PAPER • OPEN ACCESS

Norm-dependent convergence and stability of the inverse scattering series for diffuse and scalar waves

To cite this article: Srinath Mahankali and Yunan Yang 2023 *Inverse Problems* **39** 054005

You may also like

- Integrable nonlocal derivative nonlinear Schrödinger equations
  Mark J Ablowitz, Xu-Dan Luo, Ziad H Musslimani et al.

- Inverse medium scattering problems with Kalman filter techniques
  Takashi Furuya and Roland Potthast

- L^2-Sobolev space bijectivity of the inverse scattering of a 3 × 3 AKNS system
  Jiaqi Liu

Norm-dependent convergence and stability of the inverse scattering series for diffuse and scalar waves

Srinath Mahankali[1] and Yunan Yang[2,*]

[1] Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America
[2] Institute for Theoretical Studies, ETH Zürich, Zürich 8092, Switzerland

E-mail: yunan.yang@eth-its.ethz.ch

Abstract

This work analyzes the forward and inverse scattering series for scalar waves based on the Helmholtz equation and the diffuse waves from the time-independent diffusion equation, which are important partial differential equations (PDEs) in various applications. Different from previous works, which study the radius of convergence for the forward and inverse scattering series, the stability, and the approximation error of the series under the L^p norms, we study these quantities under the Sobolev H^s norm, which associates with a general class of L^2-based function spaces. The H^s norm has a natural spectral bias based on its definition in the Fourier domain: the case $s < 0$ biases towards the lower frequencies, while the case $s > 0$ biases towards the higher frequencies. We compare the stability estimates using different H^s norms for both the parameter and data domains and provide a theoretical justification for the frequency weighting techniques in practical inversion procedures. We also provide numerical inversion examples to demonstrate the differences in the inverse scattering radius of convergence under different metric spaces.

Keywords: inverse scattering, Sobolev norm, stability, convergence, Helmholtz equation, diffuse wave equation

(Some figures may appear in colour only in the online journal)

---

* Author to whom any correspondence should be addressed.

1. Introduction

We are interested in studying the forward scattering Born series and the inverse scattering Born series of both the Helmholtz equation and the diffuse wave equation. This work is motivated by the paper of Moskow and Schotland [29], where such studies were conducted under the L^p function spaces for both parameter and measurement in which the radius of convergence and the Lipschitz stability constant of the inverse problem change with respect to the choice of p. There have been many works extending this analysis to various inverse problems, including the optical diffusion tomography [28], diffuse waves [30], scalar waves [22], the electromagnetic scattering [23], the Calderón problem [2], the Schrödinger problem [6], the radiative transport equation [25] and optimal tomography on graphs [11]. In particular, studies of the radius of convergence and stability have been extended to general Banach spaces, and the parameter and data spaces do not have to be the same [31]. In this work, we set the parameter space and the measurement space as the L^2-based Sobolev space $H^{\mathfrak{a}}$ and $H^{\mathfrak{b}}$, respectively. We investigate how different choices of $\mathfrak{a}$ and $\mathfrak{b}$ could change the optimization behaviors of solving inverse problems based on the Helmholtz equation and diffuse wave equation, but similar strategies can apply to other inverse problems. The L^2-based Sobolev spaces are Hilbert spaces, and the inner products involve the Laplacian operator, thus making them easy to implement from a computational perspective. After discretization, the corresponding objective function becomes a weighted least-squares error, and the quadratic nature makes gradient computation efficient.

Practically speaking, inverse medium problems constrained by such equations are often formulated as PDE-constrained optimization problems, where well-posedness properties of the inverse problem are translated into the uniqueness of the minimizer, sensitivity to data noise, rate of convergence, and other features in an optimization framework. The choice of $\mathfrak{b}$ for the measurement space $H^{\mathfrak{b}}$ then corresponds to the choice of the objective function in the resulting nonlinear, nonconvex optimization problem. On the other hand, the choice of $\mathfrak{a}$ for the parameter space $H^{\mathfrak{a}}$ imposes *a priori* information on the regularity of the parameter one aims to reconstruct. It could also lead to different gradient flow formulations when passing the Fréchet derivative to the gradient, thus giving rise to different gradient descent algorithms for solving such nonconvex optimization problems. Both choices affect the convergence rate and potentially change the stationary points to which the iterative gradient-based algorithm converges, even with the same initial guess [34]. We will demonstrate this later in section 6.

There have been numerous works on changing the function spaces for both the parameter and the data in different applications. It is worth noting that the H^{-1} semi-norm is closely related to the quadratic Wasserstein metric from optimal transportation [36, 38]. This connection has been utilized in many imaging and inverse problem applications [13, 37, 43] and extended to the general H^s norm as an objective function for data-matching inverse problems [14, 44]. Methods based on the Sobolev gradient [33], where the gradient of a given functional with respect to the parameter function is taken with respect to the inner product induced by the underlying Sobolev norm, have demonstrated advantages in image sharpening and edge-preserving [10, 41]. In [19], it was shown that choosing H^1 gradient flow for minimizing L^1-type objective functions yields a Lipschitz constant independent of the discretization grid size, achieving an optimal rate of convergence in the gradient descent method. In [15], the impacts of changing both the data and parameter spaces were discussed in machine learning in a discrete setup.

To tackle the local minima issues due to the inherent nonconvexity and sensitivity to noise of the L^2 norm as the objective function, there is a general strategy of adding adaptive weights to different frequency components of the data, which is often referred to as *Frequency Marching* [3, 7], *Multiscale/Hierarchical Inversion* [9, 17] or *Frequency Weighting* techniques for

nonlinear inverse problems [8, 20, 21, 35]. The model parameters should be updated along directions such that both low-wavenumber and high-wavenumber structures can be appropriately resolved. One way to control the order of updating the low- and high-wavenumber model components is through the choice of $H^{\mathfrak{a}}$ parameter space. Since they are also influenced by the low- and high-frequency components of the data residual through the gradient descent update, it is also preferable to have adaptive weights regarding different data frequencies in the optimization, which can be achieved by tuning different $\mathfrak{b}$ for the data metric space $H^{\mathfrak{b}}$ in our framework.

The main novelty of this work is to utilize the natural frequency biasing features of the L^2-based Sobolev spaces, as briefly mentioned above, and prove the impact of the choice of the parameter and data function spaces on the properties of the inverse scattering problems. We also focus on the class of L^2-based Sobolev space for its particular convenience in implementation, while we believe similar conclusions could be drawn for other classes of Sobolev spaces $W^{k,p}$ with p fixed and $k \in \mathbb{Z}$. To the best of our knowledge, utilizing the spectral properties of the L^2-based Sobolev space to improve the radius of convergence and stability of the inverse Born series has yet to be considered in the literature. There are three main contributions in our work. First, we rigorously study the convergence property of the forward scattering series that maps the parameter in $H^{\mathfrak{a}}$ to the scattering data in $H^{\mathfrak{b}}$, and the inverse scattering series that maps from $H^{\mathfrak{b}}$ to $H^{\mathfrak{a}}$. We show that the radius of convergence for the forward scattering series is only affected by $\mathfrak{b}$, but the radius of convergence for the inverse scattering series is affected by both $\mathfrak{a}$ and $\mathfrak{b}$. Second, through a sequence of theorems, we demonstrate that the choice of $(\mathfrak{a}, \mathfrak{b})$ consequently changes the stability of the limit of the inverse scattering series with respect to the small changes in the observed scattering data and changes the approximation error of the inverse scattering series. Our results show that one can improve the stability by tuning $(\mathfrak{a}, \mathfrak{b})$. Third, we demonstrate the impact of the choice of these two function spaces in a few numerical inversion examples to illustrate how one can qualitatively change the performance of an inversion algorithm.

The rest of the paper is organized as follows. We first present some essential background in section 2, where we briefly review the problem setups for the Helmholtz equation and the diffuse wave equation and present assisting lemmas that will be used in the later analysis. In section 3, we analyze the radius of convergence of the forward scattering series while embedding the PDE solution into $H^{\mathfrak{b}}$ and the variable coefficient into $H^{\mathfrak{a}}$. It is followed by section 4 where we study the inverse scattering operator that maps $H^{\mathfrak{b}} \to H^{\mathfrak{a}}$ regarding its convergence, stability, and approximation error. There has been a recent result [18] deriving the inverse scattering radius of convergence using the geometric function theory under much weaker assumptions. We also incorporate analogous results here for our H^s-type function spaces. In section 5, we discuss the results from sections 3 and 4 and how the important quantities change under different choices of $(\mathfrak{a}, \mathfrak{b})$ pairs. We also compare the radius of convergence and approximation error obtained from the classic approach in [29] and the geometric approach [18]. It is followed by numerical illustrations of the physical differences between the scalar wave and the diffuse wave equations on the radius of convergence for both series. In section 6, we show some numerical inversion examples to demonstrate the improved stability and radius of convergence under proper $(\mathfrak{a}, \mathfrak{b})$ choices. Conclusion follows in section 7.

2. Preliminaries

In this section, we review some background on the forward problems and some preliminary results for the L^2-based Sobolev norms.

2.1. Diffuse and scalar waves

We consider the propagation of a scalar wave into a medium whose pressure field u satisfies the Helmholtz equation

$$\triangle_x u + \frac{\omega^2}{c(x)^2} u(x) = -S(x - x_1, \omega), \quad x \in \mathbb{R}^n, \tag{2.1}$$

where ω is the frequency, $S(x, \omega)$ is the source term, x_1 is the position of the source term, and $c(x)$ is the spatial-dependent velocity. Equation (2.1) can be derived from the scalar wave equation assuming the solution is time harmonic. It governs the propagation of time harmonic acoustic waves of small amplitude in a slowly varying inhomogenous medium [12].

We assume the inhomogenous region is contained inside a ball B_a of radius a, i.e. $c(x) = c_0$ for $x \in \mathbb{R}^n \setminus B_a$. Let the wavenumber $k = \omega/c_0$ and $\eta(x) = c_0^2/c^2(x) - 1$. The scattering problem becomes

$$\triangle_x u + k^2(1 + \eta)u = -S(x - x_1, \omega), \quad x \in \mathbb{R}^n, \tag{2.2}$$

where $u = u_i + u_s$ with the incident wavefield u_i satisfying

$$\triangle_x u_i + k^2 u_i = -S(x - x_1, \omega), \quad x \in \mathbb{R}^n.$$

Here, the scattering wavefield u_s satisfies

$$\triangle_x u_s + k^2 u_s = -k^2 \eta(x) u(x, x_1),$$

and the Sommerfeld radiation condition

$$\lim_{r \to \infty} r^{\frac{n-1}{2}} \left(\frac{\partial u_s}{\partial r} - iku_s \right) = 0, \qquad u_s(x, x_1) = u_s(r\hat{x}, x_1), \tag{2.3}$$

where $\|\hat{x}\| = 1$. Note that we assume $\eta(x) \geqslant -1$ for all $x \in B_a$ and $\eta(x) = 0$ outside B_a. The solution to equation (2.2) satisfies the Lippmann–Schwinger equation

$$u(x, x_1) = u_i(x) + k^2 \int_{\mathbb{R}^n} G(x, y, k) u(y, x_1) \eta(y) \mathrm{d}y, \tag{2.4}$$

where $u_i(x) = \int_{\mathbb{R}^n} G(x, y, k) S(y, \omega) \mathrm{d}y$ and $G(x, y, k)$ is the Green's function such that with y fixed,

$$\triangle_x G + k^2 G = -\delta(x - y).$$

Note that $G(x, y, k) = G(y, x, k)$. The expression of the Green's function depends on the dimensionality of the problem. For example, based on the outgoing Sommerfeld radiation condition equation (2.3),

$$G(x, y, k) = \begin{cases} \frac{i \exp(ik|x-y|)}{2k}, & n = 1, \\ \frac{i}{4} H_0^{(1)}(k|x-y|), & n = 2, \\ \frac{\exp(ik|x-y|)}{4\pi |x-y|}, & n = 3. \end{cases} \tag{2.5}$$

Here, $H_0^{(1)}$ denotes the Hankel function of the first kind and $|x - y|$ is the Euclidean distance between $x, y \in \mathbb{R}^n$.

We can also apply our analysis to the diffuse wave equation [26, 29],

$$\triangle_x u - k^2(1 + \eta)u(x) = -S(x - x_1, \omega), \tag{2.6}$$

where u represents the energy density and $\eta(x)$ is the absorption coefficient, again assumed to be compactly supported on B_a. We refer to [29, section 2] for details regarding the forward problem. Let $G(x,y,k)$ be the Green's function where

$$\triangle_x G - k^2 G = -\delta(x-y),$$

for a fixed y and k. The analytic form of G also depends on the dimension [27],

$$G(x,y,k) = \begin{cases} \frac{\exp(-k|x-y|)}{2k}, & n=1, \\ \frac{1}{2\pi} K_0(k|x-y|), & n=2, \\ \frac{\exp(-k|x-y|)}{4\pi|x-y|}, & n=3, \end{cases} \tag{2.7}$$

where K_0 has the following integral representations for $x > 0$

$$K_0(x) = \int_1^\infty \frac{\exp(-xt)}{(t^2-1)^{\frac{1}{2}}} \mathrm{d}t = \int_0^\infty \exp(-x\cosh t)\mathrm{d}t. \tag{2.8}$$

We can also define the incident and scattering wavefields for the diffuse wave equation equation (2.6) under similar assumptions to the case for the Helmholtz equation equation (2.2). Later, we will use the same set of notations for both the diffuse and scalar waves, except they have different Green's functions.

