COMMUN. MATH. SCI.
Vol. 20, No. 6, pp. 1551–1565

CONVERGENCE ANALYSIS ON SEISMIC TOMOGRAPHY FOR INVERSE PROBLEMS OF ACOUSTIC WAVE PROPAGATION*

HAOYU WANG†, LIHUI CHAI‡, ZHONGYI HUANG§, AND XU YANG¶

Abstract. In this paper, we provide a rigorous convergence analysis for the inverse problems of acoustic wave propagation arising from seismic tomography. Specifically, we obtain the error estimates for three cases: 1. Standard seismic tomography; 2. Seismic tomography with approximation to sensitivity kernel; 3. Seismic tomography with Tikhonov regularization.

Keywords. Seismic tomography; acoustic wave equation; convergence analysis.

AMS subject classifications. 65K99; 65M32; 86A15; 86A22.

1. Introduction

Seismic tomography has been widely used for imaging the subsurface structures of the Earth at a variety of scales, including but not limited to, plate tectonics, volcanism, and geodynamics [1, 6, 13, 16, 23]. The seismic image can be obtained by minimizing a misfit functional which measures the difference of the observed seismic data and that from synthetic simulations. The travel-time tomography based on ray theory is the main tool used at the beginning of the development of the methodology, where it assumes that the travel-time information is only determined by the infinitely thin ray paths, but this is only valid for the extremely high-frequency regime [15].

On the other hand, the wave-equation-based seismic tomography [20–22] is considered to be able to include the influence of off-ray structures that are missing in the ray theory. The adjoint method [8, 13, 21] is one of the most widely used tools in the wave-equation-based tomography. In this case, the minimization problem may be solved iteratively by the Newton method or other gradient-based methods. Nevertheless, these methods require the computation of the Fréchlet derivatives of the misfit function, or in other words, the sensitivity kernels, which are computed by solving synthetic wave equations [19]. The sensitivity kernel can be written in the form of a cross-correlation of the so-called forward and adjoint wavefields [21]. We remark that, although it provides accurate synthetic seismograms and sensitivity kernels, fully solving wave-equations in 3D is usually time demanding and even computationally unaffordable. The recent development of modern scientific computing and numerical methods for seismic wave equations, such as the pseudo-spectral method [9, 12] and hybrid method [20], make the wave-equation-based seismic tomography feasible and successfully applicable in a variety of realistic situations [11, 17, 18, 24].

In previous works [4, 5, 10], the authors have developed the frozen Gaussian approximation (FGA)-based seismic tomography method for high-frequency seismic wave imaging. The FGA is an asymptotic method which approximates the seismic wave-

*Received: April 10, 2021; Accepted (in revised form): January 16, 2022. Communicated by Gang Bao.

†Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China (wanghaoy17@mails.tsinghua.edu.cn).

‡School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China (chailihui@mail.sysu.edu.cn).

§Corresponding author. Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China (zhongyih@tsinghua.edu.cn).

¶Department of Mathematics, University of California, Santa Barbara, CA 93106, USA (xuyang@math.ucsb.edu).

field by a superposition of fixed-width Gaussian wavepackets where the Gaussians are determined by a set of decoupled ordinary differential equation (ODE) systems propagating along the classical ray paths and thus can be solved drastically fast. The rigorous asymptotic analysis [3,14] guarantees the accurate approximation of the wavefield in the high-frequency regime. Thus, the FGA may serve as an efficient and accurate forward modeling solver for seismic wave propagation and computing the sensitivity kernels. In [5,10], FGA is applied to wave-equation-based seismic tomography using the Newton method with synthetic tests on both travel-time and full-waveform as the observed data.

Despite the fact that the wave-equation-based seismic tomography has shown its successes in many realistic situations and FGA has been tested as a promising forward modeling solver for computing the wavefields and sensitivity kernels, the convergence property of seismic tomography is still unclear. First, due to the complexity of the seismic wave equations, to the best of the authors' knowledge, there are no systematic studies on the convergence rate of the iterative methods; second, numerical approximation such as FGA gives an asymptotic solution to acoustic wave propagation with an error proportional to the ratio of the wave-length over the domain size, and it is unclear how this kind of error will affect the total accuracy of seismic tomography.

In this paper, we analyze the convergence of seismic tomography for the following three cases: 1. standard setup; 2. with approximation to sensitivity kernel; 3. with Tikhonov regularization. By a careful estimate of the minimization problem, we show that the method converges linearly to a velocity model, and gives an explicit dependence of the accuracy on the error of observed data and numerical approximation.

The rest of the paper is organized as follows: In Section 2, we describe the standard setup of seismic tomography. We present the accuracy estimate for standard seismic tomography in Section 3 and seismic tomography with numerical approximation to sensitivity kernel in Section 4. In Section 5, we analyze the convergence of seismic tomography with Tikhonov regularization. We give conclusive discussions in Section 6.

2. Acoustic wave propagation and standard seismic tomography

We consider the inverse problem for the acoustic wave propagation modeled by

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} u = f(t,\boldsymbol{x}), \\ u|_{t=0} = 0, \\ \frac{\partial u}{\partial t}|_{t=0} = 0, \end{cases} \tag{2.1}$$

where c is the velocity model, u is the wavefield, f is the external forcing term and usually takes the form of $f(t,\boldsymbol{x}) = s(t)\delta_d(\boldsymbol{x}-\boldsymbol{x}_s)$ with s being the source-time function at position $\boldsymbol{x}_s$, and δ_d being the Dirac delta function. We assume that the equation in (2.1) is valid in $t \in (0,T], \boldsymbol{x} \in \Omega$, where Ω is a smooth bounded domain. As for the boundary condition, since the velocity is bounded, we can enlarge Ω so that the boundary condition will not affect the domain in concern, at least before $t = T$. We can then assume the null Dirichlet boundary, i.e. $u(t,x) = 0$, for all $t \in [0,T], x \in \partial\Omega$. [1]

Denote d_{obs} as the recorded seismic data at receivers and $d_{\text{sys}}(c)$ as the synthetic data by simulation. The inverse problem aims to seek a velocity model c to minimize the following misfit function $\chi(c)$ under a certain metric (*e.g.* the discrepancy of travel-time and full waveforms as studied in [5,10]),

$$\min_{\text{velocity } c} \|d_{\text{obs}} - d_{\text{sys}}(c)\|_{L^2}, \tag{2.2}$$

[1]We should point out that, in practical cases where Ω is not enlarged, the boundary condition does affect the whole equation, and some techniques like the Perfectly Matching Layer (PML) should be applied.

where we denote the velocity model as c, and the corresponding wavefield $u(t,\boldsymbol{x})=u(t,\boldsymbol{x};c)$. Let $d_{sys}=d_{sys}(c)$ be the synthetic data computed numerically (*e.g.*, by Frozen Gaussian Approximation [4]), which can be either the travel-time T_{sys} for travel-time adjoint tomography, or $u_{sys}(t,\boldsymbol{x}_r;c)$, for full waveform inversion. We also denote the perfect (error-free) data as $d_{exact}=d_{exact}(c)$, given by the analytical solution to (2.1).

The observed data is limited to the waveform(s) $u(t,\boldsymbol{x}_r)$ at certain receivers $\boldsymbol{x}_r$. The travel-time at $\boldsymbol{x}_r$ means the infimum of time that $u(t,\boldsymbol{x}_r)$ is nonzero. Ideally, one aims to recover the true velocity model c^* corresponding to the observation data, and $\|c-c^*\|_{L^2}$ describes the accuracy of the solution to the inverse problem (2.2). Here L^2 can be replaced by other norms of Hilbert spaces, such as $W^{3,2}(\Omega)$.

