UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF GEOSCIENCES

MASTER THESIS IN GEOPHYSICS FOR NATURAL RISKS AND RESOURCES

A Study on the Selection of Seismic Anisotropic Parameterizations in Linearized Inversion Schemes Used in Seismic Tomography

MASTER CANDIDATE

Jiahua Zhao

Student ID 2005008

SUPERVISOR

Prof. Manuele Faccenda

University of Padova

CO-SUPERVISOR

Dott. Brandon VanderBeek

University of Padova

Abstract

In recent years, researchers have focused on seismic anisotropy and developed related seismic tomography methods. One common method is inverting P-wave travel times, which provides unique constraints on the Earth's interior elastic properties. To characterize anisotropy, 21 independent parameters need to be constrained. Assuming Earth has hexagonal symmetry simplifies the problem, reducing free parameters. In this case, P-wave velocity variations are described by mean slowness, anisotropy strength, and orientation. Three representations have been proposed for these anisotropic parameters for the purpose of tomography imaging: spherical, vectoral and the ABC (a modified vectoral parameterization). Although each is equivalent in describing hexagonal anisotropy, they are nonlinear, making it unclear which is most suitable for iterative inversion schemes in seismic tomography. Therefore, we designed numerical experiments to study how different parameterizations affect the performance of linearized inversion schemes in seismic tomography, providing suggestions for parameterization selection. In this paper, we study the stability and convergence characteristics of all three parameterizations in P-wave travel time linearized inversion, and evaluate the number of iterations and errors of each parameterization in solving the problem. For the linearized inversion scheme, we explore three common methods for minimizing the objective function: gradient descent, Newton, and Levenberg-Marquardt, and select the most suitable solver based on experiments. Our numerical experiments start with a simplified case of constraining two anisotropic parameters, then extend to a full anisotropic problem with four constrained parameters. Finally, we summarize the stability and performance of various anisotropic models and initial conditions under each parameterization. For the 2-D anisotropy numerical experiments, the results show that the ABC parameterization has excellent stability and performance, with the Levenberg-Marquardt method being the best solver method; for the full anisotropy numerical experiments, the combination of ABC and spherical parameterizations becomes the best parameterization scheme choice when using the Levenberg-Marquardt method with further constrained solver parameters.

Sommario

In recent years, researchers have focused on seismic anisotropy and developed related seismic tomography methods. One common method is inverting P-wave travel times, which provides unique constraints on the Earth's interior elastic properties. To characterize anisotropy, 21 independent parameters need to be constrained. Assuming Earth has hexagonal symmetry simplifies the problem, reducing free parameters. In this case, P-wave velocity variations are described by mean slowness, anisotropy strength, and orientation. Three representations have been proposed for these anisotropic parameters for the purpose of tomography imaging: spherical, vectoral and the ABC (a modified vectoral parameterization). Although each is equivalent in describing hexagonal anisotropy, they are nonlinear, making it unclear which is most suitable for iterative inversion schemes in seismic tomography. Therefore, we designed numerical experiments to study how different parameterizations affect the performance of linearized inversion schemes in seismic tomography, providing suggestions for parameterization selection. In this paper, we study the stability and convergence characteristics of all three parameterizations in P-wave travel time linearized inversion, and evaluate the number of iterations and errors of each parameterization in solving the problem. For the linearized inversion scheme, we explore three common methods for minimizing the objective function: gradient descent, Newton, and Levenberg-Marquardt, and select the most suitable solver based on experiments. Our numerical experiments start with a simplified case of constraining two anisotropic parameters, then extend to a full anisotropic problem with four constrained parameters. Finally, we summarize the stability and performance of various anisotropic models and initial conditions under each parameterization. For the 2-D anisotropy numerical experiments, the results show that the ABC parameterization has excellent stability and performance, with the Levenberg-Marquardt method being the best solver method; for the full anisotropy numerical experiments, the combination of ABC and spherical parameterizations becomes the best parameterization scheme choice when using the Levenberg-Marquardt method with further constrained solver parameters.

Contents

List of Figures

List of Tables

1

Introduction

1.1 SEISMIC TOMOGRAPHY

Seismic tomography, as a geophysics technique, analyzes the seismic waves generated by natural earthquakes or artificial earthquakes (such as explosions or other artificial vibration sources) to obtain information on the velocity structure or other properties of the underground medium using inversion methods [1]. Since the 1970s, inspired by medical CT technology, seismic tomography has been used for the semi-quantitative study of crust and upper mantle structures, opening up new avenues in the field of geophysical research. In 1974, Aki et al. imaged the 3-D velocity anomalies of the crust and upper mantle at San Andrés Fault using teleseismic P-waves arrival times [2]. Subsequently, regional-scale tomography by Aki and Lee [3] was considered the starting point of modern seismic tomography [4], and Dziewonski [5] [6] extended the application of seismic tomography to a global scale.

In the past few decades, with the continuous increase in the number of global seismic stations, the application of digital seismic records, and the rapid development of computer technology, seismic tomography has been widely used in various fields and has achieved many important research results [7]. At the same time, the theoretical and technical methods of seismic tomography have been continuously innovated and perfected in the application process, and many new methods and technologies have emerged, such as attenuation tomography [8], finite-frequency tomography [9], ambient noise tomography [10], and the

anisotropic tomography used in this paper [11] used in this paper, etc. These new methods and technologies provide new directions for the future development of seismic tomography.

1.2 STUDY ON SEISMIC ANISOTROPY

Anisotropy is the characteristic that some or all properties of a material change with the direction. In geophysics, seismic wave anisotropy refers to the phenomenon that the velocity and particle polarization motion of seismic waves are related to the propagation direction when they propagate in an elastic medium. Seismic wave anisotropy is ubiquitous in the Earth's interior, and it is related to factors such as the composition, state, structure, and strain of materials in the Earth's deep interior [12] [13]. Therefore, seismic wave anisotropy has become an important means to study the composition, structure, and tectonics of the Earth's interior materials, as well as geodynamic processes, and has received widespread attention in recent years.

1.2.1 HISTORY AND CURRENT SITUATION OF ANISOTROPY RESEARCH

In the 19th century, some scientists began to notice that the propagation of elastic waves may have anisotropy. In 1838, Green first proposed the concept of elastic energy in his article on the propagation of elastic waves, and supported the view of 21 elastic constants; in 1856, Kelvin elaborated on the tensor form of elasticity that is independent of the coordinate system in his published article, and described the propagation of elastic waves in anisotropic media using formulas. Towards the end of the 19th century, Maurycy Pius Rudzki introduced anisotropy into the field of seismology, and conducted innovative research on seismic wave anisotropy, laying the foundation for modern seismic wave anisotropy research [14].

In the 1960s, with the study of the incompatibility of Rayleigh and Love waves (vertical anisotropy) [15], and the discovery of the variation of Pn wave velocity in the oceanic mantle with azimuth (azimuthal anisotropy) [16], the research on seismic wave anisotropy began to attract attention. Since then, the surface waves in the oceanic mantle [17] and the continental mantle [18], as well as the body waves in the continental lithosphere [19] have been found to exhibit azimuthal

anisotropy, and seismic wave anisotropy research has gradually become a research hotspot. In the 1980s, people mainly focused on the observational study of body wave anisotropy and conducted a large number of experimental studies on mineral physics and rock deformation mechanisms.

In the 21st century, in addition to measuring anisotropy in different regions and at different depths using different methods, there is a tendency to comprehensively apply various anisotropy measurement methods and constrain the presence, cause, and interpretation of anisotropy [20] [21]. After decades of efforts, rich results have been achieved in the study of seismic wave anisotropy using various means, and the main features of anisotropy in the Earth's interior and its relationship with the Earth's internal structure and dynamics processes have been revealed [22] [23] [24], greatly enhancing human understanding of the Earth's internal structure and dynamics processes.

1.2.2 THE ORIGIN OF ANISOTROPY IN THE EARTH'S INTERIOR

Seismic wave anisotropy is ubiquitous in multiple layers within the Earth's interior, such as the crust, upper mantle, mantle transition zone, D″ layer, and inner core, etc [22] [24]. The formation mechanism of seismic wave anisotropy in the Earth's interior is generally considered to be the result of the combined action of shape-preferred orientation (SPO) and lattice-preferred orientation (LPO).

The SPO refers to the dominant geometric shape distribution orientation within isotropic media. The existence of this orientation leads to different wave impedances in different directions, thereby producing seismic wave anisotropy. Normally, the velocity of elastic waves parallel to the bedding direction is higher than that perpendicular to the bedding direction; the velocity parallel to the direction of oriented fractures, faults, and lens axes is higher than the velocity perpendicular to the axis direction. Within the lithosphere, especially in the upper crust, anisotropy is mainly caused by SPO. Crampin et al. [25] believe that the formation of crustal anisotropy is mainly due to the orientation of fluid-containing fractures under stress. In a compressive environment, the fast wave direction of anisotropy is usually consistent with the orientation of major structural boundaries such as mountain belts and faults [26]; in a tensile environment, the fast wave direction is usually consistent with the extension direction [27].

Part of the cause of mantle anisotropy may be related to the cracks, lenses, or compositional lamellae filled with melt in the mantle [28] [29]. Mineral single crystals have different seismic wave velocities on different crystal axes, among which the fastest wave velocity axis is called the fast axis. When the fast axes of crystal particles are randomly distributed, the effects of fast axis wave velocity in different directions cancel each other out, and there is no overall anisotropy. However, when mineral crystals or aggregates are subjected to stress, the fast axis orientation is arranged in a certain pattern of non-randomness, forming a macroscopic fast axis orientation consistent with the relative distribution direction, which causes seismic wave anisotropy. This non-random distribution of fast axis orientation is called the LPO.

In the lower crust and upper mantle, the oriented arrangement of mineral crystals such as olivine, pyroxene, amphibole and mica can cause anisotropy. Among them, olivine, which is the main mineral component of the upper mantle, has strong seismic wave anisotropy [30], and the LPO of olivine is considered the main source of anisotropy in the upper mantle [31] [32]. Plate movement and mantle flow within a certain range of the subduction zone will cause the oriented arrangement of mineral crystals such as olivine, so the fast axis direction of the crystals is usually consistent with the direction of plate movement or mantle flow [33].

1.3 MAIN RESEARCH METHODS OF SEISMIC ANISOTROPY

Currently, the research on seismic wave anisotropy mainly focuses on the measurement of wave velocities in minerals and rocks (high temperature and high pressure) and seismology, among which the main seismological methods include shear wave splitting [34], Pn wave anisotropy, surface wave anisotropy, and P-wave anisotropy (tomography method) [35] involved in this study. Shear wave splitting refers to the phenomenon that when an S wave enters an anisotropic medium, it splits into two waves with mutually perpendicular polarization directions and different wave velocities. By extracting the polarization direction ϕ of the faster fast wave and the time difference δt between the fast and slow waves, the fast wave direction and amplitude of anisotropy in the medium can be obtained. The method of shear wave splitting is widely used in the study of anisotropy in the Earth's interior and has achieved a large number of research

results [27] [33] [36].

Both Pn waves and surface waves exhibit directional anisotropy [16] [17] [37].The method of Pn wave anisotropy research [38] divides the propagation path of Pn waves into source path section, mantle path section, and receiver path section, and divides the top of the mantle into multiple 2-D grids. Assuming that the waves are isotropic in the source path section and receiver path section, and only show directional anisotropy in the mantle path section, the ray path residual is calculated using the anisotropic wave velocity related to the ray azimuth, and then the azimuthal anisotropy of Pn wave in the top of the mantle is inverted. The method of studying the directional anisotropy of surface waves [39] is similar to that of Pn waves, except that the surface wave propagation path is frequency-dependent. Therefore, a two-step inversion method is usually adopted. First, mixed path dispersion measurement is performed to extract the ray path, and then the pure path dispersion measurement is used with the Rayleigh wave group velocity related to the ray azimuth. The anisotropic group velocity is introduced into the inversion computation to solve the directional anisotropy of the surface wave.

The method of P-wave anisotropy is based on P-wave tomography, in which the anisotropy parameters of P-wave azimuth and isotropic wave velocity anomalies are used as inversion parameters and solved simultaneously to obtain the 3-D distribution of azimuthal anisotropy and isotropic wave velocity anomalies in the study area. In recent decades, the method of P-wave anisotropy tomography has been continuously developing. Eberhart-Phillips and Henderson [11] used the P-wave anisotropy tomography method to study the Marlborough subduction zone in New Zealand, and used the results of shear wave splitting in the study as the initial values for P-wave anisotropy tomography inversion; Koulakov et al. [40] tried to use four anisotropy parameters to determine the fast wave azimuth of P-waves in 3-D space in their study of anisotropy tomography in central Java; Wang and Zhao [41] modified the TOMOG3D tomography method [42] to solve the P-wave azimuthal anisotropy and isotropic wave velocity anomalies in the 3-D model space simultaneously during tomography inversion, and further improved the method to be able to solve the radial anisotropy of P-waves. Eventually, the method was applied to the study of P-wave anisotropy tomography in northeastern Japan and Kyushu areas [43]. Overall, the method

of P-wave anisotropy tomography is increasingly being used by more and more researchers and provides critical information for exploring Earth's internal dynamics. This paper will focus on the P-wave anisotropy tomography method and discuss the influence of different seismic anisotropic parameterization on inversion results.

1.4 THE PARAMETERIZATION OF SEISMIC ANISOTROPY

In the P-wave anisotropy tomography method, if a seismic travel-time data set is given, we hope to constrain the average slowness, the strength and orientation of anisotropy. To achieve this goal, we need to parameterize these quantities appropriately so that they can be constrained in the linearized inversion scheme. Currently, three parameterization methods have been proposed, namely: (1) spherical parameterization, (2) vectoral parameterization, and (3) ABC (an improved vectoral parameterization).

In the study of teleseismic P-wave tomography to simulate anisotropic upper mantle structure by Munzarova et al. [44], the key part is deriving a clear formula to describe the P-wave velocity in weakly anisotropic media with hexagonal symmetry axes, which are usually oriented in three dimensions. They describe the derivation process of anisotropic velocity in hexagonal symmetry in detail and consider spherical parameterization. Wang and Zhao [45] also discuss the expression of P-wave anisotropic velocity in anisotropic media with inclined hexagonal symmetry axes in their research, but they adopt vectoral parameterization. VanderBeek and Faccenda [46] improved the vectoral parameterization and proposed the ABC parameterization.

In the second chapter, we will describe the above three parameterization methods in detail.

1.5 RESEARCH WORK OF THE THESIS

This paper will focus on studying the stability and convergence characteristics of three different parameterization methods in the linear inversion of P-wave travel-time of anisotropic parameters, and assessing the number of iterations

and error quantity of each parameterization in solving. For the linearized inversion scheme, we will explore three common methods for minimizing the objective function: gradient descent, Newton, and Levenberg-Marquardt, and choose the most suitable solver based on experimental results.

The main application of this paper includes a series of numerical experiments, aimed at studying how different seismic anisotropy parameterizations affect the performance of linearized inversion schemes in seismic tomography. We will first consider the simple case of constraining two anisotropy parameters, which define a uniformly anisotropic volume sampled by straight ray paths. Subsequently, we will extend the experiments to constrain four anisotropy parameters in the fully anisotropic problem and consider two ray path geometries to obtain more quantitative results. Finally, we will summarize the stability and performance of each parameterization under various anisotropy models and initial conditions, and provide recommendations for parameterization selection.

2

Theory and Methodology

2.1 PRINCIPLES AND METHODS OF SEISMIC TOMOGRAPHY

Seismic tomography can be divided into body-wave tomography and surface-wave tomography according to the different seismic phases. This paper mainly focuses on body-wave tomography. Since the 1970s, body-wave tomography has been one of the most widely used tomography methods. According to ray theory [47], the frequency of seismic waves is approximately infinite, so seismic waves can be considered as seismic rays. These rays are emitted from the source, and the travel-time to the station is determined by the length of the ray path and the velocity on the path.

We will parameterize the model space of the study region as a grid, and use the velocity or slowness perturbations of the grid points relative to the 1-D initial velocity model as the unknown variables in the inversion. Given the source and station positions, ray tracing can be used to determine the ray path length from the source to the station and the grid points the ray passes through to reach the station. The velocity along the ray path can be obtained by the velocity perturbations of the grid points the ray passes through and the 1-D initial velocity, thereby calculating the theoretical travel-time of this seismic ray. Subtracting the actual travel-time of the earthquake recorded at the station from the theoretical travel-time, the trend residual of this seismic ray can be obtained. Obviously, the trend residual is determined by the velocity perturbations of each grid point along the ray path, and thus a linear equation can be established. When there

are multiple earthquakes to multiple stations' rays passing through the model space, the linear equation can be solved simultaneously, and the equation can be represented in the following form:

$$d = Gp \qquad (2.1)$$

where d is the ray travel-time residual vector, G is the matrix of travel-time partial derivatives with respect to the model parameters, and p is the perturbations of the model parameter m relative to the 1-D velocity model. By solving the above linear equation system, we can obtain the velocity perturbations vector p at each location in the model space, and thus obtain the velocity anomaly distribution in the study region.

Next, we will introduce the body-wave tomography method from three aspects: model parameterization, ray tracing, and solution of linear equation systems.

2.1.1 MODEL PARAMETERIZATION

In order to image the velocity structure of the Earth's interior, we need to parameterize the velocity structure within the Earth as an inverse model, typically using the blocks or grids method for model parameterization. Both the blocks or grids method divides the model space into small blocks or grids to simulate fine structural anomalies and generate sparse coefficient matrices for easy inverse calculations.
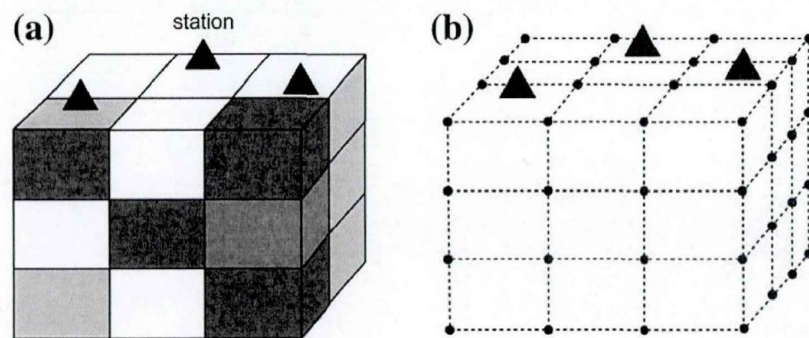
Figure 2.1: Different methods of model parameterization: (a) blocks method; (b) grids method.

The blocking method [3] divides the model space into discrete blocks (Fig.

2.1 a), forming a discrete velocity model. It is assumed that the medium within each block is uniform, and the velocity (slowness) of each point within the block is equal, i.e., the number of blocks divided is equal to the number of parameters to be determined. The grids method [48] parameterizes the model space into multiple discrete grid points (Fig. 2.1 b), each of which has a certain velocity. The velocity of any point within the model space can be obtained by linear interpolation of its surrounding 8 grid points.

2.1.2 RAY TRACING

After model parameterization, we need to determine the ray path through ray tracing, which is actually a forward problem. Since the final linear equation set is determined by the rays, the accuracy of ray tracing has a significant impact on the quality of tomography. Currently commonly used ray tracing methods include the shooting method [49], the bending method [47] and the pseudo-bending method [50], etc.

