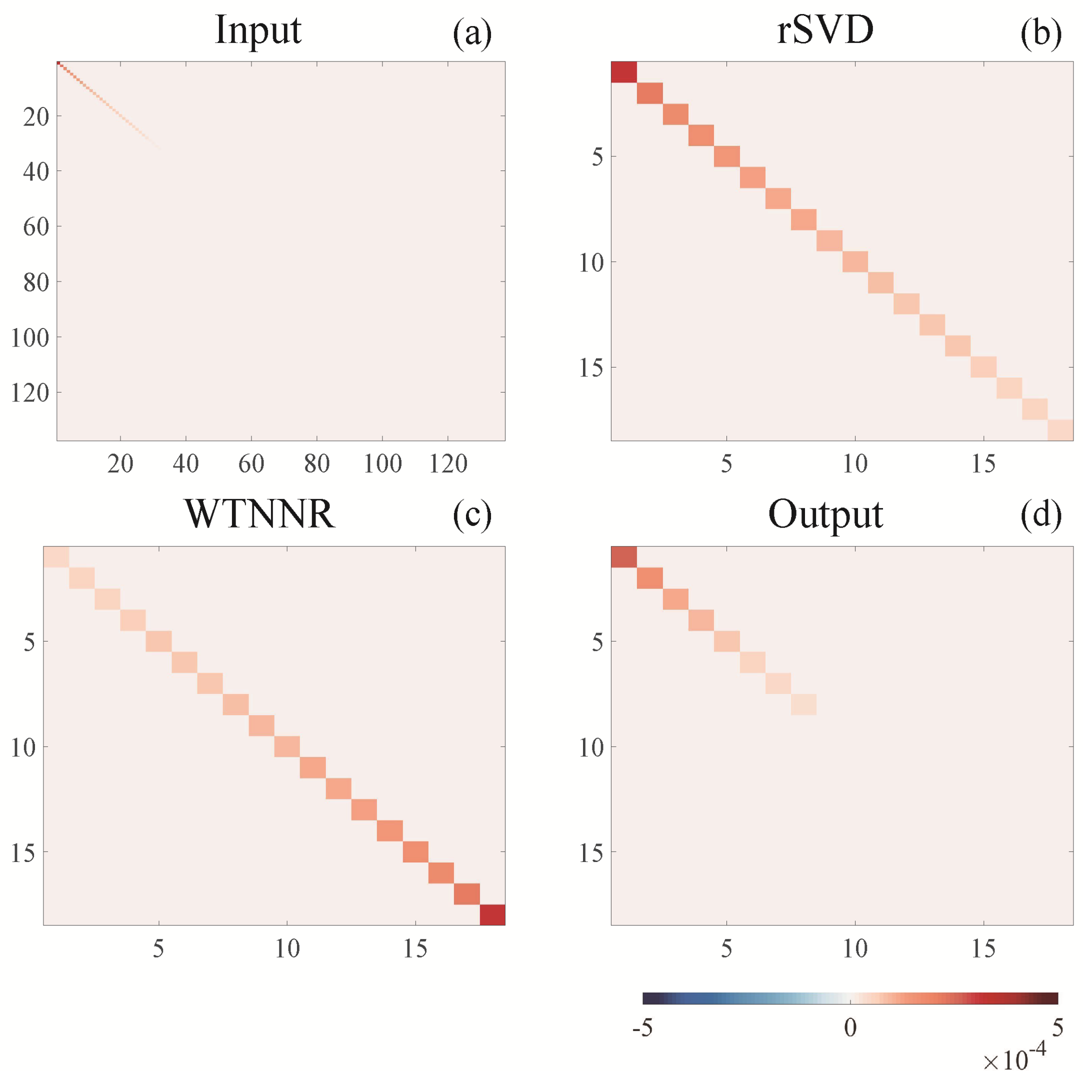

Figure 4.6. 124 × 124 singular value matrix of the reconstructed velocity increment, (a) input matrix; (b) singular value matrix after rSVD-based optimization; (c) inverse weight matrix constructed based on WTNNR; and (d) singular value matrix of the velocity increment after rSVD-WTNNR optimization (Li *et al*., 2024b).

Furthermore, to preliminarily test the effectiveness of this algorithm, Figure 4.7 presents the test results for the 2004 BP benchmark, (a-c) are the low-frequency inversion results without incorporating WTNNR under different truncation parameters, whereas (d-f) are the low-frequency inversion results with WTNNR under the same truncation parameters. The results indicate that although rSVD has a preliminary truncation effect, the WTNNR algorithm achieves superior separation of the main eigenvalues, which is manifested as the large-scale salt column features in the central part of the model being reinforced. This enhancement provides a better foundation for the subsequent high-frequency inversion.

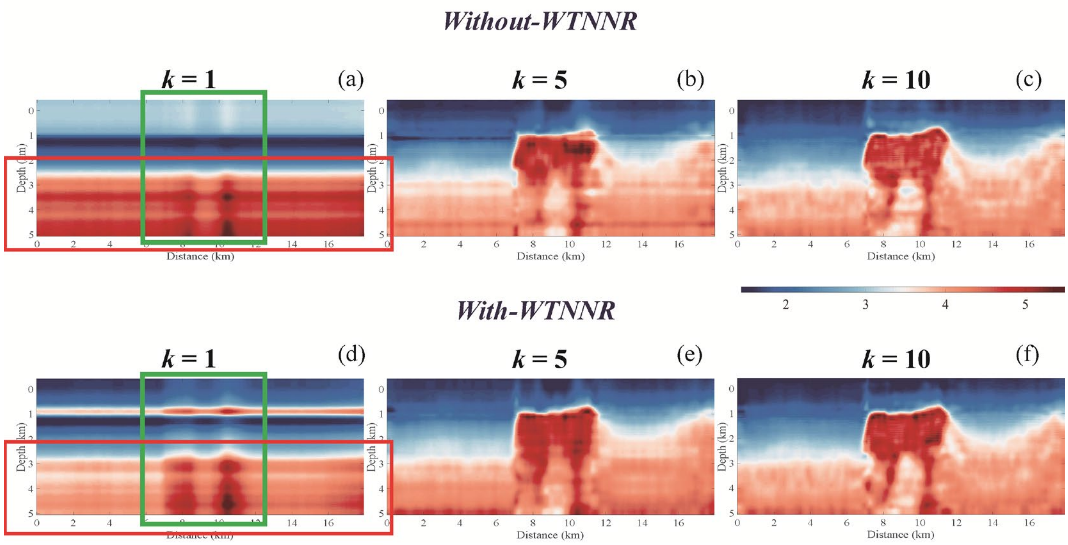

Figure 4.7. (a-c) Test results of velocity increment based on rSVD optimization without WTNNR with truncation parameters *k=1, k=5,* and *k=10*; (d-f) test results of velocity increment based on rSVD optimization with WTNNR. WTNNR can effectively enhance the features of the salt columns and slow down the disturbance of the channels on both sides (Li *et al*., 2024b).

4.4 Inexact Augmented Lagrangian Method: iALM

To make it feasible for ALM to be implemented, an accelerated method, which can enhance the efficiency of this inversion method, is introduced. In fact, in the problems based on augmented Lagrangian, the estimated parameter λ is used in the sub-iteration process. It will lead to problems such as excessive dependence on the model and poor robustness of the augmented Lagrangian method (ALM). Ideally, an iterative sequence λ_k should be used instead of the estimation value to meet the termination condition required by the algorithm in the last iteration of the minimization process. The proper setting of this multiplier vector leads to a different inversion result. Instead of a single estimation value, according to the inexact augmented Lagrangian method (iALM) algorithm, the new iterative method can reduce the iterative rate of the objective function (Kang *et al.*, 2015), and I show this conclusion in Appendix B and Appendix C. The updating process of wave-field information and model information is similar to ALM.

Finally, in the minimization process of FWI, a better way is to use the augmented Lagrangian algorithm based on as a form of constraints (van Leeuwen and Herrmann, 2013):

$$\min_{\mathbf{u},\mathbf{X}} \max_{\lambda} \mathrm{L}(\chi,\boldsymbol{\tau}) = \min_{\mathbf{u},\mathbf{X}} \max_{\lambda} \|\mathbf{AX} - \mathbf{D}\|_2^2 + (\hat{\boldsymbol{\lambda}}_j)^T \left[\mathbf{C}(\mathbf{u})\mathbf{X} - \mathbf{S}\right] + \boldsymbol{\tau}\|\mathbf{C}(\mathbf{u})\mathbf{X} - \mathbf{S}\|_2^2, \qquad \textbf{(4.14)}$$

where $\|\bullet\|_2^2$ is the Euclidean norm, $\mathbf{X} \in \mathbb{R}^{Na\times 1}$ is the model wavefield, $\mathbf{D} \in \mathbb{R}^{Ma\times 1}$ is the recorded seismic data, $\mathbf{S} \in \mathbb{R}^{Na\times 1}$ is the source term, and the linear observation operator $\mathbf{A} \in \mathbb{R}^{Ma\times Na}$ sampling $\mathbf{X}$ at the receiver positions, $\mathbf{u} \in \mathbb{R}^{Na\times 1}$ is the model parameters, which contains preliminary information about underground parameters. The $\mathbf{C}(\mathbf{u}) \in \mathbb{R}^{Na\times Na}$ is the discretized PDE, which is synergic with the $\mathbf{u}$ (Gholami *et al.*, 2022). Based on the Nesterov acceleration algorithm, adaptive parameter selection can be employed (Sahin *et al*., 2019):

$$\boldsymbol{\lambda}_j = \hat{\boldsymbol{\lambda}}_j - \tau(\mathbf{C}(\mathbf{u})\mathbf{X} - \mathbf{S}), \tag{4.15}$$

$$t_{j+1} = (1 + \sqrt{1 + 4(t_j)^2}) / 2, \tag{4.16}$$

$$\hat{\boldsymbol{\lambda}}_{j+1} = \boldsymbol{\lambda}_j + \frac{t_j - 1}{t_j + 1}(\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_{j-1}) + \frac{t_j}{t_{j+1}}(\boldsymbol{\lambda}_j - \hat{\boldsymbol{\lambda}}_j), \tag{4.17}$$

where $\hat{\boldsymbol{\lambda}}_1 = \boldsymbol{\lambda}_0$, and $t_1 = 1$. The iALM algorithm is an accelerated algorithm based on the augmented Lagrangian, which is an inexactly solved version of the original problem that does not require an exact solution for each iteration of the process, and thus it greatly reduces the computational efficiency of the:

$$\hat{\boldsymbol{\lambda}}_{j+1} = \boldsymbol{\lambda}_j + \frac{t_j - 1}{t_j + 1}(\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_{j-1}). \tag{4.18}$$

Lastly, in Algorithm 1, the flowchart of the algorithm for the rSVD-WTNNR-based FWI proposed in this chapter is given.

Algorithm 1. FWI based on the rSVD - WTNNR & iALM.

Input: Real seismic data and Initial velocity data, model size $m\times n$, minimum frequency f_{min}, maximum frequency f_{max}, parameter t, truncation parameter k, truncation step-size ς.

Initialization: $f = f_{min}$, $k = 1$, $t_1 = 1$,

for inversion frequency: $f = f_{min} : f_{max}$

Accelerated iterative optimization,

where $\lambda_j = \hat{\lambda}_j - \tau(A(\chi)u - b)$,

$t_{j+1} = (1+\sqrt{1+4(t_j)^2})/2$,

$\hat{\lambda}_{j+1} = \lambda_j + \frac{t_j - 1}{t_j + 1}(\lambda_j - \lambda_{j-1}) + \frac{t_j}{t_{j+1}}(\lambda_j - \hat{\lambda}_j)$,

for not converged **do**

$k = k + \varsigma$,

Compute rSVD: $\left[\tilde{\psi}_k, \Sigma_k, V^*_k\right] \leftarrow \mathrm{rSVD}(\tilde{\chi}^{m\times n}, T)$,

$\chi'^{m\times k} \leftarrow \tilde{\chi}^{m\times n}\ \Omega^{m\times k}$,

where Ω is Gaussian random matrix,

Economic QR Decomposition: $Q^{m\times k} \leftarrow eqr(\chi'^{m\times k})$,

$\chi_c^{k\times n} \leftarrow Q^{*k\times m}\ \tilde{\chi}^{m\times n}$,

$\tilde{\psi}_k^{k\times k}\ \Sigma_k^{k\times k} V_k^{*k\times n} \leftarrow \mathrm{SVD}(\chi_c^{k\times n})$,

where $\widetilde{\Psi} = (\psi_1, \ldots \ldots \psi_k) \in \mathbb{R}^{k\times k}$, $V^* = (V_1, \ldots \ldots V_k) \in \mathbb{R}^{k\times n}$,

Recover left singular vectors: $\tilde{\psi}^{m\times k} \leftarrow \mathrm{recover}(\tilde{\psi}^{k\times k})$,

weighted truncated nuclear norm regularization: $\Sigma_k^{k\times k} \leftarrow \Sigma_k^{k\times k} - \Sigma_{TNNR}^{k\times k}$,

$\tilde{\chi}^{m\times n} = \psi^{m\times k}\ \tilde{\Sigma}_k^{k\times k}\ V^{*k\times n}$,

end for

Model update,

Switch frequency,

Reset $k = 1$,

end for

Output: Inversion result.

Algorithm 1. FWI is based on the rSVD-WTNNR and iALM (Li *et al.*, 2024b).

In Figure 4.8, the algorithmic flowchart of the improved FWI is given.

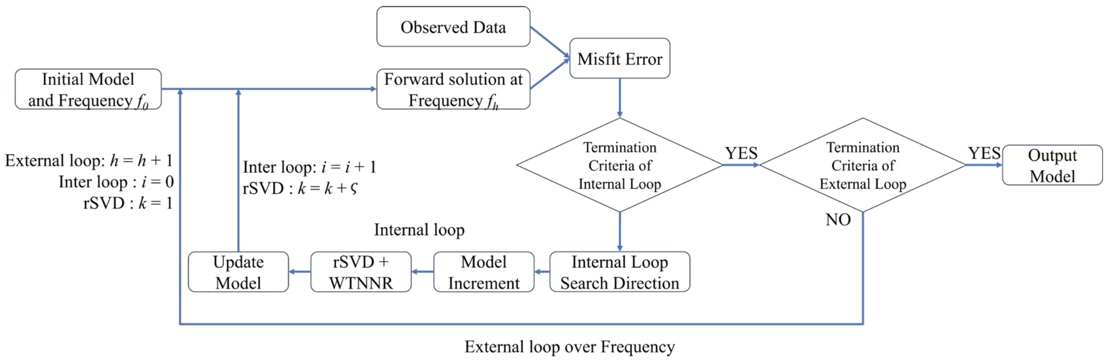

Figure 4.8. Flowchart of accelerated augmented Lagrangian full-waveform inversion based on truncated randomized singular value decomposition in the frequency domain (Li *et al.*, 2024b).

The iALM is a new iterative method for solving the linear constraint convex function minimization problem. This iterative method is an inexact version of the ALM. Since every single sub-problem needs to be solved accurately in each external iteration, the computational cost of the other variants version of ALM does not make the convergence rate faster. More importantly, the subproblems of accelerated augmentation Lagrangian do not have closed-form solutions. Therefore, the algorithm adopted allows inexact solutions to the subproblems. When solving the problem of linearly constrained convex programming, the iterative complexity of classical augmented Lagrange is $O(1/k)$, the content about the computational complexity of the ALM is in Appendix B. And the iterative complexity of the accelerated algorithm is $O(1/k^2)$ (Kang *et al.*, 2015), and the content about the computational complexity of the iALM is in Appendix C. Therefore, it is theoretically proved that my proposed iALM has a faster convergence rate than ALM, which is worth implementing for the inversion process. This point makes the iALM accelerate the convergence rate and decrease the computational cost.

4.5 Numerical Simulations on Synthetic Data

In the forthcoming section, I aim to evaluate the efficacy of the rSVD-WTNNR-based FWI using synthetic data. For this, I've chosen three intricate scenarios with diverse SNR ratios to juxtapose the inversion outcomes of this algorithm against those from the traditional Tikhonov regularization FWI. For the forward modelling aspect, I persist in employing the multi-scale frequency domain method, combined with the Perfectly Matched Layer (PML) boundary condition, and utilize model error for quantitative comparison (Bunks *et al.*, 1995):

$$\left\|\mathbf{M}_{\text{true}} - \mathbf{M}_{\text{inv}}\right\|_2 / \left\|\mathbf{M}_{\text{true}}\right\|_2 . \tag{4.19}$$

where the $\mathbf{M}_{\text{true}}$ and $\mathbf{M}_{\text{inv}}$ represent the actual model and inversion result, respectively (Komatitsch and Tromp, 2003).

To assess the performance of the algorithm under complex conditions, this chapter tests the inversion capability under three different signal-to-noise ratio conditions:

$$\text{SNR} = 20 * \text{Log}_{10}\left(\frac{\left\|\mathbf{D}\right\|_2}{\left\|\boldsymbol{\eta}\right\|_2}\right), \tag{4.20}$$

where the $\boldsymbol{\eta}$ is the noise data, the $\mathbf{D}$ is the signal data.

4.5.1 The Central Part of the 2004 BP Model

In the current chapter, I maintain the use of the central segment of the 2004 BP model. This model embodies a high-velocity salt irregularity and adjacent channels. The primary inversion difficulty of this model is in sketching the boundaries of the central high-velocity salt formation, with particular emphasis on accurately inverting the two lower salt pillars, which are concealed by potent reflective boundaries (Billette and Brandsberg-Dahl, 2005).