We consider the full waveform inversion case, i.e., $d_{exact}(t;c)=u(t,\boldsymbol{x}_r;c)$ and $d_{sys}(t;c)=u_{sys}(t,\boldsymbol{x}_r;c)$. Both d_{exact} and d_{sys} can be viewed as functionals mapping $c\in W^{3,2}(\Omega)$ to a function $(t\mapsto d(t;c))\in L^2([0,T])$, so we can define its Frechet derivatives. We suppose that, the second variation of d_{exact} and d_{sys} exists, and that

$$\begin{aligned}&\left\|\frac{\delta^2 d_{exact}}{\delta c^2}\right\|\le C_2,\quad \left\|\frac{\delta^2 d_{sys}}{\delta c^2}\right\|\le C_2;\quad \left\|\frac{\delta d_{exact}}{\delta c}\right\|\le C_1,\quad \left\|\frac{\delta d_{sys}}{\delta c}\right\|\le C_1;\\ &\left\|\left(\left(\frac{\delta d_{exact}}{\delta c}\right)^*\left(\frac{\delta d_{exact}}{\delta c}\right)\right)^{-1}\right\|\le C_{-1},\quad \left\|\left(\left(\frac{\delta d_{sys}}{\delta c}\right)^*\left(\frac{\delta d_{sys}}{\delta c}\right)\right)^{-1}\right\|\le C_{-1},\end{aligned} \tag{2.3}$$

where the $\|\cdot\|$ refers to the induced operator norm, and C_2, C_1, and C_{-1} are constants independent of c. The $*$ on an operator means the adjoint operator, i.e. $F:=\frac{\delta d}{\delta c}, F^*(\delta c_1)(\delta d_2)=(F(\delta c_1),\delta d_2)_{L^2([0,T])}$. Unless otherwise specified, for a linear operator between Hilbert spaces, $\|\cdot\|$ always refers to the induced operator norm. Assumption (2.3) is a strong assumption. It describes the well-posedness of the equation, including the continuous dependence on c, which is needed in the convergence analysis when solving c from our data. The bound C_{-1} requires that enough data should be recorded to recover c^*. For a linear operator F, F^*F maps δc_1 to $F^*F(\delta c_1)$ which satisfies $(F^*F(\delta c_1),\delta c_2)=(F(\delta c_1),F(\delta c_2))$. We will also denote F^*F as (F,F).

We shall justify the assumptions above before our analysis. Generally, those constants would not exist if the space of c were too arbitrary. It is this reason that we choose to consider a special family of the velocity models. We assume that the velocity model depends on finite number of parameters. We take $c(\boldsymbol{x})=c(\boldsymbol{x};\theta_1,\theta_2,\ldots,\theta_m)$, and we assume that c depends on them "uniformly smoothly" of order 3, in the sense that

$$\|c\|_{W^{3,\infty}(\Omega)}\le C_{m0},\quad \left\|\frac{\partial c}{\partial\theta_j}\right\|_{W^{2,\infty}(\Omega)}\le C_{m1},\quad \left\|\frac{\partial^2 c}{\partial\theta_j\partial\theta_k}\right\|_{W^{1,\infty}(\Omega)}\le C_{m2}, \tag{2.4}$$

where the constants C_{m0}, C_{m1}, C_{m2} do not depend on all θ_j. We also need to assume that the velocity should not be too high or too low, i.e. $c_{max}\ge c(\boldsymbol{x})\ge c_{min}>0$ uniformly, where c_{max} and c_{min} are known priori.

It is the nature of wave equation (or other hyperbolic equation) that the solution has poor regularity unless the initial condition, the boundary condition and the external force are all sufficiently smooth and possess high order compatibility. Thus, we replace the external force in (2.1) by f, which is assumed to be smooth in order 3 in timespace,

$$\begin{aligned}&f\in L^2\left([0,T];H^3(\Omega)\right), f_t\in L^2\left([0,T];H^2(\Omega)\right),\\ &f_{tt}\in L^2\left([0,T];H^1(\Omega)\right), f_{ttt}\in L^2\left([0,T];L^2(\Omega)\right).\end{aligned} \tag{2.5}$$

When we say $f \in H^3([0,T]\times\Omega)$, we actually refer to the four conditions above. And also, when $H^3([0,T]\times\Omega)$ is taken as a norm in an inequality, we actually mean the sum of the four norms above.

For the sake of compatibility, we also assume that the support of $f(t,\boldsymbol{x})$ is compact and lies inside $\left\{(t,\boldsymbol{x}) : t\in(0,T], \overline{B_{\boldsymbol{x}}(c_{max}T - c_{max}t)} \subset \Omega\right\}$. In this case, with null initial and null Dirichlet boundary conditions, compatibility is assured on arbitrary order, at least up to time T.

It is well known (*e.g.*, [7]) that the energy estimation

$$\sup_{t\in[0,T]}\left(\|u\|_{H^1(\Omega)} + \|u_t\|_{L^2(\Omega)}\right) \le C\|f\|_{L^2([0,T];L^2(\Omega))}, \tag{2.6}$$

and higher regularity estimation

$$\sup_{t\in[0,T]}\left(\|u\|_{H^4(\Omega)} + \dots + \|u_{tttt}\|_{L^2(\Omega)}\right) \le C\|f\|_{H^3([0,T]\times\Omega)}, \tag{2.7}$$

hold, where C is a constant that may depend on the velocity model c. Actually, C only depends on c_{min}, T, Ω, and the constants C_{m0} in (2.4), which will be shown in the appendix. We note that the left-hand side of (2.7) can be replaced by any mixed space-time partial derivative of order $n \le 4$, with the right-hand side $\|f\|_{H^{n-1}}$. Generally, the order could be arbitrarily large if enough regularity conditions on f, c and Ω were given, yet 4 is enough for our purpose.

We denote the solution u of the wave equation as $S(c,f)$. This functional is linear on f. As we have assumed, c relies on those θ_js smoothly, and we can calculate with the method of variation that

$$\frac{\partial u}{\partial \theta_j} = S\left(c, 2c(\Delta u)\frac{\partial c}{\partial \theta_j}\right), \tag{2.8}$$

$$\begin{aligned}\frac{\partial^2 u}{\partial \theta_j \partial \theta_k} = {} & S\left(c, 2c\left(\Delta\frac{\partial u}{\partial \theta_j}\right)\frac{\partial c}{\partial \theta_k}\right) + S\left(c, 2c\left(\Delta\frac{\partial u}{\partial \theta_k}\right)\frac{\partial c}{\partial \theta_j}\right) \\ & + S\left(c, 2(\Delta u)\left(\frac{\partial c}{\partial \theta_j}\frac{\partial c}{\partial \theta_k} + c\frac{\partial^2 c}{\partial \theta_j \partial \theta_k}\right)\right).\end{aligned} \tag{2.9}$$

As the inequality (2.7) implies, the norm of any fourth derivative of u is bounded by the norm of the third derivative of f up to a uniform constant C. In the following contents, the notation C may differ by a constant which may depend on the region Ω, the range of time T and the number of parameters m, but it will still be uniform over all our possible models. In order to make the relation clear, we will still use the notations C_{m0}, C_{m1}, C_{m2} in (2.4) and will not abbreviate them into C.