The shooting method starts from the source point, and traces the rays to the station direction with a certain offset angle and azimuth. By calculating the residual of the contrast travel-time, the offset angle and azimuth are corrected. The bending method fixes the station and source as the two endpoints of the ray, and gradually bends the ray according to the ray equation until the minimum travel-time path is found. In contrast, the pseudo-bending method disturbs the initial ray path using the geometric expression of the ray equation, and then calculate the minimum travel-time in segments. According to the Fermat principle, when the travel-time is minimized, the ray path is determined using the obtained solution. This method can accurately search for the ray path under 3-D models, regardless of the distance to the seismic center or the length of the ray.

2.1.3 SOLUTION OF LINEAR EQUATION SYSTEMS

After the model parameterization and ray tracing are completed, the linear system of equations (Eq. 2.1) is also established, and the Eq. 2.1 is solved to obtain the model parameter perturbation matrix p, which also leads to the tomoraphy results. Therefore, the tomography problem can ultimately be reduced to solving a linear equation set. In the field of geophysics, the linear equation set is usually a large sparse matrix, and the solution of such matrices is typically

carried out using algorithms such as ART (Algebraic Reconstruction Technique [51]), SIRT (Simultaneous Iterative Reconstruction Technique [52]) and LSQR [53]. Among them, the most widely used algorithm is LSQR, which is favored for its fast convergence speed and effective suppression of data error propagation.

As with many geophysical inversion problems, there is a need to find a balance between the roughness of the model and the misfit of the solution when solving a linear system of equations, usually by plotting the L-curve [54] to find a balance between the roughness and the misfit (Fig. 2.2).
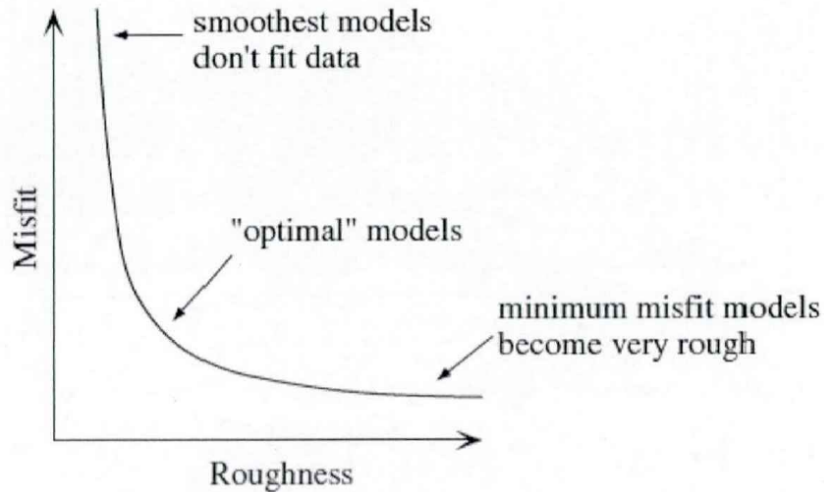
Figure 2.2: The L-curve in geophysical inversion problems.

2.2 PRINCIPLES OF P-WAVE ANISOTROPY TOMOGRAPHY

Through the introduction above, we can understand that P-wave anisotropy can be combined with body-wave tomography. The P-wave anisotropy tomography method [41] no longer regards the velocity at each point in the model space as isotropic (only setting one wave velocity as the inversion parameter), but introduces anisotropy parameters. The isotropic wave velocity and anisotropy parameters at each point in the model space are used as inversion parameters, and the calculation of travel-time residuals using anisotropic wave velocity is adopted to simultaneously solve the isotropic wave velocity anomalies and anisotropy at each point in the model space, thus realizing anisotropy tomography.

According to the research by Barclay et al. [55], the slowness of the P-wave in the 3-D ray path can be expressed as:

$$S = S_0 + Mcos(2\theta) \tag{2.2}$$

where S represents the anisotropic slowness, S_0 represents the isotropic slowness, M represents the azimuthal anisotropy strength, and θ is the angle between the 3-D P-wave ray path propagation vector and the isotropic symmetry axis. For a unit P-wave propagation vector V with ray path azimuth ϕ and angle of incidence i and a unit symmetry axis vector I with symmetry axis in the horizontal plane and azimuth ψ (Fig. 2.3):
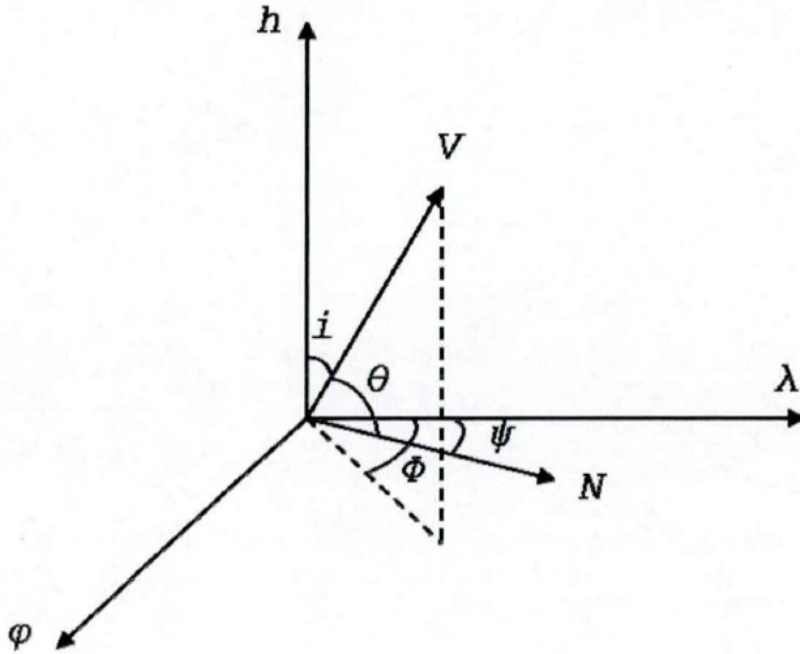
Figure 2.3: The figure of the relationship between the P-wave propagation vector and the azimuthal anisotropy symmetry axis.

They can each be expressed as:

$$V = (sinisin\phi, sinicos\phi, cosi) \tag{2.3}$$

$$I = (sin\psi, cos\psi, 0) \tag{2.4}$$

It is obvious that $cos\theta$ is the dot product of the unit vector V and I, and therefore can be obtained as follows:

$$cos\theta = sini(sin\phi sin\psi + cos\phi cos\psi) \tag{2.5}$$

Substituting Eq. 2.5 into Eq. 2.2, we can obtain:

$$S = S_0 - Mcos^2i + Msin^2i(cos2\psi cos2\phi + sin2\psi sin2\phi) \tag{2.6}$$

Setting $A = \sqrt{Mcos2\psi}$ and $B = \sqrt{Msin2\psi}$, then Eq. 2.6 can be written as:

$$S = S_0 - cos^2iM + sin^2i(Acos2\phi + Bsin2\phi) \tag{2.7}$$

Then let $M' = M/S_0, A' = A/S_0, B' = B/S_0$, Eq. 2.7 can also be expressed as:

$$V = \frac{V_0}{1 - cos^2iM' + sin^2i(A'cos2\psi + B'sin2\psi)} \tag{2.8}$$

where V represents the anisotropic velocity,, V_0 represents the isotropic velocity, A' and B' are called anisotropic parameters. Consider a small segment of the ray path with a length of d, and the average anisotropic velocity V at the midpoint is the average velocity of this small segment. Therefore, the travel-time T of this segment of the ray path can be obtained:

$$T = \frac{d}{V} = \frac{d[1 - cos^2iM' + sin^2i(A'cos2\phi + B'sin2\phi)]}{V_0} \tag{2.9}$$

In any point (φ, λ, h) within the model space, the isotropic velocity V_0 and the two anisotropic parameters A' and B' can be obtained by weighted linear interpolation of the corresponding values of the surrounding eight model grid points [48]:

$$f(\varphi, \lambda, h) = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} w_{ijk}(\varphi, \lambda, h)f(\varphi_i, \lambda_j, h_k) \tag{2.10}$$

$$w_{ijk}(\varphi, \lambda, h) = (1 - \|\frac{\varphi - \varphi_i}{\varphi_2 - \varphi_1}\|)(1 - \|\frac{\lambda - \lambda_j}{\lambda_2 - \lambda_1}\|)(1 - \|\frac{h - h_k}{h_2 - h_1}\|) \tag{2.11}$$

where $\varphi_i, \lambda_j, h_k$ represent the latitude, longitude and depth of the eight model grids points surrounding any point (φ, λ, h) in the model space, w_{ijk} represents the weighting coefficients of these eight grids points, and $f(\varphi, \lambda, h)$ and $f(\varphi_i, \lambda_j, h_k)$ respectively represent the difference value at point (φ, λ, h) and

the values at the eight surrounding model grid points.

Then, for the ith earthquake recorded by the jth station, the residual travel-time r_{ij} along ray path L_{ij} can be expressed as:

$$r_{ij} = \frac{\partial T_{ij}}{\partial \varphi_i}\Delta\varphi_i + \frac{\partial T_{ij}}{\partial \lambda_i}\Delta\lambda_i + \frac{\partial T_{ij}}{\partial h_i}\Delta_i + \Delta T_{0i} + \sum_{k=1}^{N}(\frac{\partial T_m}{\partial V_{0k}}\Delta V_{0k} + \frac{\partial T_m}{\partial A'_k}\Delta A'_k + \frac{\partial T_m}{\partial B'_k}\Delta B'_k) + E_{ij}$$
(2.12)

where T_{ij} represents the travel-time along the ray L_{ij}, T_m represents the travel-time of the mth ray path at the kth model grid, φ_i, λ_i, h_i, T_{0i} denote the latitude, longitude, depth and moment of the ith earthquake, respectively, and V_{0k}, A'_k, B'_k respectively represent the isotropic wave velocity and anisotropic parameters at the kth model grid point. The first four terms on the right-hand side of Eq. 2.12 represent the latitude, longitude, depth, and the influence of original moment of the ith earthquake on the travel-time residuals, and their partial derivatives can be obtained from the analytical solution [56]. While E_{ij} is a higher-order perturbation term that can be ignored because its influence on the travel-time residuals is very small. The next three terms represent the perturbation of the isotropic wave velocity V_{0k} at the kth grid in the model space and the effect of the perturbation of the two anisotropic parameters: A'_k and B'_k on the travel-time residuals, respectively. Their partial derivatives are given in Eq. 2.13, 2.14 and 2.15:

$$\frac{\partial T_m}{\partial V_{0k}} = -\frac{d_m w_k[1 - cos^2 i_m M'_k + sin^2 i_m(A'_k cos2\phi_m + B'_k sin2\phi_m)]}{V_{0k}^2}$$
(2.13)

$$\frac{\partial T_m}{\partial A'_k} = \frac{d_m}{V_{0k}}w_k(-cos^2 i_m \frac{A'_k}{M'_k} + sin^2 i_m cos2\phi_m)$$
(2.14)

$$\frac{\partial T_m}{\partial B'_k} = \frac{d_m}{V_{0k}}w_k(-cos^2 i_m \frac{B'_k}{M'_k} + sin^2 i_m sin2\phi_m)$$
(2.15)

Here, w_k represents the weighting coefficient of the kth grid point, d_m, i_m, and ϕ_m respectively represent the length, incidence angle, and azimuth of the mth segment of the ray L_{ij} passing through the kth grid point. The travel-time residuals of multiple earthquakes to multiple stations (Eq. 2.12) form a large linear system of equations that relates the travel-time residuals (r_{ij}) to the source

parameters (φ_i, λ_i, h_i, T_{0i}) and the medium parameters (V_{0k}, A'_k, B'_k). Since this linear system of equations is a large sparse matrix, it can be solved by the LSQR algorithm to obtain the isotropic wave velocity V_{0k} and two anisotropic parameters A'_k and B'_k at each grid point in the model space, and then the anisotropic fast wave azimuth and the anisotropic amplitude at the grids can be obtained from these two anisotropic parameters [11].

2.3 COMMON PARAMETERIZATION OF SEISMIC ANISOTROPY TOMOGRAPHY

From the previous discussion, we can understand that in order to fully describe the anisotropy of the Earth's interior, we need to constrain 21 independent elastic parameters, which is undoubtedly a complex and challenging task. In 2006, Becker et al. [57] studied the impact of several common assumptions regarding the influence of mantle convection and mineral deformation-induced LPO on the numerical prediction of upper mantle seismic anisotropy. They found that most of the anisotropy can be captured by estimating the best-fitting hexagonal symmetric tensor, and the correlation between the hexagonal anisotropy parameters of P-waves and S-waves shows a simple bilinear relationship. If we include this information in the prior information, such relationships can reduce the number of free parameters required for seismic inversion.

In order to simplify the problem and reduce the number of free parameters, we assume that the Earth's interior has hexagonal symmetry. Under this simplification, the directional dependence of P-wave velocity can be well-approximated by a periodic function of 2α and 4α terms, where α is the angle between the hexagonal symmetry axis and the direction of P-wave propagation (ray path) [58]. The 4α term of the mantle fabric is usually an order of magnitude smaller than the 2α oscillation. Under this approximated P-wave model, the travel-time computation (Eq. 2.9) can be modeled according to the equation [46]:

$$t_i = L_i u [1 + F cos(2\alpha_i)]^{-1} \tag{2.16}$$

here, for the ith travel time, L_i is the ray path length; u is the average slowness (i.e. the inverse of the velocity); F is the anisotropic strength (or in subsequent experiments called anisotropic fraction); and α_i is the angle between the ray path

and the orientation of anisotropy. Given a P-wave travel-time dataset, we want to constrain the average slowness, as well as the strength and orientation of the anisotropy. This requires appropriate parameterization of Equation 1 so that we can use a linearized inversion scheme to constrain these quantities. Currently, researchers have proposed three such parameterizations: (1) spherical, (2) vectoral, and (3) the ABC (a modified vectoral parameterization), which we will describe separately below.

2.3.1 SPHERICAL PARAMETERIZATION

In the previous section, we introduced the principle of P-wave anisotropy tomography in 3-D space. For the simplified case (i.e., hexagonal symmetry), Munzarova et al. [44] derived an explicit formula for P-wave velocity in weakly anisotropic media with hexagonal symmetry axes typically oriented in 3-D in 2018, and established a linearized relationship between data (travel-time residuals) and model parameters (spherical coordinates) describing the anisotropic medium. According to Eq. 2.16, we have already known the expression for approximating P-wave velocity in hexagonal symmetry cases. Here, we adopt the formula from Munzarova's paper to represent:

$$v = \bar{v}(1 + \frac{k}{2}cos2\alpha) \tag{2.17}$$

where $\bar{v}$ is the isotropic component of the anisotropic velocity, k is the anisotropic strength, and α is the angle between the symmetry axis and the direction of wave propagation (Fig. 2.4).

In the case of a weakly anisotropic medium with hexagonal symmetry, the cosine function (Eq. 2.17) is not sufficient to approximate the P-wave velocity of the upper mantle. For the anisotropic upper mantle approximated by peridotite (olivine) assemblages, the maximum difference between the P-wave velocity calculated by the approximation equation (Eq. 2.17) and the P-wave velocity evaluated as an exact solution of the Christoffel equation is only 0.5% [59].

In order to convert the angle α into a ray-parametric coordinate system (spherical coordinate system), we define the direction of the symmetry axis with the unit vector s and the direction of wave propagation with the unit vector n (Fig. 2.4) as:

$$s = (sin\theta sin\lambda, sin\theta cos\lambda, cos\theta) \tag{2.18}$$
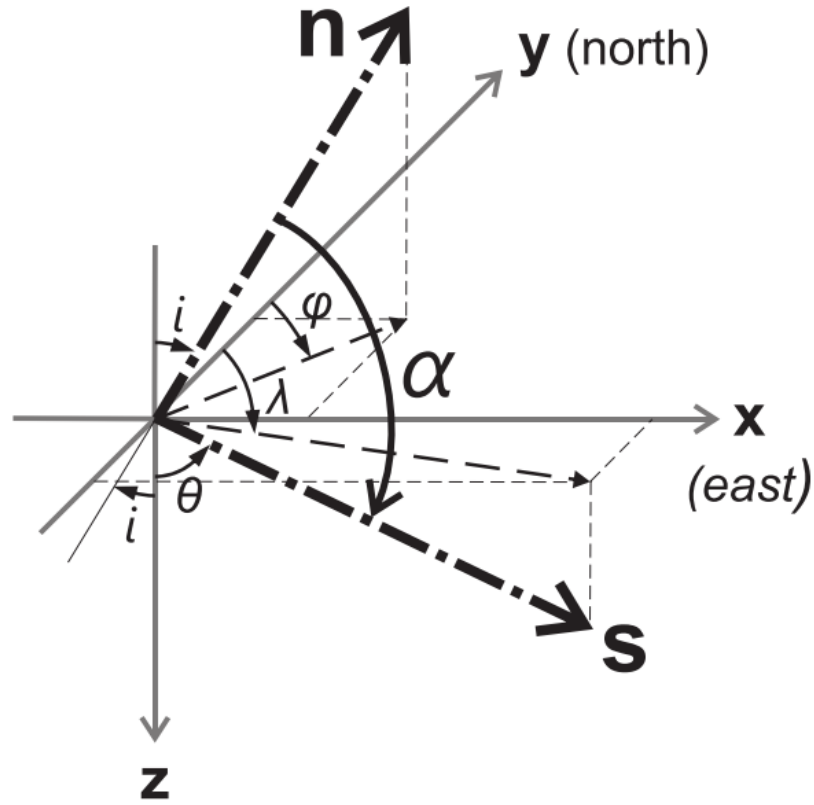
Figure 2.4: Definition of the Cartesian coordinate system used in Munzarova's paper [44]. Angles λ and θ label azimuth and inclination of the hexagonal-symmetry axis (unit vector s), respectively. Wave propagation direction is represented by unit vector n. Angles Φ and i mark backazimuth and angle of the direction of wave propagation, respectively.

where λ and θ are the azimuth and inclination angles of the symmetry axis and

$$n = (-sinisin\phi, -sinicos\phi, -cosi) \tag{2.19}$$

where ϕ and i are the back azimuth and incidence angles of wave propagation. After substituting $cos2\alpha = 2cos2\alpha - 1$ and $cos\alpha = n \cdot s$ in the Eq. 2.17, the P-wave velocity in a weakly anisotropic medium with hexagonal symmetry and 3-D direction can be expressed as:

$$v = \bar{v}\{1 + k[(sinisin\theta cos(\phi - \lambda) + cosicos\theta)^2 - \frac{1}{2}]\} \tag{2.20}$$

Eq. 2.20 relates the anisotropic velocity v of a P-wave propagating in a direction given by the angle ϕ and i to the isotropic component $\bar{v}$ of the anisotropic

velocity and variable velocity perturbation, that is the term with anisotropic strength k. We can combine and simplify Eq. 2.20 with Eq. 2.16, which finally leads to the spherical parameterization expression of this paper:

$$t_i = L_i u[1 + F(2[cos(\theta_i)cos(\gamma)cos(\psi - \phi_i) + sin(\theta_i)sin(\gamma)]^2 - 1)]^{-1} \quad (2.21)$$

In this equation, ϕ_i and θ_i represent the azimuth and elevation of the seismic ray, while ψ and γ denote the anisotropic azimuth and elevation. The Eq. 2.21 is an approximate expression for the P-wave spherical parameterization velocity.

2.3.2 VECTORAL PARAMETERIZATION

In 2021, Wang and Zhao described a new parameterization method [45]: the vectoral parameterization. From Eq. 2.2, we know the slowness expression of the P-wave in the 3-D ray path.