Figure 4.9 (a) shows the benchmark, while (b) presents the starting model derived by smoothing the benchmark.

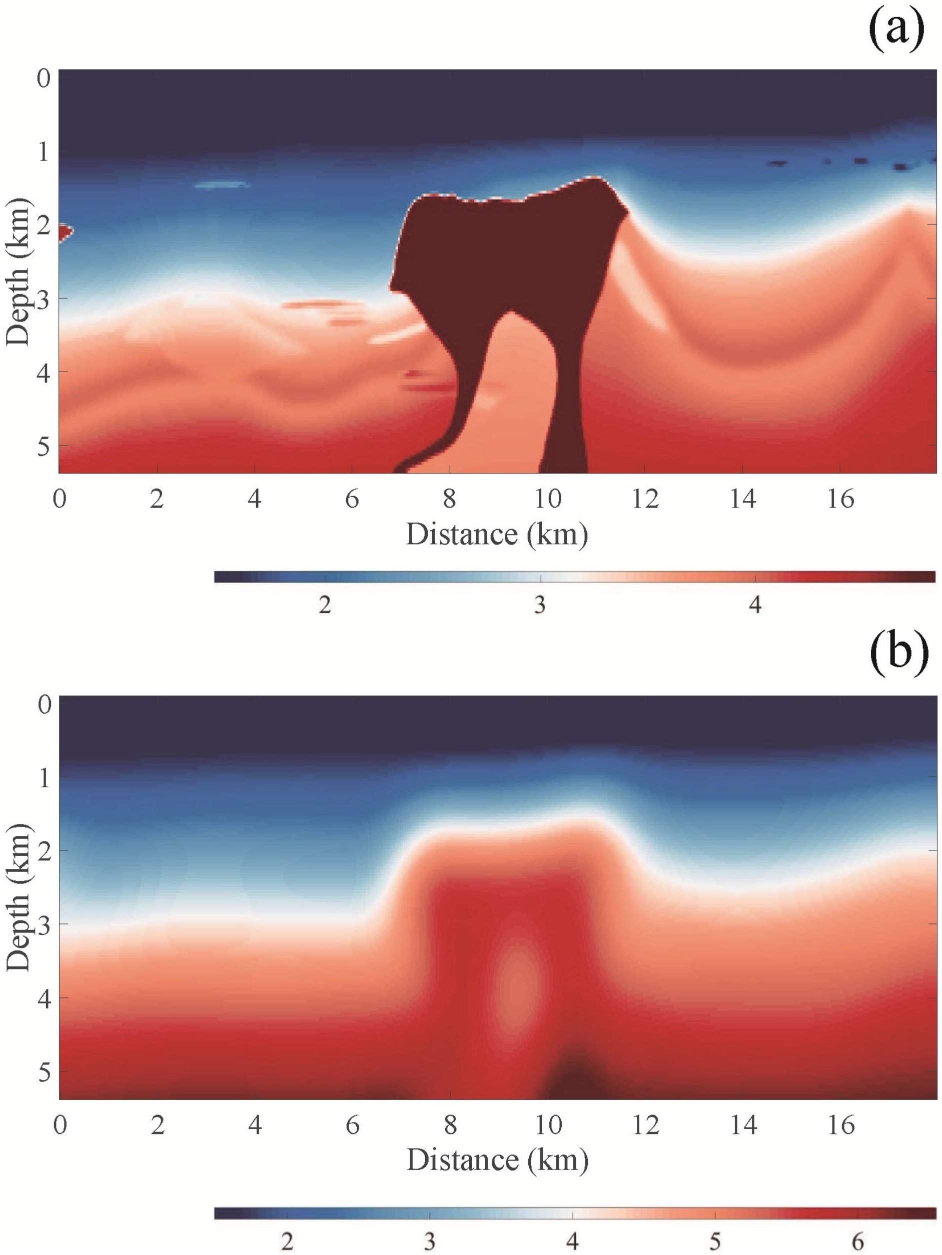

Figure 4.9. 2004 BP benchmark; (a) true benchmark; (b) starting model. The unit of the colour bar is km/s (Li *et al*., 2024b).

Figure 4.10 (a) displays the source and receiver wavefield obtained from forward modelling at 3 Hz. To test the performance under different SNR conditions, (b-d) show random noise data under three different scenarios: 8 *dB*, 12 *dB*, and 16 *dB*, these noise data are linearly superimposed on the true velocity wavefield to produce (e-g).

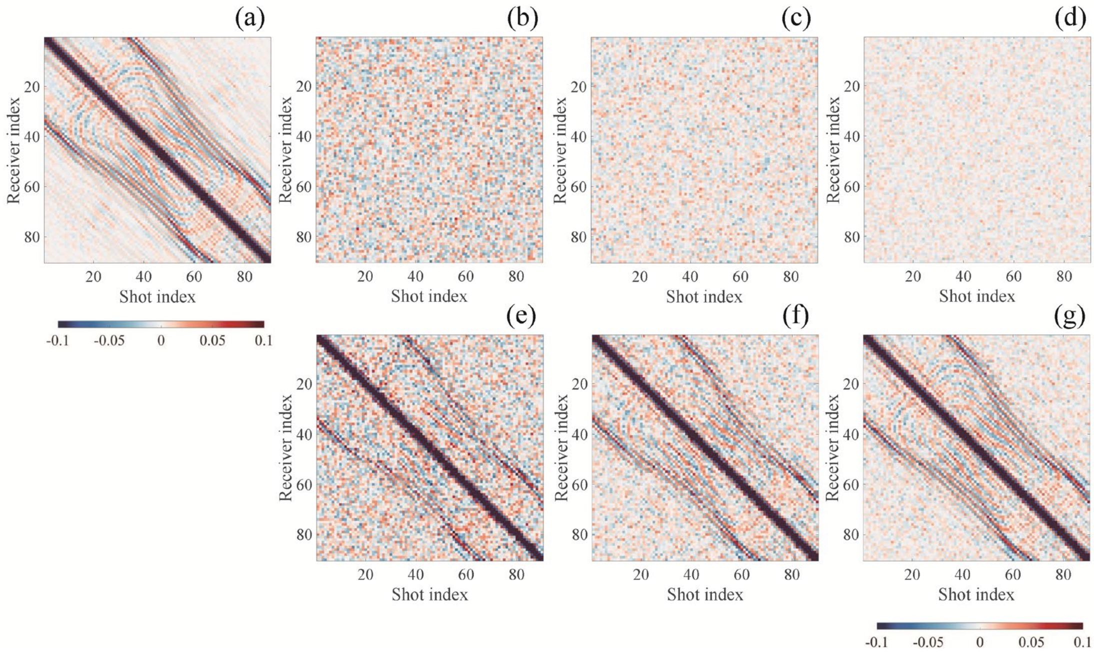

Figure 4.10. The real part of the observations and simulated data in the source-receiver domain at 3 Hz of the 2004 BP model; (a) pure data matrix; (b) 8 *dB* random noise; (c) 12 *dB* random noise; (d) 16 *dB* random noise; (e-g) noisy data (Li *et al*., 2024b).

Subsequently, Figure 4.11 compares the model increments obtained in a single iteration using conventional methods and the proposed algorithm in this chapter, where (a1-a5) represent model increments from the conventional method under various truncation parameters, while (b1-b5) represent increments obtained using the rSVD-WTNNR algorithm. Notably, the updated algorithm enhances the main features of the 2004 BP model and weakens the effect of the channels on both sides.

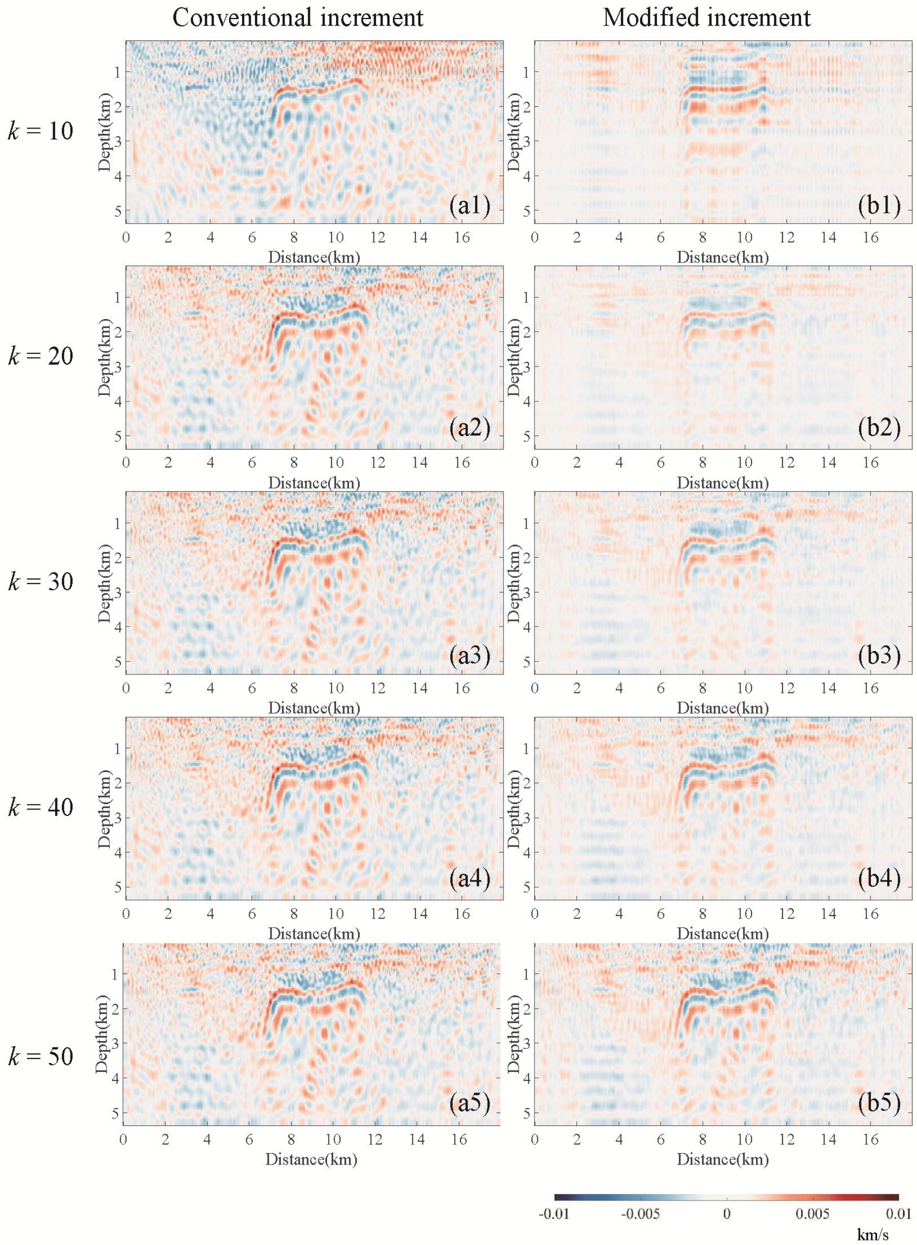

Figure 4.11. Conventional velocity increment model; (a1-a5) close to the 4.5 Hz frequency, the truncation parameter k, which increases with the rising count of internal iterations, is equal to 10, 20, 30, 40, and 50, respectively; modified velocity increment model; (b1-b5) close to the 4.5 Hz frequency, the truncation parameter k is equal to 10, 20, 30, 40, and 50, respectively (Li *et al*., 2024b).

Based on the above, Figure 4.12 displays the final inversion results of both algorithms under 8 *dB* random noise, (a1-a5) represent results based on Tikhonov regularization; (b1-b5) depict differences between the conventional method's inversion and the true benchmark, in contrast, (c1-c5) and (d1-d5) represent results and differences from the improved algorithm, respectively. It's clear that the proposed method significantly outperforms the conventional one, especially in noise removal, salt pillar reconstruction at the model bottom, and resolution.

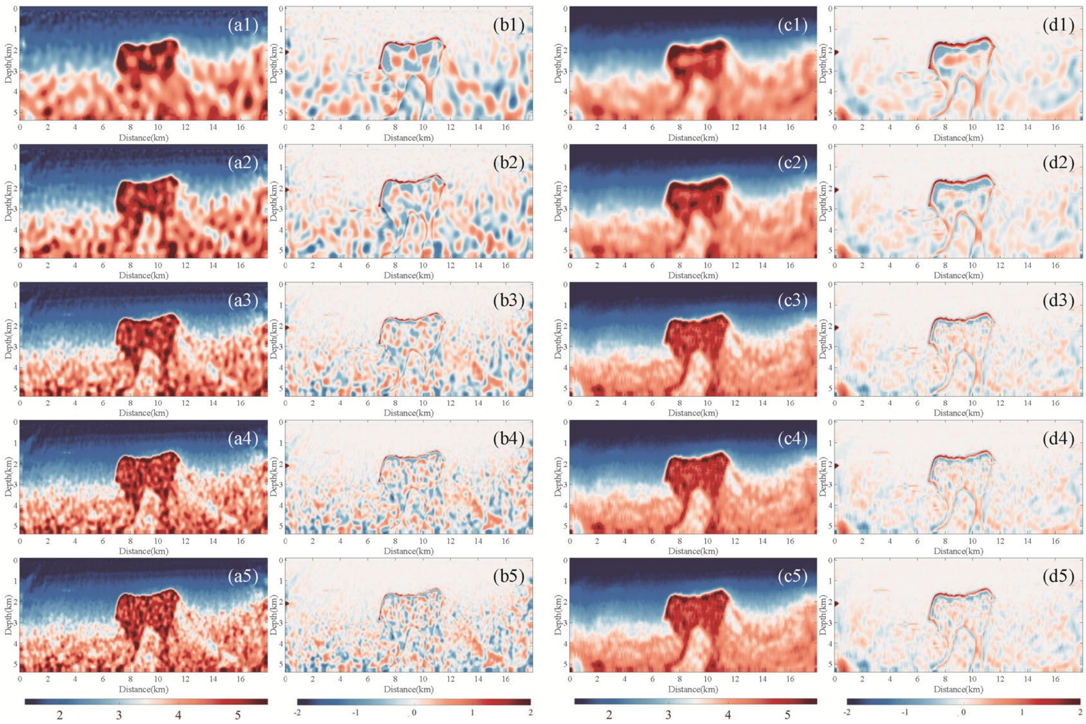

Figure 4.12. 2004 BP model with 8 *dB* random background noise; (a1-a5) conventional method results with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24 Hz, respectively; (b1-b5) velocity differences of the conventional method and the benchmark; (c1-c5) inversion results based on modified method with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24, respectively; (d1-d5) velocity differences between the inversion results based on modified method and the real benchmark (Li *et al.*, 2024b).

Figure 4.13 displays the one-dimensional velocity comparison of the two methods, with red, blue, black, and grey dashed lines representing the improved algorithm's velocity profile, the conventional method's profile, the true benchmark's profile, and the initial model's profile, respectively. Quantitative comparisons further emphasize the superiority of the improved algorithm.

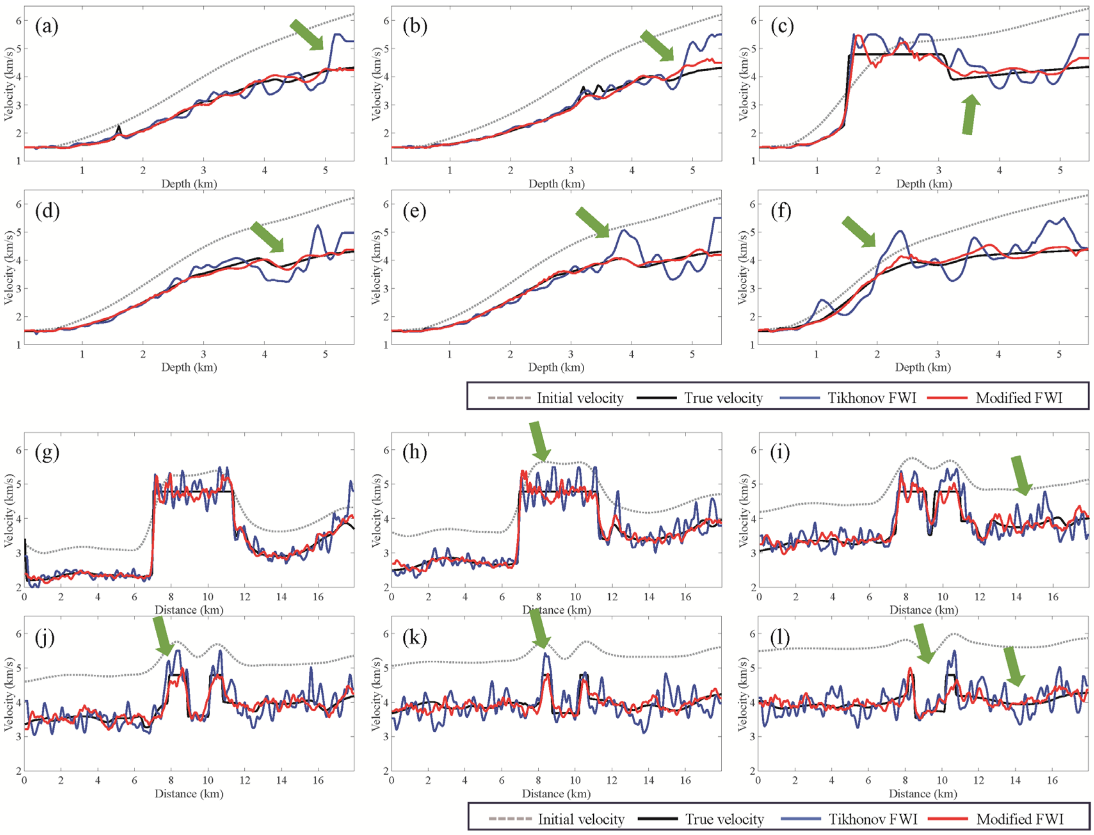

Figure 4.13. One-dimensional velocity at varying X-coordinates; for X = 2.84, 5.75, 11.08, 13.92, 15.00, and 17.80 km separately in (a-f); one-dimensional velocity at varying Y-coordinates; for Y = 2.36, 2.76, 3.36, 3.68, 4.08, and 4.56 km separately in (g-l) (Li *et al*., 2024b).