2.2. The L^2-based Sobolev space

The following definition of the Sobolev norm is based on the Sobolev space $W^{s,p}(\Omega)$ for non-negative s [1].

Definition 1 [Sobolev Space $W^{s,p}(\Omega)$]. Let $1 \leqslant p < \infty$ and s be a nonnegative integer. If a function f and its weak derivatives $D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}, |\alpha| \leqslant s$ all lie in $L^p(\Omega)$, where α is a multi-index and $|\alpha| = \sum_{i=1}^n \alpha_i$, we say $f \in W^{s,p}(\Omega)$ and define the $W^{s,p}(\Omega)$ norm of f as

$$\|f\|_{W^{s,p}(\Omega)} := \left(\sum_{|\alpha| \leqslant s} \|D^\alpha f\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}.$$

We also define the space $W_0^{s,p}(\Omega)$ as the space of functions $f \in W^{s,p}(\Omega)$ with compact support [1]. Next, we define the Sobolev norm on the boundary $\partial\Omega$, where the set Ω is bounded with sufficiently smooth boundaries. We use the following definition from [32, p 4].

Definition 2 (Bounded Domains with Continuous Boundary). Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with the boundary $\partial\Omega$. Suppose the following conditions are satisfied:

- There exist $\lambda, \gamma > 0$, and systems of local charts $(x_{r1}, x_{r2}, \ldots, x_{rn}) = (x_r', x_{rn})$ for $r \in \{1, 2, \ldots, m\}$, and continuous functions a_r defined on the closed $n-1$ dimensional cubes defined by $|x_{ri}| \leqslant \lambda$ for $i \in \{1, 2, \ldots, n-1\}$, such that every point $x \in \partial\Omega$ can be represented as $(x_r', a(x_r'))$ for at least one $r \in \{1, 2, \ldots, m\}$.
- Let $\overline{\Delta_r}$ denote the set of points x_r' such that $|x_{ri}| \leqslant \lambda$ for $i \in \{1, 2, \ldots, n-1\}$. The points (x_r', x_{rn}) where $x_r' \in \overline{\Delta_r}$ and $a_r(x_r') < x_{rn} < a_r(x_r') + \gamma$ are in Ω, while the points (x_r', x_{rn}) where $x_r' \in \overline{\Delta_r}$ and $a_r(x_r') - \gamma < x_{rn} < a_r(x_r')$ are not in $\overline{\Omega}$.

Then, the boundary $\partial\Omega$ is called *continuous*. Furthermore, if the functions a_r are all Lipschitz, then we say Ω has Lipschitz boundary $\partial\Omega$.

For example, the systems of local charts for the surface of a unit cube in $\mathbb{R}^3$ are defined on the six faces. Next, we use the definition for $\mathfrak{R}^{k,\mu}$ domains from [32, p 49].

Definition 3 [Bounded Domains of Type $\mathfrak{R}^{k,\mu}$]. Let k be a nonnegative integer or infinity, and let $0 \leqslant \mu \leqslant 1$. Consider a bounded domain Ω and suppose there exist functions a_r, with $r \in \{1, 2 \ldots, m\}$ defined on $\overline{\Delta_r}$, with continuous derivatives of order $\leqslant k$. Furthermore, suppose for all multi-indices α with $|\alpha| \leqslant k$ and for any $x'_r, y'_r \in \overline{\Delta_r}$, $|D^\alpha a_r(x'_r) - D^\alpha a_r(y'_r)| \leqslant c|x'_r - y'_r|^\mu$, for some constant $c > 0$. Then, we say $\Omega \in \mathfrak{R}^{k,\mu}$.

Using the above definitions, we can define Sobolev spaces on $\partial\Omega$; see [32, p 83].

Definition 4 (Sobolev Space $W^{s,2}(\partial\Omega)$). Let s be a positive integer and consider a domain $\Omega \in \mathfrak{R}^{s-1,1}$. We define the space $W^{s,2}(\partial\Omega)$ as the space of functions f for which $f(x'_r, a_r(x'_r)) = f_r \in W^{s,2}(\Delta_r)$. We define the $W^{s,2}(\partial\Omega)$ norm of f as

$$\|f\|_{W^{s,2}(\partial\Omega)} := \left(\sum_{r=1}^m \|f_r\|^2_{W^{s,2}(\Delta_r)} \right)^{\frac{1}{2}}.$$

The space $W^{s,2}(\partial\Omega)$ also has an inner product: for functions $f, g \in W^{s,2}(\partial\Omega)$, we define

$$\langle f, g \rangle_{W^{s,2}(\partial\Omega)} := \sum_{r=1}^m \langle f_r, g_r \rangle_{W^{s,2}(\Delta_r)} \tag{2.9}$$

where for each $r \in \{1, 2, \ldots, m\}$, $\langle f_r, g_r \rangle_{W^{s,2}(\Delta_r)} = \sum_{|\alpha| \leqslant s} \int_{\Delta_r} D^\alpha f_r(x'_r) \overline{D^\alpha g_r(x'_r)} dx'_r$.

We also extend definition 4 to the case where s is a negative integer, by defining the space $W^{-s,2}(\partial\Omega)$ as the dual of the space $W^{s,2}(\partial\Omega)$, for positive integers s.

Definition 5 (Sobolev Norm $\|\cdot\|_{W^{-s,2}(\partial\Omega)}$). Consider a domain Ω with continuous boundary $\partial\Omega$. For functions f and g on $\partial\Omega$, let $\langle f, g \rangle = \langle f, g \rangle_{W^{0,2}(\partial\Omega)}$ following equation (2.9). For a function f, we define its $W^{-s,2}(\partial\Omega)$ norm as

$$\|f\|_{W^{-s,2}(\partial\Omega)} := \sup_{g \in W^{s,2}(\partial\Omega)} \frac{\langle f, g \rangle}{\|g\|_{W^{s,2}(\partial\Omega)}}.$$

If $\|f\|_{W^{-s,2}(\partial\Omega)} < \infty$, we say $f \in W^{-s,2}(\partial\Omega)$.

For convenience, we will refer to the space $W^{s,2}$ as H^s and $W^{s,2}_0$ as H^s_0 through the rest of this paper. Next, we state the following Poincaré's inequality from [39, equation (18.1)], with a proof similar to [24, theorem 12.17].

Theorem 1 (Poincaré's Inequality in $H^1_0(B_a)$). *For any $f \in H^1_0(B_a)$, where $B_a \subset \mathbb{R}^n$ is a ball of radius a, and for any $1 \leqslant i \leqslant n$, we have*

$$\int_{B_a} |f(x)|^2 dx \leqslant 2a^2 \int_{B_a} \left| \frac{\partial}{\partial x_i} f(x) \right|^2 dx.$$

If $f \in H^s_0(B_a)$ for a positive integer $s \geqslant 1$, theorem 1 allows us to obtain a lower bound on $\|f\|_{H^s(B_a)}$ in terms of $\|f\|_{L^2(B_a)}$, which leads to the following lemma 1.

Lemma 1. *Let $f \in H_0^s(B_a)$, where s is a positive integer and $B_a \subset \mathbb{R}^n$. We define*

$$P(s,a,n) := \left(\sum_{j=0}^{s} \frac{\binom{n+j-1}{n-1}}{(2a^2)^j} \right)^{-\frac{1}{2}}. \tag{2.10}$$

Then, $\|f\|_{L^2(B_a)} \leqslant P(s,a,n)\|f\|_{H^s(B_a)}$.

Proof. Recall that

$$\|f\|_{H^s(B_a)}^2 = \sum_{|\alpha| \leqslant s} \|D^\alpha f\|_{L^2(B_a)}^2 = \sum_{j=0}^{s} \sum_{|\alpha|=j} \|D^\alpha f\|_{L^2(B_a)}^2.$$

For a particular multi-index α where $|\alpha| = j$, after applying theorem 1 for a total of j times, we have

$$\frac{1}{(2a^2)^j} \int_{B_a} |f(x)|^2 \mathrm{d}x \leqslant \int_{B_a} |D^\alpha f(x)|^2 \, \mathrm{d}x.$$

This gives

$$\|f\|_{H^s(B_a)}^2 \geqslant \sum_{j=0}^{s} \sum_{|\alpha|=j} \frac{1}{(2a^2)^j} \|f\|_{L^2(B_a)}^2 = \|f\|_{L^2(B_a)}^2 \sum_{j=0}^{s} \sum_{|\alpha|=j} \frac{1}{(2a^2)^j}.$$

The number of multi-indices α s.t. $|\alpha| = j$ is the number of n-tuples $(\alpha_1, \ldots, \alpha_n)$ of nonnegative integers $(\alpha_i \in \mathbb{N}, \forall i)$ s.t. $\alpha_1 + \ldots + \alpha_n = j$. Thus, there are $\binom{n+j-1}{n-1}$ multi-indices α with $|\alpha| = j$, i.e. $\sum_{|\alpha|=j} 1 = \binom{n+j-1}{n-1}$, which completes the proof. $\qquad\square$

We will prove the following lemma as another generalization of theorem 1.

Lemma 2. *Let $f \in H_0^{s+1}(B_a)$, where s is a positive integer and $B_a \subset \mathbb{R}^n$. Then, the following inequality holds:*

$$\frac{1}{\binom{n+s-1}{n-1}} \sum_{|\alpha|=s} \|D^\alpha f\|_{L^2(B_a)}^2 \leqslant \frac{2a^2}{\binom{n+s}{n-1}} \sum_{|\beta|=s+1} \|D^\beta f\|_{L^2(B_a)}^2.$$

Proof. Consider pairs of multi-indices α, β where $|\alpha| = s$, $|\beta| = s+1$, and $\alpha \leqslant \beta$. Thus, $\exists i \in \{1, \ldots, n\}$ where $\beta_i = \alpha_i + 1$. We define a 'weight' function $w(\alpha, \beta) := \beta_i / |\beta|$ on such pairs of multi-indices. For a fixed β where $|\beta| = s+1$, we have

$$\sum_{|\alpha|=s, \alpha \leqslant \beta} w(\alpha, \beta) = \sum_{\{i : \beta_i > 0\}} \frac{\beta_i}{s+1} = 1, \tag{2.11}$$

where the sum is taken over all valid α. Furthermore, for a fixed α where $|\alpha| = s$,

$$\sum_{|\beta|=s+1, \alpha \leqslant \beta} w(\alpha, \beta) = \sum_{i=1}^{n} \frac{\alpha_i + 1}{s+1} = \frac{n+s}{s+1}. \tag{2.12}$$

Using theorem 1 and equations (2.11) and (2.12), we have the following for $\alpha \leqslant \beta$,

$$2a^2 \sum_{|\beta|=s+1} \|D^\beta f\|_{L^2(B_a)}^2 \geqslant \sum_{|\beta|=s+1} \sum_{|\alpha|=s} w(\alpha, \beta) \|D^\alpha f\|_{L^2(B_a)}^2 = \sum_{|\alpha|=s} \frac{n+s}{s+1} \cdot \|D^\alpha f\|_{L^2(B_a)}^2.$$

This gives the desired inequality after dividing both sides by $\binom{n+s}{n-1}$. $\qquad\square$

Next, we prove the following lemma 3, which will allow us to apply definitions 4 and 5 to the Green's function of both the Helmholtz and diffuse wave equations, under the condition that $\text{dist}(\partial\Omega, B_a) = \sup_{x\in\partial\Omega, y\in B_a} |x-y| > 0$.

Lemma 3. *Let $s \in \mathbb{N}^+$ and the domain $\Omega \in \mathfrak{R}^{s-1,1}$. Suppose $\text{dist}(\partial\Omega, B_a) = \epsilon > 0$. For any C^∞ function $f : [\epsilon, \infty) \mapsto \mathbb{R}$, $\widetilde{f}(x) := f(|x-y|) \in H^s(\partial\Omega)$ for any fixed $y \in B_a$.*

Proof. We can express $\widetilde{f}$ as the composition of two functions $f \circ g$, where $g(x) = |x-y|$. First, we claim that $g \in H^s(\partial\Omega)$. For a fixed local chart Δ_r, we can write $g(x)$ as

$$g_r(x_r') = |(x_r', a_r(x_r')) - y| = \sqrt{|x_r' - y'|^2 + (a_r(x_r') - y_n)^2},$$

where $y' = (y_1, \ldots, y_{n-1})$. Since $\Omega \in \mathfrak{R}^{s-1,1}$, by definition $a_r \in C^{s-1}(\Delta_r)$. Furthermore, the derivatives of a_r of order at most $s-1$ are all Lipschitz. Thus, from Rademacher's theorem [16], the derivatives of a_r of order s exist almost everywhere and we get that $a_r \in H^s(\Delta_r)$. We also have $g_r(x_r') \in H^s(\Delta_r)$ based on the chain rule.