The unit vector along the ray propagation direction can be expressed as $l(sinpsin\lambda, sinpcos\lambda, cosp)$ (Fig. 2.5 a), while the unit vector along the hexagonal symmetry axis can be expressed as $n(sinqsin\gamma, sinqcos\gamma, cosq)$ (Fig. 2.5 b). λ and γ are the azimuth of the two vectors, while p and q are the incidence angles of the two vectors. To ensure that the direction of the hexagonal symmetry axis is unique, we assume $0 \le q < \pi$ and $0 \le \gamma < \pi$. Therefore, the cosine of the two vectors (Fig. 2.5 c) can be written as:

$$cos\theta = l \cdot n = sinpsin\lambda sinqsin\gamma + sinpcos\lambda sinqcos\gamma + cospcosq \quad (2.22)$$

According to Eq. 2.2 and Eq. 2.22, the total P-wave slowness can be rewritten as:

$$S = S_0[1 + 2(dA + eB + fC)^2 - (A^2 + B^2 + C^2)] \quad (2.23)$$

where $d = sin\lambda sinp$, $e = cos\lambda sinp$, $f = cosp$, $A = \sqrt{\frac{M}{S_0}}sinqsin\gamma$, $B = \sqrt{\frac{M}{S_0}}sinqcos\gamma$, $C = \sqrt{\frac{M}{S_0}}cosq$. Of course, combining Eq. 2.16, we can obtain the vectoral parameterization expression of this paper:

$$t_i = L_i u[1 \pm (2[r_{i1}s_1 + r_{i2}s_2 + r_{i3}s_3]^2 - s_1^2 - s_2^2 - s_3^2)]^{-1} \quad (2.24)$$

(a)

A unit ray segment

$l(\sin p \sin \lambda, \sin p \cos \lambda, \cos p)$

$0 \leq p \leq \pi$ $0 \leq \lambda \leq 2\pi$

(b)

A unit vector along symmetry axis

$n(\sin q \sin \gamma, \sin q \cos \gamma, \cos q)$

$0 \leq q < \pi$ $0 \leq \gamma < \pi$

(c)

$\cos \theta = l \cdot n =$

$\sin p \sin q \sin \lambda \sin \gamma + \sin p \sin q \cos \lambda \cos \gamma$

$+ \cos p \cos q$ $0 \leq \theta \leq \pi$

(d)

Fast velocity plane

Hexagonal symmetry axis
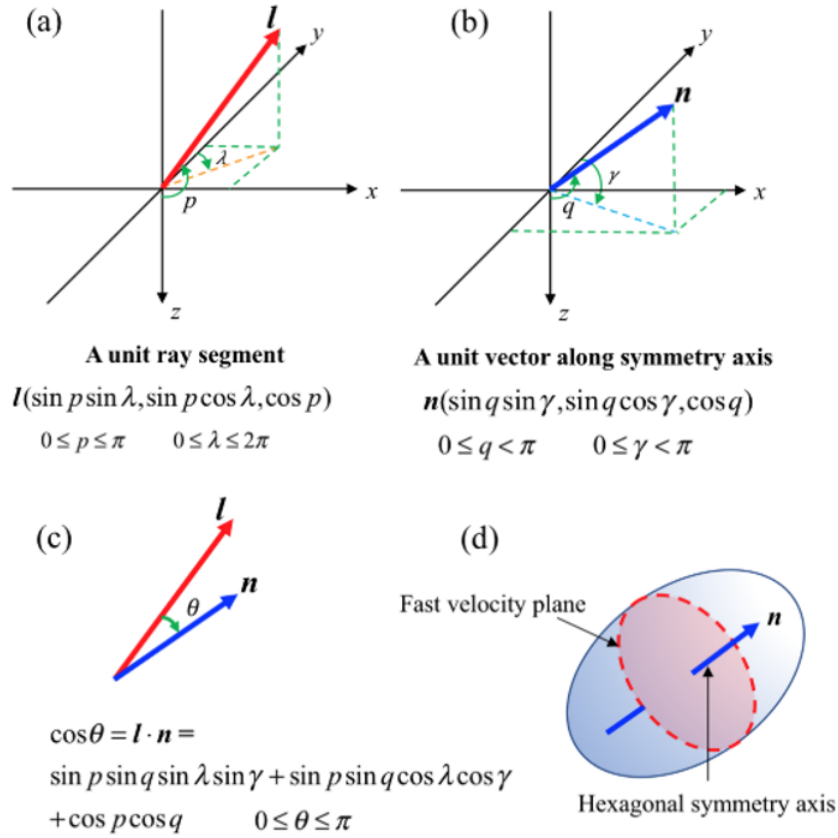
Figure 2.5: P-wave anisotropy with hexagonal symmetry adopted in this study [45]. (a), A unit vector along the ray propagation direction. (b), A unit vector along the HSA. (c), Included angle between the two vectors shown in (a) and (b). (d), Relation between the HSA and the fast velocity plane.

where r_{ij} is the $x-$, $y-$, and z-components of the unit vector pointing in the direction of the ith ray path; similarly, s_j is the component of the vector describing the orientation of anisotropy, whose squared-norm is equal to the magnitude of anisotropy. Combining Eq. 2.21 and Eq. 2.24, we can obtain the relationship between the magnitude of anisotropy (F), the azimuth (ψ) and the elevation (γ) with the vector components of the axis of symmetry as follows:

$$F = s_1^2 + s_2^2 + s_3^2 \tag{2.25}$$

$$\psi = arctan[\frac{s_2}{s_1}] \tag{2.26}$$

$$\gamma = arctan[\frac{s_3}{\sqrt{s_1^2 + s_2^2}}] \tag{2.27}$$

2.3.3 ABC PARAMETERIZATION

From the content of the previous section, we can understand that inversion requires establishing a linear relationship between travel-time and anisotropy parameters. This relationship can be obtained by directly differentiating Eq. 2.21 with respect to the spherical anisotropy parameters (i.e. F, ψ, and γ). Munzarova et al. practiced this method. However, the parameterization method in the previous text produced zero partial derivatives for the isotropic model, requiring initial anisotropy models or computing second-order partial derivatives to drive the inversion. This makes some results strongly dependent on the possibly unknown initial anisotropy model, while the calculation of second-order partial derivatives increases computational complexity. Therefore, VanderBeek and Faccenda [46] proposed a modified version of the vectoral parameterization (i.e., ABC parameterization) for imaging hexagonal anisotropy in any direction. They suggest using an alternative parameterization, which is based on the arbitrary direction hexagonal anisotropy as a combination of coefficients of the symmetric axis vector components. This modified version allows the definition of partial derivatives in the case of model isotropy, without calculating second-order terms, and has the advantage of separating variables that control the orientation and strength of the anisotropy direction from variables that control the tilt angle. The new anisotropy parameters are defined as:

$$A = s_1^2 - s_2^2 = Fcos^2(\gamma)cos(2\psi) \tag{2.28}$$

$$B = 2s_1s_2 = Fcos^2(\gamma)sin(2\psi) \tag{2.29}$$

$$C = s_3 = \sqrt{F}sin(\gamma) \tag{2.30}$$

under this parameterization, Eq. 2.16 becomes:

$$t_i = L_iu[1 \pm (Q_i - G - C^2)]^{-1} \tag{2.31}$$

where $Q_i = (r_i \cdot s)^2$ can be expressed in terms of the A, B and C parameters as:

$$Q_i = A(r_{i1}^2 - r_{i2}^2) + 2Br_{i1}r_{i2} + G(r_{i1}^2 + r_{i2}^2) + 2\sqrt{2}CY_ir_{i3} + 2C^2r_{i3}^2 \tag{2.32}$$

For the sake of convenience, the terms G and Y are defined as:

$$G = [A^2 + B^2]^{\frac{1}{2}} \tag{2.33}$$

and

$$Y = (G + A)^{\frac{1}{2}}r_{i1} \pm (G - A)^{\frac{1}{2}}r_{i2} = \sqrt{2}r_{i1}s_1 + \sqrt{2}r_{i2}s_2 \tag{2.34}$$

We can also relate A, B and C to the spherical parameters ψ, γ and F as:

$$\psi = \frac{1}{2}arctan(\frac{B}{A}) \tag{2.35}$$

$$\gamma = arctan(\frac{C}{\sqrt{G}}) \tag{2.36}$$

$$F = G + C^2 \tag{2.37}$$

This parameterization is particularly convenient because the Eq. 2.31 and bias functions reduce to the azimuthal anisotropy inversion that is widely used to model isotropy [11].

2.4 MINIMIZATION OF THE OBJECTIVE FUNCTION

In order to achieve tomography, we need to solve Eq. 2.1, which means solving for the perturbations p of the model parameters m relative to the velocity model, thus obtaining the velocity anomaly distribution in the study area. Therefore, our goal is to find the perturbations p of the model parameters m to minimize the least-squares objective function:

$$f(m,p) = \sum_{i=1}^{N} \frac{[\tau_i - t_i(m,p)]^2}{\epsilon_i^2} \tag{2.38}$$

here, τ_i and t_i are respectively the ith observed and predicted travel-time, with uncertainty ϵ_i. We will consider the following three basic solution methods to determine the minimum value of f: (1) gradient descent method, (2) Newton's method, and (3) Levenberg-Marquardt method.

2.4.1 GRADIENT DESCENT METHOD

The most direct method to minimize f is by using the gradient descent method [60]. This is a first-order optimization algorithm, which is one of the simplest and most classic methods for solving unconstrained optimization problems. The basic idea is to calculate the gradient (slope) of the objective function, and then move in the direction of the global (true) minimum value of the objective function, like "going downhill". However, this method may eventually reach a local minimum.

In particular, for an unconstrained optimization problem $min_x f(x)$, where $f(x)$ is a continuous differentiable function, if we can construct a sequence x_0, x_1, x_2,..., and satisfy the following conditions:

$$f(x_{t+1}) < f(x_t), t = 0, 1, 2, ... \tag{2.39}$$

If we can perform this process continuously until it converges to a local minimum point. Assuming we randomly choose an initial point x_1, then x_{t+1} is obtained by taking a small step Δx in a certain direction from the previous x_t ($x_{t+1} = x_t + \Delta x$). The key to this process is to ensure that $f(x_{t+1}) < f(x_t)$ while determining the direction of Δx. For a univariate function, there will be two directions: the positive direction ($\Delta x > 0$) and the negative direction ($\Delta x < 0$). In each step of direction selection, we can use the Taylor expansion formula to help us decide:

$$f(x + \Delta x) \simeq f(x) + \Delta x \nabla f(x) \tag{2.40}$$

where $\nabla f(x)$ is the gradient of the objective function. To ensure that $f(x + \Delta x) < f(x)$ (direction choice), based on the Eq. 2.40, it is clear that we need to ensure that: $\Delta x \nabla f(x) < 0$, we can choose to let:

$$\Delta x = -\eta \nabla f(x), \eta > 0 \tag{2.41}$$

where η is the step length, representing the distance the gradient moves

downwards at each iteration, hence:

$$\Delta x \nabla f(x) = -\eta (\nabla f(x))^2 \tag{2.42}$$

Due to the fact that the square of any non-zero number is greater than 0, it can be concluded that $\Delta x \nabla f(x)$ must be less than 0, and we can set $f(x + \Delta x) = f(x - \eta \nabla f(x))$, then it can be guaranteed that $f(x + \Delta x) < f(x)$, then the formula for updating x is:

$$x' \leftarrow x - \eta \nabla f(x) \tag{2.43}$$

In this way, by continuously updating x, $f(x)$ gradually moves towards the global (true) minimum value, and eventually reaches the local (or global) minimum point.

For Eq. 2.38, its gradient with respect to the model perturbation p_j is:

$$g_j = -2 \sum_{i=1}^{N} \frac{\Delta t_i}{\epsilon_i^2} \frac{\partial t_i}{\partial p_j} \tag{2.44}$$

We want to reduce the gradient (i.e., towards the minimum rather than the maximum), resulting in the following expression for the model perturbation at iteration n:

$$p_j(n + 1) = p_j(n) - \eta g_j \tag{2.45}$$

Generally, the value of η can vary with each iteration and each component of g. There are many ways to choose the value of η, but we will adopt a simple backtracking line search strategy. This means, given an initial estimate of η, we will build a new model using the perturbations in Eq. 2.45. If the value of η results in a decrease of the objective function, we accept the model and continue iterating. If the objective function does not decrease, we continue to decrease η until the objective function decreases.

2.4.2　Newton's method

The gradient descent method only uses first-order derivatives, that is, the information brought by the gradient. If we consider not only the first-order derivatives of the function, but also the information brought by the second-order derivatives, we need to introduce the Hessian matrix:

$$H(f)(x)_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(x), H_{i,j} = H_{j,i} \tag{2.46}$$

The Hessian matrix is a real symmetric matrix that can be decomposed into eigenvectors. In some cases, the gradient descent method may fail. If the current point of gradient descent is x_0, we can perform a second-order Taylor expansion at the current point:

$$f(x) \approx f(x_0) + (x - x_0)^T g + \frac{1}{2}(x - x_0)^T H(x - x_0) \tag{2.47}$$

where g is the gradient of the objective function and H is the Hessian matrix at x_0. We use gradient descent for iteration, and according to the Eq. 2.43, we can obtain a new point, substituting the above equation we get

$$f(x_0 - \eta g) \approx f(x_0) - \eta g^T g + \frac{1}{2}\eta^2 g^T H g \tag{2.48}$$

From the above equation, it can be seen that when $\frac{1}{2}\eta^2 g^T H g$ is too large, the gradient descent method may cause the value of the objective function to increase, and the following two cases may cause $\frac{1}{2}\eta^2 g^T H g$ to be too large:

1. The step length η is too large, and $x' \leftarrow x - \eta \nabla f(x)$ exceeds the x corresponding to the minimum value of $f(x)$.

2. The second-order derivative $g^T H g$ of $f(x)$ in the direction of the gradient is too large, i.e., $f(x)$ is very "steep" at x_0, so that even a small step length η causes $x' \leftarrow x - \eta \nabla f(x)$ to surpass the x corresponding to the minimum value of $f(x)$.

Therefore, we need to find a method that suits using the second derivative to guide the iteration. The simplest second-order optimization method is the Newton's method. The basic idea of the Newton's method is to approximate the objective function with a quadratic function using the first-order derivative (gradient) and the second-order derivative (Hessen matrix) at the iterative point x, and then use the minimum point of the quadratic model as the new iterative point $x + 1$, and repeat this process continuously until an approximation with satisfactory accuracy is obtained.

Compared to simply moving down the gradient directly, the Newton's method uses the curvature of the objective function to find a "more direct" path to the minimum, that is, its iteration direction at each step is along the direction of decrease in the value of the objective function. If the iterated point $x + 1$ is a minimum point, the following conditions must be satisfied:

1. $g(x + 1) = 0$ (all elements in the gradient vector are 0).

2. $H(x)$ is a positive definite matrix, i.e., it ensures that the stationary point $x + 1$ near x is a minimal point.

Once the conditions are satisfied, we can take the derivative of both sides of the formula for Eq. 2.47 and obtain:

$$g(x) \approx g(x_0) + H(x_0)(x - x_0) \tag{2.49}$$

Let $g(x) = 0$, then each step of iteration is:

$$x_{k+1} = x_k - H(x_k)^{-1} g(x_k) \tag{2.50}$$

For our objective function (Eq. 2.38), the curvature is given by $\partial g_j / \partial p_k$ which is referred to as the Hessian matrix of the objective function:

$$H_{jk} = 2 \sum_{i=1}^{N} \frac{1}{\epsilon_i^2} \frac{\partial t_i}{\partial p_j} \frac{\partial t_i}{\partial p_k} - \frac{\Delta t_i}{\epsilon_i^2} \frac{\partial^2 t_i}{\partial p_j \partial p_k} \tag{2.51}$$

The new model perturbation is derived from the second-order Taylor expansion of the objective function, leading to the following expression for updating the perturbation vector:

$$p(n + 1) = p(n) - \eta g H^{-1} \tag{2.52}$$

For general problems, we can choose to directly calculate the inverse of Hessian to update Eq. 2.52; however, for problems with a large number of parameters (i.e., high-dimensional), we do not directly calculate the inverse of Hessian, but instead use effective algorithms such as LSQR to solve the equation system $Hx = g$ to calculate the product $x = gH^{-1}$. Therefore, although the computational cost of Newton's method is larger than that of gradient descent, its convergence speed is much faster than that of gradient descent.

2.4.3 LEVENBERG-MARQUARDT METHOD

Newton's method may encounter issues in its purest form, mainly due to two reasons: first, the poor performance of the derivatives, and second, the high cost of computing the Hessian matrix in high-dimensional problems. To address these issues, the Gauss-Newton algorithm approximates the Hessian matrix of

the least squares objective function by ignoring the second-order terms (i.e. the last term in Eq. 2.50), thus reducing the computational difficulty. This method allows us to define the Hessian matrix based on the first-order derivatives of the data, also known as the data Jacobian matrix. Specifically, we can represent g and H in the form of Jacobian matrices multiplied together:

$$g = 2J_f^T f, H \approx 2J_f^T J_f \tag{2.53}$$

where J_f is the Jacobi matrix. By substituting the modified g and H into the equation Eq. 2.50, we can obtain:

$$\beta_{s+1} = \beta_s + \Delta, \Delta = -(J_f^T J_f)^{-1} J_f^T f \tag{2.54}$$

Through this approximation, we obtain the iterative update formula for the Gauss-Newton method. Therefore, for Eq. 2.52, we can represent g as $g = 2J_t^T t$, and the approximate Hessian as $H = J_t^T J_t$, where $J_{ij} = \partial t_i / \partial p_j$. Please note that the data Jacobian is a $N \times M$ matrix (N is the number of observations, and M is the number of model parameters). In the Gauss-Newton method, we find the update of the perturbation vector by solving the linear system $Jx = \Delta t$, which can be effectively done using the LSQR algorithm. The new perturbation vector is defined as:

$$p(n+1) = p(n) + \eta x, x = -H^{-1} g = -(J_t^T J_t)^{-1} J_t^T t \tag{2.55}$$

Similar to Newton's method, when the initial values are far from the minimum value, the Gauss-Newton method cannot guarantee convergence. Additionally, in cases where H is approximately singular, the Gauss-Newton method cannot converge correctly. Therefore, we adopt an improved algorithm of the Gauss-Newton method: the Levenberg-Marquardt method. This approach combines gradient descent and Gauss-Newton method through a linear combination, making the most of the advantages of both algorithms. By adding a damping factor (scaling factor) λ (λI) to the Hessian matrix, we can control the step size and direction of each iteration:

$$(H + \lambda I)\epsilon = -J_t^T t \tag{2.56}$$

here, ϵ represents the iteration direction. As λ increases, $H + \lambda I$ tends to λI, and therefore ϵ tends to $-\lambda J_t^T t$, which is the iteration direction given by gradient

27

descent; as λ decreases, $H + \lambda I$ tends to H, and ϵ tends to $-H^{-1}J_t^T t$, which is the direction given by the Gauss-Newton method. With a slight modification of the Levenberg-Marquardt method, we can add the equation in the form of a scaled identity matrix to the end of J to limit the amplitude of the perturbation in each iteration. As the scaling factor increases, the iteration follows the trend of gradient descent. Specifically, the corrected equation system we are solving for is:

$$\begin{bmatrix} J \\ \lambda I \end{bmatrix} x = \begin{bmatrix} \Delta t \\ 0 \end{bmatrix} \tag{2.57}$$

here, I is the $M \times M$ identity matrix multiplied by the proportional factor λ, and 0 is the $M \times 1$ vector of zeros. Normally, λ is unique for each row of I, but in this paper, we only focus on one scalar value.

