Gauß-Newton Full-Waveform Inversion for Acoustic Media

Master's thesis of

Kai He

at the Geophysical Institute (GPI)
KIT-Department of Physics
Karlsruhe Institute of Technology (KIT)

Date of submission:
03.02.2025

Supervisor: Prof. Dr. Thomas Bohlen
Co-supervisor: apl. Prof. Dr. Joachim Ritter

Erklärung / Statutory declaration

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

I declare truthfully that I have written this thesis by myself, that I have fully and accurately specified all aids used, that I have correctly cited everything that was taken, either unchanged or with modification, from the work of others, and that I have complied with the current version of the KIT statutes for safeguarding good scientific practice.

Karlsruhe, 03.02.2025

Signature: Kai He

Als Prüfungsexemplar genehmigt von / As examination copy approved by:

Karlsruhe, 03.02.2025

Signature: Prof. Dr. Thomas Bohlen

Abstract

This thesis investigates the performance of the Gauss-Newton (GN) method in Full-waveform inversion (FWI) in acoustic media in the time domain. FWI aims to estimate subsurface properties by iteratively minimizing the misfit between observed and synthetic seismic data. While the Steepest-descent (SD) method is commonly used in FWI, this study demonstrates that the GN method offers superior performance, achieving higher resolution and faster convergence. However, the GN method requires significant computational resources due to the calculation of the Jacobian matrix. To address these challenges, we calculate the partial derivative field for the Jacobian matrix using virtual sources and the reciprocity principle. These wavefields are explicitly obtained by convolving the forward propagated wavefields from each source with the reciprocal wavefields from each receiver. Then the Jacobian matrix is used to calculate the approximate Hessian matrix and the gradient used in the GN method. Numerical experiments reveal that the GN method outperforms the SD method across various geometries with different numbers of receivers and sources, delivering significantly higher resolution and faster convergence rates.

Contents

Chapter 1

List of Acronyms

FWI Full-waveform inversion
SD Steepest-descent
GN Gauss-Newton
FN Full Newton
PML Perfectly matched layer

Chapter 2

Introduction

2.1 Full-waveform inversion

Full-waveform inversion (FWI) is an efficient geophysical tool for seismic inversion, designed to estimate subsurface properties with high resolution. Its main challenges are the inherent non-linearities of most inversion problems. This nonlinearity arises because model responses are nonlinear functions of the model parameters, it requires an iterative process to minimize the misfit between the model response and observed data. Addressing these challenges requires advanced algorithms, careful optimization, and robust techniques to mitigate issues such as cycle skipping. In a seismic non-linear inversion, geophysicists commonly minimize the least squares residual using methods such as the Steepest-descent (SD), Gauss-Newton (GN), and Full Newton (FN) methods.

In a seminal development, Lailly (1983) and Tarantola (1984) introduced the SD method in FWI for acoustic, isotropic, elastic, and viscoelastic problems. This method avoids the need for explicit gradient computation by employing the adjoint-state technique and Born approximation. It significantly reduces computational and memory costs, making seismic waveform inversion more feasible. Subsequent developments extended this method, with numerical examples provided by Kolb et al. (1986), Gauthier et al. (1986), Mora (1987), and Pica et al. (1990), all of which were applied in the time domain. Later, Pratt et al. (1998) introduced its application in the frequency domain for acoustic and elastic wave equations. Despite its efficiency in avoiding the explicit computation of gradient direction, the SD method has notable limitations, including slow convergence rates and challenges in resolving the background velocity. These issues become particularly significant given the typically large number of model parameters in seismic inversion problems.

As forward simulation techniques advanced, researchers explored the FN and GN methods to address the limitations of the SD method. The FN method has been explored for its potential to enhance inversion accuracy in small-scale 1D problems (Santosa and Symes (1986)), it requires explicit computation and inversion of the full Hessian matrix. This makes it computationally prohibitive for large-scale seismic problems. In contrast, the GN method approximates the Hessian matrix, which strikes a balance between accuracy and computational feasibility. Initially, the GN method introduced by Tarantola (1984) was impractical at the time due to limited computational resources. However, recent advances in computational power have developed its effective application. The GN method relies on the explicit calculation of the Jacobian matrix, which represents the partial derivatives of the wavefield with respect to model parameters. This matrix captures the sensitivity of observed seismic data to small perturbations in subsurface properties, such as velocity and impedance, and plays a central role in iteratively updating the model parameters to

minimize the misfit in FWI. Explicitly calculating the Jacobian requires for each iteration as many forward simulations as there are grid points in the model. To reduce these computational demands, Rodi (1976) demonstrated that partial derivative wavefields can be obtained by solving forward problems with virtual sources placed at the perturbation locations. Pratt et al. (1998) applied this principle using virtual sources to compute partial derivative wavefields efficiently, achieving high-resolution imaging of subsurface structures. Further, Shin et al. (2001) proposed utilizing the source-receiver reciprocity principle to reduce memory requirements. This principle allows the interchange of virtual source and receiver positions (Aki and Richards (2002)), effectively shifting the computational dependency from the number of model parameters to the number of shots and receivers. Building on these innovations, the GN method has been applied in the frequency domain, using the simultaneous inversion of selected frequencies to reduce computational costs. However, the success of this approach depends heavily on the careful selection of frequencies. In contrast, Sheen (2004 and 2006) extended these ideas to the time domain. By using the full waveform information across multiple frequencies simultaneously, the time domain approach makes the GN method more robust than the discrete frequency domain approach. In this thesis, we employ the GN method in FWI, addressing the limitations of SD and avoiding the computational challenges of FN.

2.2 Main objectives

The primary objectives of this thesis are to evaluate and compare the performance of FWI using the SD method and the GN method for acoustic media in the time domain. To achieve these objectives, synthetic experiments are conducted to demonstrate the application and effectiveness of both methods. The inversion process is based on a 2D finite-difference solution of the acoustic wave equation with constant density. In the SD method, the gradient direction is implicitly derived using the adjoint-state method, avoiding the need for explicit computation. In contrast, the GN method explicitly calculates the Jacobian matrix through the use of virtual sources and the reciprocity principle. The Jacobian matrix is then used to compute the approximate Hessian matrix and gradient direction.

2.3 Thesis structure

This thesis is organized into four main chapters:

Chapter 3 provides the theory used in this thesis. It begins with an introduction to the wave equation for acoustic media and describes the finite-difference method employed for numerical simulations. This is followed by an introduction of the general inverse problem and the methods utilized in FWI, including the SD, FN, and GN methods.

Chapter 4 focuses on the implementation and computation of the GN method. It details how these computations utilize the Jacobian matrix for calculations of the approximate Hessian matrix and the gradient direction. This chapter aims to understand the computational challenges involved in the GN method thoroughly.

Chapter 5 presents synthetic tests conducted to analyze and compare the performance of the SD method and the GN method. These tests evaluate factors such as resolution and convergence, offering valuable insight into the efficiency and effectiveness of each method.

Finally, **Chapter 6** concludes the thesis by summarizing the results of synthetic experiments and discussing their implications.

Chapter 3

Methodology

This chapter provides an overview of the basic theory used in this thesis. It begins with an introduction to the wave equation for acoustic media, followed by a description of its numerical implementation using the time-domain finite-difference method. Next, the general framework of the inverse problem in FWI is discussed. Finally, different approaches to solving the inversion problem are presented, including the SD, FN, and GN methods.

3.1 Acoustic wave equation

The general wave equation is derived by combining two fundamental equations in an arbitrary medium: (i) the stress-strain relation, and (ii) the equation of motion. The acoustic wave equation describes the wave propagation in acoustic media, which represents the propagation of pressure disturbances in gases or liquids. This section focuses on the derivation of the acoustic wave equation. Note that all equations use the Einstein notation.

3.1.1 The stress-strain relation

In the generalized relation between the stress p and the strain ϵ, Hooke's law connects both physical quantities in elastic media via the elasticity tensor c_{ijkl} (Lay and Wallace, 1995):

$$p_{ij} = c_{ijkl}\epsilon_{kl} \tag{3.1}$$

where the strain ϵ_{kl} is a second-order symmetric tensor with the displacement of a particle u_k and u_l representing the l-th and k-th components, which is given by

$$\epsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right) \tag{3.2}$$

For a special case of isotropy, the stiffness tensor, c_{ijkl}, can be simplified to:

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl}) \tag{3.3}$$

where δ_{ij} the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases} \tag{3.4}$$

and the two Lamé parameters λ and μ are used, both of them are time-invariant. These two Lamé parameters are sufficient to describe the material properties of an isotropic and elastic medium. The stress-strain relations equation (3.1) can be simplified to:

$$p_{ij} = \lambda\theta\delta_{ij} + 2\mu\epsilon_{ij} \tag{3.5}$$

where cubic dilatation θ, i.e., the divergence of $\vec{u}$, is defined as $\theta = \epsilon_{11} + \epsilon_{22} + \epsilon_{33}$.

3.1.2 Equation of motion

Next, we consider the equation of motion, which is derived from Newton's second law. This law establishes a relationship between the forces acting within a medium and the resulting measurable displacements. The equation of motion is expressed as:

$$\rho \frac{\partial v_i}{\partial t} = \frac{\partial p_{ij}}{\partial x_j} + f_i \tag{3.6}$$

where ρ represents the mass density, v_i the particle velocity, t the time, x the vector in space, and f_i the external body forces.

3.1.3 Acoustic wave equation

Acoustic waves arise from localized pressure disturbances within a system. Similar to elastic waves, they are characterized by the velocity of pressure particles and their displacement from the equilibrium pressure. The velocity of pressure particles determines the kinetic energy and the displacement represents the potential energy. The key parameters in acoustic wave equation are pressure p, velocity v, and density ρ (Sahin et al., 2023).

Because of the acoustic wave propagated in a gas or liquid, the shear modulus in the stress-strain relation is zero, i.e., $\mu = 0$, and only hydrostatic pressure is left. The stress tensor p_{ij} simplifies to

$$p = \begin{pmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{pmatrix} \tag{3.7}$$

This means that seismic waves will propagate through acoustic media without a specified orientation, i.e., no anisotropy. Thus the stress-strain relation for the acoustic wave equation reduces to:

$$-p = \lambda div(\vec{u}) \tag{3.8}$$

Consequently, the external body forces in the equation of motion represented by f_i in equation (3.6) can be obsoleted, i.e., $f_i = 0$. The equation of motion in the acoustic media is given by:

$$\frac{\partial v_i}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} \tag{3.9}$$

By the stress-strain relation taking the derivative with respect to time, the first-order acoustic wave equations can be derived by equation (3.8) and (3.9). This is the so-called velocity-stress formulation of the acoustic wave equation:

$$-\frac{\partial p}{\partial t} = \lambda div(\vec{v}) \tag{3.10a}$$

$$\frac{\partial v_i}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} \tag{3.10b}$$

For the second-order acoustic wave equation, the stress-strain relation is once more derived with respect to time. Then, the equation of motion is inserted on the right side:

$$-\frac{\partial^2 p}{\partial t^2} = \lambda \left(\frac{\partial}{\partial x_1}\frac{\partial v_1}{\partial t} + \frac{\partial}{\partial x_2}\frac{\partial v_2}{\partial t} + \frac{\partial}{\partial x_3}\frac{\partial v_3}{\partial t} \right) \tag{3.11}$$

$$= \lambda \left(\frac{\partial}{\partial x_1}(-\frac{1}{\rho}\frac{\partial p}{\partial x_1}) + \frac{\partial}{\partial x_2}(-\frac{1}{\rho}\frac{\partial p}{\partial x_2}) + \frac{\partial}{\partial x_3}(-\frac{1}{\rho}\frac{\partial p}{\partial x_3}) \right) \tag{3.12}$$

$$= \lambda \sum_{3}^{i=1} \frac{\partial}{\partial x_i}\left(-\frac{1}{\rho}\frac{\partial p}{\partial x_i} \right) \tag{3.13}$$

This can be simplified by assuming the density as constant and defining the compressional velocity $c = \sqrt{\frac{\lambda}{\rho}}$, the second-order acoustic wave equation reads:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{3.14}$$

The equation (3.14) can also be written as:

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta \right) p = 0 \tag{3.15}$$

3.2 Forward modeling

The implementation of forward and inverse problems constitutes the foundation of the FWI technique. In forward modeling, the objective is to simulate wave propagation through a medium by approximating the wavefield at discrete points in time and space, thereby capturing the full seismic wavefield. Among the various numerical methods available, the finite-difference method is the most widely used and computationally efficient approach for discretizing the wave equation. This method utilizes finite-difference operators to approximate the spatial and temporal derivatives of the wave equation (Operto et al., 2007).

3.2.1 Discretization

In the finite-difference method, the wave propagation is approximated by finite-difference operators. This process involves discretizing the wave equation in both space and time on a cartesian coordinate system, where the computational discretization is divided into a finite-difference grid (Pontius, 2016; Igel, 2017).

A common and straightforward approach for discretization is to use a rectangular grid with uniform spacing, denoted as Δh. For a 2D finite-different method, the continuous spatial coordinates x and z are replaced by their discrete counterparts, expressed as:

$$x_i = i\Delta h \quad \text{and} \quad z_j = j\Delta h \tag{3.16}$$

where i and j refer to the grid points in the x and z directions, respectively. The time step Δt refers to the time sampling interval in the approximation:

$$t_n = n\Delta t \tag{3.17}$$

where n refers to a specific time step.