We have the following theorem:

THEOREM 2.1. *Under the assumption (2.4), we have*

$$\left\|\frac{\partial^2 u}{\partial \theta_j \partial \theta_k}\right\|_{L^\infty([0,T]\times\Omega)} \le \widetilde{C}\|f\|_{H^3}, \tag{2.10}$$

where the definition of the constant $\widetilde{C}$ is in the proof. It only depends on Ω, T, C_{m0}, C_{m1}, C_{m2}, c_{min} and does not depend on any particular c.

Proof. As $\|\Delta u\|_{H^2([0,T]\times\Omega)} \leq C\|f\|_{H^3([0,T]\times\Omega)}$, we can know

$$\begin{aligned}\left\|2c(\Delta u)\frac{\partial c}{\partial \theta_j}\right\|_{H^2([0,T]\times\Omega)} &\leq 2C_{m0}C_{m1}\|\Delta u\|_{H^2}\\ &\leq 2C_{m0}C_{m1}C\|f\|_{H^3},\end{aligned} \tag{2.11}$$

so with the third regularity, we get by (2.8),

$$\left\|\frac{\partial u}{\partial \theta_j}\right\|_{H^3(\Omega\times[0,T])} \leq 2C_{m0}C_{m1}C^2\|f\|_{H^3}. \tag{2.12}$$

Now we can move on the estimation of (2.9).
Since $\left\|2c\left(\Delta\frac{\partial u}{\partial\theta_j}\right)\frac{\partial c}{\partial\theta_k}\right\|_{H^1} \leq 2C_{m0}C_{m1}\left\|\frac{\partial u}{\partial\theta_j}\right\|_{H^3}$, we get

$$\begin{aligned}\left\|S\left(c,2c\left(\Delta\frac{\partial u}{\partial\theta_j}\right)\frac{\partial c}{\partial\theta_k}\right)\right\|_{L^\infty([0,T],H^2(\Omega))} &\leq C\left\|2c\left(\Delta\frac{\partial u}{\partial\theta_j}\right)\frac{\partial c}{\partial\theta_k}\right\|_{H^1}\\ &\leq 2C_{m0}C_{m1}C\left\|\frac{\partial u}{\partial\theta_j}\right\|_{H^3}\\ &\leq 4C_{m0}^2C_{m1}^2C^3\|f\|_{H^3},\end{aligned} \tag{2.13}$$

and the same holds for the second term in (2.9). Inside the third term, $\left\|2(\Delta u)\left(\frac{\partial c}{\partial\theta_j}\frac{\partial c}{\partial\theta_k}+c\frac{\partial^2 c}{\partial\theta_j\partial\theta_k}\right)\right\|_{H^1}$ does not exceed $2\left(C_{m1}^2+C_{m0}C_{m2}\right)C\|f\|_{H^3}$, so we can get

$$\left\|S\left(c,2(\Delta u)\left(\frac{\partial c}{\partial\theta_j}\frac{\partial c}{\partial\theta_k}+c\frac{\partial^2 c}{\partial\theta_j\partial\theta_k}\right)\right)\right\|_{L^\infty(H^2)} \leq 2\left(C_{m1}^2+C_{m0}C_{m2}\right)C^2\|f\|_{H^3}. \tag{2.14}$$

With (2.13) (twice) and (2.14), we finally know that

$$\left\|\frac{\partial^2 u}{\partial\theta_j\partial\theta_k}\right\|_{L^\infty([0,T],H^2(\Omega))} \leq C'\|f\|_{H^3([0,T]\times\Omega)}, \tag{2.15}$$

where $C' = \left(8C_{m0}^2C_{m1}^2C^3 + 2\left(C_{m1}^2+C_{m0}C_{m2}\right)C^2\right)$. Since $H^2(\Omega)$ can be (continuously) embedded into $L^\infty(\Omega)$ in dimension 3, we know that $\left\|\frac{\partial^2 u}{\partial\theta_j\partial\theta_k}\right\|_{L^\infty([0,T]\times\Omega)} \leq C'C_{Sobolev}\|f\|_{H^3}$. The Sobolev's constant only depends on Ω. Let $\widetilde{C} := C'C_{Sobolev}$ and then the assertion follows. ∎

For a receiver at $\boldsymbol{x}_r \in \Omega$, we should get the data $d_{exact}(t) = u(t,\boldsymbol{x}_r;c(\boldsymbol{\theta}))$ (error ignored). Obviously $\|d_{exact}\|_{L^2(0,T)} \leq \sqrt{T}\|u\|_{L^\infty}$, and similar relationship holds for their second derivative with respect to $\boldsymbol{\theta}$. In this point of view, the constant C_2 can be taken as $\widetilde{C}\sqrt{T}\|f\|_{H^3}$ under the parameter model assumption (2.3) [2], at least for d_{exact}.

[2]Variation with respect to c and w.r.t. $\boldsymbol{\theta}$ is different, and we only prove the bound for the variation (or differential) w.r.t. $\boldsymbol{\theta}$, i.e. the parameters. Considering the fact that c is determined by the parameters, we still write c in (2.3) and later paragraphs. If we want to justify (2.3) rigorously, here and in later sections, every c should be taken by $\boldsymbol{\theta}$.

3. Convergence of standard seismic tomography

First, we consider the standard seismic tomography, with the synthetic data given perfectly by the analytical solution to (2.1), where the iteration formula is given by the equation (8) in [5],

$$c_{n+1}=c_n+\left(\frac{\delta d_{exact}}{\delta c}(c_n),\frac{\delta d_{exact}}{\delta c}(c_n)\right)^{-1}\left(\frac{\delta d_{exact}}{\delta c}(c_n),d_{obs}-d_{exact}(c_n)\right), \tag{3.1}$$

where $d_{obs}=d_{exact}(c^*)$ is the (error-free) synthetic data, and c^* is the exact velocity model. We will prove that the iteration converges at the second order (near c^*), as stated in the following theorem.

THEOREM 3.1. *Under the assumption of (2.3), and suppose that c_n is sufficiently close to c^*, we have*

$$c_{n+1}-c^*=O(\|c_n-c^*\|^2), \tag{3.2}$$

i.e. $c_n\to c^*$ at the second order.

Proof. By applying Taylor's expansion at c^*,

$$d_{exact}(c_n)-d_{obs}=\frac{\delta d_{exact}}{\delta c}(c^*)(c_n-c^*)+O\left(\|c_n-c^*\|^2\right), \tag{3.3}$$

$$\frac{\delta d_{exact}}{\delta c}(c_n)=\frac{\delta d_{exact}}{\delta c}(c^*)+O(\|c_n-c^*\|), \tag{3.4}$$

we can get

$$\left(\frac{\delta d_{exact}}{\delta c}(c_n),d_{exact}(c_n)-d_{obs}\right)=\left(\frac{\delta d_{exact}}{\delta c}(c^*),\frac{\delta d_{exact}}{\delta c}(c^*)(c_n-c^*)\right)+O\left(\|c_n-c^*\|^2\right), \tag{3.5}$$

$$\left(\frac{\delta d_{exact}}{\delta c}(c_n),\frac{\delta d_{exact}}{\delta c}(c_n)\right)^{-1}=\left(\frac{\delta d_{exact}}{\delta c}(c^*),\frac{\delta d_{exact}}{\delta c}(c^*)\right)^{-1}+O(\|c_n-c^*\|). \tag{3.6}$$