3

2-D Anisotropic Numerical Experiment

In order to study how different seismic anisotropy parameterizations affect the performance of linearized inversion schemes in seismic tomography, we need to conduct a series of numerical experiments. Firstly, we focus on simple cases, where only two anisotropy parameters are constrained in 2-D cases (with two variable anisotropy parameters). At the same time, we use synthetic observed data to combine test different anisotropy parameterizations and objective function minimization schemes. In subsequent studies, we extend the 2-D anisotropy parameterization case to a full (with four variable anisotropy parameters) anisotropy parameterization case, in order to find the most suitable parameterization method for the commonly used iterative inversion schemes in seismic tomography.

3.1 Synthetic experiment preparation

3.1.1 Synthetic data generation and model parameter setting

For the generation of synthetic observed data, we first define a homogeneous anisotropic volume sampled by the straight ray path. The length of the straight ray path is $100\,\mathrm{km}$, with a uniform sampling of azimuth between $30°$ and $50°$ (the specific sampling interval depends on the number of observations), and the elevation is set to $0°$. A total of 37 observation data are included. When setting the

anisotropy model parameters, we adopt the target (true) anisotropy parameter model and the initial anisotropy parameter model. Since this chapter needs to discuss the 2-D case, we use the anisotropy fraction (F) and the anisotropy azimuth (ψ) as two variable parameters. For the target anisotropy fraction, we set it to 0.03, and the target anisotropy azimuth is set to $-21°$; for the initial anisotropy fraction, we set it to 0.01, and the initial anisotropy azimuth is set to $0°$. In addition, for both the target and initial anisotropy parameters, we set the slowness (i.e., the inverse of velocity, denoted as u) to $0.125\,\text{s/km}$.

After defining the synthetic observed data and anisotropy model parameters, we can calculate the observed and predicted travel-time using different anisotropy parameterization methods. First, we use the spherical parameterization method (Eq. 2.21) for calculation. Since the elevation of the seismic ray and the elevation of anisotropy are both $0°$, therefore:

$$cos(\theta_i) = 1, cos(\gamma) = 1, sin(\theta_i) = 0, sin(\gamma) = 0 \tag{3.1}$$

We bring in Eq. 3.1 into Eq. 2.21, and ultimately obtain a new travel-time computation formula:

$$t_i = L_i u[1 + F(2[cos(\psi - \phi_i)]^2 - 1)]^{-1} \tag{3.2}$$

Next, for the vectoral parameterization method, according to Eq. 2.24, we need to first obtain r_{ij} and s_j. Combining the content of the previous chapters, we can derive the relationship between r_{ij} and s_j and the anisotropy fraction F, azimuth ψ, and elevation γ (Eq. 2.25, Eq. 2.26 and Eq. 2.27):

$$s_1 = \sqrt{|F|}cos(\gamma)cos(\psi) \tag{3.3}$$

$$s_2 = cos(\gamma)sin(\psi) \tag{3.4}$$

$$s_3 = sin(\gamma) \tag{3.5}$$

where r_{ij} represents the $x-$, $y-$, and z-components of the unit vector pointing in the direction of the ith ray path. Combined with Eq. 2.21, we can obtain:

$$r_1 = cos(\theta_i)cos(\phi_i) \tag{3.6}$$

$$r_2 = cos(\theta_i)sin(\phi_i) \tag{3.7}$$

$$r_3 = sin(\theta_i) \tag{3.8}$$

where ϕ_i and θ_i represent the azimuth and elevation of the seismic ray. By combining Eq. 3.3 to Eq. 3.8 with Eq. 2.24, we can calculate the travel-time under the vectoral parameterization method using the values of anisotropy fraction, azimuth, and elevation. Finally, for the ABC parameterization method, we need to calculate A, B, and C according to Eq. 2.28, Eq. 2.29, and Eq. 2.30. Then, according to Eq. 2.33, Eq. 2.35, Eq. 2.36, and Eq. 2.37, we can derive the relationships between the anisotropy fraction, azimuth, elevation, and A, B, C. Finally, combining the derived anisotropy fraction, azimuth, and elevation with Eq. 2.21, we can obtain the travel-time computation formula under the ABC parameterization method:

$$t_i = L_i u[1 + (G + C^2)(2[cos(\theta_i)cos(arctan(\frac{C}{\sqrt{G}}))cos(\frac{1}{2}arctan(\frac{B}{A}) - \phi_i) \\ + sin(\theta_i)sin(arctan(\frac{C}{\sqrt{G}}))]^2 - 1)]^{-1} \tag{3.9}$$

In order to simulate the error situation of observed data in real cases, we added a Gaussian noise with a standard deviation of 0.1 to the observed travel-time data results. We can calculate the observed and predicted travel-time under different parameterization methods using Eq. 3.2, Eq. 2.24, and Eq. 3.9, and then calculate the travel-time residuals (i.e., the difference between the observed values and the predicted values). Considering our objective function is Eq. 2.38, we set the assumed data uncertainty to 0.1, and finally, the objective function value can be calculated through Eq. 2.38. Our goal is to determine the minimum value of the objective function, so we need to use some minimization methods to achieve this goal.

3.1.2 SOLVER PARAMETER SETTINGS

In this paper, we consider three basic solution methods to determine the minimum value of the objective function, which are the gradient descent method, the Newton's method, and the Levenberg-Marquardt method (also known as

the damped Gauss-Newton method or damped least-squares method). In the previous chapters, we have introduced the basic principles of these three solution methods. In this chapter, we need to set the operating parameters of the three solvers to constrain the solution process of the objective function. Here we introduce several main solver parameter settings.

In all the numerical experiments in this chapter, the maximum number of iterations for the solver is 100. The criteria for solver convergence is set to 1.01, i.e., when the average value of the objective function is less than this value (tolerance), the iterative calculation steps of the solver are exited. We will set the scale factor that modifies the size of perturbations on each iteration, i.e., set a certain step length (in all the numerical experiments in this chapter, the step length is set to 1). Since the update of the perturbations require the use of the backtracking linear search method, we can set the number of line search iterations. If the number of line search iterations is greater than 0, the solver will perform a linear search to ensure that the perturbations actually reduce the objective function. Similarly, we can also set the scale factor in the linear search iteration, so that in each iteration of the linear search, the perturbations will be scaled by this factor (in all the numerical experiments in this chapter, the scale factor of the linear search is set to 0.5). We also have specific options set for the Gauss-Newton method: damping term. If this option is enabled, the Gauss-Newton solver will include the Levenberg-Marquardt damping term. All the solver parameter settings are shown in the following items:

- Opts.SolverFunc = *. Define function handle to solver (for example: @GradientDescent).

- Opts.SolverIterations = 100. The maximum number of solver iterations.

- Opts.ObjFunTolerance = 1.01. Convergence criteria. Exit solver when the mean objective function value is less than this value. The objective function is scaled such that a value of 1 implies the data is fit to within estimated errors.

- Opts.StepLength = 1. Step length. Scale factor that modifies the size of perturbations on each iteration.

- Opts.tf_NormStep = 0. Normalize the step length. If true, length of the perturbational vector is normalized such that the norm of the step (i.e. pertubation vector) at each iteration is the same.

- Opts.LineSearchIterations = *. The number of line search iterations. If greater than 0, the solver will perform a line search to ensure that the perturbations did reduce the objective function.

- Opts.LineSearchStep = 0.5. Scale factor for linear search. On each iteration of the line search, the perturbations will be scaled by this factor.

- Opts.ForceQuadrant = 1. Force solution into the positive (1) or negative (-1) m1-coordinate. If 0, no quadrant forcing will be applied.

- Opts.MaxAbsStep = []. Option to define maximum allowed change in the model parameters.

- Opts.FigureObjFunc = 0. Option to plot iterations on existing figure.

For the Gauss-Newton solver, we also have some special solver parameter settings as follows:

- Opts.GN.Ldamp = *. Damping term. If defined as a scalar, Gauss-Newton solver will include Levenberg-Marquardt style damping.

- Opts.GN.TolLSQR = 0.000006.

- Opts.GN.MaxIterLSQR = 1000.

All of the above items indicate the names, default values, and meanings of the parameters for solvers. Please note that an asterisk (*) indicates that the parameter does not have a fixed default value, and different experiments can have different settings.

3.1.3 STABILITY INSPECTION SCHEME DESIGN

For the stability test of the 2-D anisotropic numerical experiment, we can visualize and evaluate the convergence of the objective function results by constructing a 2-D objective function surface. Simply put, we will take uniform samples of the anisotropy fraction and the azimuth of anisotropy within a certain range. Then, for different sample combinations of anisotropy fraction and anisotropy azimuth, we will calculate the travel-time residual of all combinations using Eq. 3.2, Eq. 2.24, and Eq. 3.9, and further calculate the value of objective function using Eq. 2.38. For different parameterization methods, we will calculate the relative anisotropy fraction and the anisotropy azimuth according to the transformation relationship described above. Next, we will use the anisotropy fraction as the x-axis and the anisotropy azimuth as the y-axis, and combine the objective function values of each sample point to draw the 2-D objective function surface, as shown in the Fig. 3.1.

In the Fig. 3.1, m_1 represents the anisotropy fraction and m_2 represents the anisotropy azimuth. The black contour lines show the distribution of the objective function values at the sample points, and the area enclosed by the red
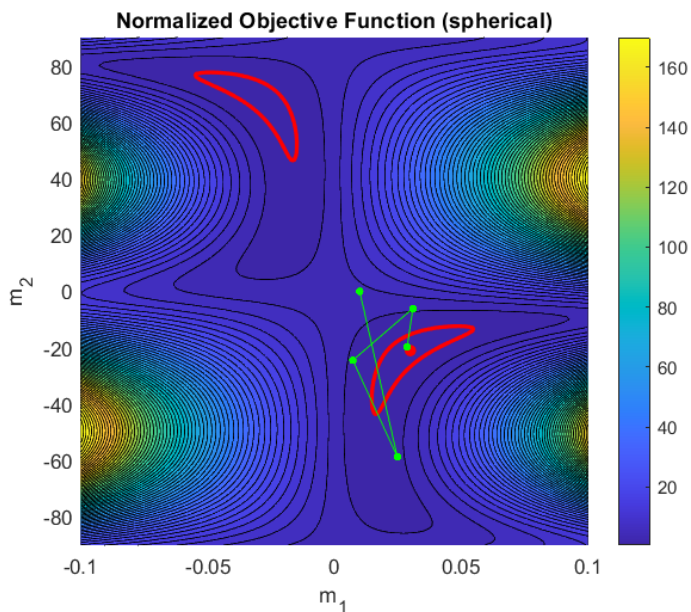
Figure 3.1: An example of a normalized objective function surface using the spherical parameterization.

contour lines represents the region where the solver convergence criteria are met (i.e., the average objective function value is less than or equal to 1.01). The red points represent the target anisotropy parameters (which are also the minimum value of the true objective function), and the color bars indicate the objective function values. Overall, we can clearly observe the distribution of the extreme values of the objective function from the objective function surface. When the solver iterates, the predicted anisotropy parameters (fraction and azimuth) obtained in each iteration are printed on the objective function surface (green points) and connected by adjacent iteration results (green lines). In this way, we can observe the effective search directions of the solver in finding the minimum value of the objective function. When the green dot is within the region enclosed by the red contour lines, it indicates that the objective function value obtained in the current iteration is less than or equal to 1.01, meeting the convergence criteria of the solver, and then the iteration is exited to complete the calculation. Eventually, we can obtain the predicted anisotropic parameters that meet the convergence criteria. From the above image, we can see that the solver's iterative calculation count is 4 times (i.e., four green lines).

Different parameterization schemes result in different objective function surfaces in visualization, as shown in the Fig. 3.2 for the vectoral and ABC param-

eterization schemes:
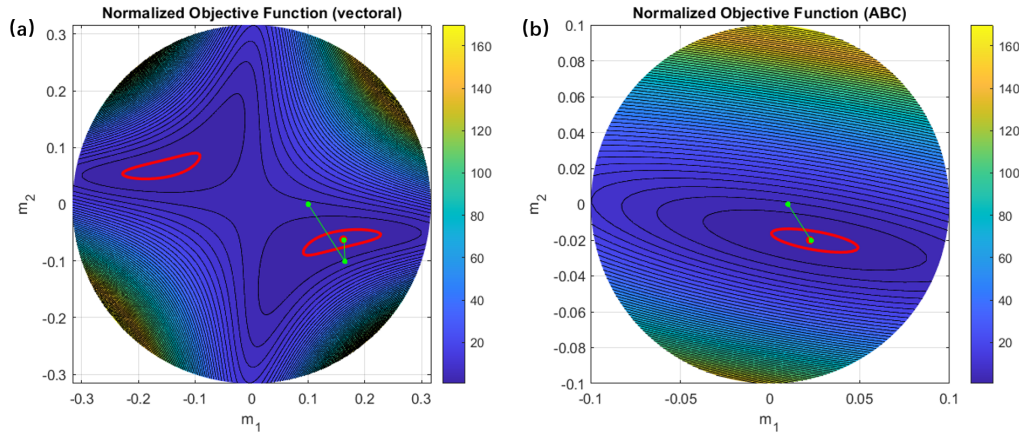
Figure 3.2: An example of a normalized objective function surface (using the same model parameters as in Figure 1). (a), vectoral parameterization scheme results; (b), ABC parameterization scheme results.

The objective function surfaces of the two parameterization schemes are circular because the anisotropic fraction term (m_1) and the azimuth term (m_2) of the two parameterization schemes need to be transformed according to the relationship described earlier with respect to the anisotropy fraction, azimuth, and elevation. However, the meaning expressed in the two figures is consistent with the Spherical parameterization, that is, the Vectoral parameterization scheme achieved the minimum value of the objective function with only two iterations, while the ABC parameterization scheme achieved the minimum value of the objective function with only one iteration.

3.2   SYNTHETIC NUMERICAL EXPERIMENT TESTING

We conducted synthetic numerical experiments using three solver methods for three parameterization schemes. Apart from the selection of solvers and parameterization schemes needing to be adjusted, all other parameters (except for the specific options for the Gauss-Newton solver) remain constant in different numerical experiments. For the settings of the observation data parameters, please refer to the Tab. 3.1:

For the target anisotropic model parameters, our settings are as shown in the Tab. 3.2:

Table 3.1: Parameter settings of the observed data in 2-D anisotropic numerical experiments.

Item	Setting
Number of observations	37
Ray path lengths	$100\,\mathrm{km}$
Ray path azimuths	30 to $50°$
Ray elevations	$0°$

Table 3.2: Settings of target anisotropic model parameters in 2-D anisotropic numerical experiments.

Item	Setting
Target slowness	$0.125\,\mathrm{s/km}$
Target anisotropic fraction	0.03
Target anisotropic azimuth	$-21°$
Target anisotropic elevation	$0°$

For the initial anisotropic model parameters, our settings are as shown in the Tab. 3.3:

Table 3.3: Initial anisotropic model parameter settings in 2-D anisotropic numerical experiments.

Item	Setting
Initial slowness	$0.125\,\mathrm{s/km}$
Initial anisotropic fraction	0.01
Initial anisotropic azimuth	$0°$
Initial anisotropic elevation	$0°$

Based on the above related parameter settings, we designed 12 2-D anisotropic parameterization case experiments (including 3 parameterization schemes and 4 solver methods, among which 1 solver method is the Gauss-Newton solver without Levenberg-Marquardt damping term).

3.2.1 EXPERIMENTS OF DIFFERENT PARAMETERIZATION SCHEMES AND GRADIENT DESCENT SOLVER

For the gradient descent solver, our solver parameter settings are as shown in the Tab. 3.4:

Table 3.4: Parameter settings of the gradient descent solver in 2-D anisotropic numerical experiments.

Parameter name	Setting content
Opts.SolverFunc	@GradientDescent
Opts.SolverIterations	100
Opts.ObjFunTolerance	1.01
Opts.StepLength	0.005
Opts.tf_NormStep	1
Opts.LineSearchIterations	1
Opts.LineSearchStep	0.01
Opts.ForceQuadrant	1
Opts.MaxAbsStep	[]
Opts.FigureObjFunc	0
Opts.GN.Ldamp	[]
Opts.GN.TolLSQR	1e-6
Opts.GN.MaxIterLSQR	1e3

We conducted numerical experiment tests for three parameterization schemes, and finally obtained the objective function surfaces of different parameterization schemes. In the 1st numerical experiment (as shown in the Fig. 3.3), the solution process was iterated a total of 100 times. From the figure, we can see that the gradient descent solver did not initially search in the correct direction. As the iteration progressed, the solution direction approached the minimum value of the objective function. However, the final predicted value of the objective function did not reach the solver's convergence criteria. In the 2nd numerical experiment (as shown in the Fig. 3.4), the solution process was iterated a total of 13 times. The initial solution search direction of the gradient descent solver did not point towards the minimum value of the objective function. However, during the iteration calculation, the solver turned and the final predicted value of the objective function reached the solver's convergence criteria. In the 3rd numerical experiment (as shown in the Fig. 3.5), the solution process was iterated a total of 3 times. The final predicted value of the objective function also reached the solver's convergence criteria. However, unlike the vectoral parameterization results, the gradient descent solver under the ABC parameterization setting can quickly find the correct direction and search, and the numerical calculation converges faster.

Figure 3.3: The normalized objective function surface of the 1st numerical experiment: Spherical parameterization method and gradient descent solver.

Figure 3.4: The normalized objective function surface of the 2nd numerical experiment: Vectoral parameterization method and gradient descent solver.

Figure 3.5: The normalized objective function surface of the 3rd numerical experiment: ABC parameterization method and gradient descent solver.

3.2.2 EXPERIMENTS OF DIFFERENT PARAMETERIZATION SCHEMES WITH NEWTON'S METHOD SOLVER

For the Newton's method solver, our solver parameters are set as shown in the Tab. 3.5:

Table 3.5: Parameter settings of the Newton's method solver in 2-D anisotropic numerical experiments.