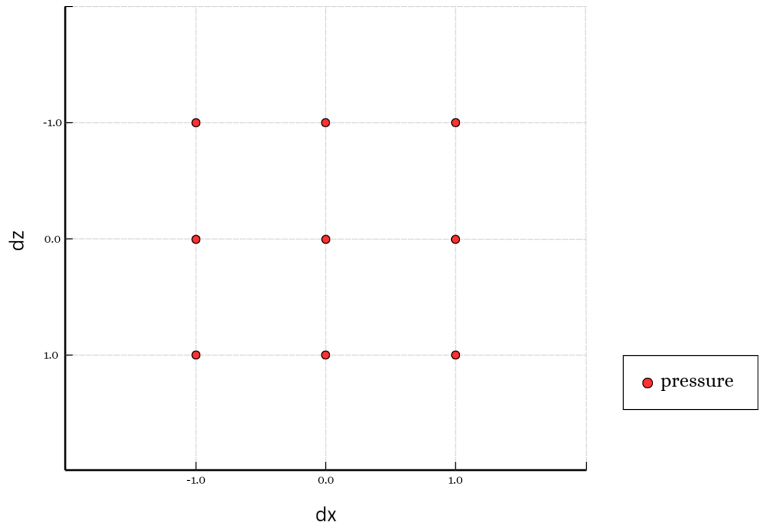


Figure 3.1: *A figure showing finite-difference grid points for pressure parameters in space domain. The space update scheme is represented by the x and z axes, which are the horizontal and vertical directions, respectively.*

Figure 3.1 shows the locations of wavefield parameters: the pressure components p_{ij} are distributed at full grid positions in the s and z directions in space. The acoustic wave equation can be both discretized on such a grid in space and time. Consequently, equation (3.14) is discretized as:

$$\partial_t^2 p(x,z,t) = c^2 \left[\partial_x^2 p(x,z,t) + \partial_z^2 p(x,z,t)\right] \tag{3.18}$$

The second-order derivatives in space and time of the pressure wavefield can be approximated by using a central difference scheme in second-order accuracy:

$$\partial_t^2 p(x,z,t) \approx \frac{p_{i,j}^{n-1} - 2p_{i,j}^{n} + p_{i,j}^{n+1}}{\Delta t^2} \tag{3.19a}$$

$$\partial_x^2 p(x,z,t) \approx \frac{p_{i-1,j}^{n} - 2p_{i,j}^{n} + p_{i+1,j}^{n}}{\Delta x^2} \tag{3.19b}$$

$$\partial_z^2 p(x,z,t) \approx \frac{p_{i,j-1}^{n} - 2p_{i,j}^{n} + p_{i,j+1}^{n}}{\Delta z^2} \tag{3.19c}$$

where p^n represents the pressure wavefield at the current time step, p^{n-1} and p^{n+1} are the wavefields at the one step before and one step after n. $p_{i,j}$ is the wavefield at the grid point (i,j), and Δx and Δz are the grid spacings in the horizontal and vertical directions. In case of uniform spacing, it is, $\Delta x = \Delta z = \Delta h$. To ensure high accuracy, the spatial derivatives with respect to x or z can be discretized using a finite difference scheme of fourth-order accuracy:

$$\partial_x^2 p(x,z,t) \approx \frac{-p_{i-2,j}^{n} + 16p_{i-1,j}^{n} - 30p_{i,j}^{n} + 16p_{i+1,j}^{n} - p_{i+2,j}^{n}}{12\Delta x^2} \tag{3.20a}$$

$$\partial_z^2 p(x,z,t) \approx \frac{-p_{i-2,j}^{n} + 16p_{i,j-1}^{n} - 30p_{i,j}^{n} + 16p_{i+1,j}^{n} - p_{i+2,j}^{n}}{12\Delta z^2} \tag{3.20b}$$

In this thesis, the acoustic wave equation (3.18) is solved by substituting the discrete approximations using a fourth-order spatial discretization and a second-order temporal

discretization scheme. Within this scheme, the pressure at time step $n+1$ at grid point i, j is:

$$p_{i,j}^{n+1} = 2p_{i,j}^{n} - p_{i,j}^{n-1} + c^2\Delta t^2\left(\frac{-p_{i-2,j}^{n} + 16p_{i-1,j}^{n} - 30p_{i,j}^{n} + 16p_{i+1,j}^{n} - p_{i+2,j}^{n}}{12\Delta x^2} + \frac{-p_{i-2,j}^{n} + 16p_{i,j-1}^{n} - 30p_{i,j}^{n} + 16p_{i+1,j}^{n} - p_{i+2,j}^{n}}{12\Delta z^2}\right) \tag{3.21}$$

3.2.2 Boundary conditions

Boundary conditions are crucial in finite-difference simulations of wave propagation to ensure the stability and accuracy of the solution. These conditions define the behavior of waves at the edges of the computational domain, ensuring the suppression of artificial reflections and preventing numerical instabilities.

3.2.2.1 Initial conditions

The initial conditions define the values of the wavefield and its time derivatives at the initial time, $t = 0$. At this moment, the medium is assumed to be at rest, with no external forces applied. Both the pressure wavefield and its time derivatives are zero across all spatial locations (Virieux, 1986). Thus, the initial conditions for the pressure wavefield at every spatial location are given as:

$$p(x, z, t = 0) = 0, \quad \dot{p}_{ij}(x, t = 0) = 0, \quad \ddot{p}_{ij}(x, t = 0) = 0 \tag{3.22}$$

3.2.2.2 Perfectly matched layer boundary condition

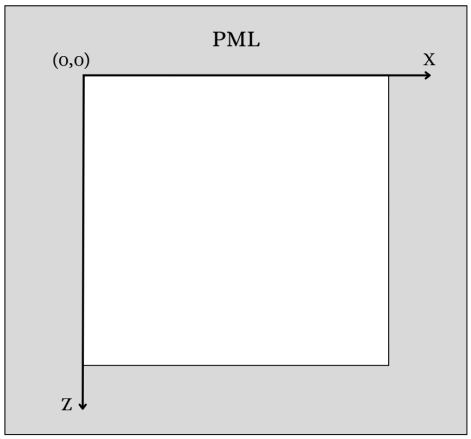


Figure 3.2: *Schematic diagram of the Perfectly Matched Layer (PML) region. The total computational domain consists of a rectangular model, with an additional PML region surrounding the model to prevent artificial reflections at the boundaries.*

Numerical simulations of wave propagation based on partial differential equations require truncating the computational domain to ensure efficient forward modeling. However, as seismic waves propagate, they eventually reach the model boundaries, where truncation can cause artificial reflections that overlap the solution and dominate the numerical error. To address this issue, it is essential to implement a boundary layer that effectively suppresses such artificial reflections. A highly efficient technique for this purpose was introduced by Berenger (1994), who proposed modifying the wave equation within a perfectly matched

layer (PML). This approach ensures that wave solutions decay exponentially without generating reflections at the interface between the regular medium and the PML. In this thesis, the PML implementation is based on the method described by Ma et al. (2014). The PML is applied to the vertical boundaries of the computational domain, as well as to the top and bottom boundaries of the model shown in Figure 3.2.

3.3 Full-Waveform Inversion

The inversion process of FWI is to estimate subsurface properties by an iteratively updating the model parameters. FWI makes use of the complete seismic waveforms, incorporating all available information, such as wave amplitude and phase, to derive an optimal model. This model is refined by minimizing the misfit between the synthetic data and the observed data. As such, the inversion problem aims to minimize the misfit, where the synthetic data is obtained through forward simulation. In this section, we introduce the SD, FN, and GN methods for solving the inversion problem.

3.3.1 Inversion problem

In general, synthetic model responses d_{syn} including the amplitude and phase information represented by model parameters m can be modeled and measured at receivers. The relationship can be expressed with the non-linear forward operator F:

$$d_{syn} = F(m) \tag{3.23}$$

The residual error Δd is defined as the difference at the receiver positions between the model responses d_{syn} and the observed data d_{obs} for each source-receiver pair of a seismic survey:

$$\Delta d = d_{syn} - d_{obs} \tag{3.24}$$
$$= F(m) - d_{obs} \tag{3.25}$$

where d_{obs} is the observed data. The data residual vector depends on the model parameter, $\Delta d = (\Delta d_1, \Delta d_2, \cdots, \Delta d_{NM})^T$, with number of the model parameters NM.

To measure the fit of the synthetic data to the observed data for non-linear inversion, we use the least-squares L_2-norm of the data residual which is easy to manipulate mathematically:

$$E(m) = \frac{1}{2} \|\Delta d\|^2 = \frac{1}{2} \Delta d^T \Delta d \tag{3.26}$$

where $E(m)$ is the data misfit function or objective function. The factor $\frac{1}{2}$ allows subsequent simplifications, and the superscript T represents the matrix transpose. In the time domain, the misfit function represents the summation of data residuals over the number of source-receiver pairs and the number of time samples in the seismograms. The inversion problem thus involves the iterative minimization of the misfit function $E(m)$ to obtain a subsurface model that best explains the observed data.

The minimum of the misfit function $E(m)$ is sought in the vicinity of the initial model, m_0. FWI is essentially a local optimization problem, where the choice of the initial model is crucial, as it serves as the starting point for the iterative model update process. Based on the Born approximation, we assume that the updated model m can be expressed as the sum of the initial model m_0 and a perturbation model Δm:

$$m = m_0 + \Delta m \tag{3.27}$$

We consider adding this model update into the development of the misfit function by a second-order Taylor expansion:

$$E(m_0+\Delta m) = E(m_0)+\Delta m\left(\frac{\partial E(m_0)}{\partial m}\right)+\frac{1}{2}\Delta m\left(\frac{\partial E(m_0)^2}{\partial m^2}\right)\Delta m^T+\mathcal{O}\left(\|\Delta m\|^3\right) \quad (3.28)$$

Note that the rest of the term including the error term $\mathcal{O}(\|\Delta m\|^3)$ in equation (3.28) can be neglected when the misfit function is quadratic. To find a minimum of this misfit function in the vicinity of the initial model, the gradient, which is the first derivative of the misfit function, is required to vanish:

$$\frac{\partial E(m)}{\partial m} = \frac{\partial E(m_0)}{\partial m}+\Delta m\left(\frac{\partial^2 E(m_0)}{\partial m^2}\right)\stackrel{!}{=}0 \quad (3.29)$$

then the perturbation model vector Δm is given by:

$$\Delta m = -\left(\frac{\partial^2 E(m_0)}{\partial m^2}\right)^{-1}\frac{\partial E(m_0)}{\partial m} = -H^{-1}\nabla_m E(m_0) \quad (3.30)$$

where the Hessian matrix H is the second-order derivative of the misfit function, which is the curvature of the misfit function m_0. The Hessian matrix reads:

$$H = \frac{\partial^2 E(m_0)}{\partial m^2} \quad (3.31)$$

Thus, the perturbation model is searched in the opposite direction of the steepest descent or gradient of the misfit function at the initial model m_0. The minimum of the misfit function in a single iteration can be determined using equations (3.27) and (3.30). Consequently, the inversion can converge towards the minimum after K iterations:

$$m_{K+1} = m_K+\Delta m_K \quad (3.32)$$
$$= m_K - H_K^{-1}\nabla_m E(m_K) \quad (3.33)$$

The gradient is obtained by taking partial derivatives of the misfit function in equation (3.26) with respect to model parameters m:

$$\nabla_m E = \frac{1}{2}\nabla_m\left[(d_{syn}-d_{obs})^T(d_{syn}-d_{obs})\right] \quad (3.34)$$
$$= \frac{1}{2}\nabla_m\left[{d_{syn}}^T d_{syn}-2{d_{syn}}^T d_{obs}-{d_{obs}}^T d_{obs}\right] \quad (3.35)$$
$$= \frac{1}{2}\left[\nabla_m({d_{syn}}^T d_{syn})-2\nabla_m({d_{syn}}^T d_{obs})\right] \quad (3.36)$$
$$= \frac{1}{2}\left[2\frac{\partial {d_{syn}}^T}{\partial m}d_{syn}-2\frac{\partial {d_{syn}}^T}{\partial m}d_{obs}\right] \quad (3.37)$$
$$= (\frac{\partial d_{syn}}{\partial m})^T(d_{syn}-d_{obs}) \quad (3.38)$$

where the partial derivative wavefield, $\frac{\partial d_{syn}}{\partial m}$, is called the Frechét derivative or Jacobian matrix J. Therefore, the gradient part can be simplified as:

$$\nabla_m E = J^T\Delta d \quad (3.39)$$

where J^T is a transpose of the Jacobian matrix. The Jacobian depends on the number of sources, receivers, model parameters, and time steps. Since the number of model parameters is typically large in FWI, directly computing the Jacobian can be computationally expensive.

This method above is known as the FN method, as it incorporates the Hessian matrix (Nocedal and Wright, 2006). However, the SD method is not necessary to explicitly compute the Jacobian and avoids also the Hessian matrix calculation. The SD method only requires several forward simulations to determine the gradient direction, which is efficiently obtained using the adjoint-state method. The details will be discussed in Section 3.3.2.

3.3.2 Steepest descent method

As outlined in the previous section, minimizing the misfit function requires the calculation of both the gradient and the Hessian matrix. To address these challenges, Tarantola (1984) and Mora (1987) introduced the SD method. This approach allows for model updates in FWI without the need for extensive matrix computations.

3.3.2.1 Model updating for Steepest descent method

The SD method minimizes the misfit function $E(m)$ by updating the model parameters in the opposite direction of the gradient of the misfit function iteratively:

$$m_{n+1} = m_n + \Delta m_n \tag{3.40}$$
$$= m_n - \alpha_n \nabla_m E(m_n) \tag{3.41}$$

where the subscripts represent iteration numbers. The step length α is in scalar value instead of the Hessian matrix, the SD method therefore neglects Hessian matrix computations. The gradient calculation in the SD method is using the adjoint-state method discussed as following to avoid additional calculation of the Jacobian matrix.

3.3.2.2 Adjoint-State method

In this thesis, the unknown model parameter vector m for the acoustic wave equation can be represented by the square of the slowness v_i at each grid point i:

$$m = \left(\frac{1}{v_1^2}, \frac{1}{v_2^2}, \cdots, \frac{1}{v_{NM}^2} \right)^T \tag{3.42}$$

Thus, the second-order acoustic wave equation can be rewritten as:

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta \right) p = f \quad \Rightarrow \quad \left(m \frac{\partial^2}{\partial t^2} - \Delta \right) p = f \tag{3.43}$$

where f is the source function. Due to the Born approximation, we assume that the model and data can be rewritten by the background information plus the perturbation, again the equation (3.26) and the data can be split up to:

$$m = m_0 + \Delta m_1 \tag{3.44a}$$

$$u = u_0 + u_{sc} \tag{3.44b}$$

where u represents pressure p in acoustic media. The background wavefields u_0 describe the wavefield propagating with the model m_0, and the scattered wavefields u_{sc} propagating with the perturabetion Δm_1. This leads to two wave equations:

$$m_0 \frac{\partial^2 u_0}{\partial t^2} - \Delta u_0 = f \tag{3.45a}$$

$$m\frac{\partial^2 u}{\partial t^2} - \Delta u = f \tag{3.45b}$$

where equation (3.45a) describes the background wavefield and equation (3.45b) the total wavefield including scattered wavefield. The scattered wavefield can be obtained by subtracting these two wave equations:

$$m_0\frac{\partial^2 u_{sc}}{\partial t^2} - \Delta u_{sc} = -\Delta m_1\frac{\partial^2 u}{\partial t^2} \tag{3.46}$$

where the term of the right-hand side, $-\Delta m_1\frac{\partial^2 u}{\partial t^2}$, represents the virtual source. We can solve for the scattered wavefield u_{sc} by using the Green's function:

$$u_{sc}(x,t) = \int_0^t \int_{\mathbb{R}^3} G_0(x,y,t-s)(-\Delta m_1\frac{\partial^2 u}{\partial t^2}(y,s))dyds \tag{3.47}$$

$$= -\Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2} \tag{3.48}$$

where $\hat{G}_0$ the short notation involves the space integral and convolution. We can rewrite the data formula by equations (3.44b) and (3.47) :

$$u = u_0 - \Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2} \tag{3.49}$$

we can re-arrange again by using the identity matrix I:

$$\left[\hat{I} - \Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}\right]u = u_0 \tag{3.50}$$

$$u = \left[\hat{I} - \Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}\right]^{-1} u_0 \tag{3.51}$$

where the inverse term represents the scattering operator. To evaluate this formula, we can utilize the Neumann formula $\hat{A} = \Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}$:

$$\left[\hat{I} + \hat{A}\right]^{-1} = \hat{I} - \hat{A} + \hat{A}^2 - \hat{A}^3 + \cdots \tag{3.52}$$

which leads to the Born series in the data formula:

$$u = u_0 - \left(\Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}\right)u_0 + \left(\Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}\right)\left(\Delta\hat{G}_0 m_1\frac{\partial^2 u}{\partial t^2}\right)u_0 + \cdots \tag{3.53}$$

$$= u_0 + \Delta u_1 + \Delta^2 u_2 + \cdots \tag{3.54}$$

The first-order term u_1 represents the single scattering and the second-order term u_2 the double scattering. The Born approximation only considers the single scattering and omits the higher-order terms. We can rewrite the equation (3.44b) with the single scattering equal to u_{sc} and neglection of the higher-order term:

$$m_0\frac{\partial^2 u_1}{\partial t^2} - \Delta u_1 = -\Delta m_1\frac{\partial^2 u_0}{\partial t^2} \tag{3.55}$$

This is the wave equation for the single scattered wavefield, which can be described by the background wavefield scattered at the model perturbation m_1, solved by:

$$u_1 = -\hat{G}_0(x_r,y,t)m_1\frac{\partial^2 u_0}{\partial t^2} \tag{3.56}$$

$$= -m_1\hat{G}_0(x_r,y,t)\frac{\partial^2}{\partial t^2}\left(\hat{G}_0(y,x_s,t)f\right) \tag{3.57}$$

Since the solution describes the linear relation between the scattered wavefield u_1 and model perturbation m_1, the partial derivative wavefield can be derived as:

$$\frac{\partial u_1}{\partial m_1} = -\hat{G_0'}(x_r, y, t)\frac{\partial^2}{\partial t^2}\left(\hat{G_0}(y, x_s, t)f\right) \tag{3.58}$$

$$= -\hat{G_0'}\frac{\partial^2}{\partial t^2}\left(\hat{G_0}f\right) \tag{3.59}$$

where the partial derivative wavefield, $\frac{\partial u}{\partial m}$, the same as in equation (3.34), is so-called the Frechét derivative or Jacobian matrix. Therefore, the partial derivative wavefield with respect to the model parameter m_1 can be interpreted as the wavefield emitted by the seismic source f at the source position and scattered by a point diffractor located at m_1.