We define $\widetilde{f}_r(x_r') = \widetilde{f}(x_r', a_r(x_r')) = f(g_r(x_r'))$ on the local chart Δ_r. Since the range of $g_r(x_r')$ is a subset of $[\epsilon, \infty)$, and $f \in C^\infty([\epsilon, \infty))$, we have $\widetilde{f}_r(x_r') \in H^s(\Delta_r)$ by the chain rule. Since for every local chart Δ_r, we have $\widetilde{f}_r(x_r') \in H^s(\Delta_r)$, we get that $\widetilde{f}(x) \in H^s(\partial\Omega)$. $\square$

Remark 1. We can apply definition 4 to the Green's function of the 2D and 3D Helmholtz equations (see equation (2.5)), since the necessary conditions to use lemma 3 apply to such $G(x,y)$. In 2D, the Green's function for the Helmholtz equation is $G(x,y) = \frac{i}{4}H_0^{(1)}(k|x-y|) = \frac{i}{4}H_0^{(1)}(kr)$, where $H_0^{(1)}(x) \in C^\infty((0,\infty))$ is the so-called Hankel function, and thus, $\frac{i}{4}H_0^{(1)}(kr) \in C^\infty([\epsilon,\infty))$ for any fixed $\epsilon > 0$. In 3D, the Green's function is $G(x,y) = f(|x-y|)$, where $f(r) = \exp(ikr)/(4\pi r)$. From the product rule, $f^{(n)}(r)$ is a polynomial in $\exp(ikr)$ and r^{-1}, with complex coefficients, $\forall n \in \mathbb{N}$. Thus, $\forall \epsilon > 0$, we also have that $f(r) \in C^\infty([\epsilon,\infty))$.

Similarly, we can also apply definition 4 to the Green's function of the 2D and 3D diffuse wave equations (see equation (2.7)). In 2D, the Green's function is $G(x,y) = \frac{1}{2\pi}K_0(k|x-y|) = \frac{1}{2\pi}K_0(kr)$, where $K_0(x)$ has the integral representation in equation (2.8) for $x > 0$. From equation (2.8), we get $K_0^{(n)}(x) = \int_0^\infty (-\cosh t)^n \exp(-x\cosh t)\mathrm{d}t$. Thus, $K_0(x) \in C^\infty((0,\infty))$. This implies that for any fixed $\epsilon > 0$, we must have $\frac{1}{2\pi}K_0(kr) \in C^\infty([\epsilon,\infty))$. Finally, in 3D, the Green's function is $G(x,y) = f(|x-y|)$, where $f(r) = \exp(-kr)/(4\pi r)$. From the product rule, $f(r) \in C^\infty([\epsilon,\infty))$ for any fixed $\epsilon > 0$.

We present lemma 4, which will be used in section 3 to obtain bounds for $\|f\|_{H^s(\partial\Omega)}$.

Lemma 4. *Given $s \in \mathbb{Z}$, let d be the total number of different multi-indices α where $|\alpha| \leqslant |s|$. There exists a linear operator*

$$\mathbf{T} : H^s(\partial\Omega) \to L^2(\partial\Omega) \times \ldots \times L^2(\partial\Omega)$$

such that $\mathbf{T}f : \partial\Omega \mapsto \mathbb{R}^d$ and $\|f\|_{H^s(\partial\Omega)} = \|\mathbf{T}f\|_{L^2}$ for all functions $f \in H^s(\partial\Omega)$.

Proof. First, if $s = 0$, then we can take $\mathbf{T}$ to be the identity. Next, consider the case where $s > 0$. We define the linear operator $\mathbf{T}$ such that for a point

$$x = (x_{r1}, x_{r2}, \ldots, x_{rn}) = (x_r', x_{rn}) = (x_r', a_r(x_r'))$$

as defined in definition 2, and a function $f \in H^s(\partial\Omega)$, $\mathbf{T}f(x)$ is the vector with components $D^\alpha f_r(x_r')$ for each multi-index $\alpha \in \mathbb{R}^{n-1}$ with $|\alpha| \leqslant s$. With this definition of $\mathbf{T}$, we note that $\mathbf{T}f$

is a piecewise function which depends on the local chart containing the input $x = (x'_r, a_r(x'_r))$. However, $\mathbf{T}$ is still a linear operator in f. From definition 4, this gives $\|f\|_{H^s(\partial\Omega)} = \|\mathbf{T}f\|_{L^2}$.

Also, if $s > 0$, $\|\cdot\|_{H^{-s}(\partial\Omega)}$ is defined following definition 5. For fixed $f \in H^{-s}(\partial\Omega)$, and any $g \in H^s(\partial\Omega)$, we have $\langle f, g \rangle \leqslant \|f\|_{H^{-s}(\partial\Omega)}\|g\|_{H^s(\partial\Omega)}$. Hence, $\langle f, g \rangle$ is a bounded linear functional on $H^s(\partial\Omega)$ for a fixed f. From the Riesz representation theorem, there is a unique $u \in H^s(\partial\Omega)$ such that for all $g \in H^s(\partial\Omega)$, $\langle f, g \rangle = \langle u, g \rangle_{H^s(\partial\Omega)}$, and moreover, $\|f\|_{H^{-s}(\partial\Omega)} = \|u\|_{H^s(\partial\Omega)}$. Since u is unique, there is an operator $\mathcal{L} : H^{-s}(\partial\Omega) \to H^s(\partial\Omega)$ such that $u = \mathcal{L}f$. It is easy to check that $\mathcal{L}$ is a linear operator and $\|f\|_{H^{-s}(\partial\Omega)} = \|\mathcal{L}f\|_{H^s(\partial\Omega)}$. From the proof for the case $s \geqslant 0$, there exists a linear operator $\mathbf{T}_1$ such that $\|u\|_{H^s(\partial\Omega)} = \|\mathbf{T}_1 u\|_{L^2(\partial\Omega)}$. Thus,

$$\|f\|_{H^{-s}(\partial\Omega)} = \|u\|_{H^s(\partial\Omega)} = \|\mathbf{T}_1 u\|_{L^2(\partial\Omega)} = \|\mathbf{T}_1\mathcal{L}f\|_{L^2(\partial\Omega)},$$

giving $\mathbf{T} = \mathbf{T}_1\mathcal{L}$ as the desired linear operator. $\square$

3. Forward scattering series

Similar to [29, equation (7)], we apply fixed point iteration beginning with u_i in equation (2.4), which gives the following infinite series:

$$u(x, x_1) = u_i(x, x_1) + k^2 \int_{B_a} G(x, y_1)\eta(y_1)u_i(y_1, x_1)\mathrm{d}y_1$$
$$+ k^4 \int_{B_a}\int_{B_a} G(x, y_1)\eta(y_1)G(y_1, y_2)\eta(y_2)u_i(y_2, x_1)\mathrm{d}y_1\mathrm{d}y_2 + \dots \quad (3.1)$$

Letting $\phi = u - u_i$, we can express the infinite series in the following manner:

$$\phi = K_1\eta + K_2\eta \otimes \eta + K_3\eta \otimes \eta \otimes \eta + \dots \quad (3.2)$$

where

$$(K_jf)(x, x_1) = k^{2j} \int_{\mathcal{B}_j} G(x, y_1)G(y_1, y_2)\dots G(y_{j-1}, y_j)u_i(y_j, x_1)f\mathrm{d}y_1\dots\mathrm{d}y_j, \quad (3.3)$$

where for $j \geqslant 1$,

$$\mathcal{B}_j = B_a \times \dots \times B_a, \quad (3.4)$$

$$f = f(y_1, \dots, y_j) = \eta(y_1)\dots\eta(y_j). \quad (3.5)$$

3.1. Bounding the $L^2 \to H^s$ norm of K_j

We will first provide an upper bound for the quantity $\|K_j\|_{L^2 \to H^s}$ for any $s \in \mathbb{Z}$. To do this, we will first generalize the operator K_j, which outputs a scalar-valued function when applied to f, to the operator A_j, which outputs a vector-valued function instead. Then, using lemma 4, we will convert the problem of bounding $\|K_jf\|_{H^s}$, for a function f, into the problem of bounding $\|A_jf\|_{L^2}$.

3.1.1. Generalizing the operator K_j.   Consider a vector-valued function

$$\mathbf{v}(x, y_1) : \partial\Omega \times B_a \to \mathbb{R}^d,$$

where d is the number of different multi-indices α such that $|\alpha| \leqslant |s|$. We define

$$A_j : L^2(\mathcal{B}) \to L^2(\partial\Omega \times \partial\Omega)$$

such that for function f in equation (3.5) and $j \geqslant 1$, we have

$$(A_j f)(x, x_1) = k^{2j} \int_{\mathcal{B}_j} \mathbf{v}(x, y_1) G(y_1, y_2) \ldots G(y_{j-1}, y_j) f\, u_i(y_j, x_1) \mathrm{d}y_1 \ldots \mathrm{d}y_j. \quad (3.6)$$

In other words, the operator A_j is a generalization of K_j obtained by replacing the scalar-valued $G(x, y_1)$ with the vector-valued $\mathbf{v}(x, y_1)$. We first determine an upper bound on the L^2 norm of the operator A_j for $j \geqslant 1$, and here,

$$\|A_j f\|_{L^2}^2 = \sum_{r=1}^m \sum_{r_1=1}^m \int_{\Delta_{r_1}} \int_{\Delta_r} |(A_j f)(x, x_1)|^2 \mathrm{d}x_r' \mathrm{d}x_{r_1}',$$

where for a pair of local charts (r, r_1), we have $x = (x_r', a_r(x_r'))$ and $x_1 = (x_{r_1}', a_{r_1}(x_{r_1}'))$. We first prove the following lemma which gives an upper bound on $\|A_j\|_{L^2}$. We define

$$C_a = |B_a|^{\frac{1}{2}} \sup_{y \in B_a} \|u_i(y, \cdot)\|_{L^2(\partial\Omega)}.$$

Lemma 5. *Let* $\mu_A := k^2 \sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(B_a)}$ *and* $\nu_A := k^2 C_a \sup_{y \in B_a} \|\mathbf{v}(\cdot, y)\|_{L^2(\partial\Omega)}$. *Then, the operator* A_j *defined by equation (3.6) satisfies*

$$\|A_j\|_{L^2} \leqslant \nu_A \mu_A^{j-1}.$$

Proof. From the Cauchy–Schwarz inequality,

$$|(A_j f)(x, x_1)|^2 \leqslant k^{4j} \|f\|_{L^2}^2 \int_{\mathcal{B}_j} |\mathbf{v}(x, y_1) G(y_1, y_2) \ldots G(y_{j-1}, y_j) u_i(y_j, x_1)|^2 \mathrm{d}y_1 \ldots \mathrm{d}y_j.$$

Let $x = (x_r', a_r(x_r'))$ and $x_1 = (x_{r_1}', a_{r_1}(x_{r_1}'))$. We first bound $\|A_1\|_{L^2}$:

$$\|A_1 \eta\|_{L^2}^2 \leqslant k^4 \|\eta\|_{L^2}^2 \sum_{r=1}^m \sum_{r_1=1}^m \int_{\Delta_r} \int_{\Delta_{r_1}} \int_{B_a} |\mathbf{v}(x, y_1) u_i(y_1, x_1)|^2 \mathrm{d}y_1 \mathrm{d}x_{r_1}' \mathrm{d}x_r'$$

$$= k^4 \|\eta\|_{L^2}^2 \int_{B_a} \|\mathbf{v}(\cdot, y_1)\|_{L^2(\partial\Omega)}^2 \|u_i(y_1, \cdot)\|_{L^2(\partial\Omega)}^2 \mathrm{d}y_1$$

$$\leqslant k^2 \|\eta\|_{L^2} C_a \sup_{y_1 \in B_a} \|\mathbf{v}(\cdot, y_1)\|_{L^2(\partial\Omega)}.$$

Thus, $\|A_1\|_{L^2} \leqslant k^2 C_a \sup_{y_1 \in B_a} \|\mathbf{v}(\cdot, y_1)\|_{L^2(\partial\Omega)}$. Now, we estimate $\|A_j\|_{L^2}$ where $j \geqslant 2$:

$$\|A_j f\|_{L^2}^2 \leqslant k^{4j} \|f\|_{L^2}^2 \int_{\mathcal{B}_j} |G(y_1, y_2) \ldots G(y_{j-1}, y_j)|^2 \mathrm{d}y_1 \ldots \mathrm{d}y_j$$

$$\sup_{y_1 \in B_a, y_j \in B_a} \sum_{r=1}^m \sum_{r_1=1}^m \int_{\Delta_r} \int_{\Delta_{r_1}} |\mathbf{v}(x, y_1) u_i(y_j, x_1)|^2 \mathrm{d}x_{r_1}' \mathrm{d}x_r'.$$

This gives

$$\|A_j\|_{L^2} \leqslant k^{2j} J_{j-1} \sup_{y \in B_a} \|u_i(y, \cdot)\|_{L^2(\partial\Omega)} \sup_{y \in B_a} \|\mathbf{v}(\cdot, y)\|_{L^2(\partial\Omega)},$$

where $J_{j-1} := \left(\int_{\mathcal{B}_j} |G(y_1, y_2) \ldots G(y_{j-1}, y_j)|^2 \mathrm{d}y_1 \ldots \mathrm{d}y_j \right)^{\frac{1}{2}}$. As shown in [29, equation (30)],

$$J_{j-1} \leqslant |B_a|^{\frac{1}{2}} \left(\sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(B_a)} \right)^{j-1},$$

which gives $\|A_j\|_{L^2} \leqslant \nu_A \mu_A^{j-1}$ based on their definitions. $\qquad \square$

Remark 2. If $d = 1$ and $\mathbf{v}(x, y) = G(x, y)$, then the operators A_j and K_j are the same. Thus, the bound proved in lemma 5 is a more general version of the bound proved in [29]. Our bound is different in that ν_A contains a factor of $\sup_{y \in B_a} \|u_i(y, \cdot)\|_{L^2(\partial\Omega)}$, which is a minor correction to [29, equation (25)].