So, considering the iteration (3.1), we have

$$\begin{aligned}c_{n+1}-c^*&=c_n-c^*+\left(\frac{\delta d_{exact}}{\delta c}(c_n),\frac{\delta d_{exact}}{\delta c}(c_n)\right)^{-1}\left(\frac{\delta d_{exact}}{\delta c}(c_n),d_{obs}-d_{exact}(c_n)\right)\\&=c_n-c^*-\left(\frac{\delta d_{exact}}{\delta c}(c^*),\frac{\delta d_{exact}}{\delta c}(c^*)\right)^{-1}\left(\frac{\delta d_{exact}}{\delta c}(c^*),\frac{\delta d_{exact}}{\delta c}(c^*)(c_n-c^*)\right)\\&\quad+O\left(\|c_n-c^*\|^2\right)\\&=O\left(\|c_n-c^*\|^2\right),\end{aligned} \tag{3.7}$$

which shows that $c_n\to c^*$ at the second order. ◻

4. Seismic tomography with approximation to sensitivity kernel

In realistic cases, the observation as well as the algorithm induce some error. We suppose that they are considerably small,

$$\|d_{sys}(c)-d_{exact}(c)\|\le e_d,\ \left\|\frac{\delta d_{sys}}{\delta c}(c)-\frac{\delta d_{exact}}{\delta c}(c)\right\|\le e_{\delta d},\ \|d_{obs}-d_{exact}(c^*)\|\le e_{obs}. \tag{4.1}$$

When applying the iteration (3.1), we will replace d_{exact} with d_{sys}, i.e.

$$c_{n+1} = c_n + \left(\frac{\delta d_{sys}}{\delta c}(c_n), \frac{\delta d_{sys}}{\delta c}(c_n)\right)^{-1} \left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right). \tag{4.2}$$

We cannot expect $c_n \to c^*$ when error exists. Instead, we are going to find c to minimize $|d_{sys}(c) - d_{obs}|^2$. Denote the minimizing c as c_e^*.

In the case with error, we have the following theorem instead of (3.1).

THEOREM 4.1. *Suppose (2.3) and (4.1), and that c_n is sufficiently close to c_e^*, we have*

$$\limsup_{n\to\infty} \frac{\|c_{n+1} - c_e^*\|}{\|c_n - c_e^*\|} \le C_{-1} C_2 (e_d + e_{obs}), \tag{4.3}$$

i.e. $c_n \to c_e^*$ linearly, as long as $C_{-1}C_2(e_d + e_{obs}) < 1$.

Proof. With the help of the method of variation, we know from the minimality of c_e^* that

$$\left(d_{sys}(c_e^*) - d_{obs}, \frac{\delta d_{sys}}{\delta c}(c_e^*)\right) = 0. \tag{4.4}$$

Now we apply the Taylor's expansion at c_e^* on (4.2).

$$\begin{aligned} c_{n+1} - c_e^* &= c_n - c_e^* + \left(\frac{\delta d_{sys}}{\delta c}(c_n), \frac{\delta d_{sys}}{\delta c}(c_n)\right)^{-1} \left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right) \\ &= c_n - c_e^* + \left(\frac{\delta d_{sys}}{\delta c}(c_n), \frac{\delta d_{sys}}{\delta c}(c_n)\right)^{-1} \cdot \\ &\quad \left(\frac{\delta d_{sys}}{\delta c}(c_e^*) + \frac{\delta^2 d_{sys}}{\delta c^2}(c_e^*)(c_n - c_e^*), d_{obs} - d_{sys}(c_n)\right) + o(\|c_n - c_e^*\|). \end{aligned} \tag{4.5}$$

By writing $d_{obs} - d_{sys}(c_n)$ as $d_{obs} - d_{sys}(c_e^*) + d_{sys}(c_e^*) - d_{sys}(c_n)$ and applying (4.4), we can get

$$\begin{aligned} &\left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right) \\ &= \left(\frac{\delta d_{sys}}{\delta c}(c_e^*) + \frac{\delta^2 d_{sys}}{\delta c^2}(c_e^*)(c_n - c_e^*), d_{obs} - d_{sys}(c_e^*)\right) \\ &\quad + \left(\frac{\delta d_{sys}}{\delta c}(c_e^*), -\frac{\delta d_{sys}}{\delta c}(c_e^*)(c_n - c_e^*)\right) + o(\|c_n - c_e^*\|) \\ &= \left(\frac{\delta^2 d_{sys}}{\delta c^2}(c_e^*)(c_n - c_e^*), d_{obs} - d_{sys}(c_e^*)\right) \\ &\quad - \left(\frac{\delta d_{sys}}{\delta c}(c_e^*), \frac{\delta d_{sys}}{\delta c}(c_e^*)(c_n - c_e^*)\right) + o(\|c_n - c_e^*\|), \end{aligned} \tag{4.6}$$

which is $O(\|c_n - c_e^*\|)$. Returning to (4.2), we know,

$$c_{n+1} - c_e^* = c_n - c_e^* + \left(\frac{\delta d_{sys}}{\delta c}(c_n), \frac{\delta d_{sys}}{\delta c}(c_n)\right)^{-1} \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right)\right)$$

$$=c_n - c_e^* + \left(\frac{\delta d_{sys}}{\delta c}(c_e^*), \frac{\delta d_{sys}}{\delta c}(c_e^*)\right)^{-1} \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right)\right) + o(\|c_n - c_e^*\|)$$
$$=\left(\frac{\delta d_{sys}}{\delta c}(c_e^*), \frac{\delta d_{sys}}{\delta c}(c_e^*)\right)^{-1} \left(\frac{\delta^2 d_{sys}}{\delta c^2}(c_e^*)(c_n - c_e^*), d_{obs} - d_{sys}(c_e^*)\right) + o(\|c_n - c_e^*\|), \quad (4.7)$$

where we apply (4.6) in the last equality.

We need to give an estimate of $(d_{obs} - d_{sys}(c_e^*))$. In fact,

$$\|d_{obs} - d_{sys}(c_e^*)\| \leq \|d_{obs} - d_{sys}(c^*)\|$$
$$\leq \|d_{obs} - d_{exact}(c^*)\| + \|d_{exact}(c^*) - d_{sys}(c^*)\| \leq e_{obs} + e_d. \quad (4.8)$$

Then we know from (4.7) that

$$\|c_{n+1} - c_e^*\| \leq C_{-1} C_2 \|c_n - c_e^*\| (e_{obs} + e_d) + o(\|c_n - c_e^*\|), \quad (4.9)$$

which shows the linear convergence $c_n \to c_e^*$ when $C_{-1} C_2 (e_{obs} + e_d) < 1$. □

The above result ensures the convergence of the iteration as long as the errors are reasonably bounded. What we also hope is that the result c_e^* can be close to c^*. In fact,

$$\|d_{exact}(c_e^*) - d_{exact}(c^*)\| \leq \|d_{exact}(c_e^*) - d_{sys}(c_e^*)\| + \|d_{sys}(c_e^*) - d_{obs}\| + \|d_{obs} - d_{exact}(c^*)\|$$
$$\leq e_d + (e_{obs} + e_d) + e_{obs} = 2(e_{obs} + e_d), \quad (4.10)$$

and also,

$$\|d_{sys}(c_e^*) - d_{sys}(c^*)\| \leq \|d_{sys}(c_e^*) - d_{obs}\| + \|d_{obs} - d_{exact}(c^*)\| + \|d_{exact}(c^*) - d_{sys}(c^*)\|$$
$$\leq (e_{obs} + e_d) + e_{obs} + e_d = 2(e_{obs} + e_d). \quad (4.11)$$

If either d_{exact} or d_{sys} has a Lipschitz-continuous inverse near c^*, $\|c_e^* - c^*\|$ will be bounded by $2L(e_{obs} + e_d)$ then.