If we first define the misfit function in the frequency domain for better explanation, in which it simplifies the mathematical treatment and enhances the clarity of the process. In that case, the misfit can be derived from the single scattered wavefield:

$$E = \frac{1}{2}\int_\omega |u_1|^2 \, d\omega \tag{3.60}$$

then the gradient can be obtained by the partial derivative wavefilelds:

$$\frac{\partial E}{\partial m} = \frac{1}{2}\frac{\partial}{\partial m}\int_\omega |u_{1R} + iu_{1I}|^2 \, d\omega \tag{3.61}$$

where the partial derivative of the complex function can be expanded as:

$$\int_\omega |u_{1R} + iu_{1I}|^2 = \frac{\partial}{\partial m}\left(u_{1R}^2 + u_{1I}^2\right) \tag{3.62}$$

$$= 2\left(u_{1R}\frac{\partial u_{1R}}{\partial m} + u_{1I}\frac{\partial u_{1I}}{\partial m}\right) \tag{3.63}$$

$$= 2\Re\left[\left(\frac{\partial u_{1R}}{\partial m} + i\frac{\partial u_{1I}}{\partial m}\right)(u_{1R} - iu_{1I})\right] \tag{3.64}$$

$$= 2\Re\left[\frac{\partial u_1}{\partial m}u_1^*\right] \tag{3.65}$$

The gradient in the equation (3.61) can be rewritten as:

$$\frac{\partial E}{\partial m} = \int_\omega \left[\frac{\partial u_1}{\partial m}u_1^*\right] d\omega \tag{3.66}$$

where u_1^* is the backpropagated wavefield. Combining these two equations (3.66) and (3.58), the gradient can be re-arrange:

$$\frac{\partial E}{\partial m} = -\omega^2 \int_\omega \Re\left[Gf_0 Gu_1^*\right] d\omega \tag{3.67}$$

where the Green's function G represents the relationship between a source and the resulting wavefield in the model. The forward propagated wavefield, i.e., Gf_0 represents the forward wavefield traveling from each source location to scattered points or grid points. The backward wavefield, i.e., Gu_1^* corresponds to the forward wavefield on the same model grid points but with the source terms replaced by the conjugate (or time-reversed) residuals propagated from all receiver positions. Therefore, in the frequency domain, the gradient is obtained by multiplying the forward wavefields with the backward wavefields.

Similarly, in the time domain, the gradient can be computed through the zero-lag cross-correlation between the forward propagated wavefields and the backpropagated wavefields:

$$\frac{\partial E}{\partial m} = \sum_{source}\int_0^t \frac{\partial^2 u_0}{\partial t^2}(x, t)\sum_r^{nr}\left[-\int_0^t G_0(x_r, t'; x, t)u_1(x_r, t')dt'\right] dt \tag{3.68}$$

where the backward (or adjoint) wavefield Ψ can be defined as:

$$\Psi(x,t) = -\sum_{r}^{nr} \int_0^t G_0(x_r, t'; x, t) u_1(x_r, t') dt' \tag{3.69}$$

Thus, the gradient represents the perturbation wavefields scattered by the missing heterogeneities in the starting model m_0, and the perturbation wavefield is built by the superposition of the wavefields scattered by each point diffractor (McMechan and Fuis, 1987; Pratt et al., 1998). This backpropagation technique in the SD method, as proposed by Lailly (1983), called the adjoint-state method, is not necessary to compute the partial derivative wavefield explicitly.

3.3.3 Full Newton method

In the SD method from section 3.3.2, the gradient part for model updating is determined only from the first-order derivatives of the misfit function in each iteration. In contrast, the FN method modifies the gradient direction by incorporating second-order derivatives of the misfit function, which is the Hessian matrix as we discussed before.

For model updating iteratively in the FN method, this is given by:

$$m_{n+1} = m_n - \alpha_n H^{-1} \nabla_m E(m) \tag{3.70}$$

where the elements of the Hessian matrix H are given by:

$$H_{ij} = \frac{\partial^2 E(m)}{\partial m_i \partial m_j}, \quad i = (1, 2, ..., m); \quad j = (1, 2, ..., m) \tag{3.71}$$

The Hessian matrix is of size M by M, where M represents the number of the model parameters, which provides more accurate estimate by preconditioning or filtering the gradient with the inverse Hessian in the FN method in equation (3.70) compared with the SD method. If model updating is applied iteratively, the inverse Hessian is effectively reconstructed during the iterative process.

Each element of the Hessian matrix can be explicitly expressed from equations (3.71) and (3.39) in differential form:

$$H_{ij} = \frac{\partial^2 E(m)}{\partial m_i \partial m_j} = \frac{\partial}{\partial m_i} \left[J^T \Delta d \right] = J^T J + \left(\frac{\partial J}{\partial m} \right)^T \Delta d \tag{3.72}$$

For simplification, the Hessian matrix reads:

$$H = H_a + R \tag{3.73}$$

where $H_a = J^T J$ is the approximate Hessian matrix, and the second term is represented by $R = \left(\frac{\partial J}{\partial m} \right)^T \Delta d$. There are M column vectors in the second term, each equal to Δd.

3.3.4 Gauss-Newton method

For the two terms of the Hessian matrix in the equation (3.73), the first term, the approximate Hessian matrix H_a, is relatively simple to compute. In contrast, the second term, R, is difficult to calculate. However, the second term is usually small and negligible, as noted by Tarantola (1987). By neglecting the second term, we obtain the GN method:

$$m_{n+1} = m_n - \alpha_n H_a^{-1} \nabla_m E(m) \tag{3.74}$$

$$= m_n - \alpha_n [J^T J]^{-1} \nabla_m E(m) \tag{3.75}$$

where the approximate Hessian matrix H_a can be derived by the partial derivative wavefields. These wavefields are generally correlated at adjacent grid points in the time domain, making the approximate Hessian matrix diagonally dominant due to auto-correlation. Consequently, the inverse of the approximate Hessian acts as a sharpening or focusing filter. The FN method only differs from the GN method in including the second term of the Hessian matrix (Levenberg, 1944; Marquardt, 1963).

Chapter 4

Gauss-Newton Method

This chapter introduces the inversion problem using the GN algorithm. Unlike the SD method, the GN method updates the model by applying the inverted approximate Hessian to the gradient. To address the associated computational challenges, this chapter introduces the use of a scaled Jacobian matrix, built by the partial derivative wavefield scaled by the source wavelet used for the forward propagation. In order to compute the Jacobian matrix, the reciprocity principle and virtual sources are employed to significantly reduce the computational cost.

4.1 Model updating for the Gauss-Newton method

In this section, we explain form of model updates applied in the GN method. Compared to the FN methods, the GN method explicitly computes the approximate Hessian matrix while avoiding the need to calculate and invert the full Hessian matrix. The approximate Hessian matrix for the acoustic case with constant density is diagonally dominant, resulting from the auto-correlation of the partial derivative wavefields (or the matrix-production of the Jacobian). The gradient term is also calculated using the Jacobian. The step length is iteratively estimated using a linearized approach using a linearized approach (Sheen (2004)).

As discussed in section 3.3.4, the GN method only inverts the first diagonal term of the full Hessian matrix while neglecting the second, smaller term R. Consequently, the model is updated as follows:

$$m_{n+1} = m_n - \alpha_n H_a^{-1} \nabla_m E(m) \tag{4.1}$$

$$= m_n - \alpha_n \left[J^T J + \lambda I^T I \right]^{-1} (J^T \Delta d) \tag{4.2}$$

The parameters m_n and m_{n+1} represent models at the current iteration (n) and subsequent iteration $(n+1)$, respectively. The step length α_n at the current iteration n is applied to the model update in the GN method. To update the model, three components must be calculated: (i) the approximate Hessian matrix, (ii) the gradient of the misfit function, and (iii) the optimal step length.

4.2 Approximate Hessian matrix

In this section, we continue explaining the GN algorithm, focusing on the explicit calculation of the approximate Hessian matrix. The approximate Hessian matrix, derived from the equation (3.72), is calculated as the auto-correlation of the partial derivative wavefields,

$J^T J$, in which here Jacobian J includes the partial derivative wavefield $\frac{\partial u}{\partial m}$. The most computationally demanding part of the GN algorithm is the calculation of the derivative wavefield or the Jacobian matrix.

4.2.1 Partial derivative wavefields

Unlike the implicit calculation in the SD method, the partial derivative wavefields in the GN method are computed explicitly. To address the huge computational demands, we use the approach combining the virtual sources with the reciprocity principle, which is given by Sheen (2006). Figure 4.1 shows how the wave propagates through the model parameters with the virtual sources and reciprocity principle in this thesis. The details are described in the following.

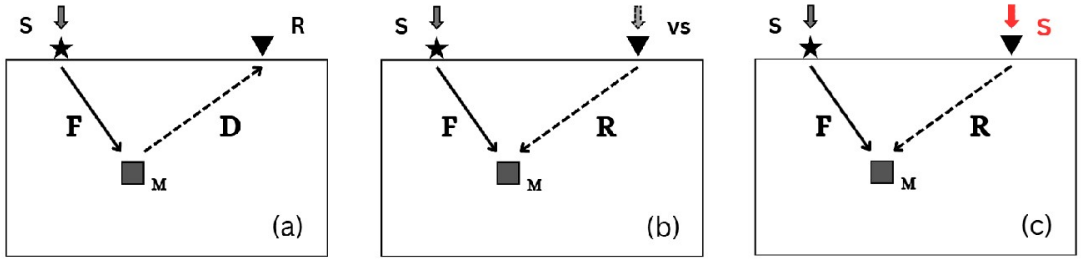


Figure 4.1: *Schematic diagram of the forward simulation for partial derivative wavefields. The partial derivative wavefield generated from a model point is shown in (a), from a reciprocal virtual source shown in (b), and from a reciprocal source shown in (c). The symbols S, R, and M denote the indices for the source, receiver, and model parameters, respectively. The vs is the virtual source, however, the S, shown in red, replaces the virtual source and propagates from the receiver. F, D and R represent the forward wavefield from the source, the differential wavefield from the virtual source, and the reciprocal wavefield from the receiver, respectively.*

4.2.1.1 Partial derivative wavefields from virtual sources

Figure 4.1 (a) shows that the wave is generated at the source location, recorded at the receiver location, and interacts with a model parameter during its propagation. The virtual sources are considered as sources representing scattering patterns of model perturbations. The forward simulation consists of two steps: First, a forward simulation with a source injected at the source position is conducted. This forward wavefield is then sampled at the grid points and used to calculate the virtual sources. Those are then injected at the model parameter positions to produce a second forward wavefield sampled at the receivers.

The virtual source is made up by the model perturbation multiplied with the partial derivative of the wave equation. In this thesis, we consider a Ricker wavelet as the source function. The wave equation is restated in equation (3.43), where the model parameter m represents the squared slowness, as shown in equation (3.42).

$$\left(m \frac{\partial^2}{\partial t^2} - \Delta \right) p = f \tag{4.3}$$

By taking the partial derivative of this equation with respect to a specific model parameter m_1, the virtual source is derived as:

$$\frac{\partial m}{\partial m_1} \frac{\partial^2 p}{\partial t^2} + m \frac{\partial^3 p}{\partial t^2 \partial m_1} = \frac{\partial \Delta p}{\partial m_1} \tag{4.4}$$

Rearranging terms gives:

$$m\frac{\partial^3 p}{\partial t^2 \partial m_1} - \frac{\partial \Delta p}{\partial m_1} = -\frac{\partial m}{\partial m_1}\frac{\partial^2 p}{\partial t^2} \tag{4.5}$$

In this formulation, the term $-\frac{\partial m}{\partial m_1}\frac{\partial^2 p}{\partial t^2}$ represents the virtual source for the partial derivative wavefield $\frac{\partial^2 p}{\partial t^2}$. A wave propagates starting from the location of the specific model parameter m_1. It is noted that $\frac{\partial m}{\partial m_1}$ is equal to 1.0 at the location of parameter m_1 and zeros for the rest of media. In other words, the virtual source is equal to $-\frac{\partial^2 p}{\partial t^2}$.

Consequently, calculating the partial derivative wavefield using the virtual source requires running forward simulations for each source location and for a perturbation in each model parameter. Specifically, one forward simulation is needed for each actual source location to obtain the virtual source at the parameter locations, and NM additional forward simulations are required to propagate the virtual sources at the parameter locations to generate responses at the receiver locations. This results in a total of $(NS + NM)$ forward simulations.

4.2.1.2 Reciprocity principle & convolution

Then, figure 4.1 (b) shows that the wave is generated by the virtual source with the reciprocity principle. The reciprocity principle allows the source-receiver locations to be interchanged. Thus, the same partial derivative wavefield at the virtual source can be obtained by propagating it from the receiver location, with the forward wavefield F recorded at the model parameter M. This is referred to as the reciprocal wavefield R. The partial derivative wavefield can be calculated by convolution of the forward wavefield F with the differential wavefield D with virtual source and also can be convolved by the forward wavefield F with the reciprocal wavefield R:

$$\frac{\partial u}{\partial m} = F_s^m * D_r^m = F_s^m * R_{vs}^m \tag{4.6}$$

where $*$ means convolution. F_s^m, D_r^m and R_{vs}^m are the forward wavefield from source location recorded at M, the differential wavefield from M recorded at receiver location, and the reciprocal wavefield with virtual source from receiver recorded at M, respectively.