Similar comparisons are provided for 12 *dB* and 16 *dB* random noise in Figures 4.14-4.16. In high SNR scenarios, although the conventional method performs exceptionally well, the improved method further refines the results.

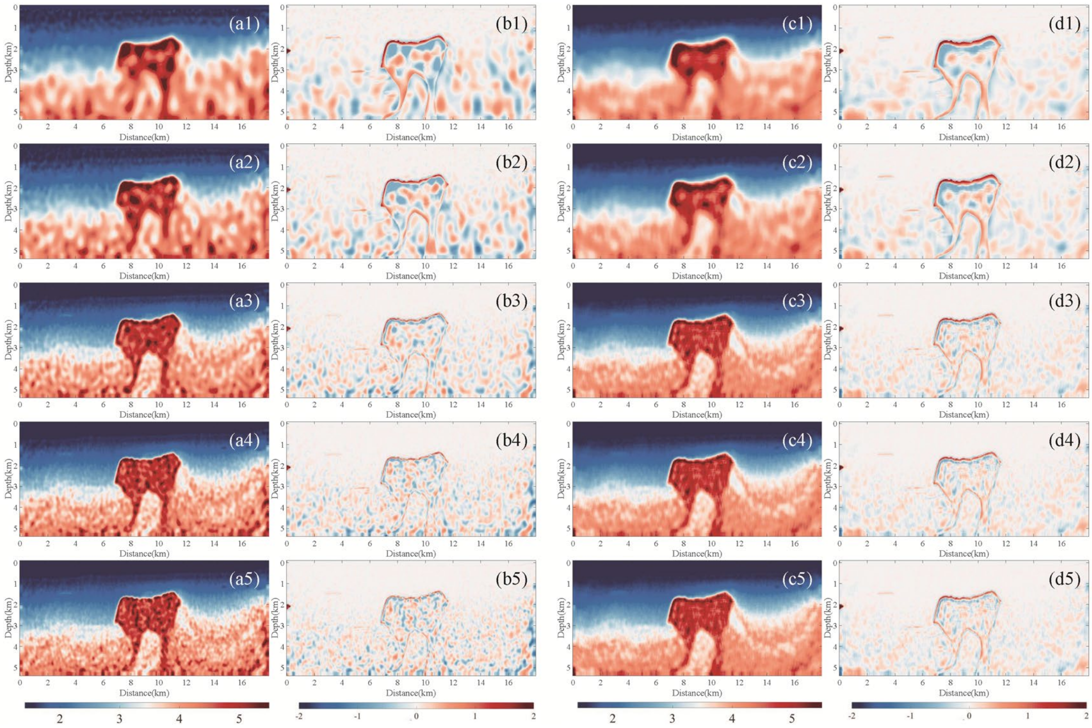

Figure 4.14. 2004 BP model with 12 *dB* random background noise; (a1-a5) conventional method results with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24 Hz, respectively; (b1-b5) velocity differences of the conventional method and the benchmark; (c1-c5) inversion results based on modified method with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24, respectively; (d1-d5) velocity differences between the inversion results based on modified FWI and the real benchmark (Li *et al*., 2024b).

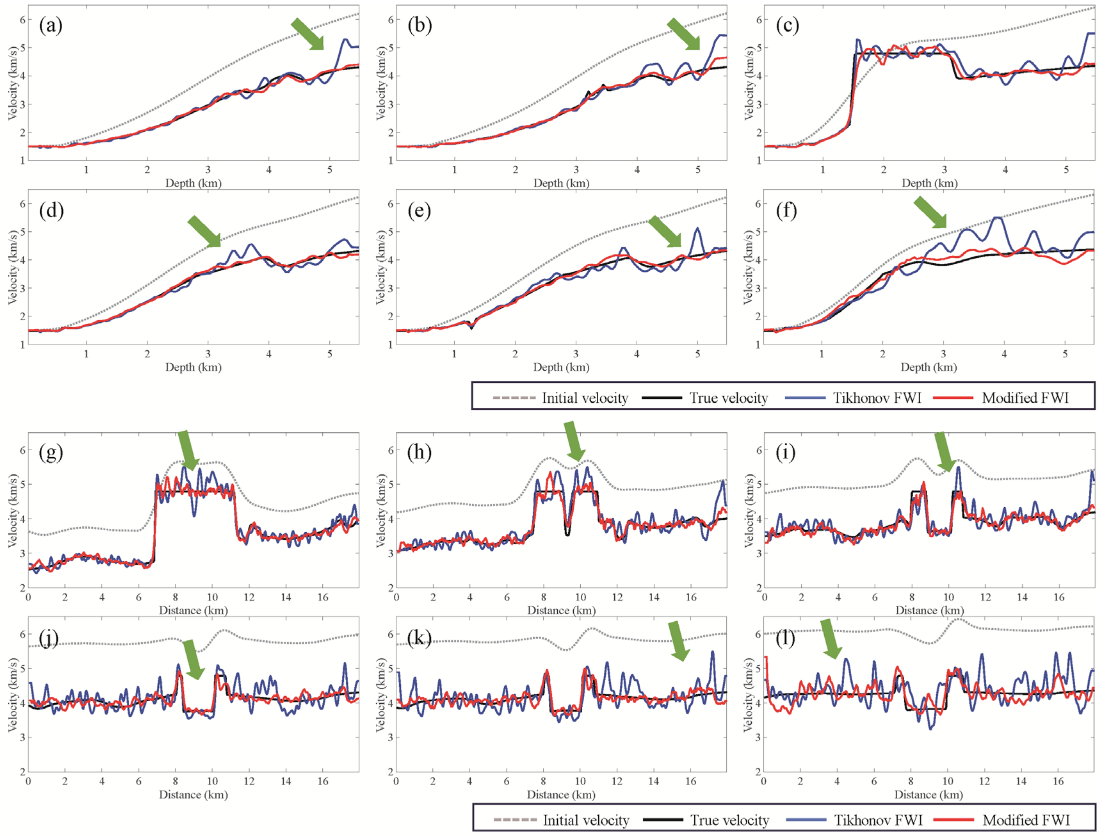

Figure 4.15. One-dimensional velocity at varying *X*-coordinates; for *X* = 4.48, 6.00, 11.08, 13.88, 14.68, and 17.76 km separately in (a-f); one-dimensional velocity at varying *Y*-coordinates; for *Y* = 2.80, 3.36, 3.80, 4.76, 4.84, and 5.28 km separately in (g-l) (Li *et al*., 2024b).

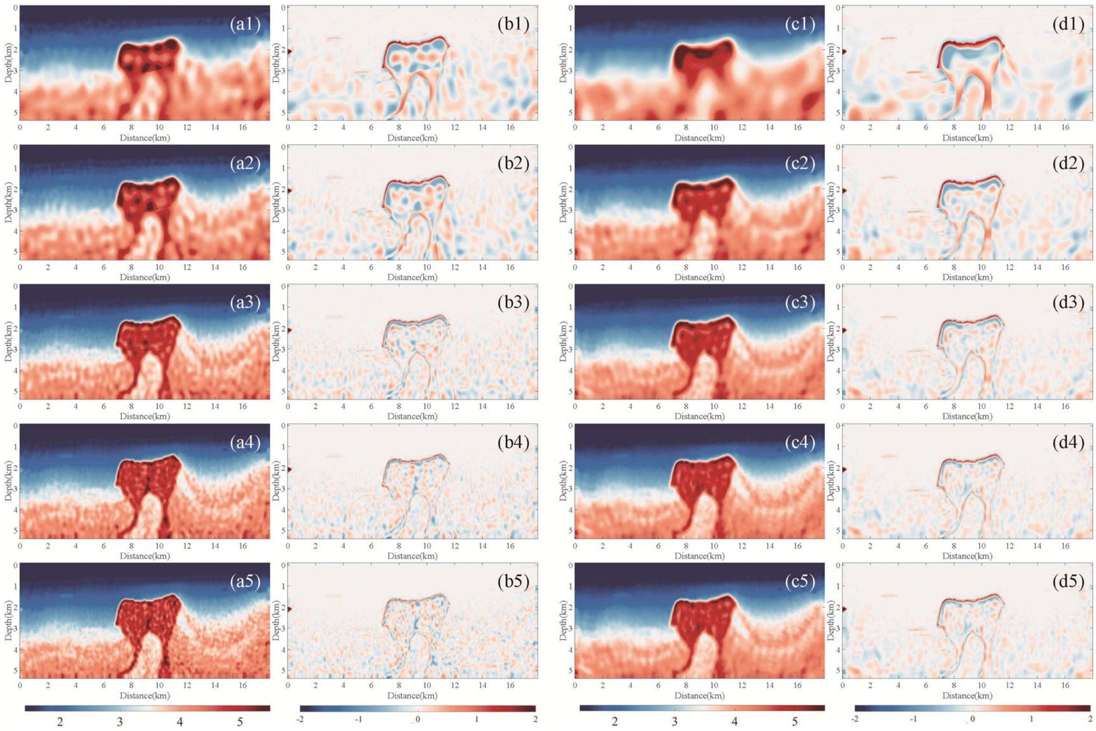

Figure 4.16. 2004 BP model with 16 *dB* random background noise; (a1-a5) conventional method results with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24 Hz, respectively; (b1-b5) velocity differences of the conventional method and the real benchmark; (c1-c5) inversion results based on modified method with frequencies of 2.15, 3.9, 5.35, 7.70, and 9.24, respectively; (d1-d5) velocity differences between the inversion results based on modified method and the real benchmark (Li *et al*., 2024b).

Figure 4.17 displays six groups of horizontal and vertical velocity comparisons, confirming that under various noise conditions, the improved algorithm results are closer to the real benchmark.

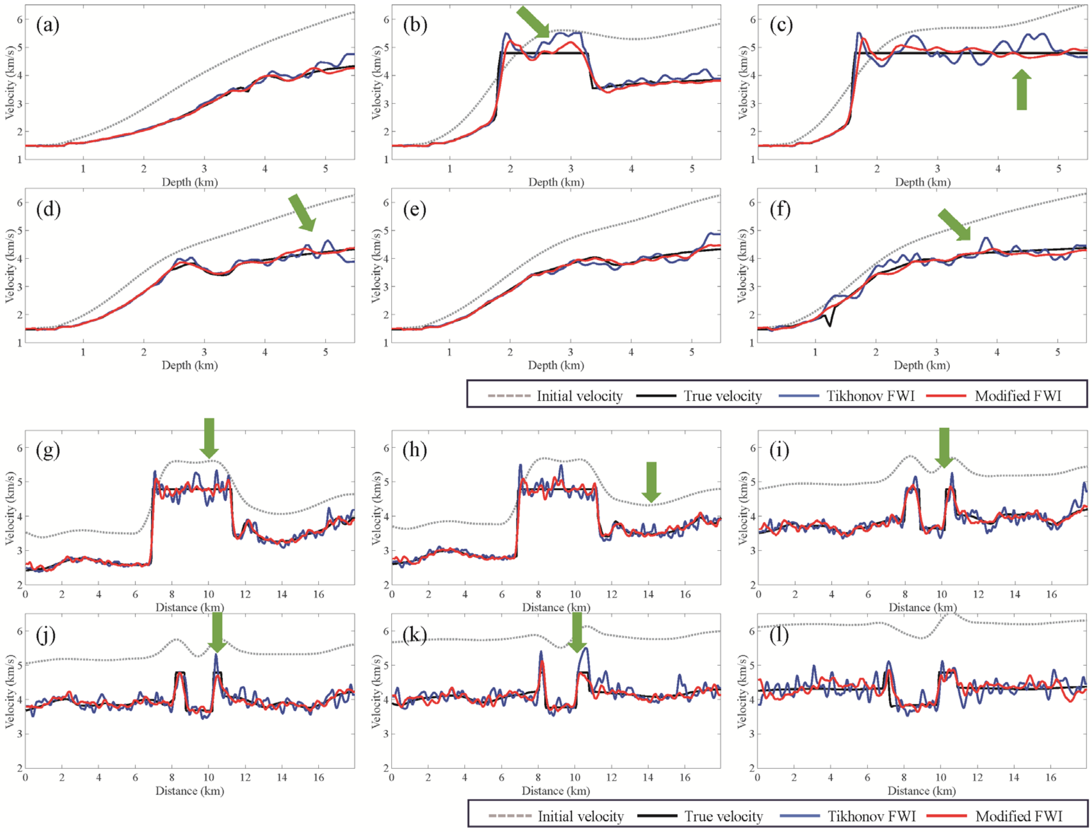

Figure 4.17. One-dimensional velocity at varying X-coordinates; for X = 6.44, 9.24, 10.44, 12.12, 16.20, and 18.00 km separately in (a-f); one-dimensional velocity at varying Y-coordinates; for Y = 2.68, 2.88, 3.84, 4.08, 4.80, and 5.44 km separately in (g-l) (Li *et al*., 2024b).

Finally, Figure 4.18 compares the convergence speeds of the mismatch function and model error under three conditions, highlighting the enhanced algorithm's faster convergence, especially under low SNR conditions.

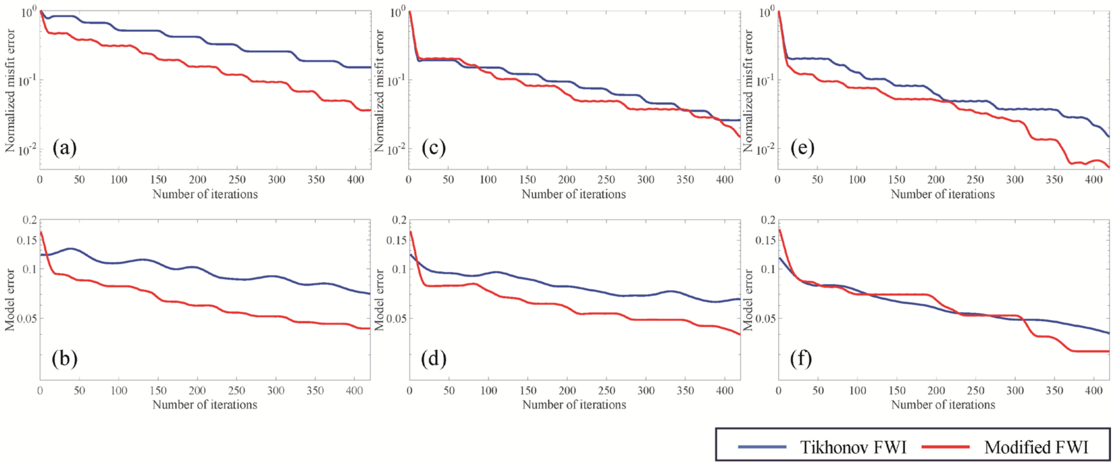

Figure 4.18. Error plots, (a-b) misfit and model error in 8 *dB* random background noise case; (c-d) misfit and model error in 12 *dB* random noise case; (e-f) misfit and model error in 16 *dB* random noise case; the solid red line is the modified method, and the blue line is the conventional method (Li *et al*., 2024b).

To better verify the efficacy of the improved algorithm in alleviating the cycle skipping issue, an additional series of tests were conducted using a suboptimal initial model for full waveform inversion. Figure 4.19 depicts the initial velocity model characterized by a one-dimensional linear increase in velocity, clearly inferior to the smooth initial model.

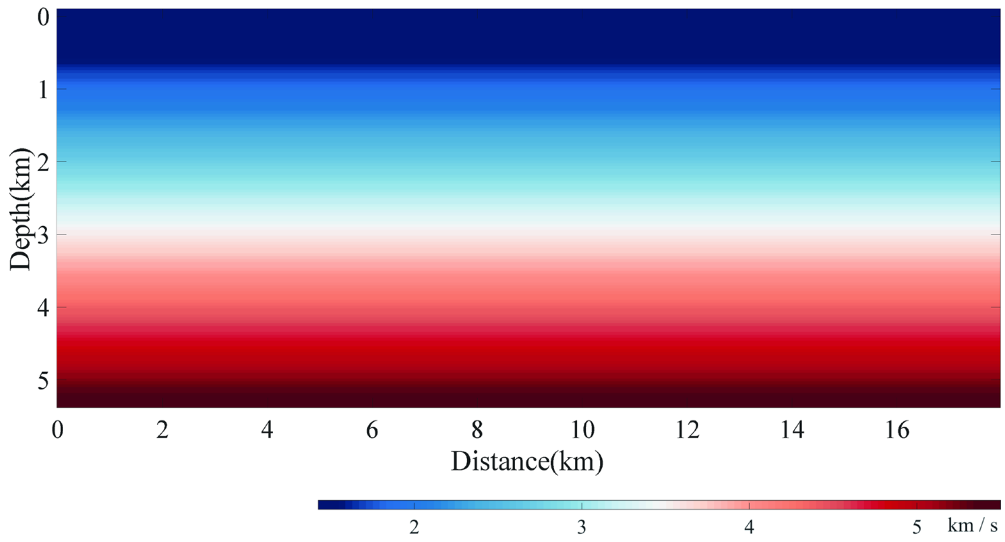

Figure 4.19. 2004 BP model, one-dimensional linearly increasing velocity initial model (Li *et al*., 2024b).