5. Seismic tomography with Tikhonov regularization

In reality, the assumption of the bound C_{-1} in (2.3) is difficult to get quantified. It is the nature of inverse problems that most of them are ill-posed, and that some techniques about regularization are needed. Here we consider the misfit loss function with a Tikhonov regularization as follows

$$\chi(c) = \frac{1}{2}\left(d_{obs} - d_{sys}(c), d_{obs} - d_{sys}(c)\right) + \varepsilon \|c - c_0\|_{L^2}^2, \quad (5.1)$$

where c_0 is a reference solution of the velocity model. We denote the minimizing c as $c_{e,\varepsilon}^*$. The iteration formula becomes

$$c_{n+1} = c_n + \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n), \frac{\delta d_{sys}}{\delta c}(c_n)\right) + 2\varepsilon id_c\right)^{-1} \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n), d_{obs} - d_{sys}(c_n)\right) - 2\varepsilon(c_n - c_0)\right). \quad (5.2)$$

Similar with (4.4), we have

$$-\left(\frac{\delta d_{sys}}{\delta c}(c_{e,\varepsilon}^*), d_{obs} - d_{sys}(c_{e,\varepsilon}^*)\right) + 2\varepsilon(c_{e,\varepsilon}^* - c_0) = 0. \quad (5.3)$$

Then, like (4.7), we obtain

$$\begin{aligned} & c_{n+1}-c^*_{e,\varepsilon} \\ =& \left(\left(\frac{\delta d_{sys}}{\delta c}(c^*_{e,\varepsilon}),\frac{\delta d_{sys}}{\delta c}(c^*_{e,\varepsilon})\right)+2\varepsilon id_c\right)^{-1}\left(\frac{\delta^2 d_{sys}}{\delta c^2}(c^*_{e,\varepsilon})(c_n-c^*_{e,\varepsilon}),d_{obs}-d_{sys}(c^*_{e,\varepsilon})\right) \\ & +O(\|c_n-c^*_{e,\varepsilon}\|^2), \end{aligned} \tag{5.4}$$

and then we gain an estimate free from C_{-1} in (2.3), namely

$$\|c_{n+1}-c^*_{e,\varepsilon}\|\leq\frac{C_2\|d_{obs}-d_{sys}(c^*_{e,\varepsilon})\|}{2\varepsilon}\|c_n-c^*_{e,\varepsilon}\|+O(\|c_n-c^*_{e,\varepsilon}\|^2). \tag{5.5}$$

We also need an estimate of $\|d_{obs}-d_{sys}(c^*_{e,\varepsilon})\|$. With minimality, we have

$$\|d_{obs}-d_{sys}(c^*_{e,\varepsilon})\|^2+2\varepsilon\|c^*_{e,\varepsilon}-c_0\|^2\leq\|d_{obs}-d_{sys}(c^*)\|^2+2\varepsilon\|c^*-c_0\|^2, \tag{5.6}$$

and we also know, as for the assumption (4.1),

$$\|d_{obs}-d_{exact}(c^*)\|\leq e_{obs},\|d_{sys}(c^*)-d_{exact}(c^*)\|\leq e_u, \tag{5.7}$$

then we know

$$\begin{aligned} \|c_{n+1}-c^*_{e,\varepsilon}\|\leq & \frac{C_2\left(\sqrt{(e_u+e_{obs})^2+2\varepsilon(\|c^*-c_0\|^2-\|c^*_{e,\varepsilon}-c_0\|^2)}\right)}{2\varepsilon}\|c_n-c^*_{e,\varepsilon}\| \\ & +O(\|c_n-c^*_{e,\varepsilon}\|^2), \end{aligned} \tag{5.8}$$

which is the regularization case of (4.9). Like (4.10) and (4.11), we also get

$$\|d_{sys}(c^*)-d_{sys}(c^*_{e,\varepsilon})\|\leq\sqrt{(e_u+e_{obs})^2+2\varepsilon(\|c^*-c_0\|^2-\|c^*_{e,\varepsilon}-c_0\|^2)}+e_u+e_{obs}, \tag{5.9}$$

and

$$\|d_{exact}(c^*)-d_{exact}(c^*_{e,\varepsilon})\|\leq\sqrt{(e_u+e_{obs})^2+2\varepsilon(\|c^*-c_0\|^2-\|c^*_{e,\varepsilon}-c_0\|^2)}+e_u+e_{obs}, \tag{5.10}$$

which also bounds the error supposing either d_{exact} or d_{sys} has Lipschitz-continuous inverse near c^*.

Unfortunately, up to our effort, there is no satisfactory a priori estimation of $\left\|d_{obs}-d_{sys}\left(c^*_{e,\varepsilon}\right)\right\|$ if the bound of $\left(\frac{\delta d_{sys}}{\delta c},\frac{\delta d_{sys}}{\delta c}\right)^{-1}$ is removed. We still need an extra assumption on $\frac{\delta d_{sys}}{\delta c}$. Let $\delta=e_{obs}+e_{sys}$ in short. According to [2], if one assumes that

$$c^*-c_0=\left(\frac{\delta d_{sys}}{\delta c},\frac{\delta d_{sys}}{\delta c}\right)^{\nu}_{c=c^*}v,\nu\geq\frac{1}{2}, \tag{5.11}$$

where v lies in the same space as c^* and c_0 and should be small enough, then it has been shown that there exist $\varepsilon(\delta)$, $N(\delta)$ such that, if the result of N iterations of the $\varepsilon(\delta)$-regularization Gauss-Newton Method is denoted as $c_{N(\delta),\varepsilon(\delta)}$, one has $\left\|c_{N(\delta),\varepsilon(\delta)}-c^*\right\|=O\left(\delta^{\frac{2\nu}{2\nu+1}}\right)$. To be rigorous, we need to replace ε in (5.2) by $\varepsilon_n=r^{-n}$ for some $r>1$ to apply the result in [2]. The stopping N is chosen such that $\delta\sim\varepsilon_N^{\nu+\frac{1}{2}}$. Also, we can let $\nu\geq 0$ in (5.11) if we assume in addition that $\frac{\delta d_{sys}}{\delta c}\equiv F'$ is nearly linear near c^*, in the sense that

$$
\begin{aligned}
&F'(c_2)=R(c,c_2)F'(c)+Q(c,c_2),\\
&\|I-R(c,c_2)\|\leq C_R,\\
&\|Q(c,c_2)\|\leq C_Q\|F'(c^*)(c-c_2)\|,
\end{aligned}
\tag{5.12}
$$

for all c, c_2 such that $\|c-c^*\|\leq 2\rho$, $\|c_2-c^*\|\leq 2\rho$, and the constants C_R, C_Q and ρ need to be sufficiently small.

For the sake of completeness, we show the result in [2] (Theorem 2.4) for the case $\nu\geq\frac{1}{2}$ in our notation.