3.1.2. Bounding $\|K_j\|_{L^2 \to H^s}$. In this section, we will consider the operator

$$K_j : L^2(\mathcal{B}_j) \to H^s(\partial\Omega) \times L^2(\partial\Omega)$$

defined in equation (3.3) for $s \in \mathbb{Z}$. Using lemmas 4 and 5, we will get an upper bound on $\|K_j\|_{L^2 \to H^s}$ for any $s \in \mathbb{Z}$. We define

$$\|K_j f\|_{H^s \times L^2}^2 := \sum_{r_1=1}^m \int_{\Delta_{r_1}} \|(K_j f)(\cdot, x_1)\|_{H^s(\partial\Omega)}^2 \mathrm{d}x'_{r_1},$$

where $x_1 = (x'_{r_1}, a_{r_1}(x'_{r_1}))$ and $\|(K_j f)(\cdot, x_1)\|_{H^s(\partial\Omega)}$ is defined through definition 4.

Lemma 6. Let $\mu_s := k^2 \sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(B_a)}$ and $\nu_s := k^2 C_a \sup_{y \in B_a} \|G(y, \cdot)\|_{H^s(\partial\Omega)}$ for $s \in \mathbb{Z}$.

Then, the operator K_j in equation (3.3) satisfies $\|K_j\|_{L^2 \to H^s} \leqslant \nu_s \mu_s^{j-1}$.

Proof. From lemma 4, there exists a linear operator $\mathbf{T}$ acting only on x such that

$$\|K_j f\|_{H^s \times L^2} = \|\mathbf{T} K_j f\|_{L^2(\partial\Omega \times \partial\Omega)}$$

where $\mathbf{T} K_j f(x, x_1) \in \mathbb{R}^d$ for some positive integer d. Then, for fixed x_1,

$$\mathbf{T} K_j f(\cdot, x_1) = \mathbf{T} \left(k^{2j} \int_{\mathcal{B}_j} G(\cdot, y_1) G(y_1, y_2) \ldots G(y_{j-1}, y_j) f(y_1, \ldots, y_j) u_i(y_j, x_1) \mathrm{d}y_1 \ldots \mathrm{d}y_j \right)$$

$$= k^{2j} \int_{\mathcal{B}_j} \mathbf{T} G(\cdot, y_1) G(y_1, y_2) \ldots G(y_{j-1}, y_j) f(y_1, \ldots, y_j) u_i(y_j, x_1) \mathrm{d}y_1 \ldots \mathrm{d}y_j,$$

where the last equation follows since $\mathbf{T}$ is linear. Letting $\mathbf{v}(\cdot, y_1) = \mathbf{T} G(\cdot, y_1)$, which is a vector-valued function in x, we get

$$(A_j f)(x, x_1) = k^{2j} \int_{\mathcal{B}_j} \mathbf{v}(x, y_1) G(y_1, y_2) \ldots G(y_{j-1}, y_j) f(y_1, \ldots, y_j) u_i(y_j, x_1) \mathrm{d}y_1 \ldots \mathrm{d}y_j.$$

From lemma 5, we have

$$\|K_j f\|_{H^s \times L^2} = \|A_j f\|_{L^2} \leqslant \nu_A \mu_A^{j-1} \|f\|_{L^2},$$

where $\mu_A = \mu_s$, and

$$\nu_A = k^2 C_a \sup_{y \in B_a} \|\mathbf{T} G(\cdot, y)\|_{L^2(\partial\Omega)} = k^2 C_a \sup_{y \in B_a} \|G(\cdot, y)\|_{H^s(\partial\Omega)} = \nu_s,$$

following from lemma 4 and the fact that G is symmetric. $\qquad \square$

11

3.2. Bounding the $H^{\mathfrak{a}} \to H^{\mathfrak{b}}$ norm of K_j

We can now derive a bound on the norm of the operator

$$K_j : H^{\mathfrak{a}}(\mathcal{B}_j) \to H^{\mathfrak{b}}(\partial\Omega) \times L^2(\partial\Omega),$$

where $\mathfrak{a}$ and $\mathfrak{b}$ are integers and $\mathfrak{a} \geqslant 0$. Consider the function f defined in equation (3.5) based on $\eta \in H_0^{\mathfrak{a}}(B_a)$. Then, based on $\mathcal{B}_j$ defined in equation (3.4), we have

$$\|f\|_{L^2}^2 = \int_{\mathcal{B}_j} |\eta(y_1)|^2 |\eta(y_2)|^2 \ldots |\eta(y_j)|^2 \mathrm{d}y_1 \mathrm{d}y_2 \ldots \mathrm{d}y_j = \|\eta\|_{L^2}^{2j}.$$

Because f is the function obtained by copying the same $\eta(y)$ for a total of j times, we can simply define

$$\|f\|_{H^{\mathfrak{a}}} := \|\eta\|_{H^{\mathfrak{a}}}^j, \tag{3.7}$$

as done in [40]. We now prove the following lemma.

Lemma 7. *Let $\mathfrak{a}, \mathfrak{b} \in \mathbb{Z}$, and $\mathfrak{a} \geqslant 0$. Define*

$$\mu_{\mathfrak{a}\mathfrak{b}} := k^2 P(\mathfrak{a}, a, n) \sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(B_a)}, \tag{3.8}$$

$$\nu_{\mathfrak{a}\mathfrak{b}} := k^2 P(\mathfrak{a}, a, n) C_a \sup_{y \in B_a} \|G(y, \cdot)\|_{H^{\mathfrak{b}}(\partial\Omega)}. \tag{3.9}$$

Then, the operator K_j defined by equation (3.3) satisfies $\|K_j\|_{H^{\mathfrak{a}} \to H^{\mathfrak{b}}} \leqslant \nu_{\mathfrak{a}\mathfrak{b}} \mu_{\mathfrak{a}\mathfrak{b}}^{j-1}$.

Proof. First, note that the following equations hold:

$$\nu_{\mathfrak{a}\mathfrak{b}} = \nu_{\mathfrak{b}} P(\mathfrak{a}, a, n), \qquad \mu_{\mathfrak{a}\mathfrak{b}} = \mu_{\mathfrak{b}} P(\mathfrak{a}, a, n). \tag{3.10}$$

We have $\|f\|_{L^2} = \|\eta\|_{L^2}^j$, and from equation (3.7), we get $\|f\|_{H^{\mathfrak{a}}} = \|\eta\|_{H^{\mathfrak{a}}}^j$. Since $\eta \in H_0^{\mathfrak{a}}(B_a)$, we can apply lemma 1 and obtain $\|\eta\|_{L^2} \leqslant \|\eta\|_{H^{\mathfrak{a}}} P(\mathfrak{a}, a, n)$. From lemma 6,

$$\|K_j f\|_{H^{\mathfrak{b}} \times L^2} \leqslant \|f\|_{L^2} \nu_{\mathfrak{b}} \mu_{\mathfrak{b}}^{j-1} \leqslant \|f\|_{H^{\mathfrak{a}}} \nu_{\mathfrak{b}} \mu_{\mathfrak{b}}^{j-1} P(\mathfrak{a}, a, n)^j = \|f\|_{H^{\mathfrak{a}}} \nu_{\mathfrak{a}\mathfrak{b}} \mu_{\mathfrak{a}\mathfrak{b}}^{j-1},$$

and the proof is completed. $\qquad \square$

Remark 3. Through a method similar to [29, proposition 2.1], one can obtain an upper bound $1/\mu_{\mathfrak{a}\mathfrak{b}}$ on the radius of convergence of the series in equation (3.1) under the $H^{\mathfrak{a}} \to H^{\mathfrak{b}}$ norm.

Remark 4. From equations (3.8) and (3.9), we see that both $\mu_{\mathfrak{a}\mathfrak{b}}$ and $\nu_{\mathfrak{a}\mathfrak{b}}$ decrease as $\mathfrak{a}$ increases while $\mathfrak{b}$ is fixed. As $\mathfrak{b}$ increases while holding $\mathfrak{a}$ constant, $\mu_{\mathfrak{a}\mathfrak{b}}$ is constant, while $\nu_{\mathfrak{a}\mathfrak{b}}$ increases.

4. Inverse scattering series

In this section, we study the properties of the inverse scattering series. In section 4.1, we prove the convergence, stability, and approximation error of the inverse scattering series under the $H^{\mathfrak{b}} \to H^{\mathfrak{a}}$ operator norm in lemma 8, theorem 2, theorem 3, and theorem 4, using the classic approach in [29]. In section 4.2, we present similar results using the geometric approach in [18], where the assumptions are less strict. Later in section 5, we will compare analogous theorems under the classic and geometric approaches in detail.

4.1. Classic approach

First, we present theorems and proofs similar to analogous results in [29] under the $L^p \mapsto L^p$ operator norm where $p \geqslant 2$, which are also derived in generality for Banach spaces in [6]. Moreover, these results have been improved in [31], and their analysis applies to our setting as well. Thus, we omit the proofs of theorem 2 and theorem 3. However, we provide sharper bounds on the approximation error in theorem 4, so we include its proof and proofs of necessary lemmas in A for completeness. This will be important as we analyze how the constants from these theorems change with respect to different choices of $\mathfrak{a}$ and $\mathfrak{b}$ in section 5.

As in [29], we assume that η can be expressed as a series based on $\phi = u - u_i$:

$$\eta = \mathcal{K}_1 \phi + \mathcal{K}_2 \phi \otimes \phi + \mathcal{K}_3 \phi \otimes \phi \otimes \phi + \dots, \tag{4.1}$$

and substituting equation (3.2) into equation (4.1) gives the solution for $\mathcal{K}_j$, as shown in [29]:

$$\mathcal{K}_1 = K_1^+,$$
$$\mathcal{K}_j = -\left(\sum_{m=1}^{j-1} \mathcal{K}_m \sum_{i_1 + \dots + i_m = j} K_{i_1} \otimes \dots \otimes K_{i_m} \right) \mathcal{K}_1 \otimes \dots \otimes \mathcal{K}_1, \tag{4.2}$$

where $\mathcal{K}_1 = K_1^+$ is the regularized pseudoinverse of the operator K_1. We will first analyze the convergence of the inverse scattering series equation (4.1) under the $H^{\mathfrak{b}} \to H^{\mathfrak{a}}$ norm, where $\mathfrak{a} \geqslant 0$ and $\mathfrak{a}, \mathfrak{b}$ are integers. Here, η, which is the perturbation from the homogeneous medium, belongs to $H_0^{\mathfrak{a}}(B_a)$ and the scattering data $\phi(x, x_1) = u(x, x_1) - u_i(x, x_1)$ belongs to $H^{\mathfrak{b}}(\partial\Omega) \times L^2(\partial\Omega)$ where x_1 denotes the source location and x represents the receiver location. We will denote

$$\|\mathcal{K}_j\|_{\mathfrak{ba}} = \|\mathcal{K}_j\|_{H^{\mathfrak{b}} \to H^{\mathfrak{a}}}$$

for convenience. Next, we will first find an upper bound on $\|\mathcal{K}_j\|_{\mathfrak{ba}}$ in lemma 8.

Lemma 8. *Suppose $j \geqslant 2$ and $(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1$. Let*

$$C := \|\mathcal{K}_1\|_{\mathfrak{ba}} \exp\left(\frac{1}{1 - (\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\|_{\mathfrak{ba}}} \right). \tag{4.3}$$

Then, the operator $\mathcal{K}_j : \left(H^{\mathfrak{b}}(\partial\Omega) \times L^2(\partial\Omega) \right)^j \to H^{\mathfrak{a}}(B_a)$ defined by equation (4.2) satisfies

$$\|\mathcal{K}_j\|_{\mathfrak{ba}} \leqslant C(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})^j \|\mathcal{K}_1\|_{\mathfrak{ba}}^j.$$

Moreover, for all $\phi \in H^{\mathfrak{b}}(\partial\Omega) \times L^2(\partial\Omega)$, we have

$$\|\mathcal{K}_j \phi \otimes \dots \otimes \phi\|_{H^{\mathfrak{a}}} \leqslant C(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})^j \|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}}^j.$$

We then present theorem 2, which gives an upper bound on the radius of convergence of the inverse scattering series and estimates the series limit $\widetilde{\eta}$.

Theorem 2. *Suppose $\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$ and $\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$. Then, the inverse scattering series converges under the $H^{\mathfrak{b}} \to H^{\mathfrak{a}}$ norm. Furthermore, if $N \geqslant 1$ and $\widetilde{\eta}$ is the limit of the inverse scattering series, the following bound holds:*

$$\left\| \widetilde{\eta} - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \dots \otimes \phi \right\|_{H^{\mathfrak{a}}} \leqslant C \frac{\left[(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}} \right]^{N+1}}{1 - (\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}}}, \tag{4.4}$$

where C is defined in equation (4.3).

Next, we present a result bounding the perturbations in η based on the perturbations in ϕ up to multiplication by an explicit stability constant in theorem 3.