The actual model is consistent with that presented in Figure 4.9 (a). Selecting a suboptimal initial model poses a test for the robustness of the improved algorithm and its ability to mitigate cycle skipping. Figure 4.20 depicts the real part of the 2.69 Hz data for the 2004 BP model.

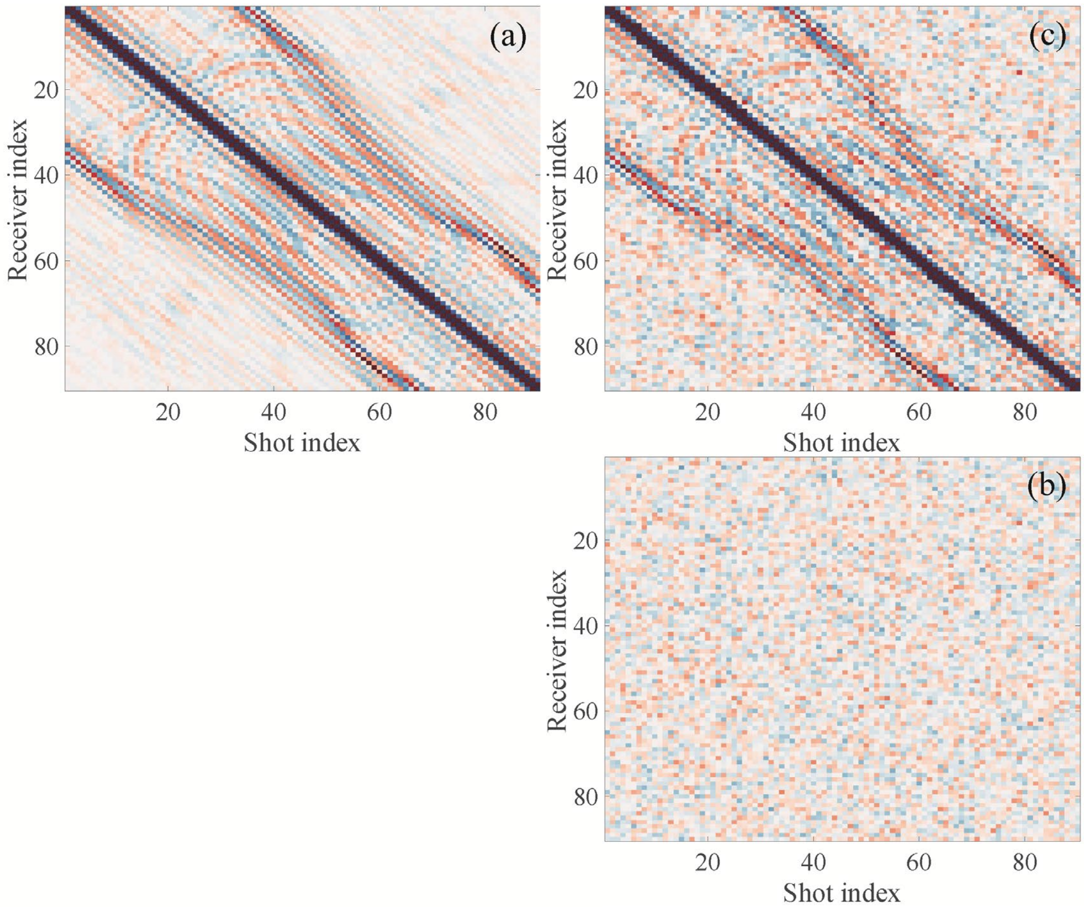

Figure 4.20. The real part of the 2.69 Hz data sets for the 2004 BP model. (a) noise-free data; (b) random noise of 12 *dB*; (c) noisy data. Figure 4.20 (a) depicts the real part of the 2.69 Hz data for the 2004 BP model without noise. Figure 4.20 (b) shows data with 12 *dB* noise, and Figure 4.20 (c) presents the real part of the 2.69 Hz data with the addition of 12 *dB* noise (Li *et al*., 2024b).

Figure 4.21 showcases the inversion results based on the suboptimal initial model, with Figures 4.21a1-4.21a5 respectively illustrating the inversion outcomes at frequencies of 2.69 Hz, 3.87 Hz, 5.57 Hz, 6.69 Hz, and 9.63 Hz, employing Tikhonov regularization.

Figures 4.21b1-4.21b5 depict the results obtained using the improved algorithm.

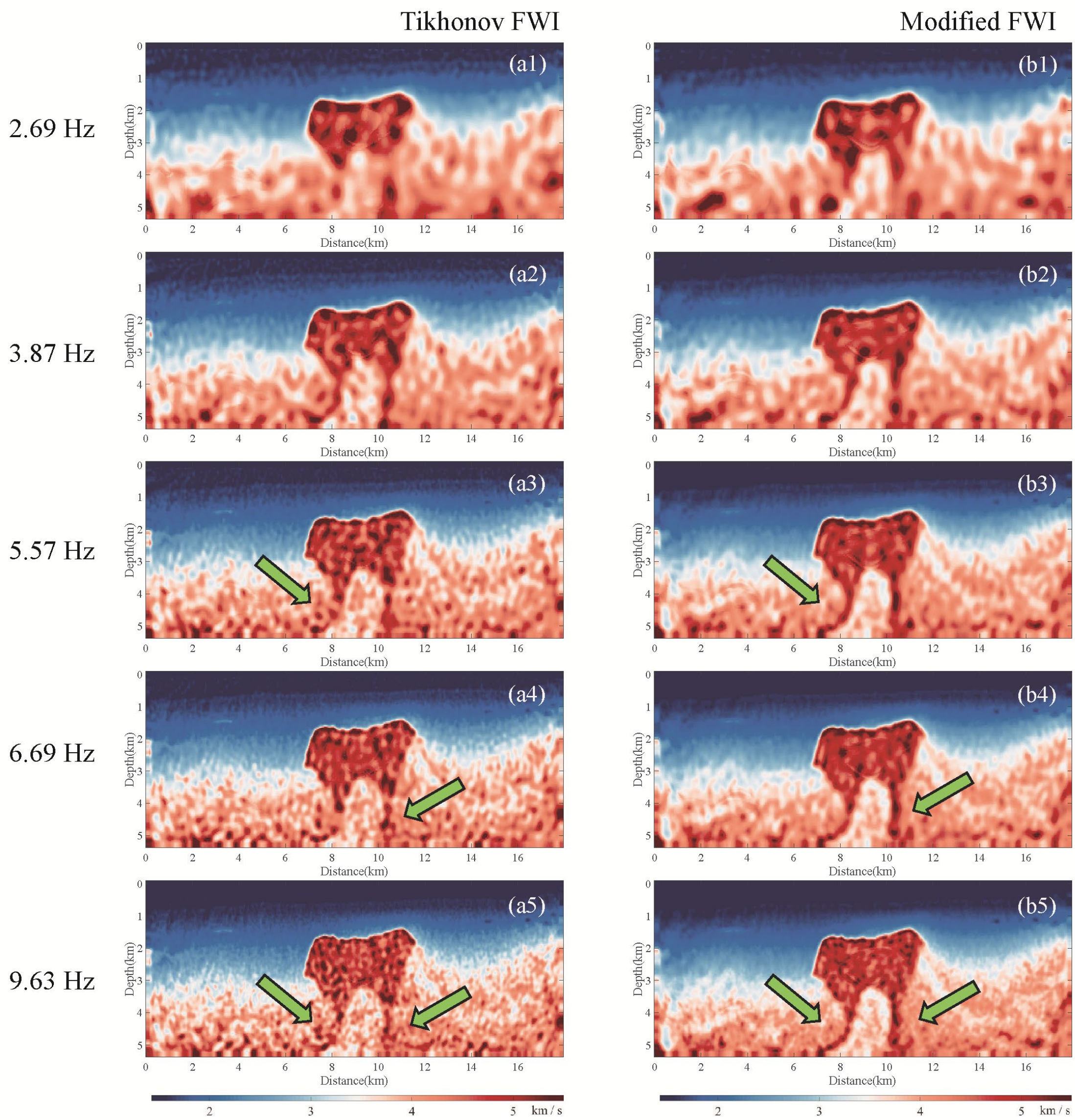

Figure 4.21. 2004 BP model with 12 *dB* random background noise, (a1-a5) inversion results based on Tikhonov regularised FWI with frequencies of 2.69 Hz, 3.87 Hz, 5.57 Hz, 6.69 Hz, and 9.63 Hz, respectively; (b1-b5) inversion results based on modified FWI with frequencies of 2.69 Hz, 3.87 Hz, 5.57 Hz, 6.69 Hz, and 9.63 Hz, respectively. The initial model is shown in Figure 4.19b (Li *et al*., 2024b).

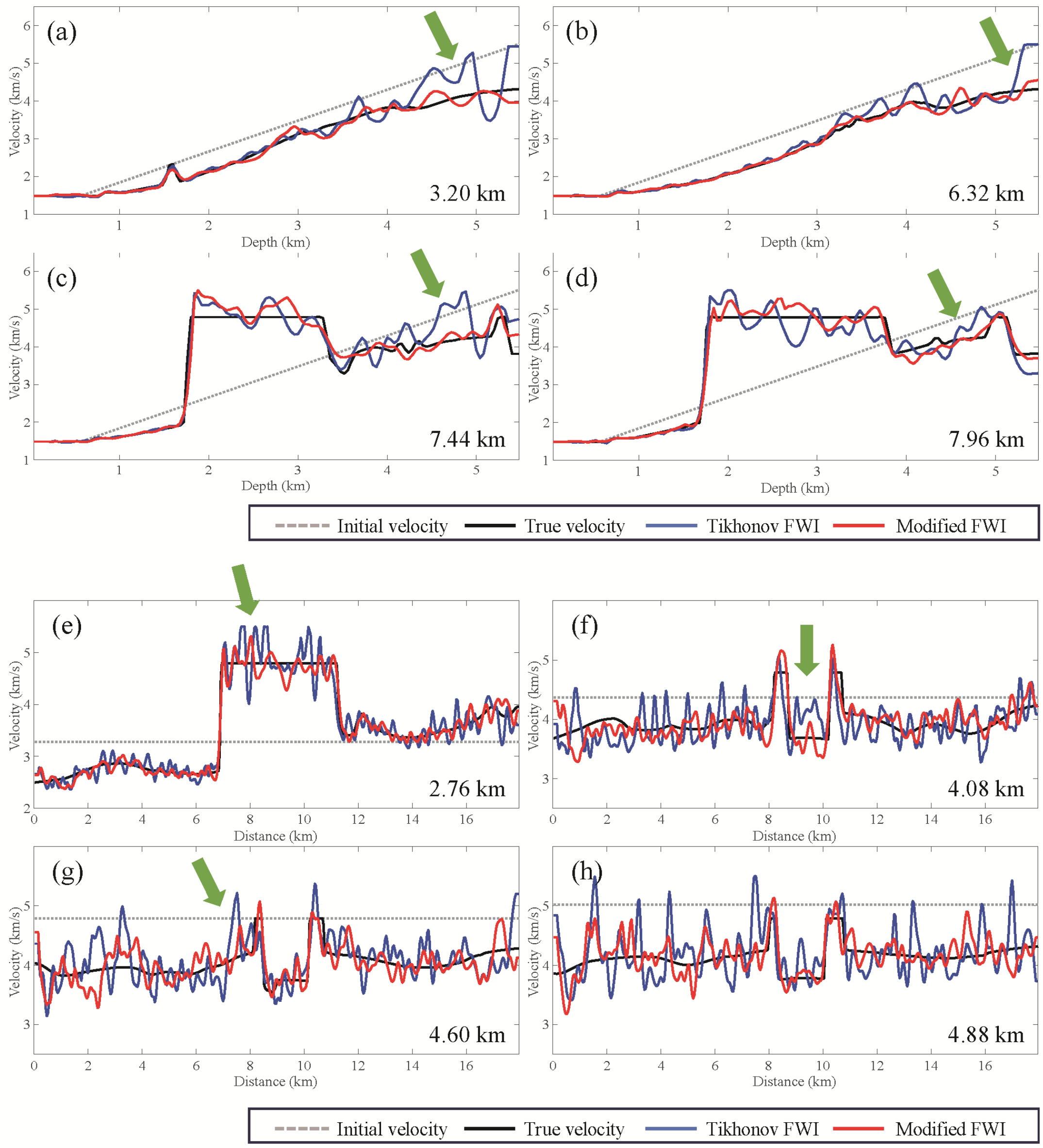

Figure 4.22. 2004 BP model with 12 *dB* random background noise, (a-d) the vertical comparison of one-dimensional velocity models at different x-positions, (a) x = 3.20 km, (b) x = 6.32 km, (c) x = 7.44 km, (d) x = 7.96 km; (e-h) the horizontal comparison of one-dimensional velocity models at different z-positions, (e) y = 2.76 km, (f) y = 4.08 km, (g) y = 4.60 km, (h) y = 4.88 km; where the actual velocity model is the solid black line, initial velocity model is a grey dotted line, the Tikhonov FWI is solid blue line, and the modified FWI is solid red line. The velocity comparisons above are based on Figure 4.21 (Li *et al.*, 2024b).

The findings reveal that the traditional algorithm's performance in the deeper regions leaves much to be desired, highlighting its vulnerability to poor initial models and noise interference. Conversely, the improved algorithm, despite the suboptimal initial model, is capable of inverting the model's deep salt structures with greater clarity, attesting to its efficacy and superiority. Figure 4.22 compares the one-dimensional velocity inversion outcomes of both algorithms under the poor initial model scenario. The comparison indicates that the velocity profile generated by the improved algorithm more closely approximates the true velocity, with green arrows highlighting areas with notable velocity discrepancies.

Lastly, results for noise-free data were demonstrated. As depicted in Figure 4.23, Figures 4.23a1-4.23a5 show the inversion results of the traditional algorithm without noise, while Figures 4.23b1-4.23b5 reflect those of the improved algorithm in a noise-free environment. Figure 4.23 illustrates the method's approach to handling noise-free data. Theoretically, in noise-free or low-noise scenarios, either a larger initial truncation value or a smaller truncation step should be set, each condition being adequate. While the improved algorithm might truncate some beneficial information, the overall resolution of the inversion results remains high in conditions of low or no noise.

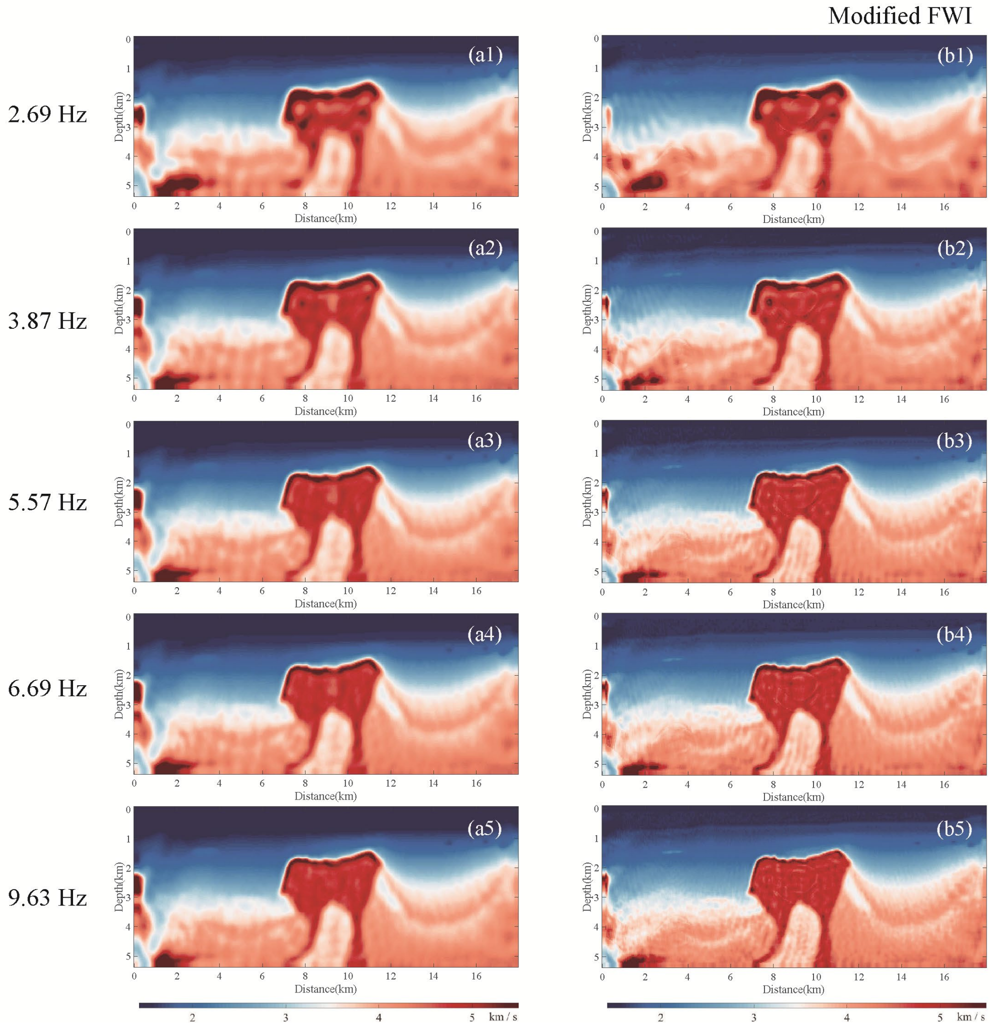

Figure 4.23. 2004 BP model without noise interference, (a1-a5) inversion results based on regularised FWI with frequencies of 2.69 Hz, 3.87 Hz, 5.57 Hz, 6.69 Hz, and 9.63 Hz, respectively; (b1-b5) inversion results based on modified FWI with frequencies of 2.69 Hz, 3.87 Hz, 5.57 Hz, 6.69 Hz, and 9.63 Hz, respectively. The initial model is shown in Figure 4.19 (Li *et al.*, 2024b).