However, it is still necessary to account for the virtual sources at each model parameter propagating from each receiver location. To reduce the number of forward simulations needed, we use source wavelets instead of virtual sources, combined with the reciprocity principle, as shown in figure 4.1 (c). The partial derivative wavefield is computed as the convolution of the virtual source, generated from the source location and recorded at the model parameters, with the reciprocal wavefield, generated from the receiver location and also recorded at the model parameters. However, the resulting partial derivative wavefield is scaled by the source function, as it involves convolution with the source function at the end. This approach significantly reduces the computational cost, requiring only $(NS + NR)$ simulations to calculate the partial derivative wavefield.

Finally, the partial derivative wavefield with respect to the model parameter for the sources and receivers can be calculated as:

$$\frac{\partial u}{\partial m} = F_s^m * R_s^m \tag{4.7}$$

where F_s^m represents the forward wavefield again, and R_s^m represents the reciprocal wavefield but with source propagating from the receiver to M.

The partial derivative wavefield used in this thesis is scaled by the source function, which is modeled as a Ricker wavelet at both the source and receiver locations. Consequently, the Jacobian matrix, rearranged from the partial derivative wavefield, is also scaled by the source function.

4.2.2 Jacobian matrix

The Jacobian matrix is constructed by arranging the partial derivative wavefields. The formula for the partial derivative is derived in equation (3.58) using the Green's function:

$$J = \frac{\partial u}{\partial m} = -\hat{G_0'}(x_r, y, t)\frac{\partial^2}{\partial t^2}\left(\hat{G_0}(y, x_s, t)f\right) \tag{4.8}$$

where $\frac{\partial^2}{\partial t^2}\left(\hat{G_0}(y, x_s, t)f\right)$ represents the virtual source though the forward simulation from the sources corresponding to the identical model parameter. $-\hat{G_0'}(x_r, y, t)$ represents another forward propagated wavefield from the receivers convolved with the virtual source.

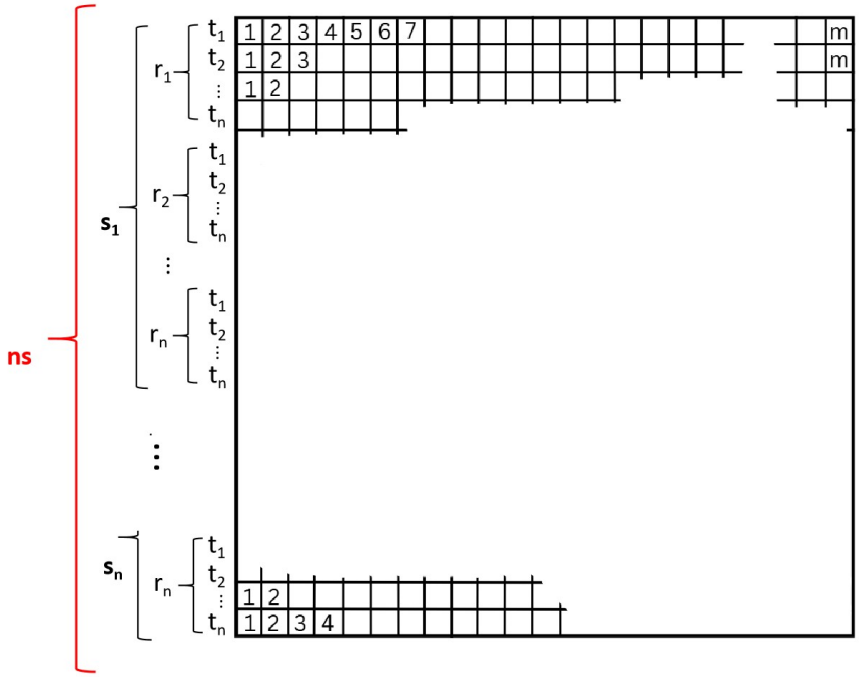


Figure 4.2: *A representation of the huge Jacobian matrix. The Jacobian matrix is a N by M matrix. Here, N is equal to $ns \times nr \times nt$, where ns, nr, and nt represent the number of sources, receivers, and time steps, respectively. M denotes the number of model parameters.*

The scaled Jacobian matrix J_s used in this thesis can be regarded as the Jacobian matrix J scaled by the source function s. Based on the formula for the Jacobian matrix, the scaled Jacobian matrix is expressed as:

$$J_s = \frac{\partial u}{\partial m} = -(G_0'(x_r, y, t) * s)\frac{\partial^2}{\partial t^2}\left(\hat{G_0}(y, x_s, t)f\right) \tag{4.9}$$

where $\frac{\partial^2}{\partial t^2}\left(\hat{G_0}(y, x_s, t)f\right)$ again represents the virtual source, but $-(G_0'(x_r, y, t) * s)$ represents the reciprocal wavefield from receivers. This formula is consistent with the previous equation (4.7). The scaled Jacobian matrix is derived still by the convolution of two wavefields: forward wavefield F_s^m and reciprocal wavefield R_s^m. Based on the above derivation,

a total of NS forward simulations are required to obtain the virtual sources at all model parameters, where NS is the number of the sources, and NR reciprocal simulations are needed to obtain the reciprocal wavefield at all model parameters, which NR is the number of the receivers. In total, $(NS + NR)$ simulations are required to compute the Jacobian matrix. The resulting Jacobian matrix is a large a $(NS \cdot NR \cdot NT)$ by NM matrix, where NT is the number of the time steps, as illustrated in figure 4.2.

4.2.3 Approximate Hessian matrix

The approximate Hessian matrix is a M by M matrix, based on the Jacobian matrix and damping term used in this thesis. The formula for approximate Hessian is given as:

$$H_a = J^T J + \lambda I^T I \tag{4.10}$$

where $\lambda I^T I$ is a damping term. It improves and stabilizes the GN method for non-linear inversion problems. It is based on the damped least-squares method introduced by Levenberg (1944) and Marquardt (1963), a small scalar regularization value λ provides a trade-off. If stronger damping is applied (i.e., $\lambda \gg 0$), the damping term dominates, causing the GN method to approach the behavior of the SD method, which significantly reduces the convergence rate. Conversely, a lower damping value provides a faster convergence but increases the risk of converging to a saddle point rather than the local minimum in the solution space. Based on several computations, this thesis uses a damping value of $\lambda = 0.001$ times the maximum element of the approximate Hessian's diagonal, equivalent to 0.1 percent of the maximum diagonal element value.

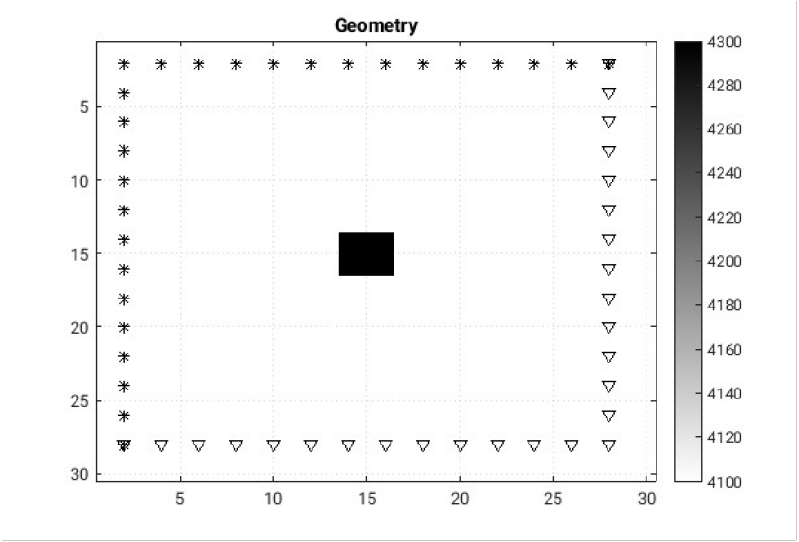


Figure 4.3: *The geometry of the true model for the approximate Hessian matrix which is shown in Figure 4.4. The source is a Ricker wavelet with a center frequency of 12 Hz. The approximate Hessian matrix corresponds to a 31 × 31 model, with 27 sources (marked by stars) and 27 receivers (marked by triangles) positioned around the boundaries. One velocity perturbation is located at the center of the model.*

Thus, the approximate Hessian matrix is diagonally dominant, due to the auto-correlations occurring on the main diagonal. Figure 4.4 shows the symmetric approximate Hessian matrix computed using the true model in Figure 4.3. The structure of the matrix naturally depends on the ordering used for the node points. The diagonal dominance of the matrix with the off-diagonal bands is clear. Additionally, the source and receiver locations influence the structure of the approximate Hessian matrix due to the geometrical spreading in the amplitudes of the partial derivative wavefields.

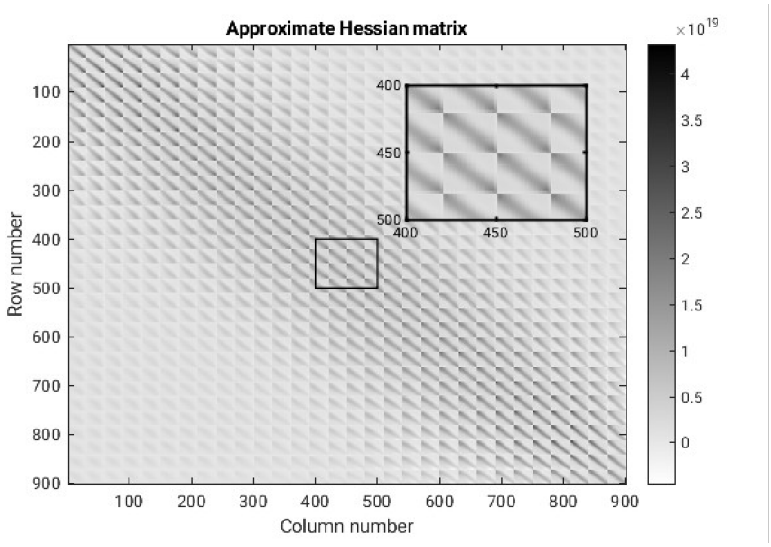


Figure 4.4: *A representation of the approximate Hessian matrix. The 961 × 961 element in the approximate Hessian matrix is primarily diagonal, but non-zero oscillatory behavior is observed off the main diagonal. Each pixel in the matrix represents the correlation between the partial derivative wavefields emerging from one node point and another.*

The approximate Hessian shown in Figure 4.4 was computed with 27 sources and 27 receivers in the true model, distributed around the boundaries of the true model. A single velocity perturbation with a value of 4300 m/s is positioned at the center of a homogeneous background model with a velocity of 4100 m/s.

4.3 Gradient calculation

The explicit calculation of the gradient requires to know the Jacobian as discussed in Section 3.3.1 and shown in equation (3.39). Compared to the SD method, which avoids the explicit computation of a large Jacobian matrix by calculation of the adjoint wavefield, the Jacobian matrix in GN method is explicitly computed. This introduces higher computational costs but also improves the accuracy of the model update. In this section, we discuss the gradient, which is computed by multiplying the transpose of the Jacobian matrix with the data residual. Since the scaled Jacobian is used in this thesis, the gradient is also scaled by the source function, distinguishing it from the gradient obtained using the SD method. We compare the two types of gradients across these methods, emphasizing their significantly different formulas and simulation requirements. Finally, we summarize the complete workflow of the methods: the GN method involves model updates using the approximate Hessian matrix and the scaled gradient, while the SD method achieves gradient computation without the explicit calculation of the Jacobian.

4.3.1 Gradient & Scaled gradient

The calculation of the gradient of the misfit function has been discussed in section 3.3.2 using the SD method. This calculation can be simplified, as shown in Equation (3.67):

$$g = \frac{\partial E}{\partial m} = -\omega^2 \int_{\omega} \Re \left[G f_0 G u_1^* \right] d\omega \tag{4.11}$$

Here, the gradient is derived from the multiplication of the forward-propagated wavefield with the source function and the backward-propagated wavefield with the data residuals

in the frequency domain, where f_0 represents the source function, which is the same as s described in the previous section, and u_1 represents the data residual. The superscript $*$ indicates time reversal.

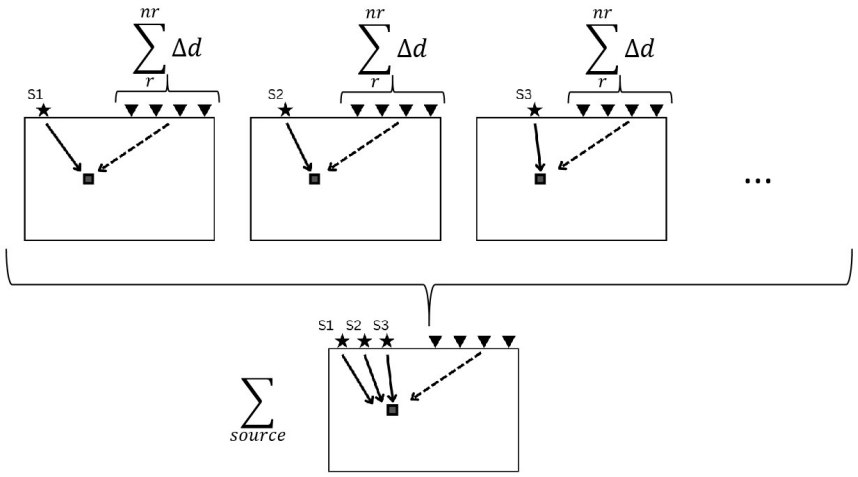


Figure 4.5: *An explanation for the gradient calculation in the SD method. s represents sources and r represents receivers. The schematic diagrams illustrate wave propagation with the data residual Δd. The gradient in the SD method is derived from both the forward and backward simulations. The forward simulation depends on each individual source, while the backward simulation is based on the data residual from all receivers. The total number of simulations required is $2NS$, where NS is the number of sources.*

A visualized explanation of the gradient in the SD method is shown in Figure 4.5. In this method, if we consider the wave propagation for each source in a synthetic simulation: the wave forward-propagates from the first source, generating data residuals from the difference between the synthetic data recorded at all receivers at once and the observed data. The partial derivative wavefield at the model parameter is obtained by multiplying two wavefields in the frequency domain or by convolving them in the time domain. Next, the partial derivative wavefield is calculated for the second source, and the results from all sources are summed up. The gradient in the SD method depends on the number of sources, denoted as NS. The total number of simulations required is $2NS$.