Parameter name	Setting content
Opts.SolverFunc	@Netwon
Opts.SolverIterations	100
Opts.ObjFunTolerance	1.01
Opts.StepLength	0.005
Opts.tf_NormStep	1
Opts.LineSearchIterations	1
Opts.LineSearchStep	0.01
Opts.ForceQuadrant	1
Opts.MaxAbsStep	[]
Opts.FigureObjFunc	0
Opts.GN.Ldamp	[]
Opts.GN.TolLSQR	1e-6
Opts.GN.MaxIterLSQR	1e3

We conducted numerical experiments for three parameterization schemes and successfully obtained the corresponding objective function surfaces. In the 4th numerical experiment (as shown in the Fig. 3.6), although the solution process was iterated a total of 100 times, the solver did not conduct a search but stayed at the initial position, which resulted in the predicted value of the objective function being far from the solver's convergence criterion. In the 5th numerical experiment (as shown in the Fig. 3.7), although the solver conducted an outward search, its initial direction did not point to the true minimum value of the objective function, and no turn occurred during the iteration process. It was not until the 6th numerical experiment (as shown in the Fig. 3.8) that the solver found an effective search direction pointing to the true minimum value of the objective function and moved quickly. Eventually, the calculation was completed in just 4 iterations, and the predicted value of the objective function met the solver's convergence criterion.
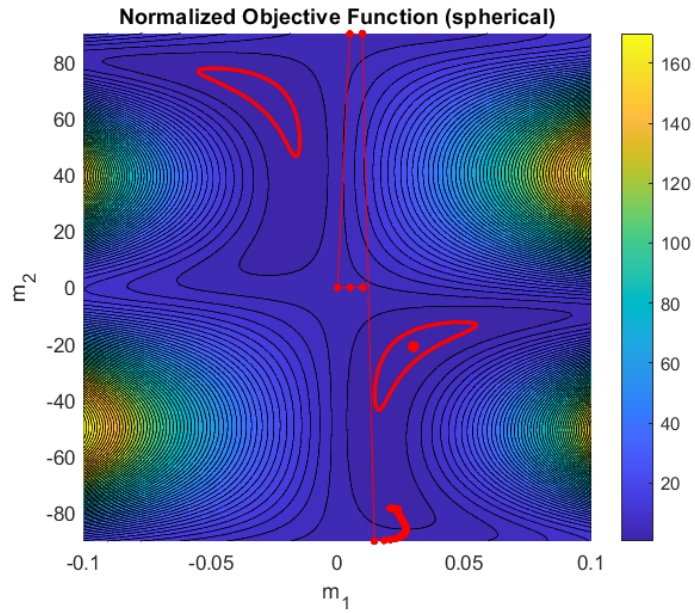
Figure 3.6: The normalized objective function surface of the 4th numerical experiment: Spherical parameterization method and Netwon's method solver.
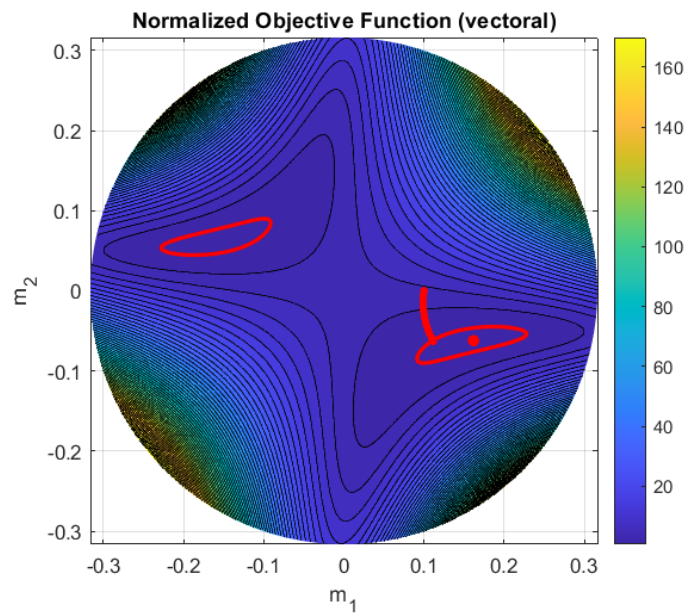

Figure 3.7: The normalized objective function surface of the 5th numerical experiment: Vectoral parameterization method and Netwon's method solver.
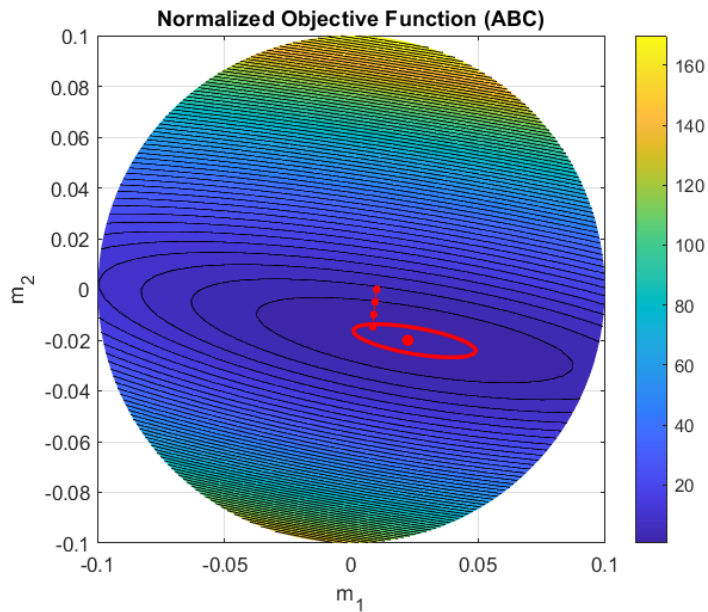
Figure 3.8: The normalized objective function surface of the 6th numerical experiment: ABC parameterization method and Netwon's method solver.

3.2.3 EXPERIMENTS OF DIFFERENT PARAMETERIZATION SCHEMES WITH THE GAUSS-NEWTON SOLVER WITHOUT LEVENBERG-MARQUARDT DAMPING TERM

For the Gauss-Newton solver without Levenberg-Marquardt damping term, our solver parameters are set as shown in the Tab. 3.6:

Table 3.6: Parameter settings of the Gauss-Newton solver (without Levenberg-Marquardt damping term) in 2-D anisotropic numerical experiments.

Parameter name	Setting content
Opts.SolverFunc	@GaussNetwon
Opts.SolverIterations	100
Opts.ObjFunTolerance	1.01
Opts.StepLength	0.005
Opts.tf_NormStep	1
Opts.LineSearchIterations	1
Opts.LineSearchStep	0.01
Opts.ForceQuadrant	1
Opts.MaxAbsStep	[]
Opts.FigureObjFunc	0
Opts.GN.Ldamp	[]
Opts.GN.TolLSQR	1e-6
Opts.GN.MaxIterLSQR	1e3

We conducted numerical experiments for three parameterization schemes and successfully obtained the corresponding objective function surfaces. In the 7th numerical experiment (as shown in the Fig. 3.9), the solution process was iterated a total of 62 times. The initial search direction of the solver pointed to the convergence direction, and the direction was continuously adjusted during the iteration calculation, eventually pointing to the true minimum value of the objective function. Meanwhile, the final predicted value of the objective function met the solver's convergence criterion. In the 8th numerical experiment (as shown in the Fig. 3.10), the solution process required only 15 iterations. The initial direction of the solver pointed directly to the true minimum value direction of the objective function, so the final predicted value of the objective function also met the solver's convergence criterion. The results of the 9th numerical experiment (as shown in the Fig. 3.11) were similar to those of the 8th numerical experiment, where the solver directly pointed to the true minimum value direction of the objective function and moved quickly. Eventually, the calculation was completed in just 4 iterations, and the final predicted value of the objective function met the solver's convergence criterion.
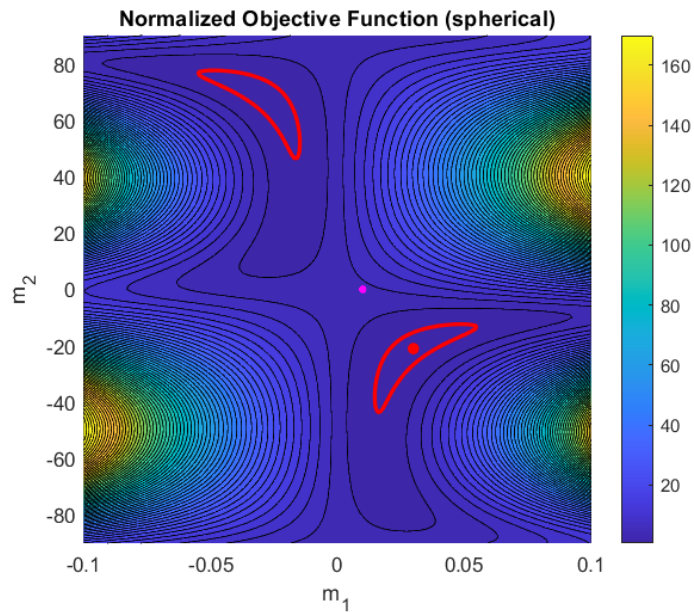
Figure 3.9: The normalized objective function surface of the 7th numerical experiment: Spherical parameterization method and Gauss-Newton solver.
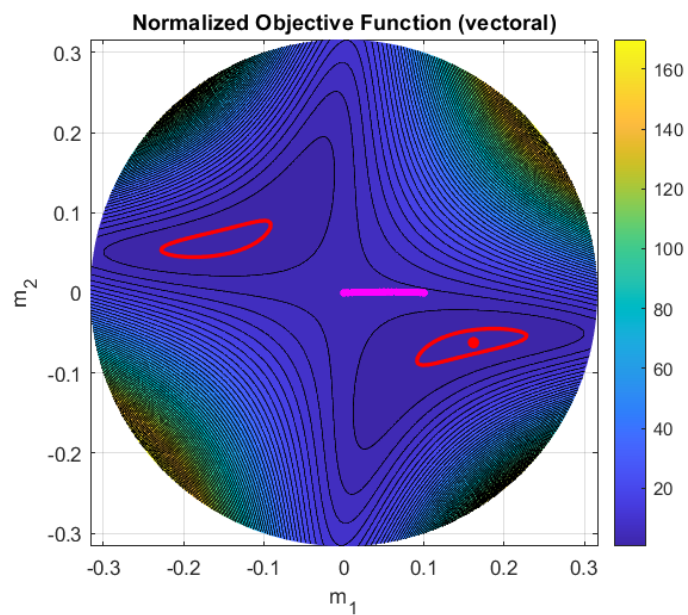

Figure 3.10: The normalized objective function surface of the 8th numerical experiment: Vectoral parameterization method and Gauss-Newton solver.
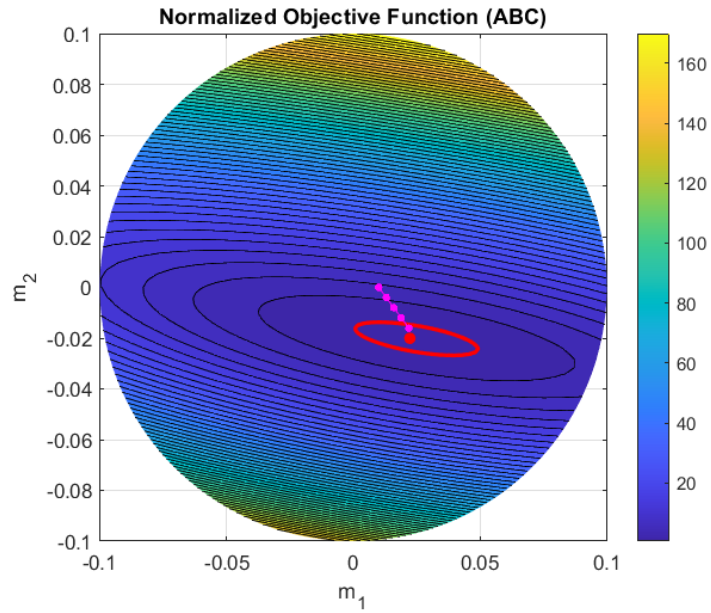
Figure 3.11: The normalized objective function surface of the 9th numerical experiment: ABC parameterization method and Gauss-Newton solver.

3.2.4 EXPERIMENTS OF DIFFERENT PARAMETERIZATION SCHEMES WITH THE LEVENBERG-MARQUARDT METHOD SOLVER

As introduced in the previous chapters, the Levenberg-Marquardt method is a Gauss-Newton method with a damping term. For the Levenberg-Marquardt method solver, the solver parameters are set as shown in the Tab. 3.7:

Table 3.7: Parameter settings of the Levenberg-Marquardt method solver in 2-D anisotropic numerical experiments.

Parameter name	Setting content
Opts.SolverFunc	@GaussNetwon
Opts.SolverIterations	100
Opts.ObjFunTolerance	1.01
Opts.StepLength	0.005
Opts.tf_NormStep	1
Opts.LineSearchIterations	1
Opts.LineSearchStep	0.01
Opts.ForceQuadrant	1
Opts.MaxAbsStep	[]
Opts.FigureObjFunc	0
Opts.GN.Ldamp	10
Opts.GN.TolLSQR	1e-6
Opts.GN.MaxIterLSQR	1e3

Here, we set the size of the damping term to 10 (Opts.GN.Ldamp=10). Then, we conducted numerical experiments for three parameterization schemes and successfully obtained the corresponding objective function surfaces. In the 10th numerical experiment (as shown in the Fig. 3.12), the solution process underwent a total of 60 iterations. Unlike the 7th numerical experiment, in this experiment, the initial search direction of the solver did not point towards the convergence direction, but continuously adjusted its direction during the iterative calculation process, eventually pointing to the true minimum value of the objective function. The final predicted value of the objective function met the convergence criteria of the solver. In the 11th numerical experiment (as shown in the Fig. 3.13), the solution process required only 13 iterations. The initial direction of the solver did not point to the true minimum direction of the objective function but to the convergence direction (marked by a red circle). Nevertheless, the final predicted value of the objective function still met the convergence criteria of the solver. In the 12th numerical experiment (as shown in the Fig. 3.14), the solver directly pointed to the convergence direction and moved quickly, completing the calculation with only 3 iterations. Moreover, the final predicted value of the objective function also met the convergence criteria of the solver.
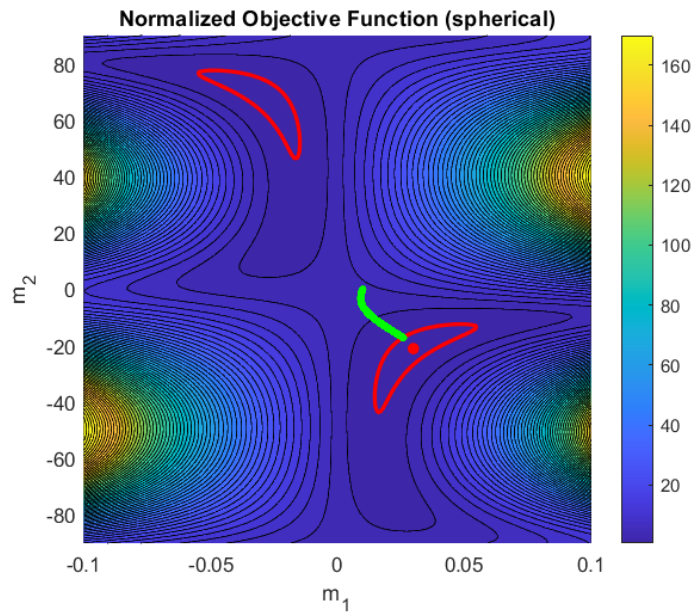
Figure 3.12: The normalized objective function surface of the 10th numerical experiment: Spherical parameterization method and Levenberg-Marquardt method solver.
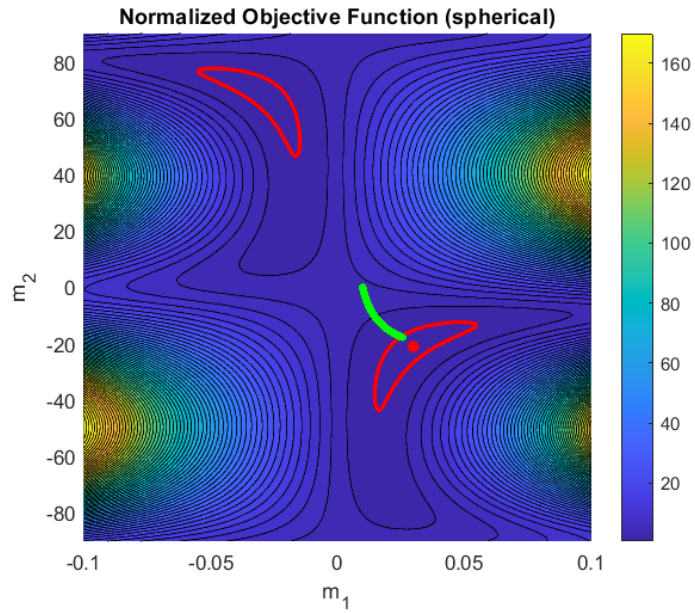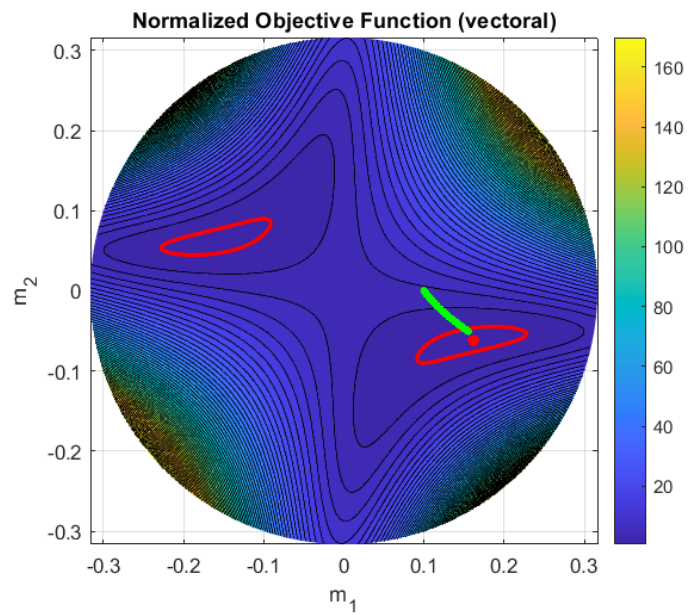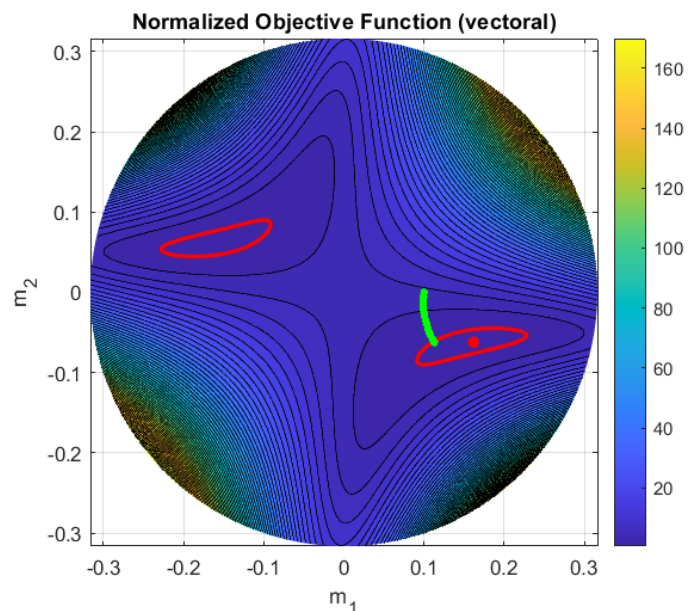

Figure 3.13: The normalized objective function surface of the 11th numerical experiment: Vectoral parameterization method and Levenberg-Marquardt method solver.
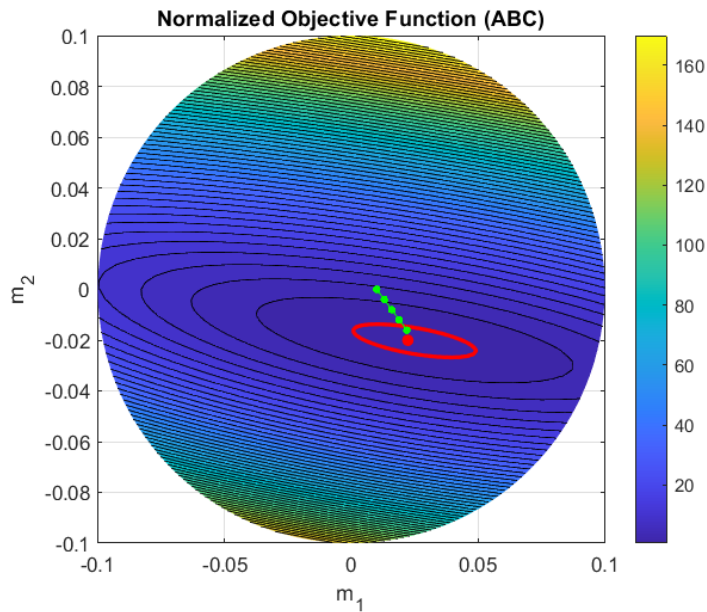
Figure 3.14: The normalized objective function surface of the 12th numerical experiment: ABC parameterization method and Levenberg-Marquardt method solver.