Across the three sets of experiments, the improved algorithm consistently outperforms under various complex conditions, suppressing noise more effectively, delivering clearer inversions of the salt pillars at the model bottom, mitigating the obscuring effects of strong reflection interfaces, and vividly outlining the detailed structural contours of the model bottom (Li *et al*., 2024b).

4.6 Chapter Discussions

While FWI is already a mature inversion technique, there are still many details that need refinement, especially when pursuing high-resolution and high-accuracy inversions. The inspiration for this chapter comes from an image processing technique, casting the denoising problem as an image processing challenge. Specifically, to validate the feasibility of the algorithm, Figure 4.24 presents a simple image processing test, I can interpret (a) as a certain model increment, through SVD, its singular value matrix (b) is derived, and then by introducing random noise (c) to the original matrix, (d) can be generated. It is evident that in comparison to the original singular value matrix, there are many superfluous singular values within the green box in (d), these smaller singular values are the noise and artefacts that need to be eliminated. By reconstructing (d), I can obtain the model in (e). Compared to the original (a), it is clear that noise data has seriously damaged the model structure, however, by following the methodology illustrated in (f), if I truncate all the singular values inside the red box and then reconstruct, I can achieve the result shown in (g). (g) reconstructed structure after truncation is notably improved compared to (e) and is closer to the initial model (a). In conclusion, this experiment validates the feasibility of the approach proposed in this chapter, and by decomposing the original matrix and truncating the smaller singular values in the singular value matrix, denoising and resolution enhancement can be achieved, which offers a new perspective for FWI optimization, providing a novel algorithm and tool for seismic exploration under complex conditions with data noise suppression and uncertainties, holding potential significant value.

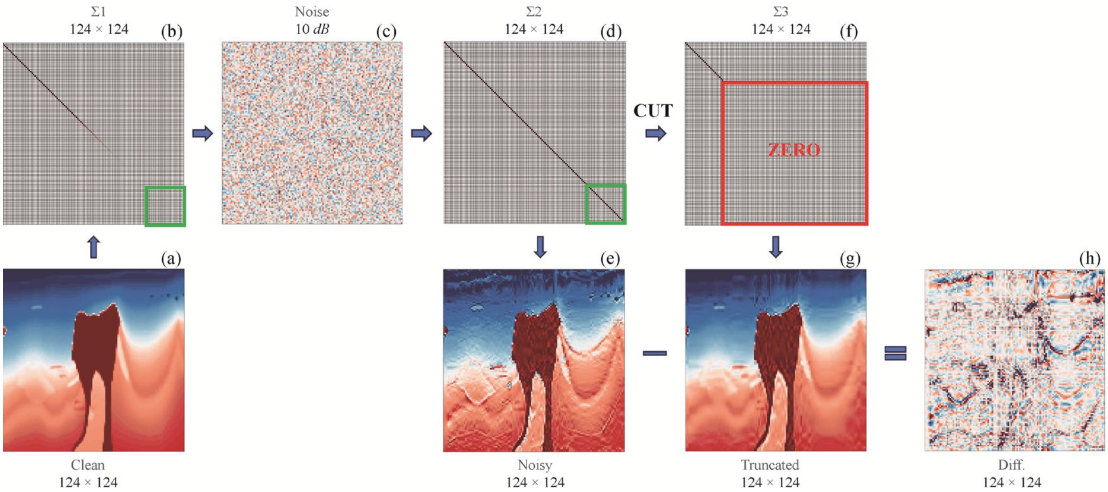

Figure 4.24. (a) The initial test velocity increment; (b) the singular value matrix for this velocity increment; (c) 10 *dB* of random noise; (d) the test velocity increment after adding the noise disturbance; (e) the velocity increment after reconstructing; (f) new matrix after simply truncating the singular value matrix (d); (g) the new velocity increment with a reconstructing of the singular value matrix (f); and (h) the difference between velocity increments (e) and (g) (Li *et al.*, 2024b).

Additionally, there is a need for further elucidation on the rSVD algorithm, this algorithm stands as a potent tool for matrix decomposition. For a more intuitive theoretical interpretation, Figure 4.25 and Figure 4.26 offer a geometric explanation. As observed in Figure 4.25, the conventional SVD decomposes a target matrix into two orthogonal matrices and a diagonal matrix, these orthogonal matrices represent two rotations in different directions, with the rotation angles denoted by the matrix values, but the diagonal matrix signifies the stretching or compression level of the singular value matrix. From Figure 4.26, it is evident that if I select only the first few larger singular values and exclude the smaller ones, I retain crucial information from the first few dimensions, while the singular values in other dimensions get compressed to zero. In essence, if the original matrix is n-dimensional, rSVD can reduce it to k dimensions, where k is the truncation parameter. Moreover, for large-scale matrices typical of seismic data, the computational efficiency of the traditional SVD is quite low, and the computational cost is substantial. Hence, the rSVD algorithm can effectively reduce memory consumption, expedite computation, and boost the overall computational efficiency of FWI.

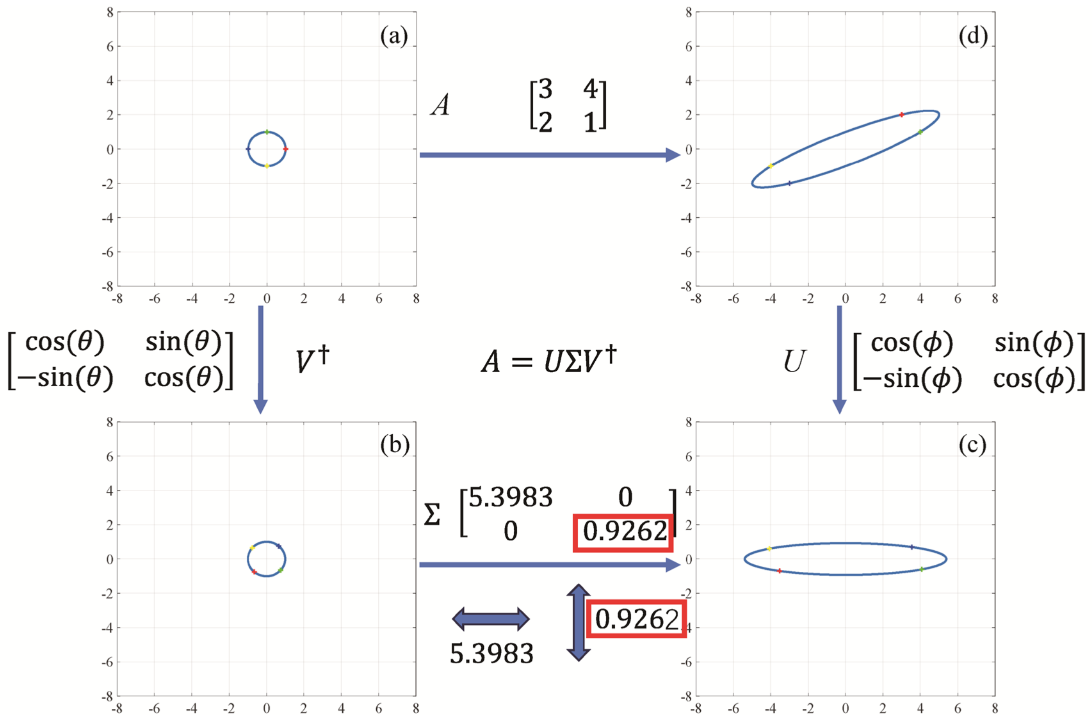

Figure 4.25. Geometric interpretation of SVD; (a) target matrix; (b) left orthogonal matrix representing rotation; (c) singular value matrix representing longitudinal and horizontal stretching; (d) right orthogonal matrix representing rotation (Li *et al*., 2024b).

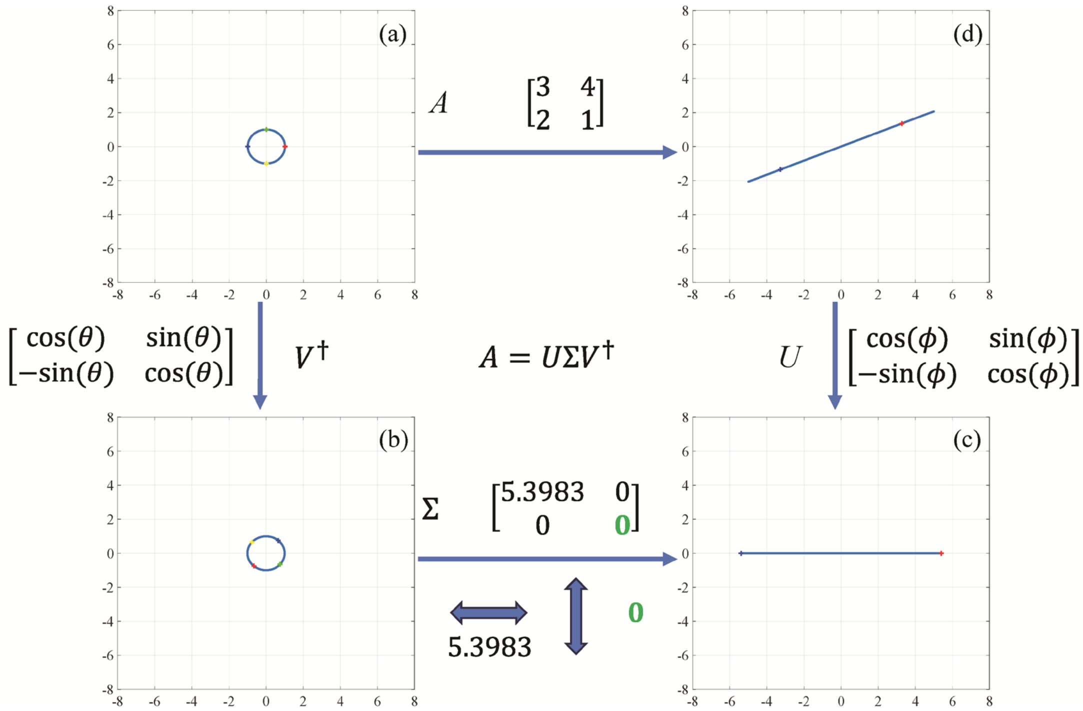

Figure 4.26. Geometric interpretation of the truncation operation of the SVD; (a) the target matrix; (b) the left orthogonal matrix represents the rotation; (c) the singular value matrix represents the longitudinal and horizontal stretching and truncation of its smallest singular value results in a longitudinal stretch of zero; (d) the right orthogonal matrix represents the rotation (Li *et al.*, 2024b).

4.7 Chapter Conclusions

In this chapter, considering the unique geological conditions of the Gulf of Mexico, I propose an rSVD-WTNNR algorithm to enhance the inversion accuracy of FWI and its noise suppression capabilities. Building on the basic FWI workflow, this chapter innovatively incorporates the matrix decomposition algorithm, rSVD, into the optimization process. Following this, I employ WTNNR to further refine the decomposed singular value matrix. Following that, the iALM algorithm is employed to hasten the overarching optimization procedure, enhancing the algorithm's rate of convergence even more. In the section dedicated to numerical experiments, I juxtapose the FWI inversion outcomes derived from Tikhonov regularization against those sourced from the rSVD-WTNNR algorithm, and from figures, it is evident that the improved algorithm introduced in this chapter offers more precise inversion accuracy under various complex signal-to-noise ratios. Specifically, it achieves accurate inversion of deep structures below the strong reflective surface of high-velocity salt bodies, accomplishing the research objective of recognizing the contour of deep high-speed salt structures in the region.

CHAPTER V:DISCUSSIONS

FWI, as a numerical fitting algorithm, has been developed for nearly 40 years. Although it has been applied on a small scale conditionally in seismic exploration and data interpretation, there is still some distance from its large-scale promotion in the industry (Li *et al*., 2024d). Here are several key issues and challenges, as well as my personal views and prospects for FWI.

Firstly, the forward modelling algorithm requires a large amount of computational memory. Whether it is single-parameter forward modelling, multi-parameter forward modelling, frequency domain, time domain, or even the Laplace domain, grid-based algorithms are almost universally used, which ensures a huge demand for computer memory. However, if I abandon grid-based algorithms, currently, there is no better forward modelling algorithm to replace them. Hence, the higher the precision of the forward modelling, the finer the grid required, and consequently, the greater the computer memory needed. One factor restraining FWI's development is not its algorithm per se but the performance of the computers themselves. Therefore, in this thesis, I propose the use of an SR3 algorithm to optimize the wavefield. This algorithm introduces a regularization term to control model complexity, thereby enhancing model stability. During the interpolation process, auxiliary matrices are utilized to estimate unknown data points, thus filling in missing values and making predictions. The SR3 algorithm can help generate smoother and more coherent data prediction results.

If we completely abandon traditional algorithms and turn to artificial intelligence, there are two approaches: 1. Utilize artificial intelligence algorithms. By drawing on a vast amount of training experience and inter-well data, one can build an initial model that conforms to big data patterns, but this approach still relies on forward computation. 2. Abandon forward computation entirely. artificial intelligence-based FWI, especially that grounded in deep learning and convolutional neural networks (CNN), does not rely on forward computation, instead, it trains a CNN to directly map seismic data to subsurface velocity models or other physical models. The advantage is evident; once trained, deep learning models can make predictions in milliseconds to seconds, making them much faster than traditional FWI. However, these models often require vast amounts of labelled data, and obtaining high-quality labelled data in geophysics is challenging. Even with ample high-quality data, training with large datasets still demands significant computational memory, bringing us back to computer performance.

Secondly, there is the issue of the misfit function in the minimization process. To enhance the denoising capability of FWI, this thesis used a K-support regularization algorithm. This algorithm combines the sparsity characteristics of TV regularization with the smoothness of Tikhonov regularization, presenting different regularization paths through an adjustable regularization parameter, thereby achieving improved denoising performance.

Furthermore, as previously mentioned, the crux of inverse problems lies in constructing and solving the misfit function. Traditional methods from least squares to quadratic penalty to augmented Lagrangian show that, regardless of constrained or unconstrained algorithms, the nature of inverse problems remains strongly nonlinear and ill-conditioned. The improvements made are essentially about further compressing the solution space; the smaller the solution space, the faster the solution speed and the less likely it is to fall into local minima. Another issue is the selection of parameters for the misfit function, an aspect that both conventional algorithms and deep learning-based construction need to address. At present, there is not a gorgeous and rigorous method to calculate the parameters for the misfit function; most rely on empirical values and heuristic thinking.

During the optimization, this thesis used an rSVD algorithm aimed at achieving background separation and low-rank denoising capabilities. By decomposing model increments, the algorithm obtains a singular value matrix and then truncates the full-rank matrix into a low-rank matrix to achieve spatial dimensionality reduction and denoising. Additionally, a multiscale inversion strategy is utilized to update the condition number, creating multiscale models. This approach continuously optimizes the model increments within each iteration, thereby achieving the goal of high-resolution inversion.

Finally, in the optimization process, the gradient methods often adopted in FWI include the conjugate gradient (CG) and the L-BFGS method. The more popular L-BFGS algorithm calculates the second derivative of the objective function to obtain the Hessian matrix. However, directly calculating and storing the Hessian matrix can be very memory-intensive, which forces FWI to adopt optimization methods based on approximate Hessian matrices. Conventional optimization algorithms have two main shortcomings: 1. They easily fall into local minima, and 2. They have slow convergence rates. Regarding this, I believe that in the future, a gorgeous and potential alternative could be the Riemannian gradient (RG) method. Firstly, when model parameters naturally exist on a manifold, one could consider defining a Riemannian manifold. A Riemannian manifold is locally like Euclidean space, but globally it might have different curvatures and topological structures, and its gradient is the tangent vector on the manifold. Once an appropriate metric and connection are determined, one can obtain the Riemannian Hessian matrix. This Riemannian Hessian matrix is a second-order differential operator, describing how a real function defined on the manifold changes. Alternatively, one can project the Euclidean gradient directly onto the tangent space of the manifold to obtain the Riemannian gradient, which can be viewed as finding the part of the Euclidean gradient closest to a given point and ensuring it is located in the tangent space of that point. Compared to conventional algorithms, or even L-BFGS, Riemannian-type algorithms can directly optimize the manifold without dealing with complicated constraints. Moreover, theoretically, Riemannian algorithms may possess better convergence properties. However, the Riemannian method has a practical issue: it can be overly complex.

In summary, although FWI was initially proposed as a Born approximation algorithm, with the introduction of newer algorithms and the development of better computer performance, it has gradually deviated from its inherent concept and has adopted more advanced algorithms and new ideas. This provides researchers with an opportunity, from an algorithmic perspective, to leverage FWI's underlying structure, proposing better optimization algorithms for inverse problems and testing the effectiveness and feasibility of new algorithms in these contexts. Inverse problems are common and crucial in engineering, so extensive research into them is profoundly significant.

CHAPTER VI: CONCLUSIONS

This thesis presents multiple solutions to common challenges faced by the FWI algorithm and difficulties encountered in practice. I developed a preprocessing multi-scale resolution sparse optimization robust FWI, and the proposed algorithm was applied and tested in three different benchmarks.