Finally the gradient can be calculated by considering the contributions from all sources as follows:

$$g = \sum_{source} \left(-\omega^2 \sum_{receiver} \int_{\omega} \Re \left[G f_0 G u_1^* \right] d\omega \right) \tag{4.12}$$

The gradient for the GN method is derived by the multiplication of transpose of the Jacobian matrix and data residual, the formula is shown in equation (3.39):

$$g = J^T \Delta d \tag{4.13}$$

Since the Jacobian matrix is scaled by the source function in the GN method, the data residual must also be scaled by the source function. Consequently, the scaled gradient of the misfit function is expressed as:

$$g_s = J_s^T (\Delta d * s) \tag{4.14}$$

A visualized explanation of the scaled gradient in the GN method is shown in Figure 4.6. In this method, wave propagation for each source and each receiver is considered

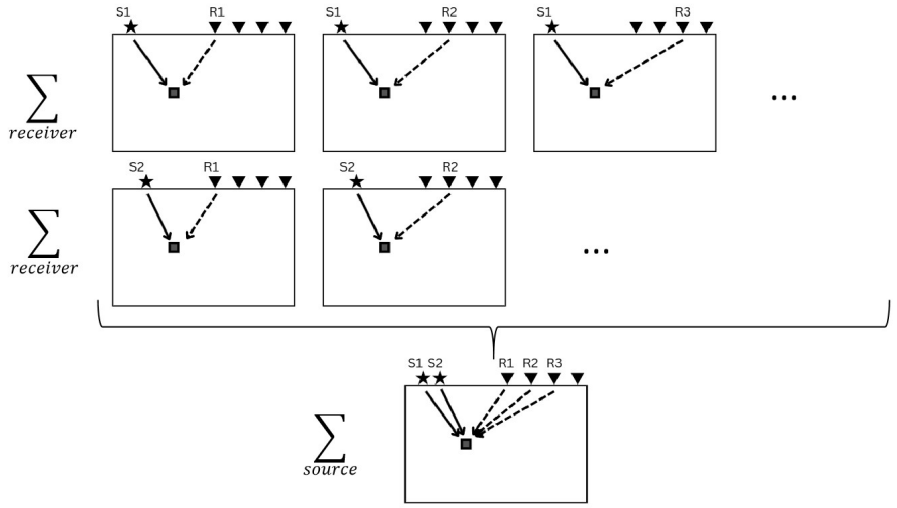


Figure 4.6: *An explanation for the gradient calculation in the GN method. The same as Figure 4.5, s represents sources and r represents receivers. The gradient in the GN method is derived from both the forward and reciprocal simulations. The forward simulation depends on each individual source, while the reciprocal simulation depends on each individual receiver. The total number of simulations required is $(NS + NR)$, where NS and NR represent the number of sources and receivers, respectively.*

as a matrix-vector multiplication with the scaled Jacobian matrix and the data residual, which is also scaled by the source function. The total number of simulations for the scaled gradient in the GN method is $(NS + NR)$.

In this method, the scaled gradient can be calculated by considering the contributions from all sources and receivers as follows:

$$g_s = \sum_{source} \left(\sum_{receiver} J_s^T (\Delta d * s) \right) \tag{4.15}$$

4.3.2 Workflow for gradient calculation

In the GN method, the procedure for computing the Jacobian matrix, the approximate Hessian matrix, and the gradient direction for multiple sources and receivers is summarized as follows:

1. Simulate all forward wave propagations from the sources to the receivers to obtain the estimated data and sample the virtual sources at the model parameters m.
2. Compute the data residuals by subtracting the estimated data obtained in step (1) from the observed data. Convolve the data residuals with the source function.
3. Simulate scaled reciprocal wave propagations from each receiver. Sample the reciprocal wavefields at the model parameter m.
4. Construct the scaled Jacobian matrix using the partial derivative wavefields obtained by convolving the virtual sources from step (1) and the reciprocal wavefields from step (3).
5. Calculate the approximate Hessian matrix through the matrix-production of the scaled Jacobian matrix obtained in step (4).
6. Compute the scaled gradient direction by multiplying the scaled Jacobian matrix from step (4) with the data residuals from step (2).

The SD method updated by the adjoint-state method without the explicit calculation of partial derivative wavefields for multiple sources and receivers is summarized as follows:

1. Simulate all forward wave propagations from the sources to the receivers to obtain the estimated data.
2. Compute the data residuals by subtracting the estimated data obtained in step (1) from the observed data.
3. Simulate backward wave propagations from all receivers. The backward source is the data residuals from step (2).
4. Construct the gradient obtained by convolving the wavefields from steps (1) and (3).

4.4 Optimal step length

The simplest way to choose the step length is to use a small constant throughout all iterations, determined by trial and error. However, in this work, an optimal step length is determined using a linearized approach based on the methods of Gauthier et al. (1986) and Pica et al. (1990). This linearized approach involves three steps:

1. Initial trial step length
   We assume a trial step length ε with a sufficiently small value for each iteration. If the maximum element of the gradient term g, i.e., $H_a^{-1}\nabla_m E(m)$, exceeds 1 percent of the maximum model value during the current iteration n, the trial step length is set to $\varepsilon = 0.5$. Otherwise, it is set to $\varepsilon = 1$. Using this trial step length, the trial model d_{trial} can be updated using the non-linear forward operator F, as described in section 3.3.1.
   $$d_{trial} = F(m_n - \varepsilon g_n) \tag{4.16}$$
2. Calculation of trial residual
   The difference between the trial and estimated data is used to calculate the optimal step length for iteration n. The trial residual Δd_{trial} which describes this difference, is scaled by the factor $\frac{1}{\varepsilon}$:
   $$\Delta d_{trial} = \frac{1}{\varepsilon}(d_{trial} - d_{syn}) = \frac{1}{\varepsilon}\left[F(m_n - \varepsilon g_n) - F(m_n)\right] \tag{4.17}$$
3. Optimal step length
   The optimal step length α_n is defined using the trial residual and the data residual, which represents the difference between the observed and estimated data for iteration n:
   $$\alpha_n = \frac{(\Delta d_{trial})^T (d_{syn} - d_{obs})}{(\Delta d_{trial})^T \Delta d_{trial}} \tag{4.18}$$

The optimal step length in this thesis based on the linearized approach is iteratively updated to adapt to the model and data characteristics. This approach enhances inversion efficiency and accuracy, ensuring the updated model converges steadily toward the desired solution over iterations.

Chapter 5

Methods Comparison

This chapter presents synthetic tests to investigate the properties of the SD method and the GN method for acoustic full-waveform inversion. In all examples, the sources, receivers, and perturbations are assumed to be known. The inversion results are compared to evaluate the resolution capabilities and efficiency of the two methods.

The chapter is organized as follows: First, we compare the gradients of the two methods, followed by an evaluation of reconstruction results based on their respective gradients. Next, we assess inversion performance for different acquisition geometries: (i) a reflection case and (ii) a transmission case. We also compare the two methods regarding the required number of sources for efficient resolution. Finally, we explore further tests of the GN method, focusing on the influence of receiver configurations.

5.1 Model setup

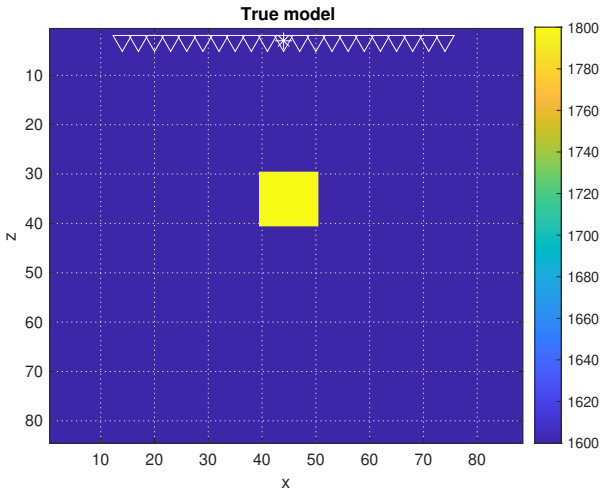


Figure 5.1: *The geometry of the true model. The source is a Ricker wavelet with a center frequency of 12 Hz. The true model has dimensions of 88×84 grid points ($ns \times nz$), with a single source (indicated by a star) located at the surface and 21 receivers (indicated by triangles) positioned along the surface. A single velocity perturbation is located at the center of the model.*

The true model consists of 88×84 grid points ($ns \times nz$), defining a rectangular computational domain, as shown in Figure 5.1. The velocity model features a homogeneous background

with a velocity of 1600 m/s, within which a single high-velocity anomaly is embedded as a perturbation. This perturbation, with a velocity of 1800 m/s, is located centrally in the model. The initial velocity model is identical to the homogeneous background, ensuring the inversion begins from a simple and unbiased assumption. A single seismic shot is positioned near the surface, accompanied by 21 symmetrically distributed receivers arranged along the surface around the source. The maximum offset between the source and the farthest receiver is 176 m. The synthetic data is generated using a Ricker wavelet with a dominant frequency of 12 Hz, with the source placed just below the surface. For wave propagation simulations, the grid spacing is set to 4 m, and the time step is 0.5 ms, ensuring sufficient spatial and temporal resolution for accurate modeling.

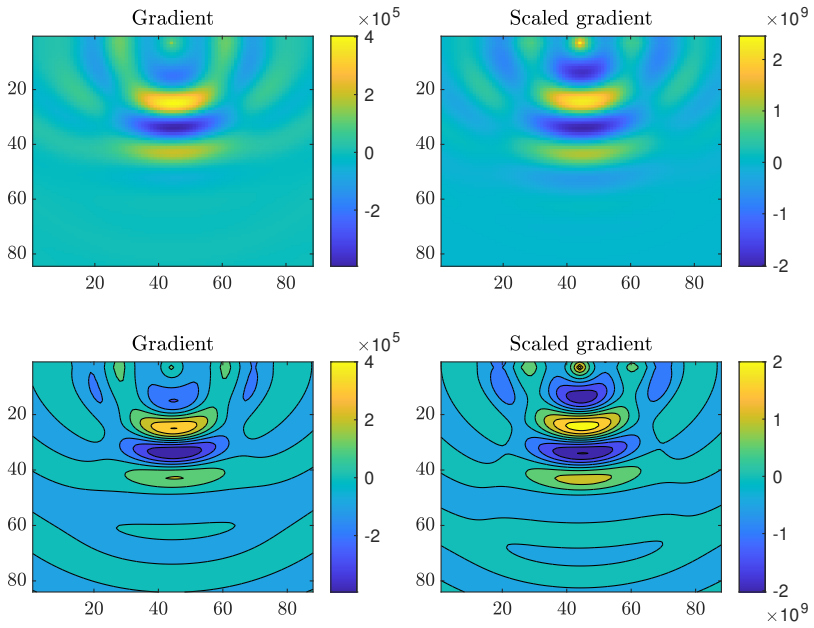


Figure 5.2: *A simple example of gradient comparison. The gradient and scaled gradient after one iteration, were computed using the SD method and the GN method, respectively. The upper panels display the comparison in smooth mode, while the lower panels present the comparison in contour mode.*

Figure 5.2 presents a comparison of the gradients obtained using the SD and GN methods. As discussed in Section 4.3.1, the gradient in the SD method is computed implicitly using the adjoint-state method, whereas, in the GN method, the gradient is scaled explicitly by the source function. The figure compares the two gradients in two modes: the upper two panels display the gradients in smooth mode, while the lower two panels use contour mode. To illustrate the difference, the scaled gradient is shown with a distinct colorbar scale compared to the gradient. Notably, the scaled gradient from the GN method exhibits more intricate artifacts, particularly around the source, receiver, and perturbation locations.

Based on the gradient calculations, the reconstructions obtained using the SD and GN methods are shown in Figure 5.3. Both methods are evaluated after only five iterations of FWI. The reconstruction from the GN method already demonstrates a distinct ability to capture the velocity perturbation's structure and trend. In contrast, the SD method exhibits a significantly limited capability in resolving the perturbation within the same number of iterations. This comparison highlights a critical advantage of the GN method: it requires significantly fewer iterations than the SD method to achieve similar or better results. To further explore this advantage, subsequent comparisons will evaluate 20 iterations in

the Gn method and 50 iterations in the SD method.

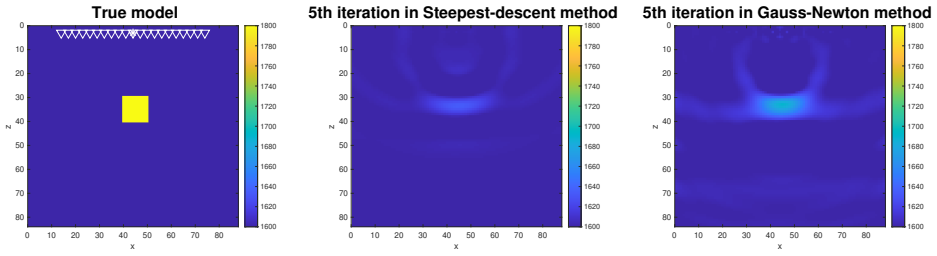


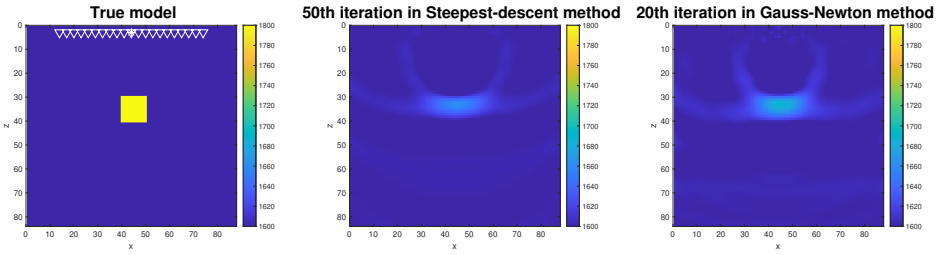
Figure 5.3: *Comparison of model updates after gradient calculation. The FWI results are obtained using the SD method (middle) and the GN method (right) compared with the true model (left) after 5 iterations.*

5.2 Multi-source test comparison

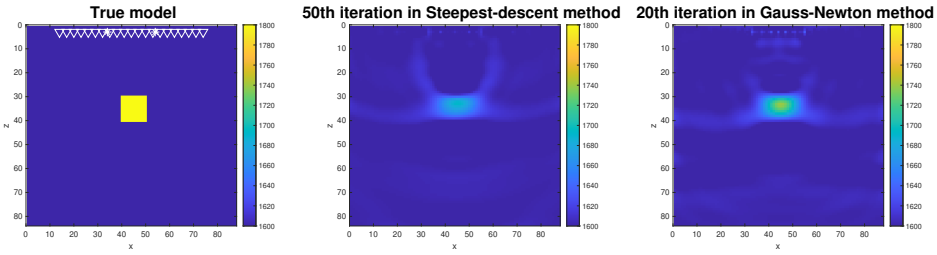
After the initial simple comparison, we focus on examining the influence of source configurations on the inversion process. Specifically, we compare two distinct scenarios: the reflection case and the transmission case. In the reflection case, both the source and receiver are positioned near the surface, with the wavefield primarily propagating in the vertical direction, reflecting off subsurface structures. In contrast, in the transmission case, the source is positioned at a deeper depth within the medium, and the seismic waves travel through the entire model.