3.3 DISCUSSION AND SUMMARY OF THIS CHAPTER

This chapter mainly introduces and discusses the impact of seismic anisotropy parameterization on the linearization inversion scheme in seismic tomography. To study this problem, we created and designed 2-D anisotropy numerical experiments, and analyzed the results of 12 experiments. Through the analysis, it was found that except for the 1st (Fig. 3.3), 4th (Fig. 3.6), and 5th (Fig. 3.7) experiment results that could not converge, the other experiment results all met the solver's convergence criteria. Specifically, on the objective function surface, the solution path (green line) eventually contacts the convergence region (region enclosed by red contour lines); during the iteration process, the solution search direction continuously approaches the convergence region or the target anisotropy parameter (red point) position.

From the perspective of parameterization schemes, regardless of which solver method is used, the ABC parameterization scheme can always complete the calculation more quickly. In comparison, the Spherical parameterization scheme has the worst computational performance, followed by the vectoral parameter-

ization scheme. This is because the vectoral parameterization method and the ABC parameterization method can simplify the objective function surface and reduce the complexity of the objective function surface, allowing the solver to quickly find the effective search direction for the true minimum value of the objective function during the solving process.

For the solver, if the Spherical parameterization scheme is used, the gradient descent solver and the Newton's method solver cannot obtain the minimum objective function value that satisfies the convergence criteria within the given number of iterations. However, the undamped Gauss-Newton solver and the Levenberg-Marquardt method solver can complete the calculation and exit the iteration within a certain number of iterations, obtaining the minimum objective function value that satisfies the convergence criteria. If the vectoral parameterization scheme or even the ABC parameterization scheme is used, the undamped Gauss-Newton solver and the Levenberg-Marquardt method solver can complete the iteration calculation more quickly.

In particular, with the addition of the damping term, the Levenberg-Marquardt method solver can obtain the computational results that satisfy the convergence criteria in the shortest time (with the fewest number of iterations). It can be said that the ability of the undamped Gauss-Newton solver and the Levenberg-Marquardt method solver to find the effective search direction for the minimum value of the objective function is stronger than that of the gradient descent solver and Newton's method solver. If the results satisfy the convergence criteria, the computational performance of the Levenberg-Marquardt method solver is stronger than that of the undamped Gauss-Newton solver.

Therefore, we believe that the combination of the Levenberg-Marquardt method solver and ABC parameterization scheme is the best approach to solve the current 2-D anisotropic travel-time tomography problem. In the subsequent numerical experiments of the full anisotropic parameters, we will also combine the above experience to conduct more in-depth exploration.

4

Full Anisotropic Numerical Experiment

Based on the results of the 2-D anisotropic numerical experiments, we found that the spherical parameterization scheme converged the slowest in all experiments, and even in the experiments using gradient descent solver and Newton's method solver, there were cases where the computation did not converge. In comparison, the ABC parameterization scheme converged the fastest in all experiments, and the final value of the objective function met the convergence criteria regardless of which solver was used.

For the solvers, the Levenberg-Marquardt method solver can complete the calculation in the shortest time while ensuring correctness (i.e., satisfying the convergence criteria) in all 2-D anisotropic numerical experiments. Therefore, we believe that the combination of ABC parameterization scheme and the Levenberg-Marquardt method solver is currently the best solution for 2-D anisotropic travel-time tomography problems.

In this chapter, we will increase the difficulty of the numerical experiments by extending the 2-D anisotropic parameters to full anisotropic parameters, and increasing the amount of observed data. We will evaluate the stability of different parameterization methods when solving all four types of anisotropic parameters (i.e., average velocity, anisotropy fraction, azimuth, and elevation), and further investigate the computational performance, correctness, and stability of the cur-

rent best solution.

4.1 SYNTHETIC EXPERIMENT PREPARATION

4.1.1 SYNTHETIC DATA GENERATION AND MODEL PARAMETER SETTING

In terms of observational data, we have considered two types of ray path geometric configurations, aimed at testing the ideal (but realistic) and biased data coverage for a variety of different target models, thereby obtaining more quantitative results. Here is the set-up we would use:

- Ideal coverage (ideal directional coverage): In this setting, the azimuth (0-360°) and elevation (0-90°) use uniform ray sampling, with a sampling interval of about $20°$.

- Biased coverage (limited directional coverage): In this setting, the azimuth (0-360°) uses uniform ray sampling with a sample spacing of about $20°$, but the elevation sampling has biases, considering only $65°$, $75°$, and $85°$, which mimics a teleseismic-like ray geometry where seismic waves propagate near-vertically under the array.

For the two different ray path geometric configurations mentioned above, we have explored different target anisotropic parameter models and initial anisotropic parameter models. For the target anisotropic parameter models, we have defined a total of 13 target models, including 12 anisotropic models and 1 isotropic model, as shown in the Tab. 4.1:

Table 4.1: Settings of target anisotropic model parameters in full anisotropic numerical experiments.

Item	Setting
Target slowness	1/7.5 s/km
Target anisotropic fraction	0 (isotropic); 0.0125; 0.025; 0.05
Target anisotropic azimuth	31°
Target anisotropic elevation	0 (isotropic); 30; 60; 90°

In order to reduce the number of target models, we do not consider different slowness values and anisotropic azimuth. We focus more on the constraint degree of anisotropic fraction and anisotropic elevation, as well as how they are balanced with slowness. There are two initial anisotropic parameter models that need to be studied: one is an isotropic model; the other is an anisotropic

model, but with a large difference from the true solution (i.e., the target and initial anisotropic azimuth are orthogonal). The setup of the initial anisotropic parameter models is shown in the Tab. 4.2:

Table 4.2: Initial anisotropic model parameter settings in full anisotropic numerical experiments.

Item	Setting
Initial slowness	$0.125\,\mathrm{s/km}$
Initial anisotropic fraction	0 (isotropic); 0.0125; 0.025; 0.05
Initial anisotropic azimuth	31°
Initial anisotropic elevation	0 (isotropic); 30; 60; 90°

Similar to the preparation part of the 2-D anisotropic parameters numerical experiment, after defining the synthetic observed data and anisotropic model parameters, we can calculate the observed travel-time and predicted travel-time under different parameterization methods according to Eq. 3.2, Eq. 2.24, and Eq. 3.9. At the same time, a Gaussian noise with a standard deviation of 0.1 is added to the observed travel-time data, then the travel-time residuals are calculated and substituted into Eq. 2.38 to combine the assumed data uncertainty (0.1), and the objective function value can be obtained. To minimize the objective function value, we need to use a solver. Considering that the Levenberg-Marquardt method solver performs best in the 2-D numerical experiment results, we will set the solver to the Levenberg-Marquardt method solver in the following numerical experiments.

4.1.2 STABILITY INSPECTION SCHEME DESIGN

In the previous 2-D anisotropic numerical experiments, we represented the convergence of the numerical calculation by constructing the objective function surface and visualizing it, which was feasible in the 2-D case. However, when the number of anisotropic parameters to be visualized increased from 2 to 4, i.e., when solving all 4 anisotropic parameters, it was not possible to generate the objective function surface (because it is difficult to display the objective function surface in high-dimensional space). Therefore, we directly used numerical indicators, such as the (final) average tolerance, the total number of iterations, and the Euclidean norm (L2 norm) of anisotropic parameters to quantify the experimental results, and produced the 2-D indicator heatmap to display the relevant

results to evaluate the stability of the parameterization scheme and solver. The introduction and calculation method of the relevant numerical indicators are as follows:

- Average tolerance: Here, we take the objective function value obtained from the last iteration (whether it converges or not), and then take the average to obtain the final average tolerance.

- Total number of iterations: The number of iterations used in the numerical calculation when it is completed can be used to evaluate the running time.

- Anisotropic L2 norm: Considering the large number of anisotropic parameters to be evaluated, we convert the other three anisotropic parameters (except for the slowness) from spherical coordinates to Cartesian coordinates, and then calculate the L2 norm (square root of the sum of squared differences) of the predicted anisotropic parameters after coordinate system conversion and the target anisotropic parameters. This helps to reduce the number of anisotropic parameters to be evaluated.

The 2-D indicator heat map is drawn based on the calculation results of the combination of target anisotropic parameters and initial anisotropic parameters. We will explain the content of the 2-D indicator heat map in detail in the subsequent results display section.

4.2 SYNTHETIC NUMERICAL EXPERIMENT TESTING

In order to thoroughly investigate the performance and stability of different parameterization schemes under these combinations of anisotropic parameters, we need to conduct a certain number of numerical experiments. Here, we designed 78 groups of numerical experiments (under one ray path geometric configuration) by combining variable target and initial anisotropic parameters and different parameterization scheme selections, as shown in the Fig. 4.1:

Solver * Parameters * True f * True gamma * Starting f * Starting psi

GN_LM	Spherical	0 (isotropic)	0	0 (isotropic)	0
	Vectoral	0.0125	30		
	ABC	0.025	60	0.025	-59
		0.05	90		

13 target structures in total (12 anisotropic models + 1 isotropic model)

2 starting models to investigate: one isotropic and the other anisotropic

1 * 3 * 13 * 2 = 78 combinations to test for each coverage conditions

Figure 4.1: Details of variable experimental parameters and related experimental times.

For two different ray path geometric configurations, a total of 156 numerical experiments will be run. In this chapter, we will also discuss the experimental results separately based on the different ray path geometric configurations. We will use the Levenberg-Marquardt method solver as the solver for each numerical experiment, and the specific settings of the solver parameters are shown in the Tab. 4.3:

Table 4.3: Parameter settings of the Levenberg-Marquardt method solver in full anisotropic numerical experiments.

Parameter name	Setting content
Opts.SolverFunc	@GaussNetwon
Opts.SolverIterations	100
Opts.ObjFunTolerance	1.01
Opts.StepLength	1
Opts.tf_NormStep	0
Opts.LineSearchIterations	0
Opts.LineSearchStep	0.5
Opts.ForceQuadrant	1
Opts.MaxAbsStep	[]
Opts.FigureObjFunc	0
Opts.GN.Ldamp	0.5
Opts.GN.TolLSQR	1e-6
Opts.GN.MaxIterLSQR	1e3

Please note that here we temporarily disabled the line search function (Opts.LineSearchIterations = 0) and set the damping term to a smaller value (Opts.GN.Ldamp = 0.5). This setting will present some "challenges" for the iterative calculation of the solver. In the following sections, we will evaluate and discuss these two settings.

4.2.1 IDEAL COVERAGE DATA

Firstly, we conducted numerical experiments on the Ideal coverage data and the numerical results of the 78 groups of experiments are shown in the Fig. 4.2:
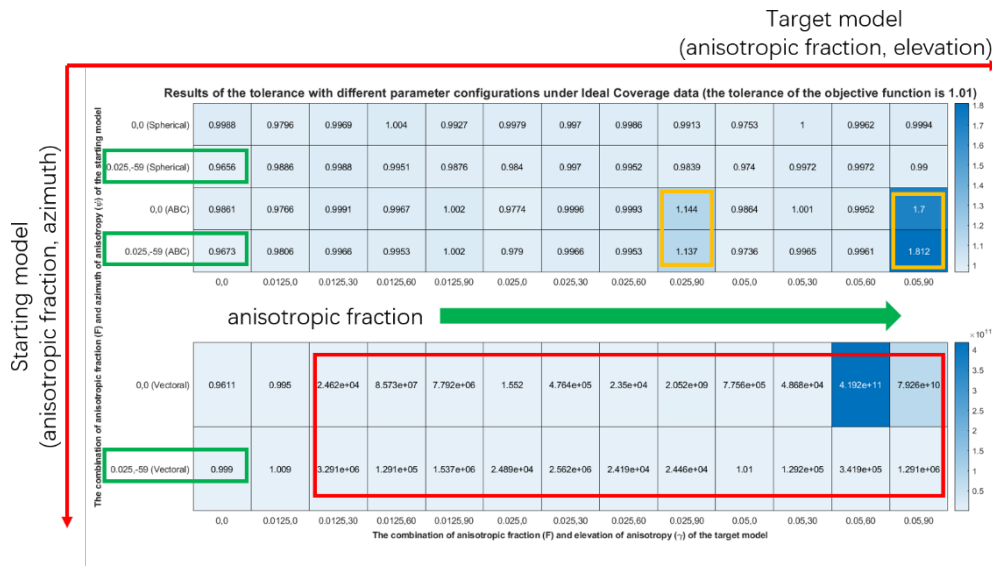
Figure 4.2: Average tolerance of 78 numerical experiments conducted on the ideal coverage data, using the Levenberg-Marquardt method solver.

The above figure shows the results of the average tolerance. The horizontal coordinate represents the variable anisotropic parameters of the target model (anisotropy fraction and anisotropy elevation), arranged in priority according to the variation of anisotropy fraction (green arrow indicates). The vertical coordinate represents the variable anisotropic parameters of the initial model (anisotropy fraction and anisotropy azimuth), and the color bar indicates the size of the average tolerance. From the figure, we can clearly observe the average tolerance of each anisotropic parameter combination experiment. We found that in most anisotropic parameter configurations, the vectoral parameterization scheme performed poorly overall, and some of the average tolerance results were even much greater than the convergence criterion (red boxed part). In comparison, the spherical parameterization scheme performed better in all anisotropic parameter combination tests, and all experimental results met the convergence criterion. For the ABC parameterization scheme, in some specific anisotropic model parameter configurations, its results were not as good as the spherical parameterization scheme (yellow boxed part), but in most anisotropic parameter configurations, the results of the ABC parameterization scheme were similar to those of the spherical parameterization scheme. However, we noticed that when the target model is isotropic and the initial model is anisotropic, the average tolerance result of the Vectoral parameterization scheme is better than the Spherical and ABC parameterization schemes (green boxed part, close to the convergence

criterion).

For the indicator of the total number of iterations, the results are shown in the Fig. 4.3:

Figure 4.3: Total number of iterations for 78 numerical experiments conducted on the ideal coverage data, using the Levenberg-Marquardt method solver.

We can observe that, for most anisotropic parameter configuration combinations, the Vectoral parameterization scheme takes the most time, reaching the maximum total number of iterations, and according to the results of the average tolerance, most of the Vectoral parameterization schemes do not converge in calculation. In comparison, the ABC parameterization scheme takes less time for numerical calculations in most cases compared to the Spherical parameterization scheme. However, for some specific anisotropic model parameter configurations, the ABC parameterization scheme has the most iterations (yellow boxed part).

We calculated the L2 norm of the velocity and the other three anisotropic parameters separately, and the results are shown in the Fig. 4.4:

Figure 4.4: L2 norm results of anisotropic parameters in 78 numerical experiments conducted on ideal coverage data. (a), L2 norm results of velocity; (b), L2 norm results of anisotropic fraction, azimuth, and elevation, using the Levenberg-Marquardt method solver.

From the results of the average tolerance in the previous section, it can be seen that the Vectoral parameterization scheme does not converge in calculation for most model parameter configurations, which leads to the poor performance of the Vectoral parameterization scheme in the L2 norm results of velocity and other three anisotropic parameters (red boxed part), and the final predicted model is far from the target model. In comparison, the Spherical parameterization scheme and the ABC parameterization scheme have better L2 norm re-

sults for the velocity and other three anisotropic parameters. Except for a few anisotropic model parameter combinations that cause the ABC parameterization scheme to be numerically unstable, these two parameterization schemes perform similarly for other model parameter configurations.

4.2.2 BIASED COVERAGE DATA

Next, we conducted another 78 numerical experiments on observation data with finite directional coverage (biased coverage). Finite directional coverage simulates the ray path geometry of earthquake waves transmitted near-vertically below the array, which is more challenging for numerical calculations compared to ideal coverage data. Based on the results of the average tolerance (as shown in the Fig. 4.5), we can directly observe that for most model parameter configurations, the results of the Vectoral parameterization scheme are still poor, and most of the numerical experiment results do not converge (red boxed part).
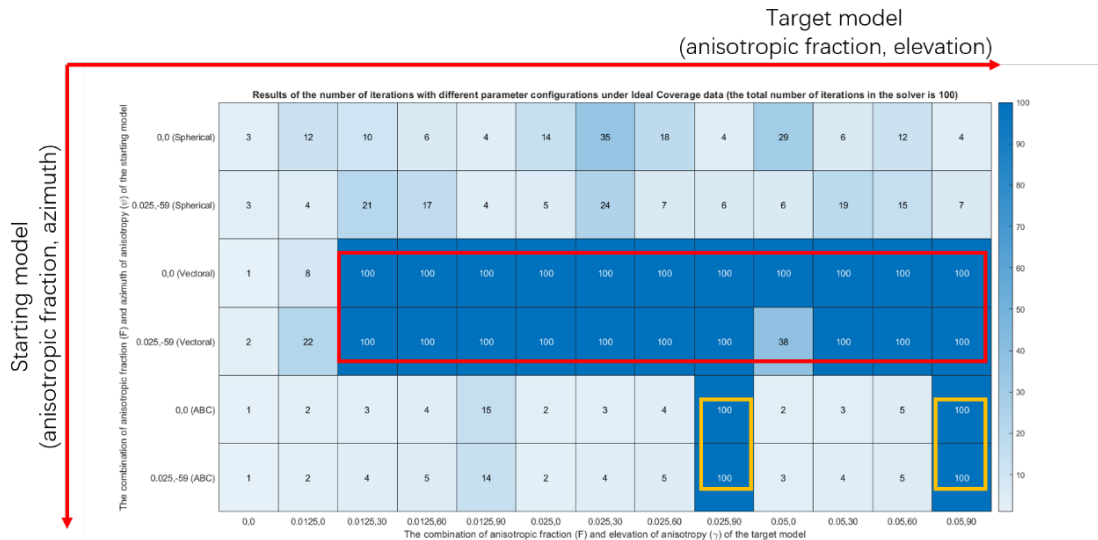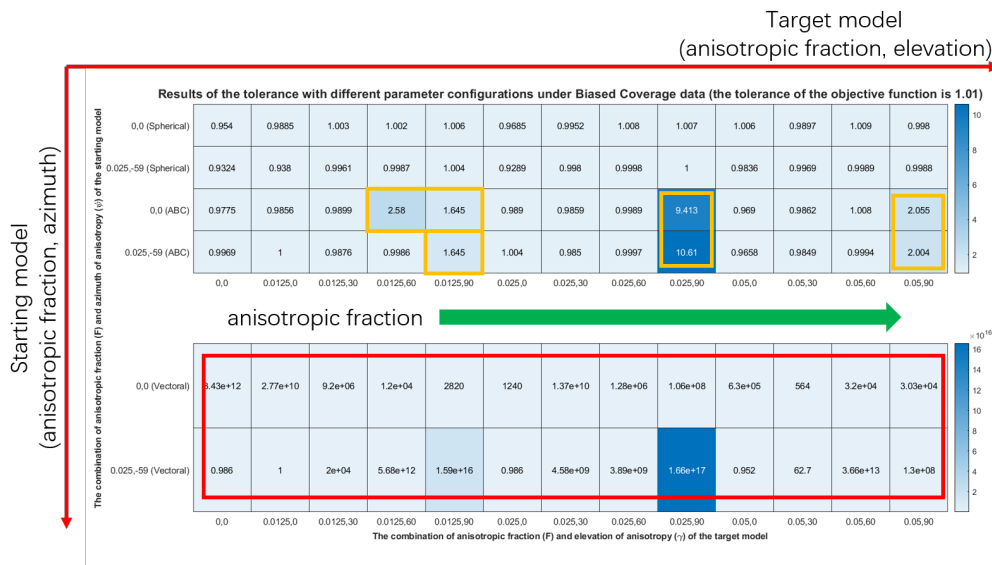

Figure 4.5: Average tolerance of 78 numerical experiments conducted on the biased coverage data, using the Levenberg-Marquardt method solver.