Theorem 3. *Consider scattering data ϕ_1 and ϕ_2, and let $M := \max\{\|\phi_1\|_{H^\flat}, \|\phi_2\|_{H^\flat}\}$. Let η_1 and η_2 denote the corresponding limits of the inverse scattering series, and let*

$$C^* := \max\left\{\frac{1}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}, C\right\}, \tag{4.5}$$

where C is defined in equation (4.3). Furthermore, let

$$\widetilde{C} := \frac{C^*(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\|_{\mathfrak{ba}}}{(1 - (\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\|_{\mathfrak{ba}}M)^2}. \tag{4.6}$$

If $\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$ and $M\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, the following bound holds:

$$\|\eta_1 - \eta_2\|_{H^\mathfrak{a}} < \widetilde{C}\|\phi_1 - \phi_2\|_{H^\flat \times L^2}.$$

While the inverse scattering series converges under the conditions of theorem 2, its limit is not equal to η in general. In theorem 4, we give an upper bound on the distance between the series limit $\widetilde{\eta}$ and η.

Theorem 4. *Let $\mathcal{M} := \max\{\|\eta\|_{H^\mathfrak{a}}, \|\mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}}\}$ and let*

$$C_{\mathfrak{ab}} := \frac{C^*(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})}{(1 - (\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\mathcal{M})^2}. \tag{4.7}$$

If $\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, $\|\mathcal{K}_1\phi\|_{H^\mathfrak{a}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, and $\mathcal{M} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, then the approximation error of the partial sum of equation (4.1) satisfies the following bound:

$$\left\|\eta - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \ldots \otimes \phi\right\|_{H^\mathfrak{a}} \leqslant C_{\mathfrak{ab}}\|(I - \mathcal{K}_1 K_1)\eta\|_{H^\mathfrak{a}} + C\frac{[(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\phi\|_{H^\mathfrak{a}}]^{N+1}}{1 - (\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})\|\mathcal{K}_1\phi\|_{H^\mathfrak{a}}}$$

where C is defined in equation (4.3).

Note that theorem 4 also implies the following by taking $N \to \infty$.

Corollary 1. *Suppose the hypotheses of theorem 4 hold. Then, the approximation error of the inverse scattering series satisfies the following bound:*

$$\|\eta - \widetilde{\eta}\|_{H^\mathfrak{a}} \leqslant C_{\mathfrak{ab}}\|\eta - \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}}, \tag{4.8}$$

where $\widetilde{\eta}$ is the limit of the inverse scattering series and $C_{\mathfrak{ab}}$ is defined in equation (4.7).

4.2. Geometric approach

Recent work [18] has improved on the results in [6, 29, 31] through a different approach, obtaining estimates on the radius of convergence and approximation error of the inverse scattering series with more relaxed assumptions. The results in [29] can easily be adapted to our setup. We thus include statements and omit the proofs which are similar to those in [29]. We will later investigate how these results change with respect to the choices of $\mathfrak{a}$ and $\mathfrak{b}$ and how they are compared with those from the classic approach.

First, we present theorem 5, which provides a radius of convergence estimate for the inverse scattering series, with different conditions from theorem 2. The main difference is that theorem 5 no longer requires an upper bound for $\|\mathcal{K}_1\|_{\mathfrak{ba}}$, which is difficult to achieve in realistic settings unless the regularization coefficient is very large [18].

Theorem 5. *Let $C_1 = \max\{2, \|\mathcal{K}_1\|_{\mathfrak{ba}} \nu_{\mathfrak{ab}}\}$. Then, the inverse scattering series converges if $\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}} < r$, where*

$$r = \frac{1}{2\mu_{\mathfrak{ab}}} \left(\sqrt{16C_1^2 + 1} - 4C_1 \right) \tag{4.9}$$

is the radius of convergence.

Next, we present theorem 6, which provides an estimate on the approximation error of the inverse scattering series with different conditions than theorem 4.

Theorem 6. *Suppose that the forward scattering series and the inverse scattering series both converge, so $\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}} < r$, where r is defined in equation (4.9). Let $\widetilde{\eta}$ be the sum of the inverse scattering series and $\mathcal{M}_1 := \max\{\|\eta\|_{H^{\mathfrak{a}}}, \|\widetilde{\eta}\|_{H^{\mathfrak{a}}}\}$. Define the constant $\widetilde{C}_{\mathfrak{ab}}$ by*

$$\widetilde{C}_{\mathfrak{ab}} := \left(1 - \frac{\nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}}}{(1 + \nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}} - \mu_{\mathfrak{ab}} \mathcal{M}_1)^2} \right)^{-1}. \tag{4.10}$$

If $\mathcal{M}_1 < \frac{1}{\mu_{\mathfrak{ab}}} \left(1 - \sqrt{\frac{\nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}}}{1 + \nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}}}} \right)$, then the approximation error can be bounded above as

$$\left\| \eta - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \ldots \otimes \phi \right\|_{H^{\mathfrak{a}}} \leqslant \widetilde{C}_{\mathfrak{ab}} \|(I - \mathcal{K}_1 K_1)\eta\|_{H^{\mathfrak{a}}}$$

$$+ \frac{2\mu_{\mathfrak{ab}}}{\sqrt{16C_1^2 + 1}} \left(\frac{\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}}}{r} \right)^{N+1} \frac{1}{1 - \frac{\|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}}}{r}}.$$

5. Discussion and summary of results

In this section, we analyze how the radius of convergence, stability, and approximation error of the inverse scattering series in section 4 change with respect to $\mathfrak{a}$ and $\mathfrak{b}$. The discussion could shed light on how to choose the metric space for the parameter η and the choice of the objective function in the corresponding inverse data-matching problems. In section 5.1, we first interpret the results in section 4.1 obtained by the classic approach. In section 5.2, we then compare the results from the classic approach and the geometric approach given in section 4.2, followed by numerical illustrations on the radius of convergence for the scalar and diffuse waves in section 5.3.

5.1. Discussions on results in section 4.1

In this subsection, we analyze and interpret the results presented in section 4.1 using the classic approach.

5.1.1. Radius of convergence and assumptions.
Based on theorem 2, the inverse scattering series will converge if

$$\|\mathcal{K}_1\|_{\mathfrak{ba}} < \frac{1}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}, \qquad \|\mathcal{K}_1 \phi\|_{H^{\mathfrak{a}}} < \frac{1}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}.$$

The radius of convergence of the inverse scattering series, $1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, controls the allowed sizes of $\|\mathcal{K}_1\|_{\mathfrak{ba}}$ and $\|\phi\|_{H^{\mathfrak{b}}}$ needed to apply our results in section 4. An increased radius of convergence is beneficial, as it indicates that the inverse scattering series will converge under a relatively large perturbation in data, $\|\phi\|_{H^{\mathfrak{b}}}$. We can thus apply the results from section 4 to

more cases. As $\mathfrak{a}$ increases, both $\mu_{\mathfrak{ab}}$ and $\nu_{\mathfrak{ab}}$ decrease, causing $1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$ to increase. As $\mathfrak{b}$ decreases, $\nu_{\mathfrak{ab}}$ will decrease and $\mu_{\mathfrak{ab}}$ remains constant, causing $1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$ to increase. To sum up, the radius of convergence of the *inverse* scattering series increases as $\mathfrak{a}$ increases and $\mathfrak{b}$ decreases. The radius of convergence of the *forward* scattering series increases as $\mathfrak{a}$ increases and does not depend on $\mathfrak{b}$.

To apply the results from section 4, we will assume that the conditions of lemma 8, theorems 2–4 hold. As the size of $\|\mathcal{K}_1\|_{\mathfrak{ba}}$ is determined by regularization [29, remark 3.2], we will assume that

$$\|\mathcal{K}_1\|_{\mathfrak{ba}} = Q/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}), \tag{5.1}$$

for a fixed constant $0 < Q < 1$ independent of $\mathfrak{a}$ and $\mathfrak{b}$.

5.1.2. Analyzing stability and approximation error.   We first analyze how the constant C in equation (4.3), and the constant C^* in equation (4.5), change with respect to $\mathfrak{a}$, $\mathfrak{b}$ under the above assumptions. Based on the assumption in equation (5.1), we have

$$C = \frac{Q \exp\left[(1-Q)^{-1}\right]}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}, \qquad C^* = \max\left\{\frac{1}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}, C\right\}, \tag{5.2}$$

implying that both C and C^* are proportional to $1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}})$, the inverse scattering series radius of convergence. Thus, under the assumption in equation (5.1), we see that C and C^* decrease as $\mathfrak{a}$ decreases and $\mathfrak{b}$ increases.

5.1.2.1. Stability Constant.   For scattering data ϕ_1 and ϕ_2 and the corresponding limits η_1 and η_2 of the inverse scattering series, we show in theorem 3 that if $M = \max\{\|\phi_1\|_{H^{\mathfrak{b}}}, \|\phi_2\|_{H^{\mathfrak{b}}}\}$ and the conditions

$$\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}), \qquad M\|\mathcal{K}_1\|_{\mathfrak{ba}} < 1/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}) \tag{5.3}$$

hold, then

$$\|\eta_1 - \eta_2\|_{H^{\mathfrak{a}}} < \widetilde{C}\|\phi_1 - \phi_2\|_{H^{\mathfrak{b}} \times L^2},$$

where $\widetilde{C}$ is defined in equation (4.6). From the assumption in equation (5.1), we get

$$\widetilde{C} = \frac{C^* Q}{(1 - QM)^2}.$$

Note that M is independent of $\mathfrak{a}$, so $\widetilde{C}$ decreases as $\mathfrak{a}$ decreases. It indicates that we obtain better stability in the inverse scattering problem if we seek the parameter in a weaker (and thus bigger) function space.

However, as $\mathfrak{b}$ increases, C^* decreases while M increases. Furthermore, the rate at which M increases is unknown, as this depends on ϕ_1 and ϕ_2, causing the change in $\widetilde{C}$ to be unclear. However, as $\mathfrak{b}$ increases, it is less likely that the second condition in equation (5.3) holds, in which case the bound on $\widetilde{C}$ may not hold as the inverse scattering series could fail to converge. To this end, it is also beneficial to consider a weaker function space for the data, i.e. to use a weaker norm as the objective function in the corresponding PDE-constrained optimization problem to invert η computationally.

5.1.2.2. Approximation Error. Under the assumptions of theorem 4, we provide a bound on the approximation error of the inverse scattering series in equation (4.8). This bound decreases with the constant $C_{\mathfrak{a}\mathfrak{b}}$, which is defined in equation (4.7). Based on equation (5.2), we get

$$C_{\mathfrak{a}\mathfrak{b}} = \frac{\max\left\{1, Q\exp\left[(1-Q)^{-1}\right]\right\}}{(1-(\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}})\mathcal{M})^2}, \tag{5.4}$$

where $\mathcal{M} = \max\{\|\eta\|_{H^{\mathfrak{a}}}, \|\mathcal{K}_1 K_1 \eta\|_{H^{\mathfrak{a}}}\}$. Since Q is a constant independent of $\mathfrak{a}$ and $\mathfrak{b}$, $C_{\mathfrak{a}\mathfrak{b}}$ increases if and only if the constant $(\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}})\mathcal{M}$ increases. Next, observe that

$$\|\mathcal{K}_1 K_1 \eta\|_{H^{\mathfrak{a}}} \leqslant \|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}} \cdot \|K_1\|_{H^{\mathfrak{a}}\to H^{\mathfrak{b}}} \cdot \|\eta\|_{H^{\mathfrak{a}}} = \frac{Q\|K_1\|_{H^{\mathfrak{a}}\to H^{\mathfrak{b}}}\|\eta\|_{H^{\mathfrak{a}}}}{\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}}} \leqslant \frac{Q\nu_{\mathfrak{a}\mathfrak{b}}\|\eta\|_{H^{\mathfrak{a}}}}{\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}}},$$

in which we applied equation (5.1) and lemma 7. Since $Q < 1$, this gives $\|\mathcal{K}_1 K_1 \eta\|_{H^{\mathfrak{a}}} \leqslant \|\eta\|_{H^{\mathfrak{a}}}$, implying that $\mathcal{M} = \|\eta\|_{H^{\mathfrak{a}}}$. Thus, $(\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}})\mathcal{M}$ increases in $\mathfrak{b}$, since $\|\eta\|_{H^{\mathfrak{a}}}$ and $\mu_{\mathfrak{a}\mathfrak{b}}$ are constant in $\mathfrak{b}$ and $\nu_{\mathfrak{a}\mathfrak{b}}$ increases in $\mathfrak{b}$.