5.2.1 Reflection case

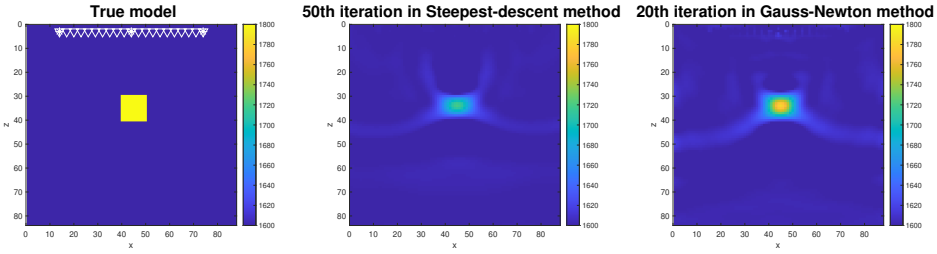
In the reflection case, we compare the reconstructions with one, two, three, and five sources. Figure 5.4 presents comparisons of model reconstruction using the SD method and the GN method, with an increasing number of sources. The source configurations vary across the figures: one source in figure 5.4 (a), two in Figure 5.4 (b), three in figure 5.4 (c), and five in figure 5.4 (c) in figure 5.4. The SD method produces broad reconstructions of the anomaly, exhibiting limited accuracy and resolution across all cases. In contrast, the GN method consistently demonstrates better reconstruction performance. With a limited number of sources, the GN results reveal a faint yet localized anomaly centered near the true location, although the resolution remains relatively low. As the number of sources increases to three and five, the reconstructions become progressively sharper and more concentrated around the true anomaly. Although the SD method shows slight improvement with more sources, its resolution remains significantly lower than that of the GN method. The velocity profiles corresponding to these results are depicted in figure 5.5. The true model, represented by the dotted line, highlights the high-velocity perturbation. In the velocity profile comparisons, the GN method demonstrates significant improvement as the number of sources increases compared with the SD method. Especially as the number of sources increases to five, the GN method shows a closer alignment with the true model, accurately capturing the location of the perturbation.



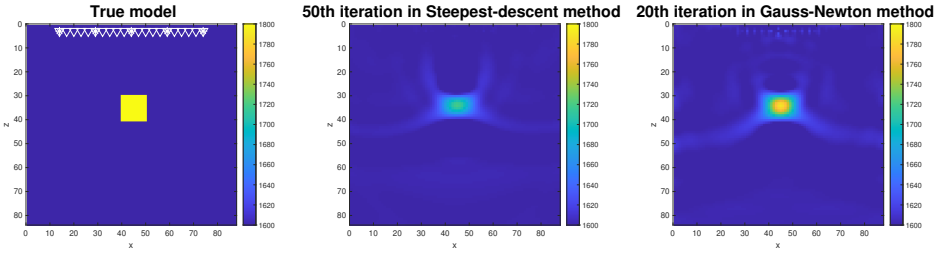
(a) Comparison of model updates with one source.



(b) Comparison of model updates with two sources.

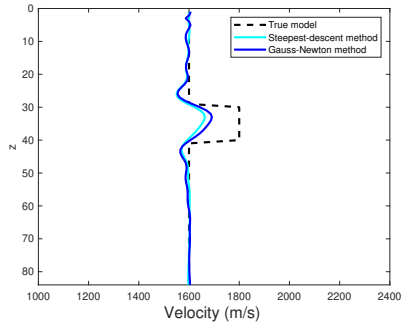


(c) Comparison of model updates with three sources.

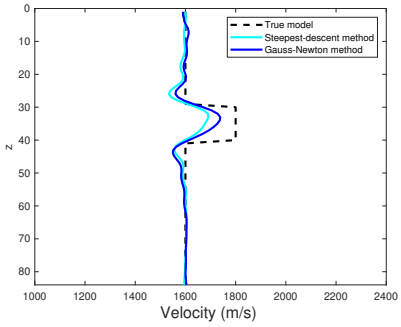


(d) Comparison of model updates with five sources.

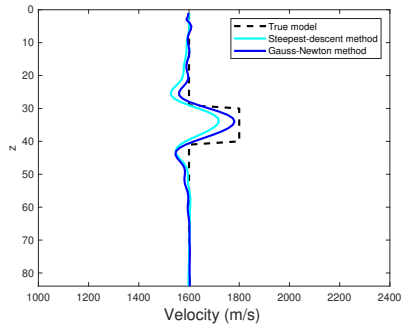
Figure 5.4: *Comparison of model updates in reflection case. The FWI results were obtained using the SD method after 50 iterations (middle) and the GN method after 20 iterations (right), compared to each true model (left).*



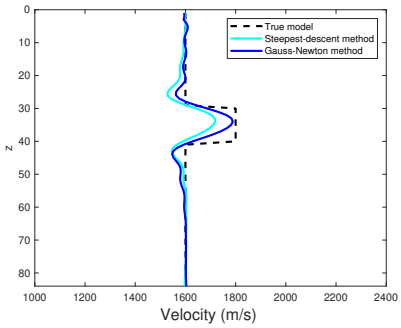
(a) Velocity profile with one source.



(b) Velocity profile with two sources.

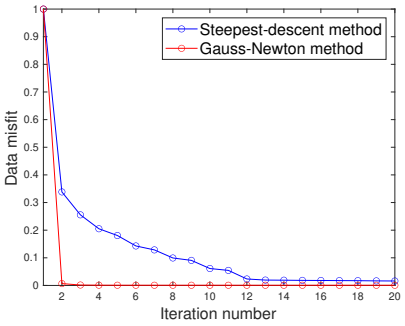


(c) Velocity profile with three sources.

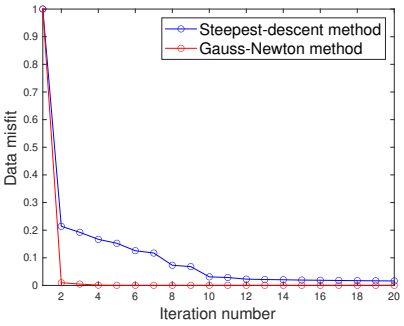


(d) Velocity profile with five sources.

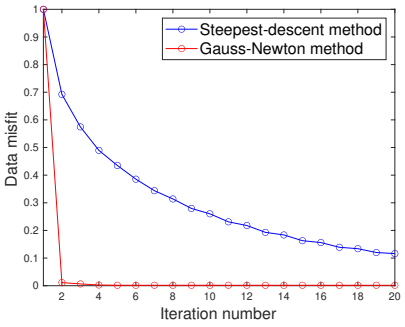
Figure 5.5: *The vertical velocity profiles in reflection case. The FWI results are produced using the SD method and the GN method, compared to the true model.*



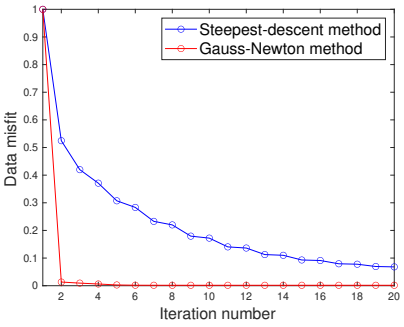
(a) Date misfit with one source.



(b) Date misfit with two sources.



(c) Date misfit with three sources.

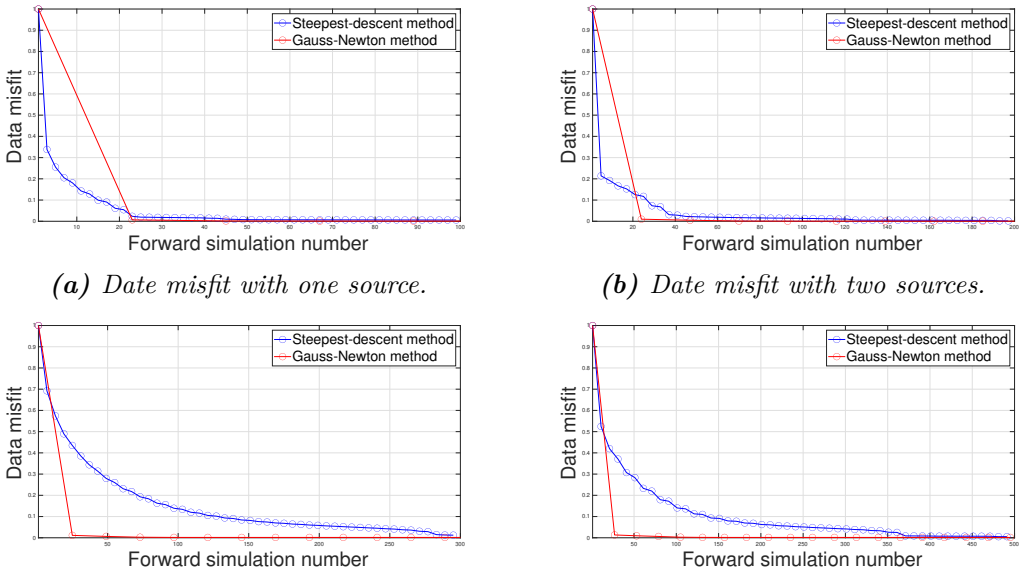


(d) Date misfit with five sources.

Figure 5.6: *Comparison of data misfit over iteration number in reflection cases. The data misfit illustrates the convergence behavior.*

After the reconstruction comparison, figures 5.6 and 5.7 show the comparison based on the data misfit, over with the iterations and forward simulations, respectively. The data misfit provides a quantitative measure of how well the predicted data matches the observed data during the inversion process. By analyzing the misfit, we can further evaluate the convergence behavior and efficiency of the GN and SD methods. Figure 5.6 presents a comparison of data misfit over iteration numbers for different multi-source configurations in reflection case. The GN method demonstrates a rapid decrease in data misfit within the first few iterations and converges to a minimal value significantly faster than the SD method in all cases. In contrast, the SD method exhibits a slower reduction in data misfit, exhibiting its less efficient convergence behavior. Note that the iterations are limited to 20, which corresponds to the maximum number of iterations used for the GN method.

However, the GN method consumes more computational time compared to the SD method. This is primarily due to the explicit calculations involved in the GN method, including the Jacobian matrix or approximate Hessian matrix. the number of forward simulations shows the computational cost for each method in figure 5.7. The two methods differ in the number of simulations required to compute their respective gradient. As discussed in Section 4.3.1, the number of forward simulations in the GN method is $(NS + NR)$ in each iteration, whereas the SD method requires $2NS$. Therefore, after 20 iterations for the GN method and 50 iterations for the SD method, the total number of simulations with a single source case amounts to 440 and 100, respectively. For the two-source case, the GN method requires a total of 460 simulations, compared to 200 simulations for the SD method. Similarly, for the three-source and five-source cases, the GN method requires 480 and 520 simulations, whereas the SD method requires 300 and 500 simulations, respectively. These comparisons show that, while the GN method exhibits a more rapid convergence rate than the SD method, it requires significantly more simulations in each iteration. Note that the number of simulations in this study is limited by the SD method.



(a) *Date misfit with one source.*

(b) *Date misfit with two sources.*

(c) *Date misfit with three sources.*

(d) *Date misfit with three sources.*

Figure 5.7: *Comparison of data misfit over the number of forward simulations. The data misfit illustrates the convergence behavior, while the number of forward simulations represents the computational cost for each method.*

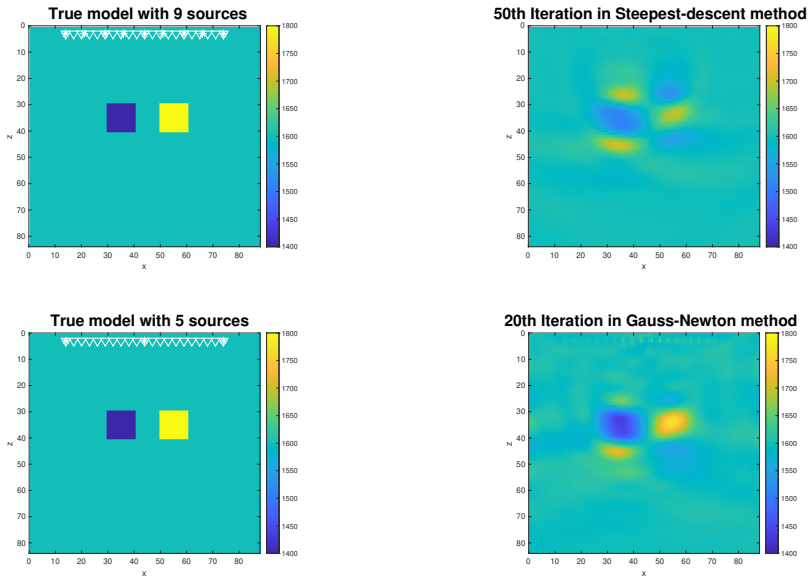


Figure 5.8: *Comparison of model updates using different source configurations with the high-velocity (yellow) and low-velocity (blue) perturbations. The SD method with nine sources after 50 iterations (upper right) is compared to the GN method with three sources after 20 iterations (lower right). The true model, featuring both high-velocity and low-velocity perturbations, is shown on the left.*

Figure 5.8 compares model updates for the two methods under different source configurations, incorporating both high-velocity and low-velocity perturbations. To achieve comparable reconstructions, the SD method uses nine sources, while the GN method requires only three sources. In the SD method, the reconstructed model exhibits significant limitations, with the two perturbations appearing blurred and surrounded by artifacts, indicating poor resolution and precision. In contrast, the GN method achieves a markedly more accurate reconstruction. The high-velocity and low-velocity anomalies are well-resolved, with sharper boundaries that closely match the true model. Corresponding velocity profiles and data misfit results are presented in figure 5.9, further supporting these observations. The comparison in velocity profiles and data misfit shows that the SD method exhibits limited accuracy and slower convergence compared with the GN method.

Notably, while the previous cases required more forward simulations for the GN method due to its computational demands, this case demonstrates that the SD method (900) is reliance on a higher number of sources and iterations results in far more forward simulations overall than the GN method (480). This indicates that the number of forward simulations is highly dependent on the specific case and configuration. Ultimately, from the results, the GN method demonstrates its ability to achieve superior resolution and accuracy, making it more efficient and reliable in this scenario.