Similarly, as in the case of the ideal coverage experimental results, under certain specific model parameter configurations, the numerical results of the ABC parameterization scheme do not converge (yellow boxed part). However, in most cases, the numerical results of the ABC parameterization scheme meet the convergence criteria. In particular, as the difficulty of the observation data increases, the number of non-converging cases of the ABC parameterization

scheme also increases accordingly, especially when the anisotropic elevation of the target model is 90 degrees (yellow boxed part). In comparison, the Spherical parameterization scheme performs stably, and the results of all model parameter combinations meet the convergence criteria.

Target model (anisotropic fraction, elevation)

Starting model (anisotropic fraction, azimuth)

Results of the number of iterations with different parameter configurations under Biased Coverage data (the total number of iterations in the solver is 100)

The combination of anisotropic fraction (F) and azimuth of anisotropy (υ) of the starting model	0,0	0.0125,0	0.0125,30	0.0125,60	0.0125,90	0.025,0	0.025,30	0.025,60	0.025,90	0.05,0	0.05,30	0.05,60	0.05,90
0,0 (Spherical)	3	2	44	9	14	11	25	9	9	5	16	6	6
0.025,-59 (Spherical)	2	2	38	15	16	3	23	11	11	4	9	24	8
0,0 (Vectoral)	100	100	100	100	100	100	100	100	100	100	100	100	100
0.025,-59 (Vectoral)	1	1	100	100	100	3	100	100	100	6	100	100	100
0,0 (ABC)	3	2	5	100	100	4	4	6	100	4	4	8	100
0.025,-59 (ABC)	1	1.006	25	11	100	3	7	7	100	7	5	6	100

The combination of anisotropic fraction (F) and elevation of anisotropy (γ) of the target model

Figure 4.6: Total number of iterations for 78 numerical experiments conducted on the biased coverage data, using the Levenberg-Marquardt method solver.

From the Fig. 4.6, it can be seen that for the model parameter configurations where the calculation results of the Vectoral and ABC parameterization schemes do not converge, the time overhead of the numerical experiments is the largest (red line and yellow boxed part), especially for the Vectoral parameterization scheme, which reaches the maximum number of iterations for most model configurations. Although the Spherical parameterization scheme has a higher computation time than the ABC parameterization scheme for some specific model parameter configurations, its performance is similar to that of the ABC parameterization scheme for other model parameter configurations, and even outperforms it in some cases.
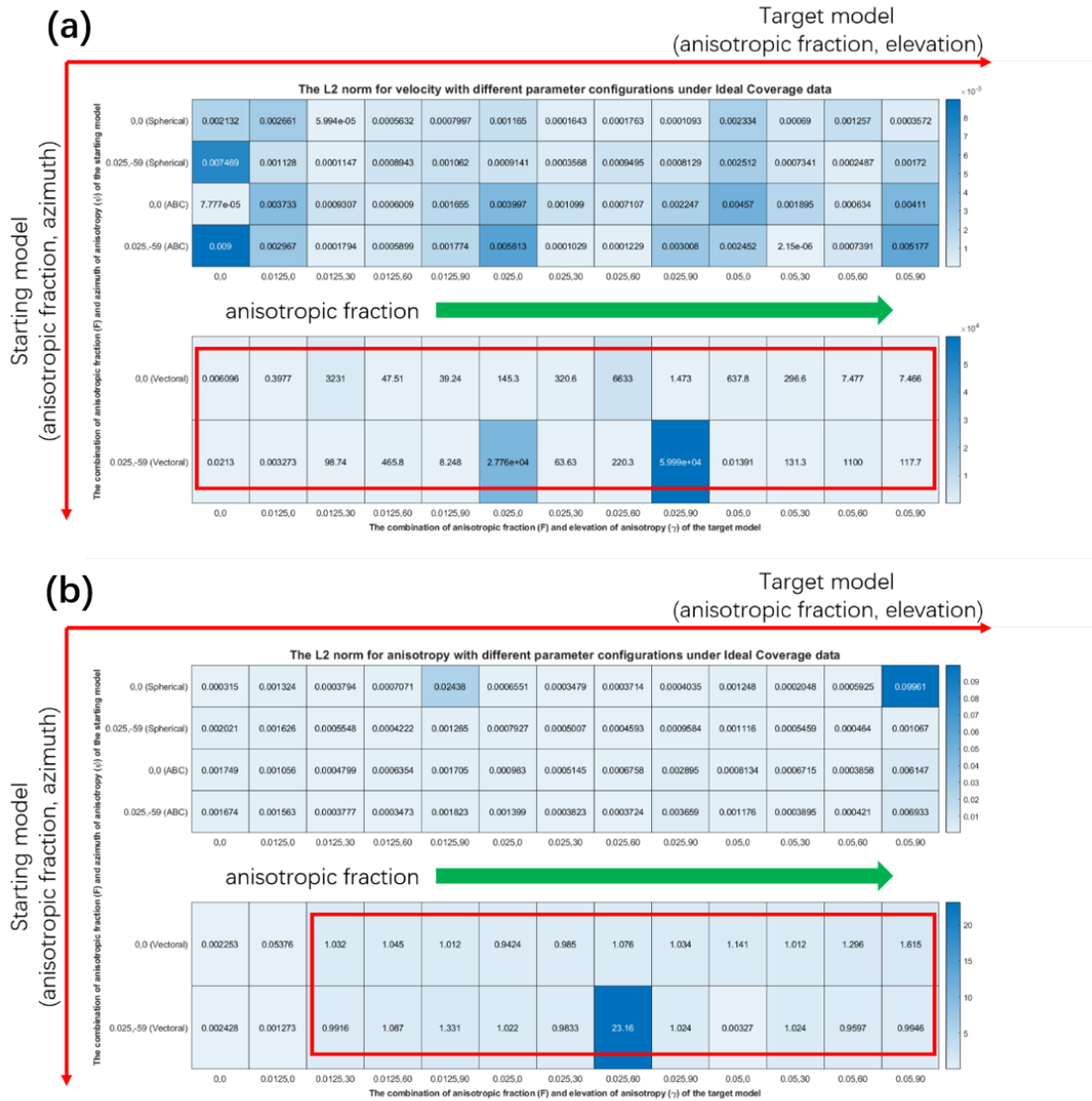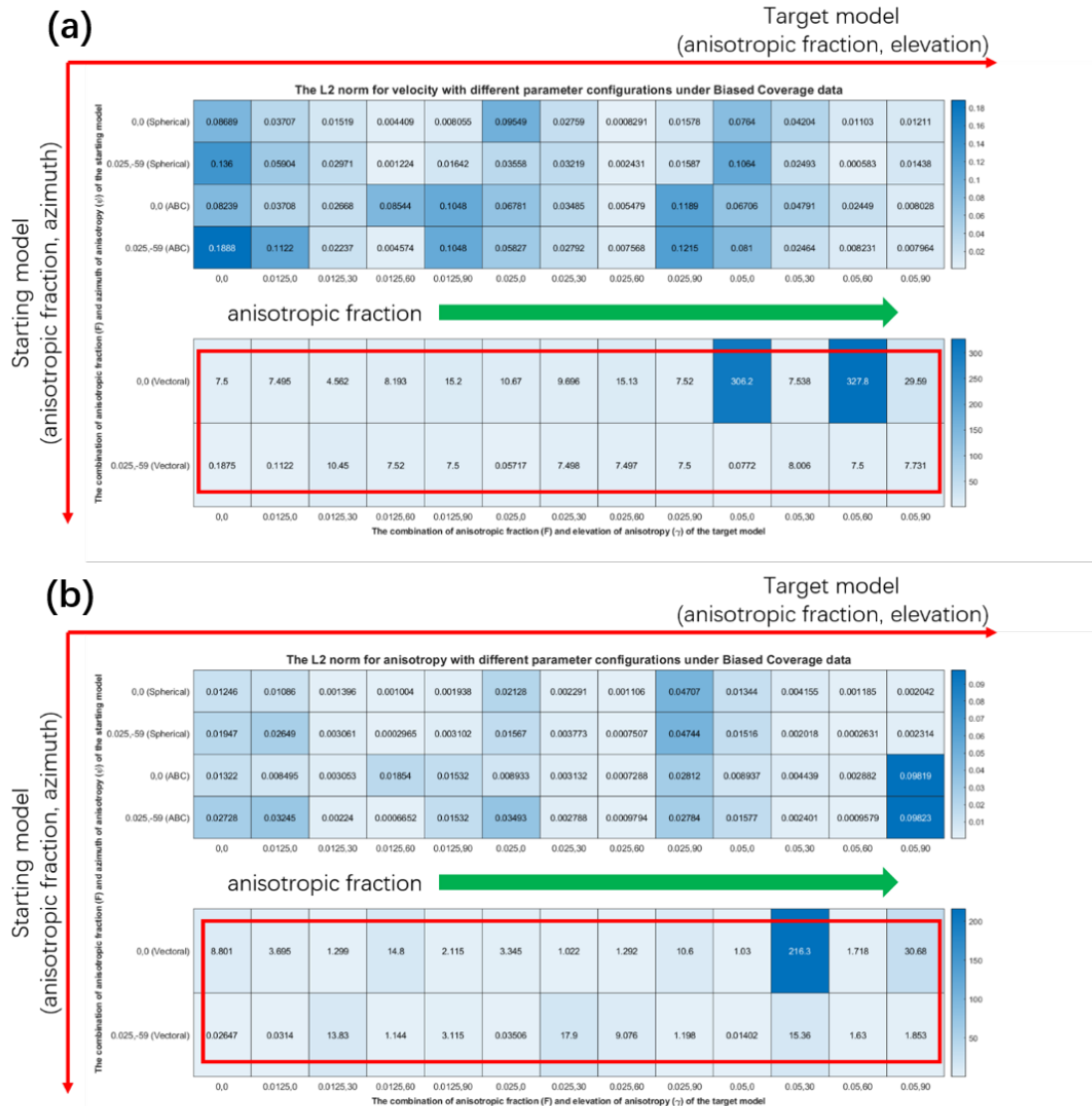
Figure 4.7: L2 norm results of anisotropic parameters in 78 numerical experiments conducted on biased coverage data. (a), L2 norm results of velocity; (b), L2 norm results of anisotropic fraction, azimuth, and elevation, using the Levenberg-Marquardt method solver.

For the L2 norm results (Fig. 4.7) of velocity and anisotropy parameters, similar to the ideal coverage experimental results, the performance of the Vectoral parameterization scheme is still poor (red boxed part), and most of the final predicted models have a large gap from the target model. In comparison, the results of the Spherical parameterization scheme and the ABC parameterization scheme are still better. However, due to the fact that the ABC parameterization scheme does not converge in more anisotropic model parameter combinations,

the performance of the Spherical parameterization scheme is better than that of the ABC parameterization scheme.

4.3 EXPLORATION OF THE SENSITIVITY OF ABC PARAMETERIZATION SCHEME TO SOLVER PARAMETERS

In the numerical experiment of 2-D anisotropic parameterization case, we found that the ABC parameterization scheme is currently the best method for solving 2-D anisotropic travel-time tomography problems. However, in the numerical experiment of the full anisotropic parameterization case, we noticed that the ABC parameterization scheme does not converge in numerical calculations under certain specific model parameter configurations. Considering that we turned off the line search function and set the damping term to a small value before the numerical experiments in this chapter, we designed a new numerical experiment: exploring the sensitivity of the ABC parameterization scheme's numerical calculation results to the solver parameter settings (line search option Opts.LineSearchIterations and damping term Opts.GN.Ldamp) under the premise of using the Levenberg-Marquardt method solver. The experimental observation data is based on ideal coverage data and biased coverage data. Besides the line search option and damping term mentioned above need to be adjusted, all other parameter settings remain unchanged and consistent with the parameter values in the previous numerical experiments.

First, we tested on the observation data of ideal coverage. According to the previous average tolerance results, we found that four model parameter combinations lead to the numerical calculation of the ABC parameterization scheme not converging (yellow boxed part).
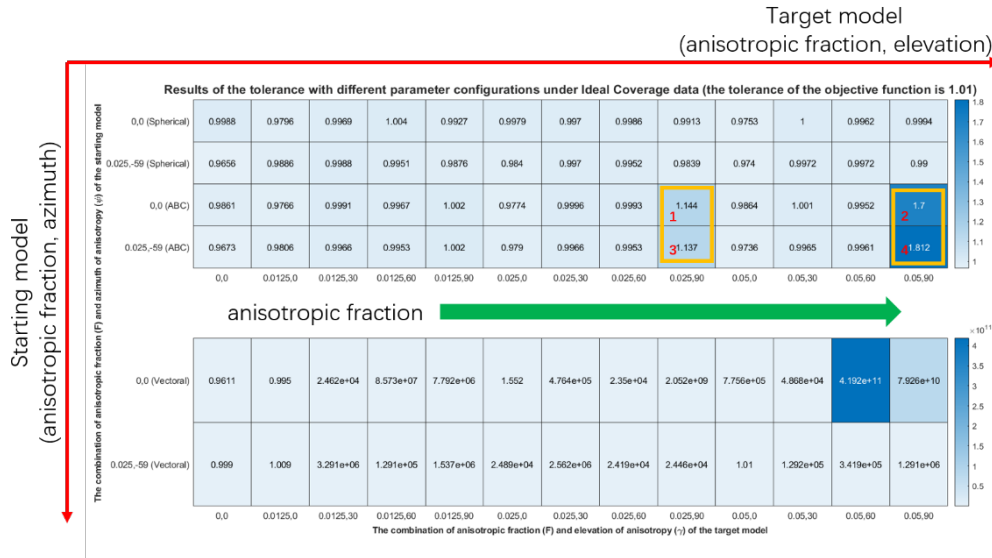
Figure 4.8: The average tolerance of 78 numerical experiments conducted on the observation data of ideal coverage, consistent with the results in the Fig. 4.2.

In the Fig. 4.8, we have numbered four model parameter combinations (red numbers) and tested different line search iteration numbers and damping sizes for each group of model parameters. We have taken uniform samples of the range of line search iteration numbers and damping sizes, where the line search iteration numbers and damping values are sampled uniformly in the range of 0-20, with a sampling interval of 1. Therefore, each group of model parameters has 441 sets of test results, as shown in the Fig. 4.9:
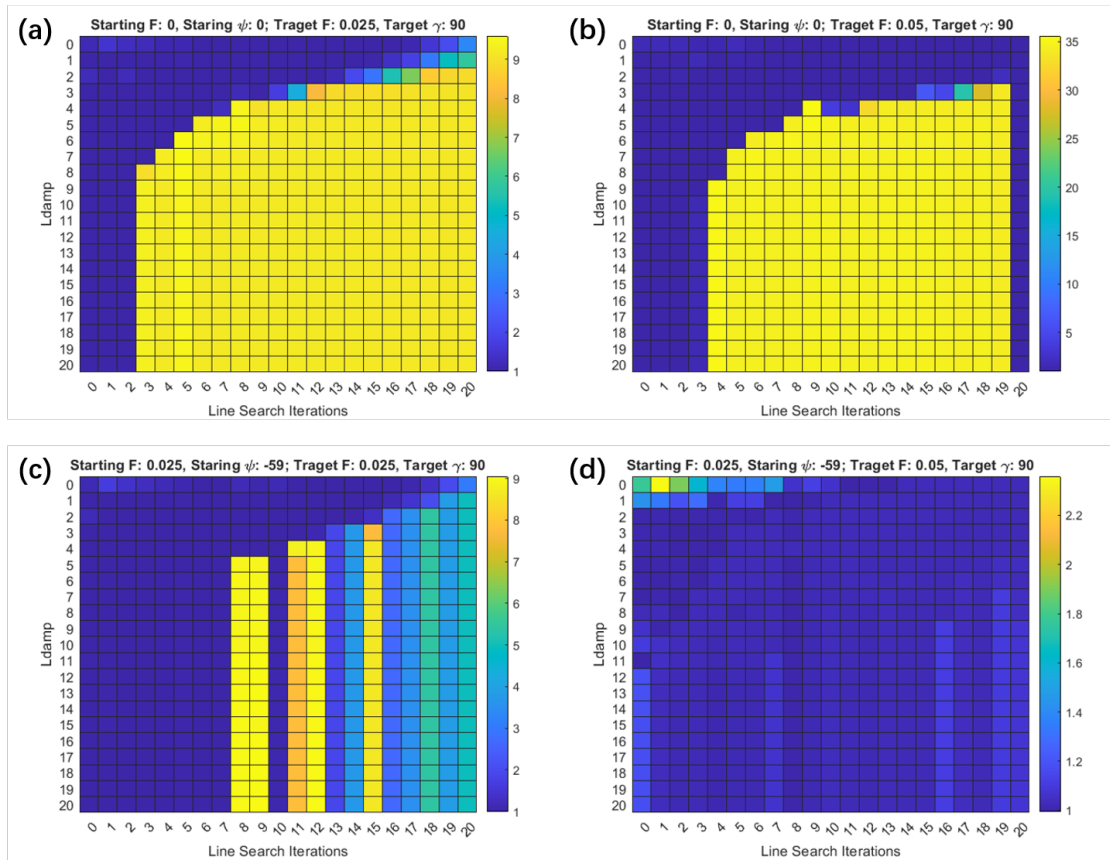
Figure 4.9: Test results of different line search iteration numbers and damping value combinations on the observation data of ideal coverage. The numerical results represent the average tolerance: (a) the result of model parameter configuration 1 in Figure 22; (b) the result of model parameter configuration 2 in Figure 22; (c) the result of model parameter configuration 3 in Figure 22; (d) the result of model parameter configuration 4 in Fig. 4.8.

We present the results of the average tolerance in the form of a 2-D heat map, as shown in the Fig. 4.9. In the figure, the horizontal coordinate represents the line search iteration number, the vertical coordinate represents the damping value, and the color bar represents the size of the average tolerance. It can be seen from the figure that the deeper the color, the closer the result of the average tolerance is to 1. In particular, when the color is dark blue, the numerical experiment results under the corresponding line search iteration number and damping value combination meet the convergence criteria. Therefore, based on the distribution of color depth, we can infer that when the settings of line search iteration number and damping term are small, the ABC parameterization scheme can obtain results that conform to the convergence criteria.

In summary, we have chosen a more suitable solver parameter combination

(Opts.LineSearchIterations, Opts.GN.Ldamp) = (2, 2) to revalidate the performance of the ABC parameterization scheme on the observation data of ideal coverage. The new result of the average tolerance is shown in the Fig. 4.10:

Figure 4.10: The result of the average tolerance after applying line search and larger damping value, using experimental data of observation data with ideal coverage, and the solver used is the Levenberg-Marquardt method solver.

In the Fig. 4.10, ABC-New represents the result of the combination of LineSearchIterations and Ldamp. From the figure, we can discover that properly setting LineSearchIterations and Ldamp can improve the stability of the ABC parameterization scheme (highlighted in the red box), especially compared to the previous four special model parameter combinations (highlighted in the yellow box), the latest result has reached the convergence criteria.