THEOREM 5.1. *Suppose $\|c^*-c_0\|\leq\rho$. We assume (4.1) and denote $\delta=e_{obs}+e_d$. Suppose (5.11) holds, and v is chosen in the perpendicular space of the kernel of $\frac{\delta d_{sys}}{\delta c}(c^*)$, and $\left\|\frac{\delta^2 d_{sys}}{\delta c^2}\right\|\leq C_2=L$, at least for c such that $\|c-c_0\|\leq 2\rho$. We also choose a sequence $\{\varepsilon_n\}$ of regularization parameters such that*

$$
\varepsilon_n>0,\quad \varepsilon_n\to 0,\quad 1<\frac{\varepsilon_n}{\varepsilon_{n+1}}\leq r,
\tag{5.13}
$$

for a constant $r>1$.

For a fixed δ, we choose $N=N(\delta)$ such that

$$
\eta\varepsilon_N^{\nu+\frac{1}{2}}<\delta\leq\eta\varepsilon_{N-1}^{\nu+\frac{1}{2}},
\tag{5.14}
$$

for some $\eta>0$. Once δ and $\{\varepsilon_n\}$ are fixed, N is unique.

If the following closeness conditions hold:

$$
\begin{aligned}
&B+2\sqrt{AC}<1,\\
&\frac{\|c_0-c^*\|}{\varepsilon_0^\nu}\leq\frac{1-B+\sqrt{(1-B)^2-4AC}}{2C},\\
&\varepsilon_0^\nu C_\gamma\leq\rho,\\
&A:=r^\nu\left(\|v\|+\frac{\eta}{2}\right),\\
&B:=r^\nu\left(\frac{\varepsilon_0^{\nu-\frac{1}{2}}}{2}\|v\|+\left\|\left(F'(c^*)^*F'(c^*)\right)^{\nu-\frac{1}{2}}v\right\|\right),\\
&C:=\frac{r^\nu L\varepsilon_0^{\nu-\frac{1}{2}}}{4},\\
&C_\gamma:=\max\left(\frac{\|c_0-c^*\|}{\varepsilon_0^\nu},\frac{2A}{1-B+\sqrt{(1-B)^2-4AC}}\right),
\end{aligned}
\tag{5.15}
$$

then we have, as $\delta\to 0$, $\left\|c_{N(\delta)}-c^*\right\|\to 0$. More accurately,

$$
\left\|c_{N(\delta)}-c^*\right\|=O\left(\delta^{\frac{2\nu}{2\nu+1}}\right),
\tag{5.16}
$$

where $c_n = c_{n,\delta}$ *is defined through*

$$c_{n+1} = c_n - \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n) \right)^* \frac{\delta d_{sys}}{\delta c}(c_n) + \varepsilon_n id \right)^{-1} \cdot \left(\left(\frac{\delta d_{sys}}{\delta c}(c_n) \right)^* (d_{sys}(c_n) - d_{obs}) + \varepsilon_n (c_n - c_0) \right). \tag{5.17}$$

6. Conclusion and discussion

In this paper, we provide a rigorous analysis for the accuracy of seismic tomography for inverse problems of acoustic wave propagation for three cases: 1. Standard seismic tomography; 2. Seismic tomography with approximation to sensitivity kernel; 3. Seismic tomography with Tikhonov regularization. Specifically, we first give a uniform bound of the second order variation of the observation data over the velocity model, and then analyze the convergence rate of the iteration. Under appropriate assumptions, we prove second order convergence in the standard seismic tomography case, and first order convergence in the case with approximation to sensitivity kernel. Moreover, when we take the Tikhonov regularization into consideration, we can show the convergence under more general assumptions. In the future work, we shall continue to look for alternatives of the uniform invertibility conditions to make the convergence hold for weaker assumptions, and analyze the convergence of seismic tomography with stochastic approximations.

Acknowledgment. Z.H. was partially supported by the NSFC Projects No. 12025104, 11871298, 81930119. L.C. was partially supported by the NSFC Project No. 11901601. X.Y. was partially supported by the NSF grants DMS-1818592 and DMS-2109116.

Appendix. Here we consider the wave equation

$$\begin{aligned} &u_{tt} - c^2(\boldsymbol{x})\Delta u = f, \quad \forall t \in [0,T],\ \boldsymbol{x} \in \mathbb{R}^3; \\ &u(0,\boldsymbol{x}) = u_t(0,\boldsymbol{x}) = 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^3, \end{aligned} \tag{A.1}$$

where c represents the velocity model. We assume that $0 < c_{min} \le c(\boldsymbol{x}) \le c_{max}$, and f is compactly supported inside $\Pi := \left\{ (t,\boldsymbol{x}) : t \in (0,T], \overline{B_{\boldsymbol{x}}(c_{max}T - c_{max}t)} \subset \Omega \right\}$ (so u vanishes outside $[0,T] \times \Omega$), where Ω is a smooth region. What we want is the regularity of u with a "uniform" constant, in the sense that it only depends on c_{min}, T, Ω and the bound of the derivative of c.

We first assume that $\|\nabla c\|_\infty \le M_1$and $f \in L^2\left([0,T]; L^2(\Omega)\right)$. We multiply the Equation (A.1) with u_t, and integrate on $\boldsymbol{x} \in \Omega$, to get

$$\frac{1}{2}\frac{\partial}{\partial t}\int_\Omega \left(u_t^2 + c^2 |\nabla u|^2 \right) d\boldsymbol{x} = \int_\Omega (f u_t - 2c u_t \nabla c \cdot \nabla u)\, d\boldsymbol{x}, \tag{A.2}$$

where $|\cdot|$ means the Euclidean norm of a vector.

Then, we integrate on t from 0 to τ, where $\tau < T$,

$$\left[\int_\Omega \left(u_t^2 + c^2 |\nabla u|^2 \right) d\boldsymbol{x} \right]_{t=\tau} = \int_0^\tau dt \left(\int_\Omega (f u_t - 2c u_t \nabla c \cdot \nabla u)\, d\boldsymbol{x} \right), \tag{A.3}$$

and with the help of the AM-GM inequality, for any $\varepsilon_1, \varepsilon_2 > 0$,

$$\left[\int_\Omega \left(u_t^2 + c^2 |\nabla u|^2 \right) d\boldsymbol{x} \right]_{t=\tau} \le \int_0^\tau dt \left(\int_\Omega \left(\frac{f^2}{\varepsilon_1} + \varepsilon_1 u_t^2 + \frac{4M_1^2}{\varepsilon_2} c^2 |\nabla u|^2 + \varepsilon_2 u_t^2 \right) d\boldsymbol{x} \right)$$

$$\leq \frac{1}{\varepsilon_1}\int_0^\tau dt\left(\int_\Omega f^2 d\boldsymbol{x}\right)$$
$$+\max\left\{\varepsilon_1+\varepsilon_2,\frac{4M_1^2}{\varepsilon_2}\right\}\int_0^\tau dt\left(\int_\Omega\left(u_t^2+c^2|\nabla u|^2\right)d\boldsymbol{x}\right). \tag{A.4}$$

With the help of Gronwall's inequality, we know

$$\left[\int_\Omega\left(u_t^2+c^2|\nabla u|^2\right)d\boldsymbol{x}\right]_{t=\tau}\leq\frac{1}{\varepsilon_1}\exp\left(\max\left\{\varepsilon_1+\varepsilon_2,\frac{4M_1^2}{\varepsilon_2}\right\}\tau\right)\|f\|_{L^2}^2$$
$$\leq\frac{1}{\varepsilon_1}\exp\left(\max\left\{\varepsilon_1+\varepsilon_2,\frac{4M_1^2}{\varepsilon_2}\right\}T\right)\|f\|_{L^2}^2, \tag{A.5}$$

thus, we can choose arbitrary $\varepsilon_1,\varepsilon_2>0$ and know that $\int_\Omega\left(u_t^2+c_{min}^2|\nabla u|^2\right)d\boldsymbol{x}$ is bounded by $C\|f\|_{L^2}^2$, where the constant C only depends on T and M_1. With the help of Poincare's inequality (where there is a constant depending on Ω), we can know

$$\left[\|u_t\|_{L^2(\Omega)}+\|u\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\leq C\|f\|_{L^2([0,T];L^2(\Omega))}, \tag{A.6}$$

where the constant C here only depends on T,M_1,c_{min} and Ω, not on concrete $c(\boldsymbol{x})$. Thus, the first order regularity is proved. We will apply the inequality (A.6) for several times in the proof of the higher order regularity.