Next, note that $\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}} = (\mu_{\mathfrak{b}}+\nu_{\mathfrak{b}})P(\mathfrak{a},a,n)$ based on equation (3.10). Thus, as $\mathfrak{a}$ changes, $C_{\mathfrak{a}\mathfrak{b}}$ increases if and only if $P(\mathfrak{a},a,n)\|\eta\|_{H^{\mathfrak{a}}}$ increases. From lemma 2, we have

$$\frac{1}{\binom{n-1}{n-1}}\|\eta\|^2_{L^2(B_a)} \leqslant \frac{2a^2}{\binom{n}{n-1}}\sum_{|\alpha|=1}\|D^{\alpha}\eta\|^2_{L^2(B_a)} \leqslant \cdots \leqslant \frac{(2a^2)^{\mathfrak{a}+1}}{\binom{n+\mathfrak{a}}{n-1}}\sum_{|\alpha|=\mathfrak{a}+1}\|D^{\alpha}\eta\|^2_{L^2(B_a)}.$$

This implies

$$\frac{\sum_{|\alpha|\leqslant\mathfrak{a}}\|D^{\alpha}f\|^2_{L^2(\Omega)}}{\sum_{j=0}^{\mathfrak{a}}\frac{\binom{n+j-1}{n-1}}{(2a^2)^j}} \leqslant \frac{\sum_{|\alpha|\leqslant\mathfrak{a}+1}\|D^{\alpha}f\|^2_{L^2(\Omega)}}{\sum_{j=0}^{\mathfrak{a}+1}\frac{\binom{n+j-1}{n-1}}{(2a^2)^j}}.$$

Thus, $P(\mathfrak{a},a,n)\|\eta\|_{H^{\mathfrak{a}}}$ is increasing in $\mathfrak{a}$. To sum up, $C_{\mathfrak{a}\mathfrak{b}}$ increases in both $\mathfrak{a}$ and $\mathfrak{b}$.

5.2. Comparing classic and geometric approaches

Next, we compare the results from section 4.1 obtained from a classic approach with similar results from section 4.2 obtained using geometric function theory. Certainly, theorem 5 and theorem 6, obtained through geometric function theory, are improvements in the sense that the assumption on $\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}$ is removed. However, for this comparison, we use the same assumptions so that all theorems based on both approaches are valid.

5.2.1. Radius of convergence. We begin by analyzing the radius of convergence of the inverse scattering series. The classic approach, shown in theorem 2, yields a radius of convergence of $r = 1/(\mu_{\mathfrak{a}\mathfrak{b}}+\nu_{\mathfrak{a}\mathfrak{b}})$. On the other hand, the assumption equation (5.1) implies that $2 > \nu_{\mathfrak{a}\mathfrak{b}}\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}$, which means theorem 5 yields a radius of convergence of $r = (\sqrt{65}-8)/(2\mu_{\mathfrak{a}\mathfrak{b}})$. Comparing the two, we see that theorem 5 yields a larger radius of convergence whenever $\nu_{\mathfrak{a}\mathfrak{b}} > \mu_{\mathfrak{a}\mathfrak{b}}(2\sqrt{65}+15)$. This occurs when $\mathfrak{b}$ is sufficiently large, based on equations (3.8) and (3.9). Based on our numerical setup in section 5.3, the threshold varies for the particular PDE forward model and occurs when $\mathfrak{b} \approx 1$ for the scalar wave and $\mathfrak{b} \approx 4$ for the diffuse wave, as shown in figure 2(a). In this figure, we also plot the radius of convergence of the inverse scattering series under our assumptions given by both the classic and the geometric approaches for both diffuse and scalar waves as $\mathfrak{b}$ increases. The radius of convergence given by the geometric approach is constant in $\mathfrak{b}$. Although the $\mu_{\mathfrak{a}\mathfrak{b}}$ values are very different between the diffuse and scalar waves, their geometric radii of convergence are visually close due to the multiplication by the constant $(\sqrt{65}-8)/2 \approx 0.03$.

5.2.2. Approximation error.   Next, we compare theorems 4 and 6, which both analyze the approximation error of the inverse scattering series. As they are proven using different approaches, they lead to different bounds and conditions. For this analysis, we assume that the conditions of both theorems hold, and that

$$\mathcal{M}_1 = \mathcal{M} = \|\eta\|_{H^{\mathfrak{a}}}, \quad and \quad \mathcal{M} = Q_2/(\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}),$$

for a fixed $0 < Q_2 < 1$, similar to the assumption in equation (5.1). Under these assumptions, the conditions of theorem 4 are satisfied. On the other hand, the condition of theorem 6 can be expressed in terms of Q and Q_2 as:

$$Q_2 < \frac{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}}}{\mu_{\mathfrak{ab}}} \left(1 - \sqrt{\frac{\nu_{\mathfrak{ab}} Q}{\mu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}} + \nu_{\mathfrak{ab}} Q}} \right) = (1 + R) \left(1 - \sqrt{\frac{RQ}{1 + R + RQ}} \right), \qquad (5.5)$$

where $R = \nu_{\mathfrak{ab}}/\mu_{\mathfrak{ab}}$, constant in $\mathfrak{a}$ and decreasing as $\mathfrak{b}$ increases. Since we know $0 < Q < 1$ and $R > 0$, the last term in equation (5.5) has a global minimum in this region, and the minimum value is around 0.737. Furthermore, for sufficiently large R, Q_2 can be larger than 1, which is an improvement from theorem 4 which requires $0 < Q_2 < 1$. To sum up, theorem 4 requires $0 < Q, Q_2 < 1$ while theorem 6 does not have these strict assumptions. On the other hand, theorem 6 poses a condition on the relationship among R, Q and Q_2; see equation (5.5).

Next, we compare the bounds on the approximation error of the limit of the inverse scattering series. For theorem 4, this bound is controlled by the constant $C_{\mathfrak{ab}}$ given in equation (5.4), while the bound in theorem 6 is controlled by the constant

$$\widetilde{C}_{\mathfrak{ab}} = \left(1 - \frac{\nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}}}{(1 + \nu_{\mathfrak{ab}} \|\mathcal{K}_1\|_{\mathfrak{ba}} - \mu_{\mathfrak{ab}} \mathcal{M}_1)^2} \right)^{-1} = \left(1 - \frac{QR(1 + R)}{(1 + R + RQ - Q_2)^2} \right)^{-1},$$

under all the assumptions above. As Q approaches 1, the constant $C_{\mathfrak{ab}}$ rapidly blows up. However, there is no such issue with $\widetilde{C}_{\mathfrak{ab}}$, provided that the constraint equation (5.5) is met. Also, under equation (5.5), $\widetilde{C}_{\mathfrak{ab}}$ does not blow up even if Q_2 approaches 1. However, $C_{\mathfrak{ab}} \to +\infty$ again in this case. From this perspective, theorem 6 is an improvement over theorem 4. Note that the second error terms of the bound in theorem 4 and theorem 6 are controlled by the radius of convergence of the inverse scattering series. The larger the radius, the smaller the error obtained from truncating the series, and as discussed previously, whether an approach leads to an increased radius of convergence depends on $\mathfrak{b}$.

5.3. Numerical illustrations

In this section, we compare the radius of convergence for both the forward and inverse scattering series in $\mathbb{R}^3$.

5.3.1. Sobolev space on a sphere.   Given $s \in \mathbb{N}$, we define the spherical Sobolev space [5]

$$H^s(\mathcal{S}^{n-1}) = \left\{ f \in L^2(\mathcal{S}^{n-1}) : \|f\|_{H^s(\mathcal{S}^{n-1})}^2 = \|f\|_{L^2(\mathcal{S}^{n-1})}^2 + \|(-\Delta_{\mathcal{S}^{n-1}})^{s/2} f\|_{L^2(\mathcal{S}^{n-1})}^2 < \infty \right\},$$

where $\Delta_{\mathcal{S}^{n-1}}$ is the Laplace–Beltrami operator on the unit sphere in $\mathbb{R}^n$. We equip the space H^s with the Sobolev norm $\|\cdot\|_{H^s(\mathcal{S}^{n-1})}$. Similar to the definition of the $H^s(\mathbb{R}^n)$ norm through the Fourier transform [1], the spherical Sobolev spaces have an equivalent definition using the spherical harmonic transform, which allows us to extend the definition to the case where s is a real number [4]. More precisely, the space of spherical harmonics of degree ℓ on $\mathcal{S}^{n-1}$ has

an orthonormal basis $Y_{\ell k}$ for $1 \leqslant k \leqslant N(n,\ell)$, where

$$N(n,0) = 1, \qquad N(n,\ell) = \frac{(2\ell + n - 2)\Gamma(\ell + n - 2)}{\Gamma(\ell + 1)\Gamma(n - 1)}, \quad \ell \geqslant 1.$$

Every $f \in L^2(\mathcal{S}^{n-1})$ can be expressed in a spherical harmonic expansion of the form

$$f = \sum_{\ell=0}^{\infty} \sum_{k=1}^{N(n,\ell)} \hat{f}_{\ell k} Y_{\ell k}, \qquad \hat{f}_{\ell k} = \int_{\mathbb{S}^{n-1}} f \cdot \overline{Y_{\ell k}} \, \mathrm{d}\sigma.$$

The Sobolev space $H^s(\mathcal{S}^{n-1})$ with a real number s is defined by

$$H^s(\mathcal{S}^{n-1}) = \left\{ f \in \mathcal{D}'(\mathcal{S}^{n-1}) : \|f\|'_{H^s(\mathcal{S}^{n-1})} = \left(\sum_{\ell=0}^{\infty} \sum_{k=1}^{N(n,\ell)} (1 + \ell)^{2s} |\hat{f}_{\ell k}|^2 \right)^{\frac{1}{2}} < \infty \right\} \tag{5.6}$$

where $\mathcal{D}'(\mathcal{S}^{n-1})$ is the space of distributions on $\mathcal{S}^{n-1}$. We remark that the norms $\|\cdot\|_{H^s(\mathcal{S}^{n-1})}$ and $\|\cdot\|'_{H^s(\mathcal{S}^{n-1})}$ are norm-equivalent but not the same.

5.3.2. Forward scattering series.   We compute $\mu_{\mathfrak{ab}}$ in $\mathbb{R}^3$ for varying a (the radius of B_a) and $\mathfrak{a}$ (which controls the parameter space). First, recall that

$$\mu_{\mathfrak{ab}} = k^2 P(\mathfrak{a}, a, n) \sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(B_a)}$$

and note that $\mu_{\mathfrak{ab}}$ does not depend on $\mathfrak{b}$. In the case of the Helmholtz equation equation (2.2), we have $G(x,y) = \exp(ik|x-y|)/(4\pi|x-y|)$. Then, we have

$$\mu_{\mathfrak{ab}} = k^2 P(\mathfrak{a}, a, n) \sup_{y \in B_a} \left(\int_{B_a} \frac{1}{16\pi^2 |x-y|^2} \mathrm{d}x \right)^{\frac{1}{2}}$$

$$= \frac{k^2}{4\pi} P(\mathfrak{a}, a, n) \left(\int_{B_a} \frac{1}{|x|^2} \mathrm{d}x \right)^{\frac{1}{2}} = k^2 P(\mathfrak{a}, a, n) \left(\frac{a}{4\pi} \right)^{\frac{1}{2}}.$$

For the diffuse wave equation (2.6), we have $G(x,y) = \exp(-k|x-y|)/(4\pi|x-y|)$, which gives

$$\mu_{\mathfrak{ab}} = k^2 P(\mathfrak{a}, a, n) \sup_{y \in B_a} \left(\int_{B_a} \frac{\exp(-2k|x-y|)}{16\pi^2 |x-y|^2} \mathrm{d}x \right)^{\frac{1}{2}}$$

$$= \frac{k^2}{4\pi} P(\mathfrak{a}, a, n) \left(\int_{B_a} \frac{\exp(-2k|x|)}{|x|^2} \mathrm{d}x \right)^{\frac{1}{2}}$$

$$= k^2 \exp\left(-\frac{ka}{2} \right) P(\mathfrak{a}, a, n) \left(\frac{\sinh(ka)}{4\pi k} \right)^{\frac{1}{2}}.$$

Again, we can see that $\mu_{\mathfrak{ab}}$ does not depend on $H^{\mathfrak{b}}$ for both equations.

We compare the values of $1/\mu_{\mathfrak{ab}}$, the radius of convergence of the forward scattering series, for the Helmholtz equation and the diffuse wave in figure 1, where we set $k = 1$. First, we let a increase, with $\mathfrak{a} \in \{0, 1\}$ fixed. As shown in figure 1(a), while the radius of convergence for the diffuse wave equation is bounded below as a increases, the radius of convergence for the Helmholtz equation goes to 0 as a increases. Also, note that for both the Helmholtz and diffuse wave equations, $1/\mu_{\mathfrak{ab}}$ increases as $\mathfrak{a}$ increases, which aligns with our analysis in section 5.1.2. The amount of increase becomes more significant as a becomes smaller (i.e. the support of the
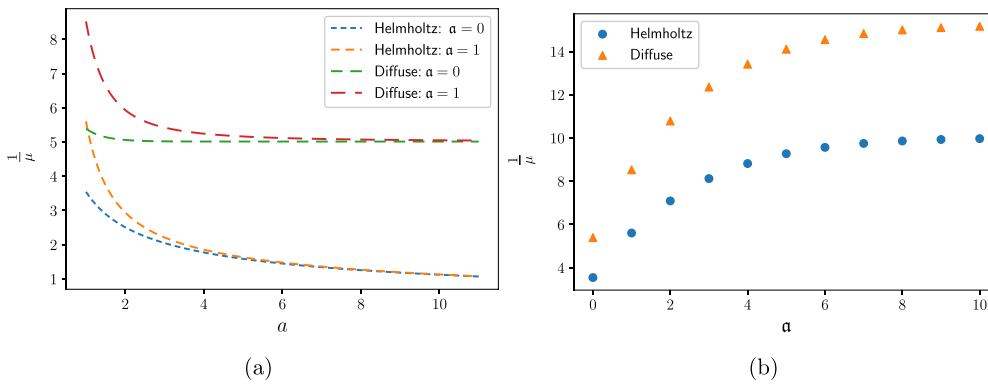
Figure 1. The radius of convergence of the forward scattering series (a) for Helmholtz and diffuse wave equations under the $H^{\mathfrak{a}} \to H^{\mathfrak{b}}$ norm as a (the radius of ball B_a) ranges from 1 to 11, and (b) as $\mathfrak{a}$ ranges from 0 to 10.

parameter perturbation η becomes smaller), as a result that $P(\mathfrak{a}, a, n)$ defined in equation (2.10) decreases in a.