Specifically, first, in response to the FWI forward problem being affected by low-density acquisition and the background noise of low-frequency components, Chapter II introduces a sparse regression regularization algorithm, SR3. By constructing auxiliary terms, it interpolates and reconstructs the seismic source-receiver data sets obtained in the frequency domain, enhancing the resolution of the wideband forward wavefield. This method achieves low-frequency interpolation denoising and high-frequency anti-aliasing effects, effectively addressing the wavefield spatial discontinuity caused by low-density acquisition and potential spatial aliasing effects. In the numerical experiments, this algorithm was tested on both single-layer and double-layer uniform media and applied to more complex synthetic data, proving the effectiveness of the proposed method.

Subsequently, in Chapter III, the misfit function, which is critical in inverse problems like FWI, was optimized. Conventional least-squares have strong non-linearity, easily falling into local minima and being sensitive to noise. Therefore, this chapter adopts a new regularization method that combines norms, specifically the K-support norm. By efficiently combining the ℓ_1 and ℓ_2 norms and utilizing the quadratic penalty method and ADMM algorithm enhances function convexity while emphasizing noise suppression. In the fitting experiment, this chapter uses random and coherent noise to test the robustness of the algorithm in various complex situations. Applications in three different synthetic datasets modelling different underground structures and real scenarios demonstrated the excellent property of the K-support norm.

Finally, focusing on the research hotspot of this thesis - the high-precision deep imaging of the abundant high-speed salt bodies and salt pillars structures in the Gulf of Mexico area, Chapter IV proposes a new rSVD-WTNNR algorithm to optimize the FWI inversion process. By integrating matrix and image processing techniques, the algorithm performs multi-resolution deep optimization on the velocity increment during the iterative process. By extracting singular values, it emphasizes the main features in the deep parts, avoiding the shielding effects of strong salt body reflections on deeper layers and the imprecise boundary recognition caused by strong velocity contrasts. In the experiments, I implemented the proposed approach on the 2004 BP model, which mirrors the eastern Gulf of Mexico and offshore Angola. The outcomes distinctly highlighted the efficiency and preeminence of the modified algorithm.

In summary, this thesis provides a comprehensive, multi-faceted enhancement and optimization of traditional FWI. Testing across multiple benchmarks has demonstrated that the preprocessing multi-scale resolution sparse optimization robust FWI proposed in this thesis is a promising approach, fulfilling the research goals and objectives outlined in this study.

APPENDICES

Appendix A. Wavefield-Reconstruction Inversion: WRI

Van Leeuwen and Herrmann proposed a wave equation FWI algorithm based on the penalty function in 2015 (van Leeuwen & Herrmann, 2015):

The regularization term of PDE constrained form can be expressed as follows:

$$\min_{\mathbf{M},\mathbf{X}} \|\mathbf{AX}-\mathbf{D}\|_2^2, \quad \text{subject to. } \mathbf{C}(\mathbf{M})\mathbf{X}=\mathbf{S}, \tag{A.1}$$

where $\|\bullet\|_2^2$ is the ℓ_2 norm, and the $\mathbf{M}\in\mathbb{R}^{\mathcal{N}\times 1}$ is the model parameter; $\mathbb{R}$ is the regularization function, contains preliminary information about model parameters, $\mathbf{X}\in\mathbb{R}^{\mathcal{N}\times 1}$ represents the wavefield, $\mathbf{D}\in\mathbb{R}^{\mathcal{M}\times 1}$ is the recorded seismic data, $\mathbf{S}\in\mathbb{R}^{\mathcal{N}\times 1}$ is the source term, and the linear observation operator $\mathbf{A}\in\mathbb{R}^{\mathcal{M}\times\mathcal{N}}$ sampling $\mathbf{X}$ at the receiver positions.

From the mathematical point of view, the most appropriate way to solve this type of constrained optimization problem is in the form of a Lagrangian function:

$$\min_{\mathbf{M},\mathbf{X}} \max_{\mathbf{V}} \mathrm{F}(\mathbf{M},\mathbf{X},\mathbf{V}) = \min_{\mathbf{M},\mathbf{X}} \max_{\mathbf{V}} \|\mathbf{AX}-\mathbf{D}\|_2^2 + \mathbf{V}^T\left[\mathbf{C}(\mathbf{M})\mathbf{X}-\mathbf{S}\right], \tag{A.2}$$

where $\mathbf{V}=\left[\mathbf{V}_1;\mathbf{V}_2;\ldots\right]$. The $\mathbf{V}$ represents the Lagrange multiplier. The advantage of using the Lagrange function is that after iterative optimization, new fitting data, forward results, and an adjoint matrix will be obtained simultaneously, which avoids the need for an explicit solution in the optimization process of conventional FWI, which means that the wavefield reconstruction inversion algorithm obtains a solution consistent with augmented wave equation term.

Since the model parameter $\mathbf{C}$ based on the PDE operator may converge to an approximate minimum value when the start model is not ideal, therefore, van Leeuwen and Herrmann redefine the primitive constrained problem into a quadratic penalty problem in 2013:

$$\min_{\mathbf{M},\mathbf{X}} \mathrm{F}(\mathbf{M},\mathbf{X}) = \min_{\mathbf{M},\mathbf{X}} \left\|\mathbf{AX}-\mathbf{D}\right\|_2^2 + \lambda \left\|\mathbf{C}(\mathbf{M})\mathbf{X}-\mathbf{S}\right\|_2^2 . \tag{A.3}$$

With the form of a quadratic penalty term, the original full space-constrained form can be turned into a double penalty term form, reducing the complexity of the algorithm while improving the ability of FWI to converge to the more exact minimum value when the initial model is poor.

Appendix B. Iteration-Complexity of the Classical Augmented Lagrangian Method: ALM

The iterative complexity of ALM is $O(1/k)$, and the proof as follows (He *et al*., 2015; Kang *et al*., 2015; Yurtsever *et al*., 2019):

Lemma 1.1 For the convex function minimization problem under the constraints of linear equations, it can be expressed as:

$$\min\left\{\mathrm{f}(\mathbf{x})\middle|\mathbf{Ax}=\mathbf{b},\mathbf{x}\in\mathbf{X}\right\}, \tag{B.1}$$

where $\mathrm{f}(\mathbf{x}):\Re^n\to\Re$ is a convex function, $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{X}$ are convex closed sets in $\Re^n$. According to equation B.1, $(\mathbf{x}^{k+1},\lambda^{k+1})$ could be generated. For each feasible solution $(\mathbf{x},\lambda)$, I could get:

$$\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1})-\mathrm{L}(\mathbf{x},\lambda)\geq\left\|\lambda^k-\lambda^{k+1}\right\|_{\tau_k^{-1}}^2+(\lambda-\lambda^k)^T\tau_k^{-1}(\lambda-\lambda^{k+1}). \tag{B.2}$$

Proof. The convexity of f can be used:

$$\begin{aligned}\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1})-\mathrm{L}(\mathbf{x},\lambda)&=\mathrm{f}(\mathbf{x}^{k+1})-\mathrm{f}(\mathbf{x})+\lambda^{\mathrm{T}}(\mathbf{Ax}-\mathbf{b})-(\lambda^{k+1})^{\mathrm{T}}(\mathbf{Ax}^{k+1}-\mathbf{b})\\&\geq(\mathbf{x}^{k+1}-\mathbf{x})^{\mathrm{T}}\nabla\mathrm{f}(\mathbf{x})+\lambda^{\mathrm{T}}(\mathbf{Ax}-\mathbf{b})-(\lambda^{k+1})^{\mathrm{T}}(\mathbf{Ax}^{k+1}-\mathbf{b}),\end{aligned} \tag{B.3}$$

where $(\mathbf{x},\lambda)$ is the feasible solution as know, I can get:

$$(\mathbf{x}^{k+1}-\mathbf{x})^{\mathrm{T}}\nabla\mathrm{f}(\mathbf{x})\geq(\mathbf{x}^{k+1}-\mathbf{x})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\lambda=\lambda^{\mathrm{T}}\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}). \tag{B.4}$$

Substitute inequality B.4 into inequality B.3, I obtain:

$$
\begin{aligned}
\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1})-\mathrm{L}(\mathbf{x},\lambda) &\geq \lambda^{\mathrm{T}}\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x})+\lambda^{\mathrm{T}}(\mathbf{Ax}-\mathbf{b})-(\lambda^{k+1})^{\mathrm{T}}(\mathbf{Ax}^{k+1}-\mathbf{b}) \\
&=(\lambda-\lambda^{k+1})^{\mathrm{T}}(\mathbf{Ax}^{k+1}-\mathbf{b}) \\
&=\left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2}+(\lambda-\lambda^{k})^{\mathrm{T}}\tau_k^{-1}(\lambda^{k}-\lambda^{k+1}).
\end{aligned}
\tag{B.5}
$$

The lemma 1.1 is proved.

Lemma 1.2. For a given λ^k**,** λ^{k+1} can be generated, and I can get:

$$
\begin{aligned}
\left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2} \leq \left\|\lambda^{k}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2}-\left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2} \\
-2(\mathrm{L}(\mathbf{x}^{*},\lambda^{*})-\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1})),\forall(\mathbf{x}^{*},\lambda^{*})\in\mathbf{X}^{*}\times\mathbf{\Lambda}^{*}.
\end{aligned}
\tag{B.6}
$$

Proof. $(\mathbf{x}^*,\lambda^*)$, which in inequality B.6 is dual feasible, when I set $(\mathbf{x},\lambda)=(\mathbf{x}^*,\lambda^*)$ in B.2, I can obtain:

$$
(\lambda^{k}-\lambda^{*})^{\mathrm{T}}\tau_k^{-1}(\lambda^{k}-\lambda^{k+1}) \geq \left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2}+(\mathrm{L}(\mathbf{x}^{*},\lambda^{*})-\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1})).
\tag{B.7}
$$

According to inequality B.7, I obtain:

$$
\begin{aligned}
\left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2} &= \left\|(\lambda^{k}-\lambda^{*})-(\lambda^{k}-\lambda^{k+1})\right\|_{\tau_k^{-1}}^{2} \\
&=\left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2}-2(\lambda^{k+1}-\lambda^{*})^{\mathrm{T}}\tau_k^{-1}(\lambda^{k}-\lambda^{k+1})+\left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2} \\
&\leq \left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2}-\left\|\lambda^{k}-\lambda^{k+1}\right\|-2(\mathrm{L}(\mathbf{x}^{*},\lambda^{*})-\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1}).
\end{aligned}
\tag{B.8}
$$

The lemma 1.2 is proved.

Theorem 1.3. Substitute $(\mathbf{x}^{k+1},\lambda^{k+1})$ into ALM, I can obtain:

$$
\begin{cases}
\mathrm{L}(\mathbf{x}^{k+1},\lambda^{k+1}) \geq \mathrm{L}(\mathbf{x}^{k},\lambda^{k})+\left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2}, \\
\left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2} \leq \left\|\lambda^{k}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2}-\left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2},
\end{cases}
\tag{B.9}
$$

if $\tau_k \equiv \tau$, I obtain:

$$\left\|\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}\right\|_{\tau}^{2} \leq \left\|\mathbf{A}\mathbf{x}^{k}-\mathbf{b}\right\|_{\tau}^{2} - \left\|\mathbf{A}(\mathbf{x}^{k}-\mathbf{x}^{k+1})\right\|_{\tau}^{2}. \tag{B.10}$$

Proof. From inequalities B.6 and B.9, because $\mathrm{L}(\mathbf{x}^{k+1}, \lambda^{k+1}) \leq \mathrm{L}(\mathbf{x}^{*}, \lambda^{*})$, I obtain:

$$\left\|\lambda^{k+1}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2} \leq \left\|\lambda^{k}-\lambda^{*}\right\|_{\tau_k^{-1}}^{2} - \left\|\lambda^{k}-\lambda^{k+1}\right\|_{\tau_k^{-1}}^{2}. \tag{B.11}$$

I know that $\tau_{k+1}^{-1} \preceq \tau_k^{-1}$, so B.9 can be obtained by B.11.

If $\mathbf{x}=\mathbf{x}^{k}$ in B.4, I obtain:

$$\begin{cases} (\mathbf{x}^{k}-\mathbf{x}^{k+1})^{\mathrm{T}}(\nabla \mathrm{f}(\mathbf{x}^{k+1})-\mathbf{A}^{\mathrm{T}}\lambda^{k+1}) \geq 0, \\ (\mathbf{x}^{k+1}-\mathbf{x}^{k})^{\mathrm{T}}(\nabla \mathrm{f}(\mathbf{x}^{k})-\mathbf{A}^{\mathrm{T}}\lambda^{k}) \geq 0. \end{cases} \tag{B.12}$$

By adding the above two inequalities, I can obtain:

$$(\mathbf{x}^{k}-\mathbf{x}^{k+1})\mathbf{A}^{\mathrm{T}}(\lambda^{k}-\lambda^{k+1}) \geq 0. \tag{B.13}$$

If $\tau_k \equiv \tau$, and $\lambda^{k+1}=\lambda^{k}-\tau_k(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b})$, the above inequality could be expressed as:

$$(\mathbf{x}^{k}-\mathbf{x}^{k+1})\mathbf{A}^{\mathrm{T}}\tau(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}) \geq 0, \tag{B.14}$$

and then, I obtain:

$$\left\|\mathbf{A}\mathbf{x}^{k}-\mathbf{b}\right\|_{\tau}^{2} = \left\|\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}\right\|_{\tau}^{2} + \left\|\mathbf{A}(\mathbf{x}^{k}-\mathbf{x}^{k+1})\right\|_{\tau}^{2} + 2(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b})^{T}\tau\mathbf{A}(\mathbf{x}^{k}-\mathbf{x}^{k+1}), \tag{B.15}$$

and:

$$\left\|\mathbf{A}\mathbf{x}^{k}-\mathbf{b}\right\|_{\tau}^{2} \geq \left\|\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}\right\|_{\tau}^{2} + \left\|\mathbf{A}(\mathbf{x}^{k}-\mathbf{x}^{k+1})\right\|_{\tau}^{2}. \tag{B.16}$$

The inequality B.10 is proved.

Theorem 1.4. For any $k \geq 1$, under ALM, I can obtain:

$$\mathrm{L}(\mathbf{x}^*,\lambda^*)-\mathrm{L}(\mathbf{x}^k,\lambda^k)\le\frac{\left\|\lambda^0-\lambda^*\right\|_{\tau_0^{-1}}^2}{2k},\forall(\mathbf{x}^*,\lambda^*)\in\mathbf{X}^*\times\mathbf{\Lambda}^*. \tag{B.17}$$

Proof. Because $\tau_{k+1}^{-1}\preceq\tau_k^{-1}$, I can obtain the following inequality from B.4:

$$\begin{aligned}&2(\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1})-\mathrm{L}(\mathbf{x}^*,\lambda^*))\\&\ge\left\|\lambda^{j+1}-\lambda^*\right\|_{\tau_{j+1}^{-1}}^2-\left\|\lambda^j-\lambda^*\right\|_{\tau_j^{-1}}^2+\left\|\lambda^j-\lambda^{j+1}\right\|_{\tau_j^{-1}}^2,\forall(\mathbf{x}^*,\lambda^*)\in\mathbf{X}^*\times\mathbf{\Lambda}^*.\end{aligned} \tag{B.18}$$

As I know $\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1})-\mathrm{L}(\mathbf{x}^*,\lambda^*)\le 0$, I obtain the following inequality from B.18:

$$\begin{aligned}&2(\sum_{j=0}^{k-1}\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1})-k\,\mathrm{L}(\mathbf{x}^*,\lambda^*))\\&\ge\left\|\lambda^k-\lambda^*\right\|_{\tau_{j+1}^{-1}}^2-\left\|\lambda^0-\lambda^*\right\|_{\tau_j^{-1}}^2+\sum_{j=1}^{k-1}\left\|\lambda^j-\lambda^{j+1}\right\|_{\tau_j^{-1}}^2.\end{aligned} \tag{B.19}$$

If $k=j$ and $(\mathbf{x},\lambda)=(\mathbf{x}^j,\lambda^j)$, according to inequality B.2, I obtain:

$$\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1})-\mathrm{L}(\mathbf{x}^j,\lambda^j)\ge\left\|\lambda^j-\lambda^{j+1}\right\|_{\tau_j^{-1}}^2, \tag{B.20}$$

and:

$$2\sum_{j=1}^{k-1}((j+1)\,\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1})-j\,\mathrm{L}(\mathbf{x}^j,\lambda^j)-\mathrm{L}(\mathbf{x}^{j+1},\lambda^{j+1}))\ge\sum_{j=1}^{k-1}2j\left\|\lambda^j-\lambda^{j+1}\right\|_{\tau_j^{-1}}^2. \tag{B.21}$$

Adding B.19 and B.21, then I obtain:

$$\begin{aligned}&2k(\mathrm{L}(\mathbf{x}^k,\lambda^k)-\mathrm{L}(\mathbf{x}^*,\lambda^*))\\&\ge\left\|\lambda^k-\lambda^*\right\|_{\tau_k^{-1}}^2-\left\|\lambda^0-\lambda^*\right\|_{\tau_0^{-1}}^2+\sum_{j=0}^{k-1}(2j+1)\left\|\lambda^j-\lambda^{j+1}\right\|_{\tau_j^{-1}}^2.\end{aligned} \tag{B.22}$$

Therefore, I obtain:

$$\mathrm{L}(\mathbf{x}^*, \lambda^*) - \mathrm{L}(\mathbf{x}^k, \lambda^k) \le \frac{\left\| \lambda^0 - \lambda^* \right\|^2_{\tau_0^{-1}}}{2k}. \tag{B.23}$$

According to the above derivation process, I can conclude that the iterative complexity of ALM is $O(1/k)$, and the proof ends.