Now we assume that $f\in L^2\left([0,T];H^1(\Omega)\right)$ and $f_t\in L^2\left([0,T];L^2(\Omega)\right)$. By taking the partial derivative $\frac{\partial}{\partial t}$ on (A.1), we know

$$(u_t)_{tt}-c^2\Delta u_t=f_t, \tag{A.7}$$

thus we can apply (A.6) to get

$$\left[\|u_{tt}\|_{L^2(\Omega)}+\|u_t\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\leq C\|f_t\|_{L^2([0,T];L^2(\Omega))}. \tag{A.8}$$

Now we need a uniform estimation of $\|u\|_{H^2(\Omega)}$. It suffice to bound $\|u_x\|_{H^1(\Omega)}$, since the estimation for $\|u_y\|_{H^1(\Omega)}$ and $\|u_z\|_{H^1(\Omega)}$ should be the same. Now we take $\frac{\partial}{\partial x}$ on (A.1),

$$(u_x)_{tt}-c^2\Delta(u_x)=f_x+2cc_x\Delta u, \tag{A.9}$$

here we solve Δu out of the Equation (A.1) to get

$$(u_x)_{tt}-c^2\Delta(u_x)=f_x-\frac{2c_x}{c}f+\frac{2c_x}{c}u_{tt}\triangleq g, \tag{A.10}$$

and we apply (A.6) again

$$\left[\|u_x\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\leq C\|g\|_{L^2([0,T];L^2(\Omega))}$$
$$\leq C\left(\|f_x\|+\frac{2M_1}{c_{min}}\|f\|+\frac{2M_1}{c_{min}}\|u_{tt}\|\right), \tag{A.11}$$

where the norm in the last inequality is also $L^2\left([0,T];L^2(\Omega)\right)$. We see in (A.8) that we have already bounded $\|u_{tt}\|$ in the norm of $L^\infty\left([0,T];L^2(\Omega)\right)$, which is stronger, in the sense that $\|u_{tt}\|_{L^2}\le\sqrt{T}\|u_{tt}\|_{L^\infty}$. Hence, we get

$$\left[\|u_x\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\le C\left(1+\frac{2M_1}{c_{min}}\right)\|f\|_{L^2([0,T];H^1(\Omega))}+\frac{2C^2M_1\sqrt{T}}{c_{min}}\|f_t\|_{L^2([0,T];L^2(\Omega))}. \tag{A.12}$$

With the result (A.8) and (A.12), we conclude that

$$\begin{aligned}&\left[\|u_{tt}\|_{L^2(\Omega)}+\|u_t\|_{H^1(\Omega)}+\|u\|_{H^2(\Omega)}\right]_{L^\infty([0,T])}\\ \le& C^{(2)}\left(\|f\|_{L^2([0,T];H^1(\Omega))}+\|f_t\|_{L^2([0,T];L^2(\Omega))}\right),\end{aligned} \tag{A.13}$$

where the constant $C^{(2)}$ only depends on M_1,T,c_{min} and Ω. Note that we sometimes use the notation $\|f\|_{H^1([0,T]\times\Omega)}$ for the sum of the two norms on the right-hand side of (A.13), for the sake of brevity.

Thus, the uniform estimation of the second order derivative has been built.

For the third order, we need the assumption $\left\|\nabla^2 c\right\|_\infty\le M_2$ and $f\in H^2([0,T]\times\Omega)$ in addition. Here $f_t\in H^1([0,T]\times\Omega)$, so by (A.7) and (A.13), replacing f with f_t and u with u_t, we have

$$\begin{aligned}&\left[\|u_{ttt}\|_{L^2(\Omega)}+\|u_{tt}\|_{H^1(\Omega)}+\|u_t\|_{H^2(\Omega)}\right]_{L^\infty([0,T])}\\ \le& C^{(2)}\left(\|f_t\|_{L^2([0,T];H^1(\Omega))}+\|f_{tt}\|_{L^2([0,T];L^2(\Omega))}\right).\end{aligned} \tag{A.14}$$

To build an estimation of $\|u\|_{H^3(\Omega)}$, we consider to take $\frac{\partial}{\partial y}$ on (A.10). Here y can be replaced by z, or even x itself.

$$\begin{aligned}(u_{xy})_{tt}-c^2\Delta(u_{xy})=&2cc_y\Delta u_x+f_{xy}-\frac{2c_{xy}}{c}f+\frac{2c_xc_y}{c^2}f-\frac{2c_x}{c}f_y\\ &+\frac{2c_{xy}}{c}u_{tt}-\frac{2c_xc_y}{c^2}u_{tt}+\frac{2c_x}{c}u_{tty},\end{aligned} \tag{A.15}$$

and we can solve Δu_x out of (A.10). Hence

$$\begin{aligned}(u_{xy})_{tt}-c^2\Delta(u_{xy})=&2c_y\frac{u_{xtt}-f_x+\frac{2c_x}{c}f-\frac{2c_x}{c}u_{tt}}{c}+f_{xy}\\ &-\frac{2c_{xy}}{c}f+\frac{2c_xc_y}{c^2}f-\frac{2c_x}{c}f_y+\frac{2c_{xy}}{c}u_{tt}-\frac{2c_xc_y}{c^2}u_{tt}+\frac{2c_x}{c}u_{tty},\end{aligned} \tag{A.16}$$

and by checking every term, we know their $L^2([0,T]\times\Omega)$-norms are all bounded by $K\|f\|_{H^2([0,T]\times\Omega)}$ and $K\|u_t\|_{H^2([0,T]\times\Omega)}$ for some $K=K(M_1,M_2,c_{min})$. However, (A.14) says that $\|u_t\|_{H^2([0,T]\times\Omega)}\le\sqrt{T}C^{(2)}\|f\|_{H^2([0,T]\times\Omega)}$, so the right-hand side of (A.16) is bounded by $\|f\|_{H^2}$ up to a constant $\widetilde{K}$, which only depends on M_1,M_2,c_{min},T and Ω. With the help of (A.6) again, we get $\left[\|u_{xy}\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\le C\widetilde{K}\|f\|_{H^2}$. With (A.14), we conclude

$$\left[\|u_{ttt}\|_{L^2(\Omega)}+\|u_{tt}\|_{H^1(\Omega)}+\|u_t\|_{H^2(\Omega)}+\|u\|_{H^3(\Omega)}\right]_{L^\infty([0,T])}\le C^{(3)}\left(\|f\|_{H^2([0,T]\times\Omega)}\right), \tag{A.17}$$

where $C^{(3)}=C^{(3)}(M_1,M_2,c_{min},T,\Omega)$.