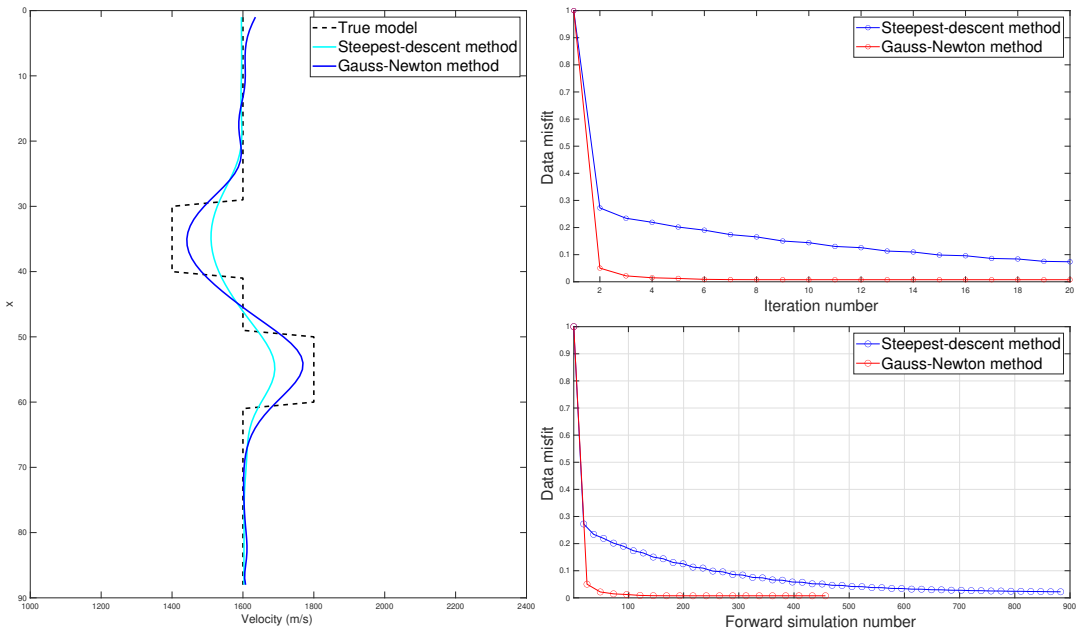
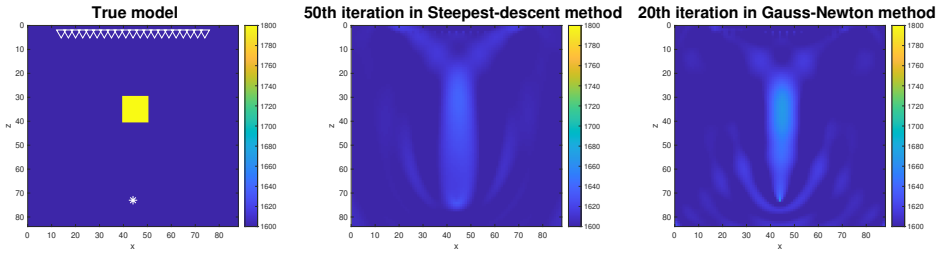


Figure 5.9: *Comparison of the horizontal velocity profiles (left) and of the data misfit (right) in different source case. The data misfit over iteration numbers shown in the upper right and over number of the forward simulations in the lower right.*

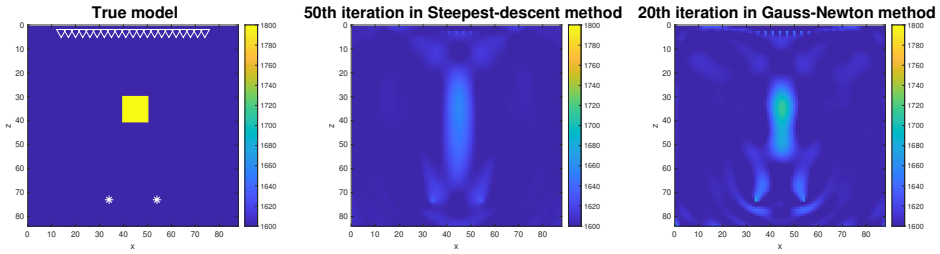
5.2.2 Transmission case

In the transmission case, the receivers are positioned near the surface, while the sources are strategically placed at a deeper depth beneath the velocity perturbations. This configuration allows the transmitted waves to travel directly through the subsurface anomalies, providing critical information about the internal velocity structure. Similar to the reflection case, we compare the reconstruction with increasing the source numbers. The source configurations include one source in figure 5.10 (a) two in figure 5.10 (b), three in figure 5.10 (c), and five in figure 5.10 (d). As the number of sources increases, the GN method consistently provides more precise and sharper reconstructions than the SD method. However, unlike the reflection case, the reconstructions for both methods in the transmission case exhibit more artifacts, likely caused by the placement of sources below the perturbations and the propagation dynamics of transmitted waves. The corresponding velocity profiles in figure 5.11 confirm these observations, demonstrating that the GN method achieves more accurate reconstructions than the SD method, even in the presence of transmission-specific artifacts. When the number of sources increases to five in the GN method, the reconstruction accurately identifies the perturbation's location. However, this improvement is accompanied by a significant increase in artifacts throughout the results, particularly around the source and receiver positions.

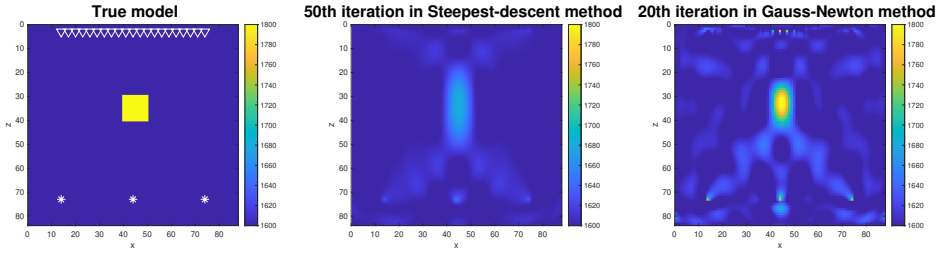
The data misfit in the transmission case is compared in terms of both iterations and forward simulations, as shown in figures 5.12 and 5.13, respectively. The GN method demonstrates faster convergence in both comparisons, with a rapid reduction in data misfit during the initial iterations than the SD method. However, due to the presence of artifacts, the final data misfit achieved by the SD method is slightly better than that of the GN method. Nevertheless, combining both the reconstruction and data misfit comparisons, the inversion results from the GN method are closer to the true model.



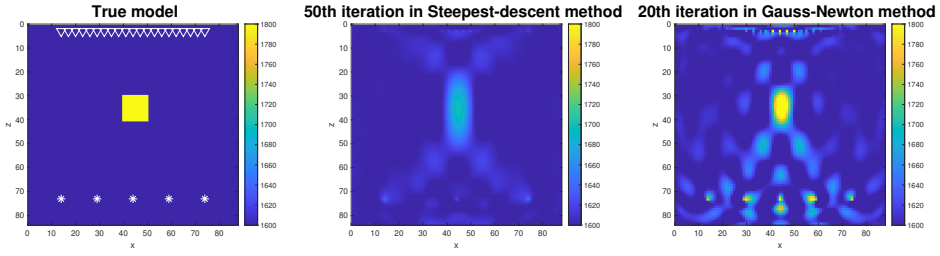
(a) Comparison of model updates with one source.



(b) Comparison of model updates with two sources.

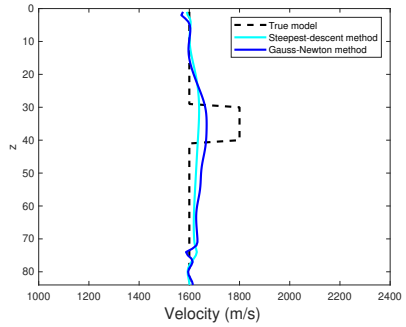


(c) Comparison of model updates with three sources.

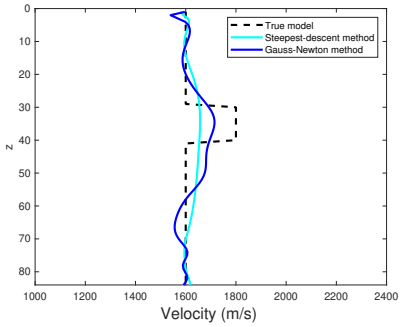


(d) Comparison of model updates with five sources.

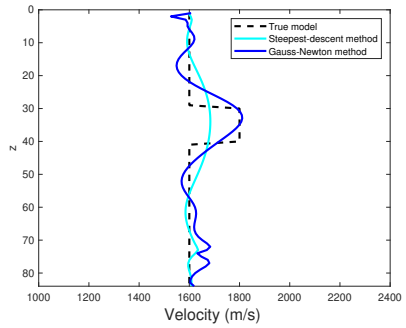
Figure 5.10: *Comparison of model updates in transmission case.*



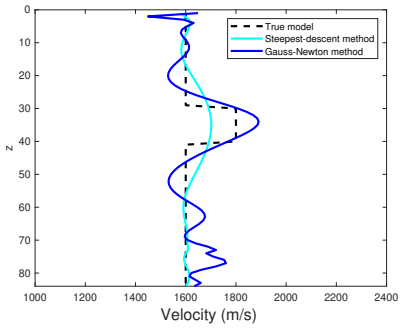
(a) Velocity profile with one source.



(b) Velocity profile with two sources.

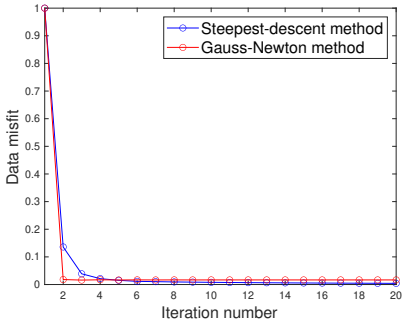


(c) Velocity profile with three sources.

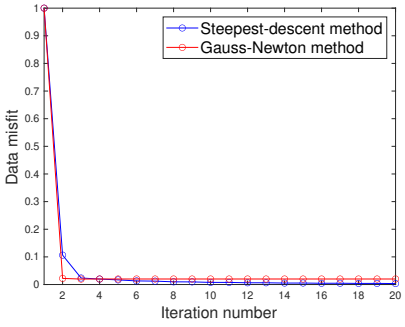


(d) Velocity profile with five sources.

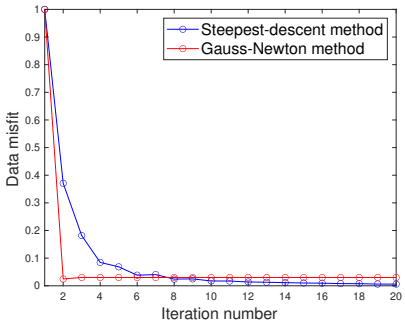
Figure 5.11: *The vertical velocity profiles in transmission case.*



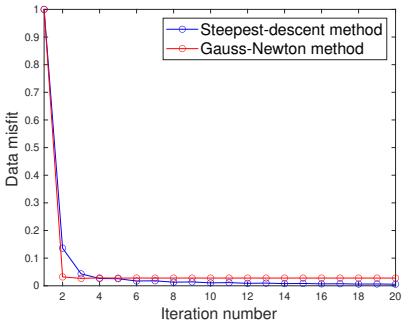
(a) Data misfit with one source.



(b) Data misfit with two sources.

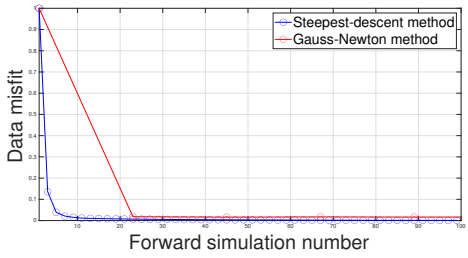


(c) Data misfit with three sources.

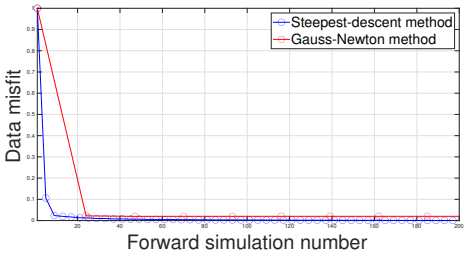


(d) Data misfit with five sources.

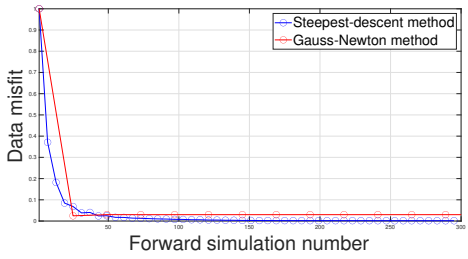
Figure 5.12: *Comparison of data misfit over iteration number in transmission case.*



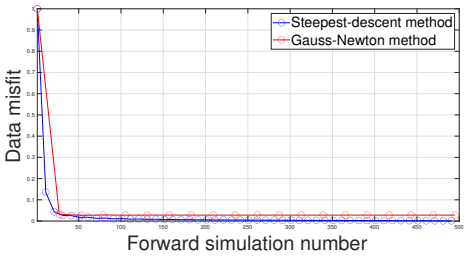
(a) Data misfit with one source.



(b) Data misfit with two sources.



(c) Data misfit with three sources.



(d) Data misfit with five sources.

Figure 5.13: *Comparison of data misfit over the number of forward simulations in transmission case.*

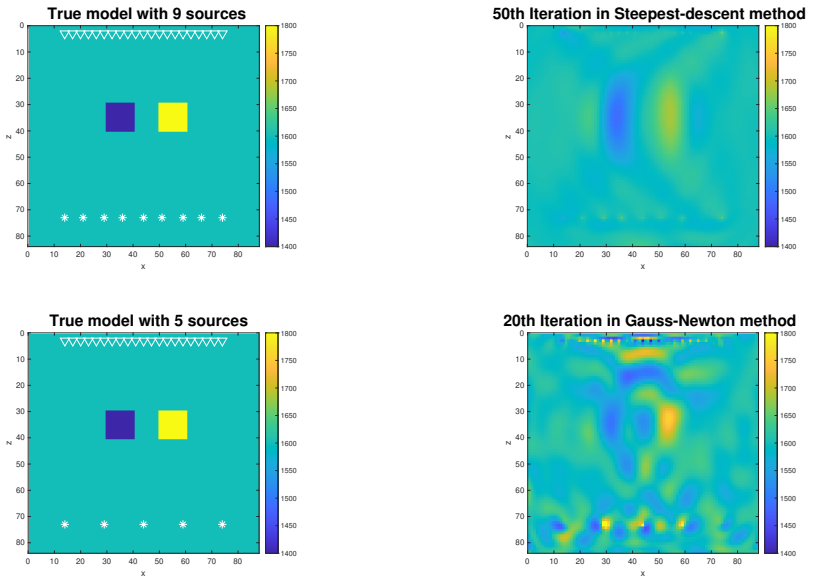


Figure 5.14: *Comparison of model updates using different source configurations.*

Figure 5.14 compares the model updates for the two methods, considering both high-velocity and low-velocity perturbations. The SD method uses nine sources, while the GN method uses only five. The corresponding velocity profiles and data misfit are shown in Figure 5.15. These results indicate that while the GN method exhibits a higher data misfit compared to the SD method, this is mainly due to the artifacts. In terms of reconstruction and velocity profiles, the GN method provides a clearer and more accurate result, particularly in regions with high-velocity anomalies, where the resolution is notably superior. However, in contrast, the low-velocity regions are significantly affected by artifacts, which diminish the overall accuracy in these areas. As observed in the reflection case, the SD method requires more forward simulations due to its higher number of sources and iterations, whereas the GN method converges faster. This demonstrates that the GN method achieves higher resolution and faster convergence, though at the cost of increased artifacts.