Appendix C. Iteration-Complexity of the Inexact ALM: iALM

The iterative complexity of iALM is $O(1/k^2)$, and the proof as follows (Kang *et al.*, 2015; Sahin *et al.*, 2019; Nedelcu *et al.*, 2014):

Similar to ALM, in iALM, I have:

$$\begin{cases} \tilde{\mathbf{x}}^k = \arg\min\left\{ \mathrm{f}(\mathbf{x}) - (\lambda^k)^{\mathrm{T}}(\mathbf{A}x - \mathbf{b}) + \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_{\tau_k}^2 \,|\mathbf{x} \in \mathbf{X} \right\}, \\ \tilde{\lambda}^k = \lambda^k - \tau_k(\mathbf{A}\tilde{x}^k - \mathbf{b}). \end{cases} \tag{C.1}$$

Similar to inequality B.6 in ALM, in iALM, I have:

$$\begin{aligned} \left\|\tilde{\lambda}^k - \lambda^*\right\|_{\tau_k^{-1}}^2 \le \left\|\lambda^k - \lambda^*\right\|_{\tau_k^{-1}}^2 - \left\|\lambda^k - \tilde{\lambda}^k\right\|_{\tau_k^{-1}}^2 \\ -2(\mathrm{L}(\mathbf{x}^*, \lambda^*) - L(\tilde{\mathbf{x}}^k, \tilde{\lambda}^k)), \forall(\mathbf{x}^*, \lambda^*) \in \mathbf{X}^* \times \mathbf{\Lambda}^*. \end{aligned} \tag{C.2}$$

Lemma 2.1. When $t^1 = 1$, it satisfies:

$$t^k \ge (k+1)/2, \forall k \ge 1. \tag{C.3}$$

Proof. I can know the following equations by induction:

$$\begin{cases} \mathbf{v}^k = \mathrm{L}(\mathbf{x}^*, \lambda^*) - \mathrm{L}(\tilde{\mathbf{x}}^k, \tilde{\lambda}^k), \\ \mathbf{u}^k = t^k(2\tilde{\lambda}^k - \lambda^k - \tilde{\lambda}^{k-1}) + \tilde{\lambda}^{k-1} - \lambda^*. \end{cases} \tag{C.4}$$

Lemma 2.2. According to C.4, λ^k and $\tilde{\lambda}^k$ in iALM satisfied:

$$4(t^k)^2\mathbf{v}^k - 4(t^{k+1})^2\mathbf{v}^{k+1} \ge \left\|\mathbf{u}^{k+1}\right\|_{\tau_{k+1}^{-1}}^2 - \left\|\mathbf{u}^k\right\|_{\tau_{k+1}^{-1}}^2, \forall k \ge 1. \tag{C.5}$$

Proof. According to Lemma 2.1, if $(\mathbf{x}, \lambda) = (\tilde{\mathbf{x}}^k, \lambda^k)$ and $(\mathbf{x}, \lambda) = (\mathbf{x}^*, \lambda^*)$, I obtain:

$$\begin{cases} \mathrm{L}(\tilde{\mathbf{x}}^{k+1}, \tilde{\lambda}^{k+1}) - \mathrm{L}(\tilde{\mathbf{x}}^{k}, \tilde{\lambda}^{k}) \geq \left\| \lambda^{k+1} - \tilde{\lambda}^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} + (\tilde{\lambda}^{k} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}), \\ \mathrm{L}(\tilde{\mathbf{x}}^{k+1}, \tilde{\lambda}^{k+1}) - \mathrm{L}(\mathbf{x}^{*}, \lambda^{*}) \geq \left\| \lambda^{k+1} - \tilde{\lambda}^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} + (\lambda^{*} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}). \end{cases} \tag{C.6}$$

Substituting C.4 into C.6, I can get:

$$\begin{cases} \mathbf{v}^{k} - \mathbf{v}^{k+1} \geq \left\| \lambda^{k+1} - \tilde{\lambda}^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} + (\tilde{\lambda}^{k} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}), \\ -\mathbf{v}^{k+1} \geq \left\| \lambda^{k+1} - \tilde{\lambda}^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} + (\lambda^{*} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (\lambda^{k+1} - \tilde{\lambda}^{k+1}), \end{cases} \tag{C.7}$$

and:

$$\begin{aligned} &((t^{k+1} - 1)\mathbf{v}^{k} - t^{k+1}\mathbf{v}^{k+1}) \\ &\quad \geq t^{k+1} \left\| \tilde{\lambda}^{k+1} - \lambda^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} + (\tilde{\lambda}^{k+1} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (t^{k+1}\lambda^{k+1} - (t^{k+1} - 1)\tilde{\lambda}^{k} - \lambda^{*}). \end{aligned} \tag{C.8}$$

According to $t_k^2 = t_{k+1}^2 - t_{k+1}$, I obtain:

$$\begin{aligned} &(t_k^2 \mathbf{v}^{k} - t_{k+1}^2 \mathbf{v}^{k+1}) \\ &\geq \left\| t^{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1}) \right\|_{\tau_{k+1}^{-1}}^{2} + t^{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1})^{T} \tau_{k+1}^{-1} (t^{k+1}\lambda^{k+1} - (t^{k+1} - 1)\tilde{\lambda}^{k} - \lambda^{*}) \\ &= (t^{k+1}(\tilde{\lambda}^{k+1} - \lambda^{k+1}))^{T} \tau_{k+1}^{-1} (t^{k+1}\tilde{\lambda}^{k+1} - (t^{k+1} - 1)\tilde{\lambda}^{k} - \lambda^{*}), \end{aligned} \tag{C.9}$$

and:

$$\begin{aligned} (t_k^2 \mathbf{v}^{k} - t_{k+1}^2 \mathbf{v}^{k+1}) \geq &\frac{1}{4} \left\| t^{k+1}(2\tilde{\lambda}^{k+1} - \lambda^{k+1} - \tilde{\lambda}^{k}) + \tilde{\lambda}^{k} - \lambda^{*} \right\|_{\tau_{k+1}^{-1}}^{2} \\ &- \frac{1}{4} \left\| t^{k+1}(\lambda^{k+1} - \tilde{\lambda}^{k}) + \tilde{\lambda}^{k} - \lambda^{*} \right\|_{\tau_{k+1}^{-1}}^{2}. \end{aligned} \tag{C.10}$$

According to C.4, the last inequality can be expressed as:

$$4(t_k^2 \mathbf{v}^{k} - t_{k+1}^2 \mathbf{v}^{k+1}) \geq \left\| \mathbf{u}^{k+1} \right\|_{\tau_{k+1}^{-1}}^{2} - \left\| t^{k+1}(\lambda^{k+1} - \tilde{\lambda}^{k}) + \tilde{\lambda}^{k} - \lambda^{*} \right\|_{\tau_{k+1}^{-1}}^{2}. \tag{C.11}$$

If I set:

$$t^{k+1}(\lambda^{k+1}-\tilde{\lambda}^k)+\tilde{\lambda}^k-\lambda^* = t^k(2\tilde{\lambda}^k-\lambda^k-\tilde{\lambda}^{k-1})+\tilde{\lambda}^{k-1}-\lambda^*, \tag{C.12}$$

from C.12 I can get:

$$\lambda^{k+1}=\tilde{\lambda}^k+(\frac{t^k-1}{t^{k+1}})(\tilde{\lambda}^k-\tilde{\lambda}^{k-1})+(\frac{t^k}{t^{k+1}})(\tilde{\lambda}^k-\lambda^k). \tag{C.13}$$

The Lemma is proven.

Corollary 2.3. According to C.4, I have:

$$4t_k^2\mathbf{v}^k \le 4t_1^2\mathbf{v}^1+\left\|\mathbf{u}^1\right\|_{\tau_1^{-1}}^2, \forall k\ge 1. \tag{C.14}$$

Proof. As I used before, that $\tau_{k+1}^{-1}\preceq\tau_k^{-1}$, and from C.5, I can obtain:

$$4(t_k^2\mathbf{v}^k-t_{k+1}^2\mathbf{v}^{k+1})\ge\left\|\mathbf{u}^{k+1}\right\|_{\tau_{k+1}^{-1}}^2-\left\|\mathbf{u}^k\right\|_{\tau_k^{-1}}^2. \tag{C.15}$$

Theorem 2.4 In iALM, when $k\ge 1$, I obtain:

$$\mathrm{L}(\mathbf{x}^*,\lambda^*)-\mathrm{L}(\tilde{\mathbf{x}}^k,\tilde{\lambda}^k)\le\frac{\left\|\lambda^1-\lambda^*\right\|_{\tau_1^{-1}}^2}{(k+1)^2}, \forall(\mathbf{x}^*,\lambda^*)\in\mathbf{X}^*\times\mathbf{\Lambda}^*. \tag{C.16}$$

Proof. According to C.4, I obtain:

$$\mathrm{L}(\mathbf{x}^*,\lambda^*)-L(\tilde{\mathbf{x}}^k,\tilde{\lambda}^k)=\mathbf{v}^k\le\frac{4t_1^2\mathbf{v}^1+\left\|\mathbf{u}^1\right\|_{\tau_1^{-1}}^2}{4t_k^2}. \tag{C.17}$$

Combining the above inequality with inequality C.3, I can get:

$$\mathrm{L}(\mathbf{x}^*,\lambda^*)-\mathrm{L}(\tilde{\mathbf{x}}^k,\tilde{\lambda}^k)\le\frac{4t_1^2\mathbf{v}^1+\left\|\mathbf{u}^1\right\|_{\tau_1^{-1}}^2}{(k+1)^2}. \tag{C.18}$$

Because $t^1=1$, and according to C.2 and C.4, I can get:

$$4(\mathrm{L}(\mathbf{x}^*,\lambda^*)-\mathrm{L}(\tilde{\mathbf{x}}^1,\tilde{\lambda}^1))\le 2\left\|\lambda^1-\lambda^*\right\|_{\tau_1^{-1}}^2-2\left\|\tilde{\lambda}^1-\lambda^*\right\|_{\tau_1^{-1}}^2-2\left\|\tilde{\lambda}^1-\lambda^1\right\|_{\tau_1^{-1}}^2. \tag{C.19}$$

Similar to the derivation process from C.9 to C.11, I obtain:

$$4(\mathrm{L}(\mathbf{x}^*,\lambda^*)-\mathrm{L}(\tilde{\mathbf{x}}^1,\tilde{\lambda}^1))\le \left\|\lambda^1-\lambda^*\right\|_{\tau_1^{-1}}^2-\left\|2\tilde{\lambda}^1-\lambda^1-\lambda^*\right\|_{\tau_1^{-1}}^2. \tag{C.20}$$

From C.19 and C.20, I have:

$$4t_1^2\mathbf{v}^1+\left\|\mathbf{u}^1\right\|_{\tau_1^{-1}}^2\le\left\|\lambda^1-\lambda^*\right\|_{\tau_1^{-1}}^2. \tag{C.21}$$

Combining C.21 with C.18, I proved that the iterative complexity of iALM is $O(1/k^2)$, and the proof ends.

REFERENCES

Abubakar, A. & Berg, P.M. van den, 2002. The contrast source inversion method for location and shape reconstructions. *Inverse Problems*, **18**, 495. DOI:10.1088/0266-5611/18/2/313

Aghamiry, H S, Gholami, A. & Operto, S., 2021. Full waveform inversion by proximal Newton method using adaptive regularization. *Geophysical Journal International*, **224**, 169–180. DOI:10.1093/gji/ggaa434

Aghamiry, H S, Gholami, A., Operto, S. & Malcolm, A., 2022. ADMM-based full-waveform inversion for microseismic imaging. *Geophysical Journal International*, **228**, 259–274. DOI:10.1093/gji/ggab332

Aghamiry, H S, Mamfoumbi-Ozoumet, F.W., Gholami, A. & Operto, S., 2021. Efficient extended-search space full-waveform inversion with unknown source signatures. *Geophysical Journal International*, **227**, 257–274. DOI:10.1093/gji/ggab202

Aghamiry, Hossein S., Gholami, A. & Operto, S., 2019. Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers. *Geophysics*, **84**, R139–R162. DOI:10.1190/geo2018-0093.1

Aghamiry, Hossein S., Gholami, A. & Operto, S., 2020. Compound Regularization of Full-Waveform Inversion for Imaging Piecewise Media. *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 1192–1204. DOI:10.1109/TGRS.2019.2944464

Aghamiry, Hossein S., Gholami, A. & Operto, S., 2020. Multiparameter wavefield reconstruction inversion for wavespeed and attenuation with bound constraints and total variation regularization. *Geophysics*, **85**, R381–R396. DOI:10.1190/geo2019-0596.1

Aghamiry, Hossein S., Gholami, A. & Operto, S., 2021. On efficient frequency-domain full-waveform inversion with extended search space. *Geophysics*, **86**, R237–R252. DOI:10.1190/geo2020-0478.1

Aghazade, K., Gholami, A., Aghamiry, H.S. & Operto, S., 2022. Anderson-accelerated augmented Lagrangian for extended waveform inversion. *Geophysics*, **87**, R79–R91. DOI:10.1190/geo2021-0409.1

Andrychowicz, M., Denil, M., Gómez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., *et al.*, 2016. Learning to learn by gradient descent by gradient descent. *Advances in Neural Information Processing Systems*, **29**, Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/hash/fb87582825f9d28a8d42c5e5e5e8b23d-Abstract.html

Aravkin, A., Kumar, R., Mansour, H., Recht, B. & Herrmann, F.J., 2014. Fast Methods for Denoising Matrix Completion Formulations, with Applications to Robust Seismic Data Interpolation. *SIAM Journal on Scientific Computing*, **36**, S237–S266. DOI:10.1137/130919210

Argyriou, A., Foygel, R. & Srebro, N., 2012. Sparse Prediction with the k-Support Norm. *Advances in Neural Information Processing Systems*, **25**, Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/hash/99bcfcd754a98ce89cb86f73acc04645-Abstract.html

Bai, L. & Liang, J., 2020. Sparse Subspace Clustering with Entropy-Norm. Proceedings of the 37th *International Conference on Machine Learning*, pp. 561–568, Presented at the International Conference on Machine Learning, PMLR. https://proceedings.mlr.press/v119/bai20b.html

Belilovsky, E., Argyriou, A., Varoquaux, G. & Blaschko, M., 2015. Convex relaxations of penalties for sparse correlated variables with bounded total variation. *Machine Learning*, **100**, 533–553. DOI:10.1007/s10994-015-5511-2

Billette, F.J. & Brandsberg-Dahl, S., 2005. The 2004 BP Velocity Benchmark, p. cp, Presented at the *67th EAGE Conference & Exhibition*. DOI:10.3997/2214-4609-pdb.1.B035

Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J., 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers.

Foundations and Trends® in Machine learning, **3**, 1–122, Now Publishers, Inc. DOI:10.1561/2200000016

Boyd, S.P. & Vandenberghe, L., 2004. Convex Optimization, Cambridge University Press. https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

Brossier, R., Operto, S. & Virieux, J., 2009. Robust elastic frequency-domain full-waveform inversion using the L1 norm. *Geophysical Research Letters*, **36**. DOI:10.1029/2009GL039458

Brossier, Romain, Etienne, V., Operto, S., Virieux, J., Brossier, R., Etienne, V., Operto, S., *et al.*, 2010. Frequency-Domain Numerical Modelling of Visco-Acoustic Waves Based on Finite-Difference and Finite-Element Discontinuous Galerkin Methods. in *Acoustic Waves*, IntechOpen. DOI:10.5772/9714

Brossier, Romain, Operto, S. & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, **75**, R37–R46. DOI:10.1190/1.3379323

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion. *Geophysics*, **60**, 1457–1473. DOI:10.1190/1.1443880

Chang, S.G., Yu, B. & Vetterli, M., 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, **9**, 1532–1546. DOI:10.1109/83.862633

Chi, B., Dong, L. & Liu, Y., 2014. Full waveform inversion method using envelope objective function without low frequency data. *Journal of Applied Geophysics*, **109**, 36–46. DOI:10.1016/j.jappgeo.2014.07.010

Choi, Y. & Alkhalifah, T., 2012. Application of multi-source waveform inversion to marine streamer data using the global correlation norm. *Geophysical Prospecting*, **60**, 748–758. DOI:10.1111/j.1365-2478.2012.01079.x

Deng, H., Tao, J., Song, X. & Zhang, C., 2020. Estimation of the parameters of a weighted nuclear norm model and its application in image denoising. *Information Sciences*, **528**, 246–264. DOI:10.1016/j.ins.2020.04.028

Dong, W., Shi, G. & Li, X., 2013. Nonlocal Image Restoration With Bilateral Variance Estimation: A Low-Rank Approach. *IEEE Transactions on Image Processing*, **22**, 700–711. DOI:10.1109/TIP.2012.2221729

Donoho, D.L., 2006. For most large underdetermined systems of linear equations the minimal $\ell 1$-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, **59**, 797–829. DOI:10.1002/cpa.20132

Gambella, C. & Simonetto, A., 2020. Multiblock ADMM Heuristics for Mixed-Binary Optimization on Classical and Quantum Computers. *IEEE Transactions on Quantum Engineering*, **1**, 1–22. DOI:10.1109/TQE.2020.3033139

Gao, K. & Huang, L., 2019. Acoustic- and elastic-waveform inversion with total generalized p-variation regularization. *Geophysical Journal International*, **218**, 933–957. DOI:10.1093/gji/ggz203

Gholami, A., Aghamiry, H.S. & Operto, S., 2021. Extended-space full-waveform inversion in the time domain with the augmented Lagrangian method. *Geophysics*, **87**, R63–R77. DOI:10.1190/geo2021-0186.1