Next, we apply the above test scheme to the observation data of biased coverage, and the solver parameter combination remains consistent with the settings in the above experiment, i.e., using (Opts.LineSearchIterations, Opts.GN.Ldamp) = (2, 2). Finally, the new results of the average tolerance are shown in the Fig. 4.11:
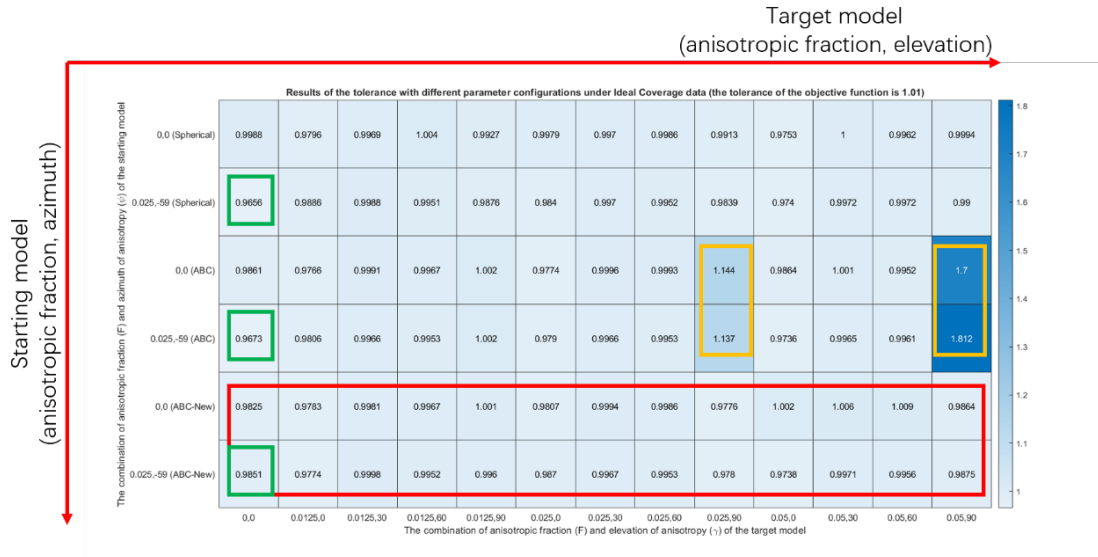
Figure 4.11: The result of the average tolerance after applying line search and larger damping value, using experimental data of observation data with biased coverage, and the solver used is the Levenberg-Marquardt method solver.

From the Fig. 4.11, we can also observe that properly setting LineSearchIterations and Ldamp can improve the stability of the ABC parameterization scheme (highlighted in the red box) on the biased coverage data, especially compared to the previous results of some special model parameter combinations (highlighted in the yellow box), the latest results have reached the convergence criteria. However, except for the part highlighted in the green box, the results still do not meet the convergence criteria, but compared to the previous results, there has been an improvement in performance.

4.4 DISCUSSION AND SUMMARY OF THIS CHAPTER

This chapter mainly introduces the numerical experiments and experimental results of the full anisotropic parameterization case, as well as the optimal solution. Through numerical experiments, we evaluated the stability and computational performance of three anisotropic parameterization methods. On the observation data, we considered two types of ray path geometric configurations and tested the ideal and biased coverage data of various different target models to obtain more quantitative results. At the same time, we also designed a new stability test scheme to determine whether the numerical results converge in high-dimensional space. This chapter also explores the relationship between

the ABC parameterization scheme and the solver parameters. We found that the different constraint degrees of the solver parameters also have different effects on the convergence speed and correctness of the ABC parameterization scheme.

Overall, in the numerical experiments of this chapter, the vectoral parameterization scheme performed poorly. In comparison, the spherical parameterization scheme and the ABC parameterization scheme converge to the required tolerance in most numerical experiments. Among them, the ABC parameterization scheme has an advantage in computational speed (fewer iterations compared to the Spherical parameterization scheme). However, under specific model parameter configurations, the results of the ABC parameterization scheme are not as good as those of the Spherical parameterization scheme.

After imposing constraints on the solver parameters (line search iteration number and damping value), the ABC parameterization scheme performs very well on both ideal coverage data and bias coverage data. In only a very few special cases, the performance of the ABC parameterization scheme after imposing solver parameter constraints is still not as good as that of the Spherical parameterization scheme.

Overall, the ABC parameterization scheme is an excellent method, especially when the related parameters of solver are constrained, and it can achieve the best results in most numerical experiment cases. However, in a few special experimental cases, we will also consider using the Spherical parameterization scheme, even though it takes longer to compute, but the correctness and stability of the results are more important. Although the Spherical parameterization scheme takes longer to compute, its results are stable, and therefore, the role of the spherical parameterization scheme should not be ignored.

Finally, we believe that the ABC parameterization scheme with constraints on the solver parameters is suitable for most experimental cases and should be used for computation as a priority. Only when the computation with the ABC parameterization scheme combined with constraints on the solver parameters cannot converge will we use the Spherical parameterization scheme for computation to ensure the stability of the results. Therefore, the combination of the Levenberg-Marquardt method solver and the ABC & spherical parameterization scheme is

currently the best solution for solving full anisotropic travel-time tomography problems.

5

Conclusions

This paper first elaborates on the research background, methods, and applications of seismic tomography, anisotropy, and anisotropic parameterization methods. In order to investigate which of the three commonly used parameterizations (spherical, vectoral, and ABC) is most suitable for iterative inversion schemes in seismic tomography, we then designed a series of numerical experiments, considering the simple case of constraining two anisotropy parameters, and expanded the experiments to the full anisotropy problem with four constrained anisotropy parameters. We also studied the stability and convergence characteristics of the three parameterizations in P-wave travel-time linearization inversion for both anisotropy problems. Moreover, we explored three methods for minimizing the objective function: gradient descent, Newton, and Levenberg-Marquardt, and selected the most suitable solver based on experimental results. Finally, we summarized the stability and performance of parameterizations under various anisotropy models and different initial conditions.

The results of the 2-D anisotropic numerical experiments showed that except for a few experiments that did not converge, the other experiments all met the solver's convergence criteria. Among the parameterization schemes, the ABC parameterization had the best computational performance, while the spherical parameterization had the worst. For solvers, the undamped Gauss-Newton solver and the Levenberg-Marquardt method solver were able to complete the calculations within a certain number of iterations and meet the convergence criteria. The Levenberg-Marquardt method solver showed stronger computational

performance under the premise of meeting the convergence criteria. Therefore, we believe that the combination of the Levenberg-Marquardt method solver and ABC parameterization is the best solution for 2-D anisotropic travel-time tomography problems.

The experimental results of the full anisotropic numerical experiments showed that the vectoral parameterization performed poorly, while the spherical parameterization and ABC parameterization converged to the required tolerance in most experiments. The ABC parameterization had an advantage in computational speed, but under specific model parameter configurations, the results were not as good as those of the spherical parameterization. When the solver parameters were constrained, the ABC parameterization performed well in terms of ideal and bias coverage rates. However, in a few special cases, the performance of the constrained ABC parameterization was still worse than that of the spherical parameterization. In summary, the ABC parameterization is an excellent method that can obtain the best results in most numerical experiment cases when the solver-related parameters are constrained. However, in a few special experimental cases, the spherical parameterization will also be considered. Therefore, the combination of the Levenberg-Marquardt method solver and the ABC & spherical parameterization is currently the best solution for solving full anisotropic travel-time tomography problems. In practical applications, it is recommended to use the constrained ABC parameterization for calculations first, and only use the spherical parameterization when convergence fails to ensure stability in the results.

For the future works, there is significant potential for optimization and enhancement of anisotropic numerical experiments. This particularly pertains to identifying the reasons behind the failure (non-convergence) of certain cases in partially anisotropic parameterization applications, and devising and testing additional strategies to overcome these challenges. For instance, adopting uncertainty estimation methods such as perturbational uncertainty estimation or gradient-based uncertainty estimation could help assess the robustness of the tomography model and pinpoint the causes of results non-convergence in anisotropic parameterization applications. Moreover, efforts can be made to refine the current parameterization method selection scheme by further optimizing the solver parameters and the parameterization selection process itself.

References

1. Zhao, D. New advances of seismic tomography and its applications to sub-duction zones and earthquake fault zones: A review. *Island Arc* **10,** 68–84 (2001).

2. Aki, K., Christoffersson, A., Husebye, E. & Powell, C. Three-dimensional seismic velocity anomalies in the crust and upper-mantle under the USGS, California seismic array. *Eos Trans. AGU* **56,** 1145 (1974).

3. Aki, K. & Lee, W. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model. *Journal of Geophysical research* **81,** 4381–4399 (1976).

4. Wapenaar, K. A breviary of seismic tomography: imaging the interior of the earth and the sun. *Journal of the Acoustical Society of America* **126,** 2129–2130 (2008).

5. Dziewonski, A. M., Hager, B. H. & O'Connell, R. J. Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research* **82,** 239–255 (1977).

6. Dziewonski, A. M. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6. *Journal of Geophysical Research: Solid Earth* **89,** 5929–5952 (1984).

7. Zhao, D. Multiscale seismic tomography and mantle dynamics. *Gondwana Research* **15,** 297–323 (2009).

8. Roth, E. G., Wiens, D. A. & Zhao, D. An empirical relationship between seismic attenuation and velocity anomalies in the upper mantle. *Geophysical Research Letters* **27,** 601–604 (2000).

9. Dahlen, F., Hung, S.-H. & Nolet, G. Fréchet kernels for finite-frequency traveltimes—I. Theory. *Geophysical Journal International* **141,** 157–174 (2000).

10. Shapiro, N. M., Campillo, M., Stehly, L. & Ritzwoller, M. H. High-resolution surface-wave tomography from ambient seismic noise. *Science* **307,** 1615–1618 (2005).

11. Eberhart-Phillips, D. & Mark Henderson, C. Including anisotropy in 3-D velocity inversion and application to Marlborough, New Zealand. *Geophysical journal international* **156,** 237–254 (2004).

12. Mainprice, D. 2.16 Seismic anisotropy of the deep Earth from a mineral and rock physics perspective. *Treatise on Geophysics, Volume 2: Mineral Physics,* 437 (2010).

13. Karato, S.-i., Jung, H., Katayama, I., Skemer, P., *et al.* Geodynamic significance of seismic anisotropy of the upper mantle: New insights from laboratory studies. *Annual Review of Earth and Planetary Sciences* **36,** 59–95 (2008).

14. Helbig, K. & Thomsen, L. 75-plus years of anisotropy in exploration and reservoir seismics: A historical review of concepts and methods. *Geophysics* **70,** 9ND–23ND (2005).

15. Anderson, D. L. Elastic wave propagation in layered anisotropic media. *Journal of Geophysical Research* **66,** 2953–2963 (1961).

16. Backus, G. E. Possible forms of seismic anisotropy of the uppermost mantle under oceans. *Journal of Geophysical Research* **70,** 3429–3439 (1965).

17. Forsyth, D. W. The early structural evolution and anisotropy of the oceanic upper mantle. *Geophysical Journal International* **43,** 103–162 (1975).

18. Crampin, S. & King, D. W. Evidence for anisotropy in the upper mantle beneath Eurasia from the polarization of higher mode seismic surface waves. *Geophysical Journal International* **49,** 59–85 (1977).

19. Bamford, D. Pn velocity anisotropy in a continental upper mantle. *Geophysical Journal International* **49,** 29–48 (1977).

20. Simons, F. J., van der Hilst, R. D. & Zuber, M. T. Spatiospectral localization of isostatic coherence anisotropy in Australia and its relation to seismic anisotropy: Implications for lithospheric deformation. *Journal of Geophysical Research: Solid Earth* **108** (2003).

21. Audet, P. & Mareschal, J.-C. Anisotropy of the flexural response of the lithosphere in the Canadian Shield. *Geophysical Research Letters* **31** (2004).

22. Fouch, M. J. & Rondenay, S. Seismic anisotropy beneath stable continental interiors. *Physics of the Earth and Planetary Interiors* **158,** 292–320 (2006).

23. Maupin, V., Park, J., Romanowicz, B. & Dziewonski, A. Theory and observations—wave propagation in anisotropic media. *Seismology and the Structure of the Earth. Treatise on Geophysics* **1,** 289–321 (2007).

24. Long, M. D. & Becker, T. W. Mantle dynamics and seismic anisotropy. *Earth and Planetary Science Letters* **297,** 341–354 (2010).

25. Bamford, D. & Crampin, S. Seismic anisotropy–the state of the art. *Geophysical Journal International* **49,** 1–8 (1977).

26. Silver, P. G. & Chan, W. W. Shear wave splitting and subcontinental mantle deformation. *Journal of Geophysical Research: Solid Earth* **96,** 16429–16454 (1991).

27. Silver, P. G. Seismic anisotropy beneath the continents: Probing the depths of geology. *Annual review of earth and planetary sciences* **24,** 385–432 (1996).

28. Zimmerman, M. E., Zhang, S., Kohlstedt, D. L. & Karato, S.-i. Melt distribution in mantle rocks deformed in shear. *Geophysical Research Letters* **26,** 1505–1508 (1999).

29. Walker, K. T., Nyblade, A. A., Klemperer, S. L., Bokelmann, G. H. & Owens, T. J. On the relationship between extension and anisotropy: Constraints from shear wave splitting across the East African Plateau. *Journal of Geophysical Research: Solid Earth* **109** (2004).

30. Ji, S., Wang, Q. & Xia, B. *Handbook of seismic properties of minerals, rocks and ores* (Presses inter Polytechnique, 2002).

31. Mainprice, D. & Silver, P. G. Interpretation of SKS-waves using samples from the subcontinental lithosphere. *Physics of the Earth and Planetary Interiors* **78,** 257–280 (1993).

32. Holtzman, B. *et al.* Melt segregation and strain partitioning: Implications for seismic anisotropy and mantle flow. *Science* **301,** 1227–1230 (2003).

33. Long, M. D. & Silver, P. G. Shear wave splitting and mantle anisotropy: Measurements, interpretations, and new directions. *Surveys in Geophysics* **30,** 407–461 (2009).

34. Ando, M., Ishikawa, Y. & Yamazaki, F. Shear wave polarization anisotropy in the upper mantle beneath Honshu, Japan. *Journal of Geophysical Research: Solid Earth* **88,** 5850–5864 (1983).

35. Hirahara, K. Detection of three-dimensional velocity anisotropy. *Physics of the earth and planetary interiors* **51,** 71–85 (1988).

36. Savage, M. Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting? *Reviews of Geophysics* **37,** 65–106 (1999).

37. Hess, H. Seismic anisotropy of the uppermost mantle under oceans. *Nature* **203,** 629–631 (1964).

38. Hearn, T. M. Anisotropic Pn tomography in the western United States. *Journal of Geophysical Research: Solid Earth* **101,** 8403–8414 (1996).

39. Montagner, J.-P. & Nataf, H.-C. A simple method for inverting the azimuthal anisotropy of surface waves. *Journal of Geophysical Research: Solid Earth* **91,** 511–520 (1986).

40. Koulakov, I., Jakovlev, A. & Luehr, B. G. Anisotropic structure beneath central Java from local earthquake tomography. *Geochemistry, Geophysics, Geosystems* **10** (2009).

41. Wang, J. & Zhao, D. P-wave anisotropic tomography beneath Northeast Japan. *Physics of the Earth and Planetary Interiors* **170,** 115–133 (2008).

42. Zhao, D., Hasegawa, A. & Horiuchi, S. Tomographic imaging of P and S wave velocity structure beneath northeastern Japan. *Journal of Geophysical Research: Solid Earth* **97,** 19909–19928 (1992).

43. Wang, J. & Zhao, D. P-wave tomography for 3-D radial and azimuthal anisotropy of Tohoku and Kyushu subduction zones. *Geophysical Journal International* **193,** 1166–1181 (2013).

44. Munzarová, H., Plomerová, J. & Kissling, E. Novel anisotropic teleseismic body-wave tomography code AniTomo to illuminate heterogeneous anisotropic upper mantle: Part I—Theory and inversion tuning with realistic synthetic data. *Geophysical Journal International* **215,** 524–545 (2018).

45. Wang, Z. & Zhao, D. 3D anisotropic structure of the Japan subduction zone. *Science advances* **7,** eabc9620 (2021).

46. VanderBeek, B. P. & Faccenda, M. Imaging upper mantle anisotropy with teleseismic P-wave delays: insights from tomographic reconstructions of subduction simulations. *Geophysical Journal International* **225,** 2097–2119 (2021).

47. Cervenỳ, V. *Seismic ray theory* (Cambridge university press Cambridge, 2001).

48. Thurber, C. H. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California. *Journal of Geophysical Research: Solid Earth* **88,** 8226–8236 (1983).

49. Julian, B., Gubbins, D., *et al.* Three-dimensional seismic ray tracing. *Journal of Geophysics* **43,** 95–113 (1977).

50. Um, J. & Thurber, C. A fast algorithm for two-point seismic ray tracing. *Bulletin of the Seismological Society of America* **77,** 972–986 (1987).

51. Gordon, R., Bender, R. & Herman, G. T. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology* **29,** 471–481 (1970).

52. Trampert, J. & Leveque, J.-J. Simultaneous iterative reconstruction technique: physical interpretation based on the generalized least squares solution. *Journal of Geophysical Research: Solid Earth* **95,** 12553–12559 (1990).

53. Paige, C. C. & Saunders, M. A. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)* **8,** 43–71 (1982).

54. Shearer, P. M. *Introduction to seismology* (Cambridge university press, 2019).

55. Barclay, A. H., Toomey, D. R. & Solomon, S. C. Seismic structure and crustal magmatism at the Mid-Atlantic Ridge, 35° N. *Journal of Geophysical Research: Solid Earth* **103,** 17827–17844 (1998).

56. Engdahl, E. & Lee, W. Relocation of local earthquakes by seismic ray tracing. *Journal of Geophysical Research* **81,** 4400–4406 (1976).

57. Becker, T. W., Chevrot, S., Schulte-Pelkum, V. & Blackman, D. K. Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models. *Journal of Geophysical Research: Solid Earth* **111** (2006).

58. Thomsen, L. Weak elastic anisotropy. *Geophysics* **51,** 1954–1966 (1986).

59. Babuška, V. & Plomerová, J. European mantle lithosphere assembled from rigid microplates with inherited seismic anisotropy. *Physics of the Earth and Planetary Interiors* **158,** 264–280 (2006).

60. Avriel, M. *Nonlinear programming: analysis and methods* (Courier Corporation, 2003).

Acknowledgments

As I complete my master's thesis, I would first like to express my sincere gratitude to my academic advisor, Professor Manuele Faccenda. Before starting my postgraduate studies in early 2020, I contacted Professor Manuele Faccenda and became deeply interested in his research direction. During my postgraduate studies, I communicated with Professor Manuele Faccenda frequently and attended his courses. His rigorous academic style and broad knowledge have benefited me greatly. He not only provides me with meticulous guidance in professional studies, but also offers a lot of care and help in my daily life and PhD application. Here, I would like to express my heartfelt thanks to you! Under your teaching and planning, I have come to love the field of numerical computation in geoscience and firmly embarked on the road of scientific research.

Secondly, I would like to thank Dott. Brandon VanderBeek. Under his guidance, I gradually became familiar with and mastered anisotropy tomography methods and numerical optimization methods, and gained a deeper understanding of mathematical derivations and code implementation. During the completion of my master's thesis, I often communicated with him to discuss issues. He always managed to explain complex problems clearly and concisely with his simple and lucid language, which was of great help to me.

In addition, I would like to thank the teachers and classmates of the Department of Geosciences, University of Padua, for accompanying me in my learning and growth. It is your company that has allowed me to experience a wonderful study life abroad.

Finally, I would like to thank my parents and friends for their care and support. No matter where I am overseas, whenever I encounter difficulties in life, they

always lend a helping hand in time. Their concern has allowed me to devote myself wholeheartedly to course learning and master's thesis completion.

Although I am about to leave the University of Padua and Italy, this does not mean the end of my learning and research journey. On the contrary, my research journey has just begun. In my future doctoral career, I will continue to work hard and not let everyone's help and care down. In the new studying and living environment, I will also miss the wonderful times spent in Padua!