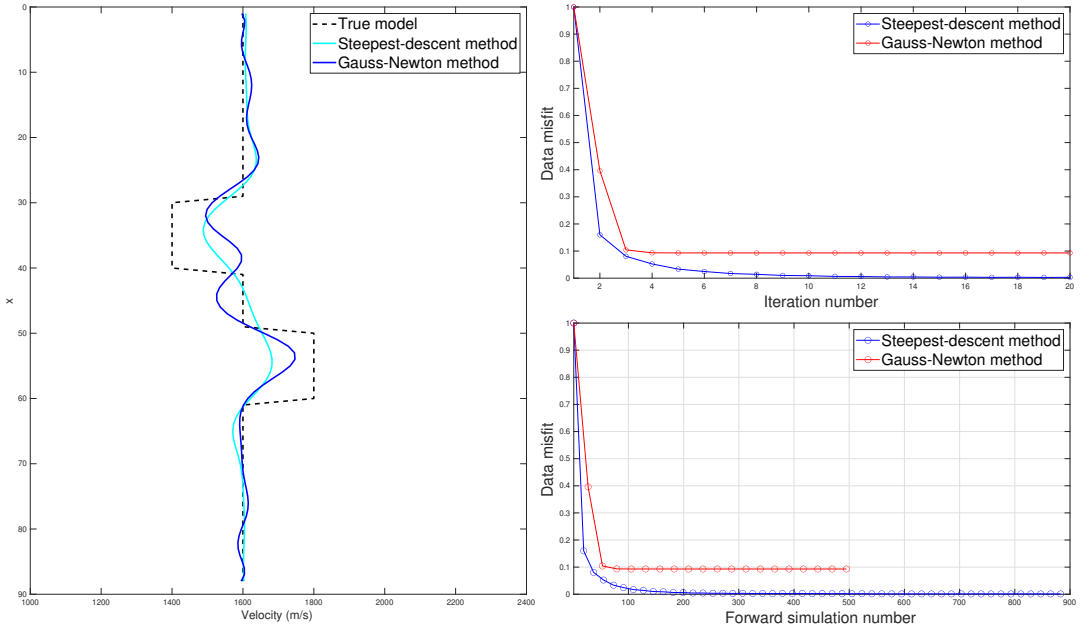


Figure 5.15: *Comparison of the horizontal velocity profiles (left) and of the data misfit (right) in different source case. The data misfit over iteration numbers shown in the upper right and over number of the forward simulations in the lower right.*

5.2.3 Discussion

The GN method achieves accurate model reconstructions with significantly fewer iterations and sources compared to the SD method. Specifically, the GN method converges in just 20 iterations, while the SD method requires 50 iterations to reach comparable results. In different source cases, the SD method needs nine sources to produce results similar to those obtained with only three or five sources in the GN method. Despite the presence of more artifacts in the transmission case compared to the reflection case, the GN method still demonstrates a clear advantage in terms of reconstruction quality. This indicates that the GN method delivers higher accuracy and is better suited to capturing subsurface features.

In terms of data misfit over iterations, the GN method exhibits a faster convergence rate, achieving substantial reductions in data misfit within the initial iterations. When considering the total number of forward simulations, the GN method's gradient calculation is based on $(NS + NR)$, while the SD method is based on $2NS$. As a result, the GN method requires more forward simulations than the SD method. However, in order to obtain similar inversion results in two methods, the SD method typically uses more sources and iteration numbers, leading to a significantly higher number of forward simulations compared to the GN method.

5.3 Further Gauss-Newton test

Finally, we conduct additional tests for the GN method, focusing on the influence of receiver configurations on the inversion results. This is because the gradient calculation in the GN method depends not only on the number of sources but also on the number of receivers. To evaluate this effect, we compare the inversion results using the same simple true model as in the multi-source case, but with varying numbers of receivers: 3, 6, 9, 11, 21, and 41.

5.3.1 Receiver test

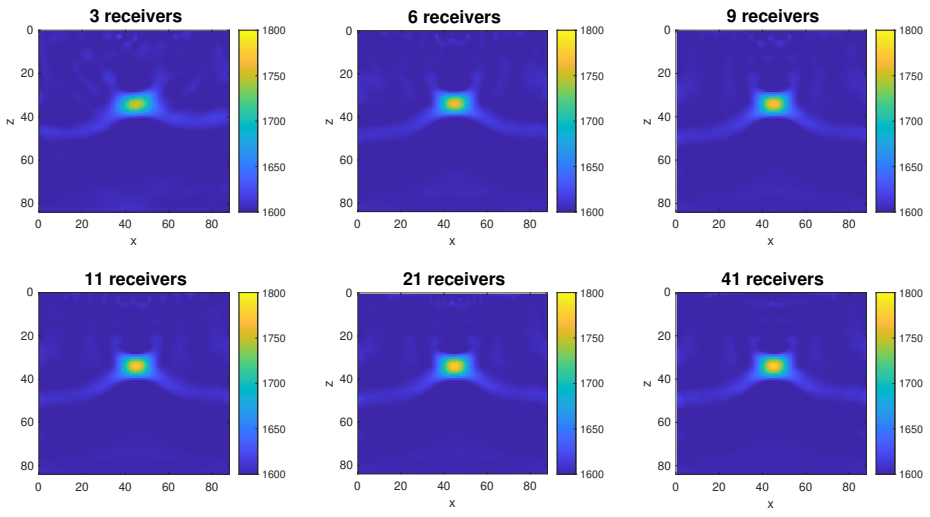


Figure 5.16: *Comparison of model updates using three sources and varying numbers of receivers.*

Figure 5.16 displays reconstructions after 20 iterations, using three sources and varying numbers of receivers. The true model contains a single high-velocity perturbation at the center shown in figure 5.1. Figure 5.17 further illustrates the velocity profiles at the model's

center, providing a detailed comparison of the reconstructions.

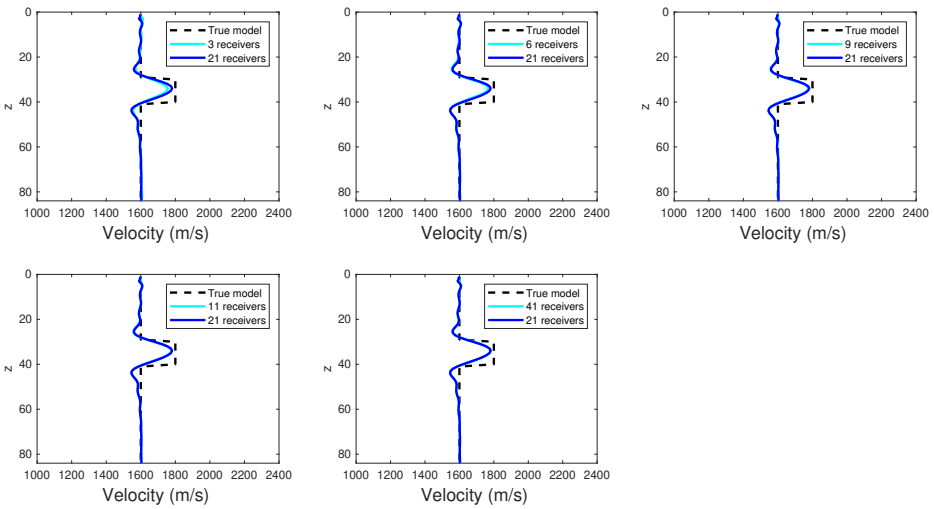


Figure 5.17: *Comparison of velocity profiles for different receiver configurations.*

Both the model updates and velocity profiles show that increasing the number of receivers enhances the accuracy and resolution of the reconstructed model. For configurations with fewer receivers (e.g., three and six), the reconstruction has low resolution and inaccurately represents the high-velocity anomaly. With nine receivers, the reconstruction improves significantly, capturing the anomaly's shape and position more accurately, though it still falls short of the resolution achieved with 21 receivers, as shown in the velocity profiles. Increasing the number of receivers to 41 yields only a slight resolution improvement over the 21-receiver case. Thus, using 11 or 21 receivers strikes a good balance between accuracy and computational efficiency, providing a well-resolved and accurate reconstruction of the high-velocity perturbation.

5.3.2 Discussion

Receiver configuration is crucial in the GN method for obtaining accurate inversion results, as the gradient calculation depends not only on the number of sources but also on the number of receivers. As shown in the receiver tests, increasing the number of receivers enhances the resolution and accuracy of the reconstructed model. However, a balance between accuracy and computational cost is necessary. In the simple model test, 21 receivers provided an optimal trade-off, delivering sufficient accuracy and resolution without unnecessary computational overhead.

Chapter 6

Conclusion

This study has successfully achieved the objectives by analyzing and comparing the performance of the SD method and the GN method for FWI. Using a 2D finite-difference forward simulation, we explored acoustic wave propagation in a medium with constant density to investigate the higher resolution and faster convergence rates of the GN method in inversion results compared to the SD method.

The GN method requires the explicit computation of the Jacobian matrix, followed by the derivation of the approximate Hessian matrix and gradient calculation. To address the computational burden, we utilized virtual sources with the reciprocity principle to obtain the scaled gradient. While the SD method avoids extensive matrix calculations, it offers limited resolution.

Through synthetic examples, the GN method consistently demonstrated superior performance over the SD method in both reflection and transmission scenarios. It achieved higher resolution with fewer iterations for the GN method (20) compared to the SD method (50). Moreover, the GN method exhibited a significantly faster convergence rate, reducing data misfit substantially during initial iterations. However, this advantage involves increased computational costs, particularly influenced by the number of forward simulations per iteration. For the GN method, this number is proportional to both the number of sources and receivers, whereas for the SD method, it depends only on the sources. However, to achieve similar inversion results, the SD method typically requires a greater number of sources and iterations, resulting in a higher total number of forward simulations than the GN method. A receiver configuration test identified that 11 or 21 receivers achieved a good trade-off in this synthetic case, offering adequate accuracy and resolution, while additional receivers yielded diminishing returns.

In conclusion, while the gradient method is commonly employed in FWI, the GN method showed superior performance by providing higher resolution and faster convergence. However, the significant computational memory and processing demands associated with the computation of the Jacobian and approximate Hessian limit its primary advantage to acoustic media and simple synthetic models. Future research should consider extending the GN method to more complex and realistic media, such as elastic and viscoelastic models, to expand its applicability and impact in practical scenarios.

Bibliography

Aki, K. and Richards, P. G. (2002). *Quantitative seismology*.

Berenger, J.-P. (1994). "A perfectly matched layer for the absorption of electromagnetic waves". *Journal of computational physics* 114.2, pp. 185–200.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). "Two-dimensional nonlinear inversion of seismic waveforms: Numerical results". *geophysics* 51.7, pp. 1387–1403.

Igel, H. (2017). *Computational seismology: a practical introduction*. Oxford University Press.

Kolb, P., Collino, F., and Lailly, P. (1986). "Pre-stack inversion of a 1-D medium". *Proceedings of the IEEE* 74.3, pp. 498–508.

Lailly, P. (1983). "The seismic inverse problem as a sequence of before stack migrations". In: *Conference on inverse scattering: theory and application*. Vol. 1983. Philadelphia, Pa, pp. 206–220.

Lay, T. and Wallace, T. C. (1995). *Modern global seismology*. Elsevier.

Levenberg, K. (1944). "A method for the solution of certain non-linear problems in least squares". *Quarterly of applied mathematics* 2.2, pp. 164–168.

Ma, Y., Yu, J., and Wang, Y. (2014). "A novel unsplit perfectly matched layer for the second-order acoustic wave equation". *Ultrasonics* 54.6, pp. 1568–1574.

Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters". *Journal of the society for Industrial and Applied Mathematics* 11.2, pp. 431–441.

McMechan, G. A. and Fuis, G. S. (1987). "Ray equation migration of wide-angle reflections from southern Alaska". *Journal of Geophysical Research: Solid Earth* 92.B1, pp. 407–420.

Mora, P. (1987). "Nonlinear two-dimensional elastic inversion of multioffset seismic data". *Geophysics* 52.9, pp. 1211–1228.

Nocedal, J. and Wright, S. J. (2006). "Large-scale unconstrained optimization". *Numerical optimization*, pp. 164–192.

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J.-Y., Giraud, L., and Ali, H. B. H. (2007). "3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study". *Geophysics* 72.5, SM195–SM211.

Pica, A., Diet, J., and Tarantola, A. (1990). "Nonlinear inversion of seismic reflection data in a laterally invariant medium". *Geophysics* 55.3, pp. 284–292.

Pontius, M. (2016). "Joint acoustic full-waveform and gravity inversion-development and synthetic application to a salt dome". PhD thesis. Karlsruher Institut für Technologie (KIT).

Pratt, R. G., Shin, C., and Hick, G. (1998). "Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion". *Geophysical journal international* 133.2, pp. 341–362.

Rodi, W. L. (1976). "A technique for improving the accuracy of finite element solutions for magnetotelluric data". *Geophysical Journal International* 44.2, pp. 483–506.

Sahin, M. A., Ali, M., Park, J., and Destgeer, G. (2023). "Fundamentals of Acoustic Wave Generation and Propagation". *Acoustic Technologies in Biology and Medicine*, pp. 1–36.

Santosa, F. and Symes, W. W. (1986). "Linear inversion of band-limited reflection seismograms". *SIAM journal on scientific and statistical computing* 7.4, pp. 1307–1330.

Sheen, D.-H., Tuncay, K., Baag, C.-E., and Ortoleva, P. J. (2006). "Time domain Gauss—Newton seismic waveform inversion in elastic media". *Geophysical Journal International* 167.3, pp. 1373–1384.

Sheen, D.-H., Tuncay, K., Ortoleva, P. J., and Baag, C.-E. (2004). "Efficient finite difference calculation of partial derivative seismic wavefield using reciprocity and convolution". In: *SEG International Exposition and Annual Meeting*. SEG, SEG–2004.

Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., Chung, S., and Shin, S. (2001). "Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion". *Geophysics* 66.6, pp. 1856–1863.

Tarantola, A. (1984). "Inversion of seismic reflection data in the acoustic approximation". *Geophysics* 49.8, pp. 1259–1266. DOI: 10.1190/1.1441754.

Virieux, J. (1986). "P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method". *Geophysics* 51.4, pp. 889–901.

List of Figures

Acknowledgments

I would like to first thank Prof. Dr. Thomas Bohlen for providing a fascinating topic. The monthly meetings, filled with insightful advice and constructive feedback, helped me overcome challenges and improve the quality of my work. Also thank Prof. Dr. Joachim Ritter for kindly agreeing to be my co-supervisor. Then so many thanks to Sarah and Arash for their patient guidance and support, taking so much time to proofread my thesis. I truly appreciate that I could always approach them directly at their office with any questions. Finally, I would like to extend my thanks to all the professors and tutors at GPI, especially Dr. Thomas Hertweck, Dr. Lars Houpt, Sonia, Daniel, Fiona, Marina, as well as my tutor, Annika, for their invaluable support throughout my Master's studies. Without their guidance and assistance, this journey would have been incredibly difficult for me.

Lastly, I would like to mention my family and friends, especially Yumeng, who have been my best companions not only throughout my Master's journey but also for the rest of my life.