Gholami, A., Aghamiry, H.S. & Operto, S., 2022. Extended-space full-waveform inversion in the time domain with the augmented Lagrangian method. *Geophysics*, **87**, R63–R77. DOI:10.1190/geo2021-0186.1

Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X. & Zhang, L., 2017. Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision. *International journal of computer vision*, **121**, 183–208. DOI:10.1007/s11263-016-0930-5

Guo, Y., Huang, J.-P., Chao, C., Li, Z.-C., Li, Q.-Y. & Wei, W., 2020. Sparse constrained encoding multi-source full waveform inversion method based on K-SVD dictionary learning. *Applied Geophysics*, **17**, 111–123. DOI:10.1007/s11770-019-0797-7

Halko, N., Martinsson, P.G. & Tropp, J.A., 2011. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM review*, **53**, 217–288. DOI:10.1137/090771806

Hastie, T., Friedman, J. & Tibshirani, R., 2001. The Elements of Statistical Learning. Springer Series in Statistics, Springer. DOI:10.1007/978-0-387-21606-5

He, B., Hou, L., & Yuan, X., 2015. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *SIAM Journal on Optimization*, **25**, 2274-2312. DOI:10.1137/130922793

He, X., Hu, R. & Fang, Y.-P., 2022. Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. *Numerical Algorithms*, **90**, 1669–1690. DOI:10.1007/s11075-021-01246-y

Horst, R., Pardalos, P.M. & Thoai, N.V., 2000. Introduction to Global Optimization, Springer Science & Business Media. Introduction to Global Optimization | SpringerLink

Jo, C. H., Shin, C., & Suh, J. H., 1996. An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator. *Geophysics*, **61**, 529-537. DOI: 10.1190/1.1443979

Kang, M., Kang, M., & Jung, M., 2015. Inexact accelerated augmented Lagrangian methods. *Computational Optimization and Applications*, **62**, 373-404. DOI:10.1007/s10589-015-9742-8

Koh, K., Kim, S.-J. & Boyd, S., 2007. An Interior-Point Method for Large-Scale l1-Regularized Logistic Regression. *Journal of Machine learning research*, **8**, 1519–1555. An Interior-Point Method for Large-Scale l1-Regularized Logistic Regression | The Journal of Machine Learning Research (acm.org)

Komatitsch, D. & Tromp, J., 2003. A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation. *Geophysical Journal International*, **154**, 146–153. DOI:10.1046/j.1365-246X.2003.01950.x

Kwak, N., 2008. Principal Component Analysis Based on L1-Norm Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, 1672–1680. DOI:10.1109/TPAMI.2008.114

Lai, H., Pan, Y., Lu, C., Tang, Y. & Yan, S., 2014. Efficient k-Support Matrix Pursuit. *Computer Vision – ECCV 2014* Lecture Notes in Computer Science, pp. 617–631, eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T., Springer International Publishing. DOI:10.1007/978-3-319-10605-2_40

Leeuwen, T. van & Herrmann, F.J., 2015. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, **32**, 015007. DOI:10.1088/0266-5611/32/1/015007

Leeuwen, Tristan van & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**, 661–667. DOI:10.1093/gji/ggt258

Li, D. & Harris, J.M., 2018. Full waveform inversion with nonlocal similarity and model-derivative domain adaptive sparsity-promoting regularization. *Geophysical Journal International*, **215**, 1841–1864. DOI:10.1093/gji/ggy380

Li, J., Mikada, H., & Takekawa, J., 2024a. Improved full-waveform inversion for seismic data in the presence of noise based on the K-support norm. *Pure and Applied Geophysics*. DOI:10.1007/s00024-024-03449-5

Li, J., Mikada, H., & Takekawa, J., 2024b. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomised Singular Value Decomposition. *Journal of Geophysics and Engineering*, gxae015. DOI:10.1093/jge/gxae015

Li, J., Mikada, H., & Takekawa, J., 2024c. Improving Full-waveform Inversion Based on Sparse Regularisation for Geophysical Data. *Journal of Geophysics and Engineering*. DOI:10.1093/jge/gxae036

Li, J., Takekawa, J., Kurihara, K., Halim, K. D., Kazuhiko, M., & Yasuyuki, M., 2024d. Enhanced Geological Prediction for Tunnel Excavation Using Full Waveform Inversion Integrating Sobolev Space Regularization with a Quadratic Penalty Method. *arXiv preprint*, arXiv:2405.16812. DOI:10.48550/arXiv.2405.16812

Li, J., Mikada, H., & Takekawa, J., 2023a. Inexact Augmented Lagrangian Method-Based Full-waveform Inversion with Randomized Singular Value Decomposition. *arXiv preprint,* arXiv:2309.13871. DOI:10.48550/arXiv.2309.13871

Li, J., Mikada, H., & Takekawa, J., 2023b. Accelerated augmented Lagrangian full-waveform inversion based on truncated randomized singular value decomposition. In *Third International Meeting for Applied Geoscience & Energy*, pp. 645-649. Society of Exploration Geophysicists and American Association of Petroleum Geologists. DOI:10.1190/image2023-3910542.1

Li, J., Mikada, H., & Takekawa, J., 2023c. Improved full-waveform inversion based on sparse relaxed regularized regression processing of seismic data. In *Seventh International Conference on Engineering Geophysics*, *Al Ain, UAE, pp 262–6.* Society of Exploration Geophysicists. DOI:10.1190/iceg2023-059.1

Li, J., Mikada, H., & Takekawa, J., 2023d. Improving full-waveform inversion based on sparse regularisation for geophysical data. *arXiv preprint,* arXiv:2311.01688. DOI:10.48550/arXiv.2309.13871

Li, J., Mikada, H., & Takekawa, J., 2022. Improved full-waveform inversion for seismic data in the presence of noise based on the k-support norm. *arXiv preprint,* arXiv:2212.07074. DOI:10.48550/arXiv.2212.07074

Li, J., & Xu, S., 2022. The modified full waveform inversion with the K-Support norm for the noise existing data. In *83rd EAGE Annual Conference & Exhibition*, Vol. 2022, No. 1, pp. 1-5. European Association of Geoscientists & Engineers. DOI:10.3997/2214-4609.202210322

Li, Y. & Demanet, L., 2017. Extrapolated full-waveform inversion: An image-space approach, Presented at the 2017 SEG International Exposition and Annual Meeting, OnePetro. DOI:10.1190/segam2017-17725126.1

Liu, Q., Beller, S., Lei, W., Peter, D. & Tromp, J., 2022. Pre-conditioned BFGS-based uncertainty quantification in elastic full-waveform inversion. *Geophysical Journal International*, **228**, 796–815. DOI:10.1093/gji/ggab375

Liu, Q. & Peter, D., 2020. Square-Root Variable Metric-Based Nullspace Shuttle: A Characterization of the Nonuniqueness in Elastic Full-Waveform Inversion. *Journal of Geophysical Research: Solid Earth*, **125**, e2019JB018687. DOI:10.1029/2019JB018687

Loris, I., Nolet, G., Daubechies, I. & Dahlen, F.A., 2007. Tomographic inversion using $\ell 1$-norm regularization of wavelet coefficients. *Geophysical Journal International*, **170**, 359–370. DOI:10.1111/j.1365-246X.2007.03409.x

Lu, J., Wang, G. & Zhou, J., 2017. Simultaneous Feature and Dictionary Learning for Image Set Based Face Recognition. *IEEE Transactions on Image Processing*, **26**, 4042–4054, Presented at the IEEE Transactions on Image Processing. DOI:10.1109/TIP.2017.2713940

Martinsson, P.-G. & Voronin, S., 2016. A Randomized Blocked Algorithm for Efficiently Computing Rank-revealing Factorizations of Matrices. *SIAM Journal on Scientific Computing*, **38**, S485–S507. DOI:10.1137/15M1026080

McDonald, A.M., Pontil, M. & Stamos, D., 2014. Spectral k-Support Norm Regularization. *Advances in Neural Information Processing Systems*, Vol. 27, Curran Associates, Inc. Spectral k-Support Norm Regularization (neurips. cc)

Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full Waveform Inversion and the Truncated Newton Method. *SIAM Journal on Scientific Computing*, **35**, B401–B437. DOI:10.1137/120877854

Moghaddam, P.P. & Herrmann, F.J., 2010. Randomized full waveform inversion: a dimenstionality-reduction approach. in *SEG Technical Program Expanded Abstracts 2010*, pp. 977–982. DOI:10.1190/1.3513940

Nedelcu, V., Necoara, I., & Tran-Dinh, Q., 2014. Computational complexity of inexact gradient augmented Lagrangian methods: application to constrained MPC. *SIAM Journal on Control and Optimization*, **52**, 3109-3134. DOI: 10.1137/120897547

O'Brien, J. J., & Lerche, I., 1987. Modelling of the deformation and faulting of formations overlying an uprising salt dome. In *Dynamical Geology of Salt and Related Structures*, 419-455. DOI: 10.1016/B978-0-12-444170-5.50014-2

Operto, S. & Miniussi, A., 2018. On the role of density and attenuation in three-dimensional multiparameter viscoacoustic VTI frequency-domain FWI: an OBC case study from the North Sea. *Geophysical Journal International*, **213**, 2037–2059. DOI:10.1093/gji/ggy103

Pan, W., Innanen, K.A., Margrave, G.F., Fehler, M.C., Fang, X. & Li, J., 2016. Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion. *Geophysics*, **81**, R275–R291. DOI:10.1190/geo2015-0594.1

Plessix, R.-E., Baeten, G., Maag, J.W. de, Klaassen, M., Rujie, Z. & Zhifei, T., 2010. Application of Acoustic Full Waveform Inversion to a Low-frequency Large-offset Land Data Set, Presented at *the 2010 SEG Annual Meeting*. https://onepetro.org/SEGAM/proceedings-abstract/SEG10/All-SEG10/96119

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics*, **64**, 888–901. DOI:10.1190/1.1444597

Qu, S., Verschuur, E. & Chen, Y., 2019. Full-waveform inversion and joint migration inversion with an automatic directional total variation constraint. *Geophysics*, **84**, R175–R183. DOI:10.1190/geo2018-0085.1

Rabalais, N. N., Turner, R. E., & Wiseman Jr, W. J., 2001. Hypoxia in the Gulf of Mexico. *Journal of environmental quality*, **30**, 320-329. DOI:10.2134/jeq2001.302320x

Ramani, S. & Fessler, J.A., 2011. Parallel MR Image Reconstruction Using Augmented Lagrangian Methods. *IEEE Transactions on Medical Imaging*, **30**, 694–706. DOI:10.1109/TMI.2010.2093536

Ridder, S. de & Dellinger, J., 2011. Ambient seismic noise eikonal tomography for near-surface imaging at Valhall. *The Leading Edge*, **30**, 506–512. DOI:10.1190/1.3589108

Sahin, M.F., eftekhari, A., Alacaoglu, A., Latorre, F. & Cevher, V., 2019. An Inexact Augmented Lagrangian Framework for Nonconvex Optimization with Nonlinear Constraints. *Advances in Neural Information Processing Systems*, Vol. 32, Curran Associates, Inc. An Inexact Augmented Lagrangian Framework for Nonconvex Optimization with Nonlinear Constraints (neurips. cc)

Schmidt, M., 2005. Least Squares Optimization with L1-Norm Regularization. *CS542B Project Report*, 504, 195-221. DOI: https://people.duke.edu/~hpgavin/SystemID/References/Schmidt-LASSO-2005.pdf

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, **69**, 231–248. DOI:10.1190/1.1649391

Song, C. & Alkhalifah, T.A., 2020. Efficient Wavefield Inversion With Outer Iterations and Total Variation Constraint. *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 5836–5846. DOI:10.1109/TGRS.2020.2971697

Song, P., Trzasko, J.D., Manduca, A., Qiang, B., Kadirvel, R., Kallmes, D.F. & Chen, S., 2017. Accelerated Singular Value-Based Ultrasound Blood Flow Clutter Filtering With Randomized Singular Value Decomposition and Randomized Spatial Downsampling. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **64**, 706–716. DOI:10.1109/TUFFC.2017.2665342

Sun, B. & Alkhalifah, T., 2018. Mitigate Cycle Skipping in FWI: a Generalized Instanatenous Travel-Time Approach, Vol. 2018, pp. 1–5, In *80th EAGE Conference and Exhibition 2018*. DOI:10.3997/2214-4609.201801028

Sun, H., & Demanet, L., 2020. Extrapolated full-waveform inversion with deep learning. *Geophysics*, **85**, R275-R288. DOI: 10.1190/geo2019-0195.1

Turner, R. E., & Rabalais, N. N., 2019. The Gulf of Mexico. In World seas: An environmental evaluation, pp. 445-464, Academic Press. DOI:10.1016/B978-0-12-805068-2.00022-X

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. & Zhou, W., 2014. 6. An introduction to full waveform inversion. in *Encyclopedia of Exploration Geophysics*, pp. R1-1. DOI:10.1190/1.9781560803027.entry6

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**, WCC1–WCC26. DOI:10.1190/1.3238367

Wang, H., Chen, W., Zhang, Q., Liu, X., Zu, S. & Chen, Y., 2021. Fast Dictionary Learning for High-Dimensional Seismic Reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, **59**, 7098–7108. DOI:10.1109/TGRS.2020.3030740

Wang, R. & Herrmann, F., 2016. Frequency down extrapolation with TV norm minimization. in *SEG Technical Program Expanded Abstracts 2016* SEG Technical

Program Expanded Abstracts, pp. 1380–1384, Society of Exploration Geophysicists. DOI:10.1190/segam2016-13879674.1

Wang, Y., Yin, W. & Zeng, J., 2019. Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. *Journal of Scientific Computing*, **78**, 29–63. DOI:10.1007/s10915-018-0757-z

Warner, M. & Guasch, L., 2016. Adaptive waveform inversion: Theory. *Geophysics*, **81**, R429–R445. DOI:10.1190/geo2015-0387.1

Wong, R.T., 1984. A dual ascent approach for steiner tree problems on a directed graph. *Mathematical Programming*, **28**, 271–287. DOI:10.1007/BF02612335

Wu, R.-S., Luo, J. & Wu, B., 2014. Seismic envelope inversion and modulation signal model. *GEOPHYSICS*, **79**, WA13–WA24, Society of Exploration Geophysicists. DOI:10.1190/geo2013-0294.1

Xu, H., Caramanis, C. & Mannor, S., 2012. Sparse Algorithms Are Not Stable: A No-Free-Lunch Theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 187–193. DOI:10.1109/TPAMI.2011.177

Yurtsever, A., Fercoq, O., & Cevher, V., 2019. A conditional-gradient-based augmented lagrangian framework. In *International Conference on Machine Learning*, pp. 7272-7281. PMLR. https://proceedings.mlr.press/v97/yurtsever19a.html

Zheng, P., Askham, T., Brunton, S.L., Kutz, J.N. & Aravkin, A.Y., 2018. A Unified Framework for Sparse Relaxed Regularized Regression: SR3. *IEEE Access*, **7**, 1404–1423. DOI:10.1109/ACCESS.2018.2886528

Zhong, F. & Zhang, J., 2013. Linear Discriminant Analysis Based on L1-Norm Maximization. *IEEE Transactions on Image Processing*, **22**, 3018–3027. DOI:10.1109/TIP.2013.2253476

Zhou, W., Brossier, R., Operto, S. & Virieux, J., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, **202**, 1535–1554. DOI:10.1093/gji/ggv228

Zhou, X., Huo, S., Wang, H., Dong, S., Liang, Y. & Cao, J., 2023. Model parameter design for modeling surface topography in VTI elastic finite-difference near-surface simulations. *Geophysics*, **88**, C33–C52. DOI:10.1190/geo2022-0027.1

Zhu, W., Xu, K., Darve, E., Biondi, B. & Beroza, G.C., 2021. Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. *Geophysics*, **87**, R93–R109. DOI:10.1190/geo2020-0933.1

ACKNOWLEDGMENTS & DEDICATION

Peace[1,2,3,4].

1. Peace & love
2. Peace be with you
3. Peace out & good-bye
4. Peace lights filter cigarettes

謝辞

本研究を進め，博士論文を執筆するにあたり，大変多くの方々にご指導，ご協力を賜りました。お世話になった皆様への感謝を申し上げます。

京都大学大学院工学研究科社会基盤工学専攻名誉教授三ケ田均博士, 同社会基盤工学専攻准教授武川順一博士には, 日頃から本質的なご指導ご鞭撻を頂きました。

本研究を遂行するにあたり，常に有益な議論·適切なご助言を頂きました応用地球物理学分野研究室の大学院生ならびに卒業生，及び多くの諸先輩方，様々な面において支えてくださった研究室秘書亀井志織氏に対し感謝の意を表します。

末筆ながら，改めてお世話になったすべての皆様への感謝の意を述べ，御礼の言葉とさせていただきます。

なお、本研究の一部は以下の研究助成を受けて実施されたものです。ここに記し、感謝の意を表します。

・日本政府（文部科学省）留学生（国費外国人留学生）奨学金

致謝

連編累牍，筆落難継。遙離塵囂，涙灑紙際。

娟娟師恩，銘記三心。願歲併謝，与友長兮。

哀哀父母，生我劬力。欲報之徳，昊天罔極。

愛隔山海，山海可移。恵而好我，不棄已心。

凡諸過往，皆為序起。前路漫漫，其修遠兮。

天地立心，生生不息。継之不息，希冀可期。

李嘉杭 謹上

二零二四年三月

甲辰龍年丁卯月

於　京都・桂