For the fourth order, we assume $\left\|\nabla^3 c\right\|_\infty \le M_3$ and $f\in H^3([0,T]\times\Omega)$. By (A.7) and (A.17), replacing f with f_t and u with u_t, we have

$$\left[\|u_{tttt}\|_{L^2(\Omega)}+\|u_{ttt}\|_{H^1(\Omega)}+\|u_{tt}\|_{H^2(\Omega)}+\|u_t\|_{H^3(\Omega)}\right]_{L^\infty([0,T])}\le C^{(3)}\left(\|f\|_{H^3([0,T]\times\Omega)}\right), \tag{A.18}$$

hence we only need to bound $\|u\|_{H^4(\Omega)}$. This comes from taking $\frac{\partial}{\partial z}$ on (A.16) and solving $\Delta(u_{xy})$ out of it, we get

$$(u_{xyz})_{tt}-c^2\Delta(u_{xyz})=\ldots\triangleq h, \tag{A.19}$$

where h is the sum of a lot of terms, which concern f's derivatives (up to order 4) and u_t's derivatives (up to order 3). For example, from the last term in (A.16), namely $\frac{2c_x}{c}u_{tty}$, we get $\frac{2c_{xz}}{c}u_{tty}$, $-\frac{2c_xc_z}{c^2}u_{tty}$, and $\frac{2c_x}{c}u_{ttyz}$. Thus, the $L^2([0,T]\times\Omega)$-norm of h is bounded by $K\|f\|_{H^3([0,T]\times\Omega)}$ and $K\|u_t\|_{H^3([0,T]\times\Omega)}$, where $K=K(M_1,M_2,M_3,c_{min})$. Since (A.17) shows $\|u_t\|_{H^3([0,T]\times\Omega)}\le\sqrt{T}C^{(3)}\|f\|_{H^3([0,T]\times\Omega)}$, we know that $\|h\|_{L^2([0,T]\times\Omega)}$ is bounded by $\|f\|_{H^3}$ up to a constant $\widetilde{K}$, which only depends on M_1,M_2,M_3,c_{min},T and Ω. Again, with (A.6), we know $\left[\|u_{xyz}\|_{H^1(\Omega)}\right]_{L^\infty([0,T])}\le C\widetilde{K}\|f\|_{H^3}$. Thus, we finally conclude the fourth order uniform estimation

$$\begin{aligned}&\left[\|u_{tttt}\|_{L^2(\Omega)}+\|u_{ttt}\|_{H^1(\Omega)}+\|u_{tt}\|_{H^2(\Omega)}+\|u_t\|_{H^3(\Omega)}+\|u\|_{H^4(\Omega)}\right]_{L^\infty([0,T])}\\ \le&C^{(4)}\left(\|f\|_{H^3([0,T]\times\Omega)}\right),\end{aligned} \tag{A.20}$$

where $C^{(4)}=C^{(4)}(M_1,M_2,M_3,c_{min},T,\Omega)$.

Actually, we can continue this process to prove that, as long as c's derivatives are uniformly bounded up to order $(m-1)$, we have the estimation with a uniform constant up to order m. However, the fourth regularity is enough for our purpose.

REFERENCES

[1] K. Aki and W.H.K. Lee, *Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous intial model*, J. Geophys. Res., 81:4381–4399, 1976. 1

[2] B. Barbara, *On convergence rates for the iteratively regularized Gauss-Newton method*, IMA J. Numer. Anal., 17:421–436, 1997. 5, 5, 5

[3] L. Chai, J.C. Hateley, E. Lorin, and X. Yang, *On the convergence of frozen Gaussian approximation for linear non-strictly hyperbolic systems*, Commun. Math. Sci., 19(3):585–606, 2021. 1

[4] L. Chai, P. Tong, and X. Yang, *Frozen Gaussian approximation for 3-D seismic wave propagation*, Geophys. J. Int., 208(1):59–74, 2017. 1, 2

[5] L. Chai, P. Tong, and X. Yang, *Frozen Gaussian approximation for 3-D seismic tomography*, Inverse Probl., 34(5):055004, 2018. 1, 2, 3

[6] A.M. Dziewonski, B.H. Hager, and R.J. O'Connell, *Large-scale heterogeneities in the lower mantle*, J. Geophys. Res., 82:239–255, 1977. 1

[7] L.C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, 1998. 2

[8] A. Fichtner, H.P. Bunge, and H. Igel, *The adjoint method in seismology I. Theory*, Phys. Earth Planet. Inter., 157:86–104, 2006. 1

[9] A. Fichtner, H. Igel, H.P. Bunge, and B.L.N. Kennett, *Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method*, J. Numer. Anal. Indust. Appl. Math., 4:11–22, 2009. 1

[10] J.C. Hateley, L. Chai, P. Tong, and X. Yang, *Frozen Gaussian approximation for 3-D elastic wave equation and seismic tomography*, Geophys. J. Int., 216(2):1394–1412, 2018. 1, 2

[11] X. Huang, D. Yang, P. Tong, J. Badal, and Q. Liu, *Wave equation based reflection tomography of the 1992 Landers earthquake area*, Geophys. Res. Lett., 43(5):1884–1892, 2016. 1

[12] D. Komatitsch and J. Tromp, *Spectral-element simulations of global seismic wave propagation I. Validation*, Geophys. J. Int., 149(2):390–412, 2002. 1

[13] Q. Liu and Y.J. Gu, *Seismic imaging: From classical to adjoint tomography*, Tectonophysics, 566-567:31–66, 2012. 1

[14] J. Lu and X. Yang, *Convergence of frozen Gaussian approximation for high frequency wave propagation*, Commun. Pure Appl. Math., 65:759–789, 2012. 1

[15] N. Rawlinson, S. Pozgay, and S. Fishwick, *Seismic tomography: A window into deep Earth*, Phys. Earth Planet. Inter., 178(3-4):101–135, 2010. 1

[16] B. Romanowicz, *Seismic tomography of the Earth's mantle*, Annu. Rev. Earth Planet. Sci., 19:77–99, 1991. 1

[17] C. Tape, Q. Liu, A. Maggi, and J. Tromp, *Adjoint tomography of the southern California crust*, Science, 325:988–992, 2009. 1

[18] C. Tape, Q. Liu, A. Maggi, and J. Tromp, *Seismic tomography of the southern California crust based on spectral-element and adjoint methods*, Geophys. J. Int., 180(1):433–462, 2010. 1

[19] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation*, Geophysics, 49(8):1259–1266, 1984. 1

[20] P. Tong, C.-W. Chen, D. Komatitsch, P. Basini, and Q. Liu, *High-resolution seismic array imaging based on an SEM-FK hybrid method*, Geophys. J. Int., 197(1):369–395, 2014. 1

[21] J. Tromp, C. Tape, and Q. Liu, *Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels*, Geophys. J. Int., 160(1):195–216, 2005. 1

[22] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, Geophysics, 74:WCC1–WCC26, 2009. 1

[23] D. Zhao, *Tomography and dynamics of Western-Pacific subduction zones*, Monogr. Environ. Earth Planets, 1:1–70, 2012. 1

[24] H. Zhu, E. Bozdag, D. Peter, and J. Tromp, *Structure of the European upper mantle revealed by adjoint tomography*, Nature Geosci., 5:493–498, 2012. 1