Next, in figure 1(b), we plot the radius of convergence for $\mathfrak{a} \in \{0, 1, 2, \ldots, 10\}$ for both the Helmholtz and diffuse wave equations. We set $a = 1$ and $k = 1$. Since $P(\mathfrak{a}, a, n)$ increases in $\mathfrak{a}$, we see that $1/\mu_{\mathfrak{a}\mathfrak{b}}$ increases in $\mathfrak{a}$, which is also shown in figure 1(b). The value of $1/\mu_{\mathfrak{a}\mathfrak{b}}$ appears to converge for both the Helmholtz and diffuse wave equations as $\mathfrak{a}$ increases when $a = 1$. For $a \leqslant 1/\sqrt{2}$, the series in equation (2.10) does not converge, so for such a, the value of $1/\mu_{\mathfrak{a}\mathfrak{b}}$ would grow arbitrarily large as $\mathfrak{a}$ increases. The specific constant $1/\sqrt{2}$ is due to the coefficient of 2 in theorem 1, which is not sharp and can be improved with more precise estimates of the Poincaré constant.

5.3.3. Inverse scattering series. We compute $\mu_{\mathfrak{a}\mathfrak{b}}$ and $\nu_{\mathfrak{a}\mathfrak{b}}$ assuming that Ω is a ball of radius 100 centered at zero and B_a is centered at $(98, 0, 0)$ with radius $a = 1$ unless otherwise specified. We also assume that $u_i(x, x_1) = G(x, x_1)$ by setting the source to be a delta function. Since $u_i = G$, we get

$$\nu_{\mathfrak{a}\mathfrak{b}} = k^2 P(\mathfrak{a}, a, n) |B_a|^{\frac{1}{2}} \left(\sup_{y \in B_a} \|G(y, \cdot)\|_{L^2(\partial\Omega)} \right) \left(\sup_{y \in B_a} \|G(y, \cdot)\|_{H^{\mathfrak{b}}(\partial\Omega)} \right).$$

Since we are considering the case where Ω is a ball, we use the definition from equation (5.6) to compute $\|G(y, \cdot)\|_{L^2(\partial\Omega)}$ and $\|G(y, \cdot)\|_{H^{\mathfrak{b}}(\partial\Omega)}$ for any $y \in B_a$ and $\mathfrak{b} \in \mathbb{R}$. We use the software SHTools [42] to perform spherical transforms. The radius of convergence of the inverse scattering series $(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^{-1}$, for both the Helmholtz and diffuse wave equations, are plotted in figure 2. We set $a = 1$ and $k = 1$. In figure 2(a), we set $\mathfrak{a} = 0$ and compute $(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^{-1}$ as $\mathfrak{b}$ ranges from -2 to 6. Due to the factor $\|G(y, \cdot)\|_{H^{\mathfrak{b}}(\partial\Omega)}$, which grows exponentially based on equation (5.6), $\nu_{\mathfrak{a}\mathfrak{b}}$ increases exponentially in $\mathfrak{b}$. Thus, the radius of convergence for the inverse scattering series decreases exponentially in $\mathfrak{b}$ for both the Helmholtz and diffuse wave equations. Furthermore, since $(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^{-1}$ is bounded above by $1/\mu_{\mathfrak{a}\mathfrak{b}}$, and the latter is constant in $\mathfrak{b}$, it explains why the decay of $(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^{-1}$ is similar to that of a sigmoid function. In figure 2(a), we also plot the radius of convergence obtained using the geometric approach in theorem 5, which is constant in $\mathfrak{b}$ under the same assumptions of theorem 2.
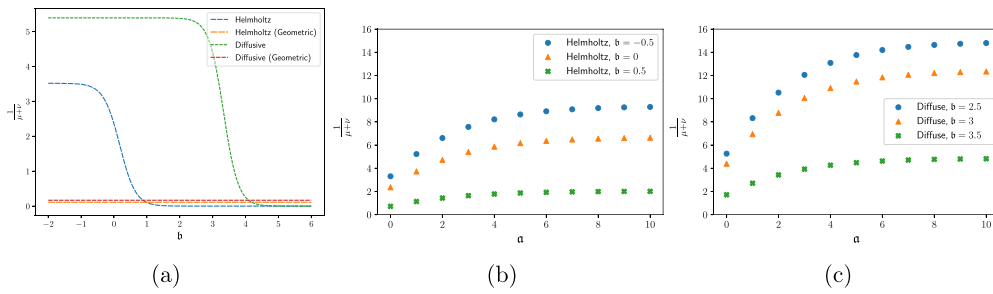
Figure 2. The radius of convergence of the inverse scattering series, (a) as $\mathfrak{b}$ ranges from -2 to 5, (b) for the Helmholtz equation as $\mathfrak{a}$ ranges from 0 to 10 and (c) for the diffuse wave equation as $\mathfrak{a}$ ranges from 0 to 10. The radius of convergence obtained using the geometric approach is also shown in (a), under the assumption that the conditions of both theorems 2 and 5 hold. We set $a = 1$ and $k = 1$ in all three plots.
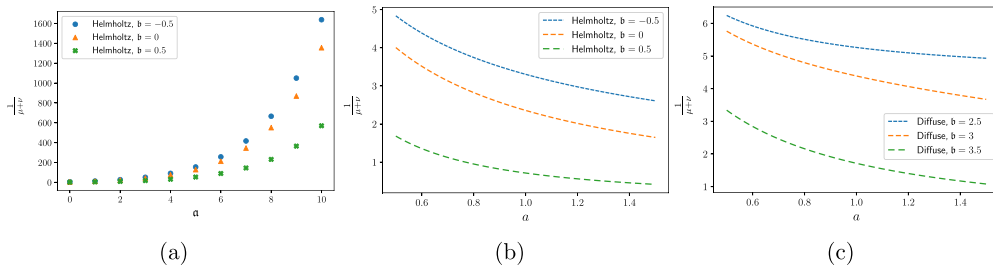
Figure 3. The radius of convergence of the inverse scattering series (a) as $\mathfrak{a}$ ranges from 0 to 10 when $a = 0.5$, (b) for the Helmholtz equation as a ranges from 0.5 to 1.5, and (c) for the diffuse wave equation as a ranges from 0.5 to 1.5. We set $k = 1$ in all plots.

Also, the radius of convergence for the inverse scattering series is much less sensitive to increases in $\mathfrak{b}$ for the diffuse wave equation than in the case of the Helmholtz equation. This is due to the exponential decay of the Green's function for the diffuse wave equation, as shown in equation (2.7). Since the radius of the domain is relatively large, the terms $\|G(y, \cdot)\|_{L^2(\partial\Omega)}$ and $\|G(y, \cdot)\|_{H^{\mathfrak{b}}(\partial\Omega)}$ are small due to this exponential decay. This causes $\nu_{\mathfrak{a}\mathfrak{b}}$ to be far smaller, only having a visible effect on the radius of convergence when $\mathfrak{b} > 2$. We do not see much difference in the radius obtained from the geometric approach due to the multiplication by a small constant; see section 5.2.2 for details.

In figures 2(b) and (c), we fix $\mathfrak{b}$ and compute the radius of convergence of the inverse scattering series as $\mathfrak{a}$ increases from 0 to 10. Note that using different choices of $\mathfrak{b}$ causes the radius of convergence for the Helmholtz equation to change significantly. On the other hand, for the diffuse wave equation, $\mathfrak{b}$ can be safely increased to 2.5 without significantly decreasing the radius of convergence. Due to the factor of $P(\mathfrak{a}, a, n)$ in both $\mu_{\mathfrak{a}\mathfrak{b}}$ and $\nu_{\mathfrak{a}\mathfrak{b}}$, the radius of convergence increases as $\mathfrak{a}$ increases, similar to figure 1(b). Similarly, the behavior of the radius of convergence as $\mathfrak{a}$ grows large depends on a. When $a \leqslant 1/\sqrt{2}$, $P(\mathfrak{a}, a, n)$ is unbounded as $\mathfrak{a}$ increases.

In figure 3(a), we demonstrate how the radius of convergence behaves as $\mathfrak{a}$ increases for smaller a (we set $a = 0.5$). Since $P(\mathfrak{a}, a, n)$ is unbounded as $\mathfrak{a}$ increases if a is sufficiently small, the radius of convergence of the inverse scattering series will grow arbitrarily large as $\mathfrak{a}$ increases. In figures 3(b) and (c), we plot the radius of convergence for both the Helmholtz and

diffuse wave equation as *a* ranges from 0.5 to 1.5. We set $\mathfrak{a} = 0$ since the only effect of $\mathfrak{a}$ on $\mu_{\mathfrak{ab}}$ and $\nu_{\mathfrak{ab}}$ is an extra multiplicative factor of $P(\mathfrak{a}, a, n)$. The radius of convergence decreases as *a* increases, and while changing $\mathfrak{b}$ has a significant effect on the Helmholtz equation, $\mathfrak{b}$ needs to be at least 2.5 to impact the radius of convergence in the case of the diffuse wave equation.

6. Numerical inversion examples

In this section, we present some numerical examples to show the impact of choosing different $H^{\mathfrak{a}}$ for the parameter space and $H^{\mathfrak{b}}$ for the data space in an optimization framework to solve the inverse scattering problems.

6.1. The gradient formulation

To begin with, we explain first how the chosen pair $(\mathfrak{a}, \mathfrak{b})$ changes the gradient calculation in the PDE-constrained optimization framework. Consider the optimization problem

$$\min_{\eta \in H^{\mathfrak{a}}} J_{\mathfrak{b}}(\eta) = \frac{1}{2} \|\mathcal{F}(\eta) - \phi\|_{H^{\mathfrak{b}}}^2, \tag{6.1}$$

where ϕ is the observed data, $\mathcal{F}$ is the forward operator that maps the parameter η to the data, which can be represented in the form of equations (3.1) and (3.2). We denote by $\mathcal{G}_{\mathfrak{ab}}(\eta)$ the gradient of the objective function $J_{\mathfrak{b}}(\eta)$.

By the Riesz representation theorem, $\forall \eta'$, we have

$$(\mathcal{G}_{\mathfrak{ab}}(\eta), \eta')_{H^{\mathfrak{a}}} = \lim_{\epsilon \to 0} \frac{J_{\mathfrak{b}}(\eta + \epsilon \eta') - J_{\mathfrak{b}}(\eta)}{\epsilon} = \int_{\Omega} \frac{\delta J_{\mathfrak{b}}}{\delta \eta} \eta' \mathrm{d}x = (\mathcal{G}_{0\mathfrak{b}}(\eta), \eta')_{L^2},$$

where $\mathcal{G}_{0\mathfrak{b}}(\eta) = \frac{\delta J_{\mathfrak{b}}}{\delta \eta}$ denotes the gradient assuming $\eta \in L^2$. We then have the relation

$$\mathcal{G}_{\mathfrak{ab}}(\eta) = (I - \Delta)^{-\mathfrak{a}} \mathcal{G}_{0\mathfrak{b}}(\eta).$$

As a result, for the same objective function, the gradient with the assumption that $\eta \in H^1$ (i.e. η has a higher regularity) is smoother than assuming that $\eta \in L^2$. Similarly, the L^2 gradient $\frac{\delta J_{\mathfrak{b}}}{\delta \eta}$ depends on the choice of $\mathfrak{b}$. Recall that K_1 is the linearization of the nonlinear forward operator $\mathcal{F}$. We then have

$$\frac{\delta J_{\mathfrak{b}}}{\delta \eta} = K_1^* (I - \Delta)^{\mathfrak{b}} (\mathcal{F}(\eta) - \phi),$$

where K_1^* denotes the adjoint operator of K_1. If we use $\mathcal{K}_1$ instead of K_1^*, it yields the Gauss–Newton method. Combining the above equations, we have

$$\mathcal{G}_{\mathfrak{ab}}(\eta) = (I - \Delta)^{-\mathfrak{a}} K_1^* (I - \Delta)^{\mathfrak{b}} (\mathcal{F}(\eta) - \phi). \tag{6.2}$$

We remark that the two Laplacian operators Δ in equation (6.2) above do not act on the same domain. The first one acts on the model parameter space, while the second one is defined over the data space. Nevertheless, it is evident from equation (6.2) that both $\mathfrak{a}$ and $\mathfrak{b}$ directly change the gradient in the PDE-constrained optimization.

Similarly, to calculate the linear action of $\mathcal{K}_1$, the pseudo-inverse of K_1 defined in equation (4.2), it is equivalent to another optimization problem

$$\min_{\eta \in H^{\mathfrak{a}}} \frac{1}{2} \|K_1 \eta - \phi\|_{H^{\mathfrak{b}}}^2. \tag{6.3}$$

Again, the solution changes with respect to the chosen $(\mathfrak{a}, \mathfrak{b})$, especially when the data is noisy. After discretization, the two norms are reflected in two weight matrices in a least-squares
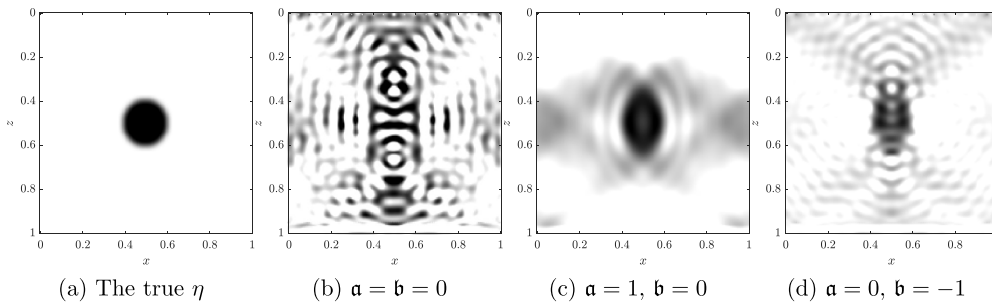
Figure 4. (Setting One) (a): The true scatterer η. (b)–(d): The reconstructed results after 100 iterations of L-BFGS algorithms under different $(\mathfrak{a}, \mathfrak{b})$ choices.
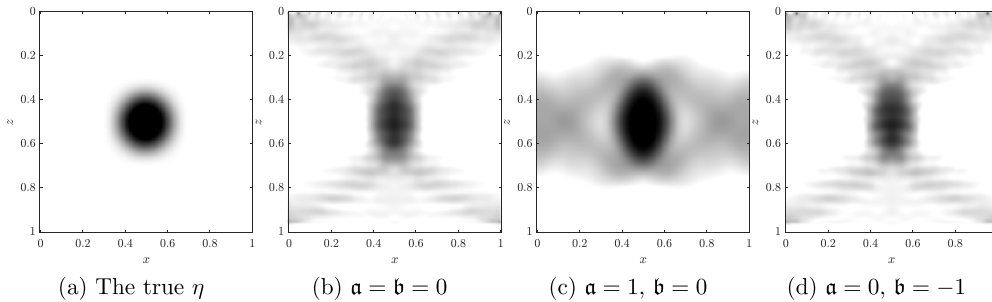
Figure 5. (Setting Two) (a): The true scatterer η. (b)–(d): The reconstructed results after 100 iterations of L-BFGS algorithms under different $(\mathfrak{a}, \mathfrak{b})$ choices.

problem. We refer to [15] where the impact of the weights in the resulting discrete problem is rigorously analyzed under different assumptions on K_1.

6.2. Inversion examples

Next, we present a few inversion examples based on the scalar wave equation equation (2.1), where the forward problem is not too smoothing, and the phenomena of using different norms are easy to observe.

We consider a square domain $\Omega = [0,1]^2$ discretized with the spatial spacing $dx = dz = 0.005$. The spatial unit is kilometer (km). We use point sources placed on the top of the domain at depth $z_s = 0.1$ and horizontal location $x_s \in \{0, 0.05, 0.1, \ldots, 1\}$. The measured data are recorded from the bottom of the domain at the depth $z_r = 0.95$ and horizontal location $x_r \in \{0, 2dx, 4dx, \ldots, 1\}$. We use a single frequency $\omega = 21$ Hz and the background velocity $c_0 = 2.5$ km s^{-1}, so we have $k = \omega/c_0 = 8.4$; see equation (2.2).

We consider two different true scatterer η, which are shown in figure 4(a) (Setting One) and figure 5(a) (Setting Two), respectively. Note that both scatterers are C^∞, sharing a similar circular structure, but the one in figure 5(a) is a smoothed version of the scatterer in figure 4(a) through a Gaussian filter. We choose these two examples to demonstrate the different radius of convergence for the inverse scattering series under different $(\mathfrak{a}, \mathfrak{b})$ pairs. The inversion results under Setting One are shown in figure 4 for three cases: $(\mathfrak{a} = 0, \mathfrak{b} = 0)$, $(\mathfrak{a} = 1, \mathfrak{b} = 0)$, and $(\mathfrak{a} = 0, \mathfrak{b} = -1)$. Similarly, inversion results under the same three cases for Setting Two are shown in the rest of figure 5.

(a) Noisy data   (b) $\mathfrak{a} = \mathfrak{b} = 0$ (c) $\mathfrak{a} = 1, \mathfrak{b} = 0$ (d) $\mathfrak{a} = 0, \mathfrak{b} = -1$
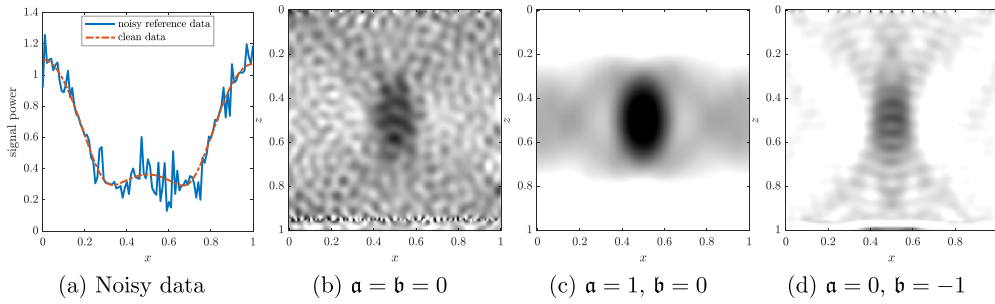
Figure 6. (Setting Two with noise) (a): Noisy data from one of the point sources. (b)–(d): The reconstructed results after 100 iterations of L-BFGS algorithms under different $(\mathfrak{a}, \mathfrak{b})$ choices while the reference data is polluted by white noise as shown in (a).

From figure 4(b), we observe that the less smooth scatterer shown in figure 4(a) is *outside* the radius of convergence by considering the inverse scattering as a map from L^2 to L^2. Figure 4(b) is full of wrong features and shares few similarities with the ground truth. In contrast, figure 4(a) is *within* the radius of convergence under the setups of $(\mathfrak{a} = 0, \mathfrak{b} = -1)$, and $(\mathfrak{a} = 1, \mathfrak{b} = 0)$ since figures 4(c) and (d) both recover the location and concentration of the scatterer, which are the main features of figure 4(a). We remark that this phenomenon matches our theoretical results in section 4 and our follow-up discussions in section 5 where we conclude that both increasing $\mathfrak{a}$ and decreasing $\mathfrak{b}$ will increase the radius of convergence for the inverse scattering series.

For comparison, we also perform inversion where the ground truth is the smoother scatterer shown in figure 5(a). When $\mathfrak{a} = \mathfrak{b} = 0$, the inverted result in figure 5(b) also recovers the primary features of the ground truth, such as the location, concentration, and contrast of the scatterer. This shows that the η shown in figure 5(a) is *within* the radius of convergence for the case $\mathfrak{a} = \mathfrak{b} = 0$ while the one in figure 4(a) is not. The inversion result in figure 5(c) is the closest to the ground truth after the same number of iterations since the choice $\mathfrak{a} = 1$ imposes an *a priori* assumption on the regularity of the scatterer, we aim to invert, and meanwhile, our ground truth is indeed a smooth function. Note that we do not have an explicit regularization term here, but the choice of $\mathfrak{b}$ plays a role in enforcing an *implicit* regularization for the inversion. On the other hand, the recovery in figure 5(d) using $\mathfrak{b} = -1$ is visually more oscillatory than the case $\mathfrak{b} = 0$ in figure 5(b) while both set $\mathfrak{a} = 0$. This is because the squared H^{-1} norm as the objective function has small weights on the high-wavenumber components of the data misfit and thus overlooked the differences, resulting in slow convergence on the high-wavenumber components of the parameter. We refer interested readers to [14, 15, 44] for a detailed discussion on the trade-offs of using the H^s norms. We also note that in practice, varying $\mathfrak{a}$ and $\mathfrak{b}$ can improve the reconstructed results as it can ensure that η is in the radius of convergence early on in optimization (large $\mathfrak{a}$ or small $\mathfrak{b}$) while later fall into regimes with better data sensitivity to improve resolution (small $\mathfrak{a}$ or large $\mathfrak{b}$).

Having illustrated the different radius of convergence, we show an example regarding the stability with respect to data noises. The setup remains the same as Setting Two except that the reference data is now polluted with white noise; see figure 6(a) for an illustration. The inversion results under the same three cases are shown in the rest of figure 6. We do not have a regularization term in the objective function. The $L^2 \mapsto L^2$ setting in figure 4(b) overfit

the noise (see figure 6(b)), while the *a priori* assumption that $\eta \in H^1$ prevented the reconstruction in the case $(\mathfrak{a} = 1, \mathfrak{b} = 0)$ from being polluted by the noisy data (see figure 6(c)). Finally, the inversion result by embedding the noisy data into the weaker H^{-1} space also helped mitigate the high-frequency noise due to the natural smoothing property of the weak norm (see figure 6(d)).

7. Conclusion

In this work, we generalize the results in [29] and consider a different class of function spaces to analyze the inverse scattering series for Helmholtz and diffuse wave equations. Specifically, we analyze the convergence property of the forward scattering series that maps the parameter in $H^{\mathfrak{a}}$ to the scattering data in $H^{\mathfrak{b}}$ and the convergence, stability, and approximation error of the inverse scattering series. Since we express all the results and constants explicitly in terms of $\mathfrak{a}$ and $\mathfrak{b}$, we could analyze how the choices of $H^{\mathfrak{a}}$ and $H^{\mathfrak{b}}$ for the parameter and the scattering data, respectively, impact certain properties of the inverse scattering problem. For example, we can increase the radius of convergence for the inverse scattering series by using a stronger function space $H^{\mathfrak{a}}$ for the parameter while using a weaker metric space $H^{\mathfrak{b}}$ for the scattering data, which corresponds to better convexity when using weaker H^s norm as the objective function in practice [44]. Also, we can obtain a smaller stability constant by considering a weaker space for the parameter $H^{\mathfrak{a}}$, but it is at the cost of obtaining a reconstruction with less regularity.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

Appendix. Proofs of lemmas and theorems in section 4

Here, we provide the proof of lemma 8 and theorem 4. The proofs are based on results and proof techniques in [29, section 3]. We include the proofs simply for completeness, as we explicitly state the values of all constants in section 4.

Proof of lemma 8

We provide a bound on the quantity $\|\mathcal{K}_j\|_{\mathfrak{b}\mathfrak{a}}$ in lemma 8.

Proof of lemma 8. Consider $j \geqslant 2$. From equation (4.2), we first have the bound

$$
\|\mathcal{K}_j\|_{\mathfrak{b}\mathfrak{a}} \leqslant \left(\sum_{m=1}^{j-1} \|\mathcal{K}_m\|_{\mathfrak{b}\mathfrak{a}} \sum_{i_1+\ldots+i_m=j} \|K_{i_1}\|_{H^\mathfrak{a} \to H^\mathfrak{b}} \ldots \|K_{i_m}\|_{H^\mathfrak{a} \to H^\mathfrak{b}} \right) \|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}^j
$$

$$
\leqslant \|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}^j \left(\sum_{m=1}^{j-1} \|\mathcal{K}_m\|_{\mathfrak{b}\mathfrak{a}} \sum_{i_1+\ldots+i_m=j} \nu_{\mathfrak{a}\mathfrak{b}} \mu_{\mathfrak{a}\mathfrak{b}}^{i_1-1} \ldots \nu_{\mathfrak{a}\mathfrak{b}} \mu_{\mathfrak{a}\mathfrak{b}}^{i_m-1} \right), \qquad \text{(A.1)}
$$

where we have used lemma 7 to obtain the second inequality. Using equation (A.1), the rest of the proof follows similarly as in [31, lemma 2.1]. $\qquad \square$

Proof of theorem 4

While the inverse scattering series converges under the conditions of theorem 2, its limit is not equal to η in general. In theorem 4, we give a bound on the distance between this limit and η.

Proof of theorem 4. The first part of the proof uses the following derivations from [31, theorem 2.1]. Since $\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}} < 1/(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})$ and $\|\mathcal{K}_1 \phi\|_{H^\mathfrak{a} \times L^2} < 1/(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})$, the inverse scattering series

$$
\widetilde{\eta} = \sum_j \mathcal{K}_j \phi \otimes \ldots \otimes \phi \qquad \text{(A.2)}
$$

converges. Also, since $\mathcal{M} < 1/(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})$, the forward scattering series

$$
\phi = \sum_j K_j \eta \otimes \ldots \otimes \eta
$$

converges as well. Substituting the forward scattering series for ϕ into equation (A.2) gives

$$
\widetilde{\eta} = \sum_j \widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta
$$

where $\widetilde{\mathcal{K}}_1 = \mathcal{K}_1 K_1$ and for $j \geqslant 2$,

$$
\widetilde{\mathcal{K}}_j = \left(\sum_{m=1}^{j-1} \mathcal{K}_m \sum_{i_1+\ldots+i_m=j} K_{i_1} \otimes \ldots \otimes K_{i_m} \right) + \mathcal{K}_j K_1 \otimes \ldots \otimes K_1,
$$

as defined in [29, equation (67)]. Using the formula for $\mathcal{K}_j$ in equation (4.2), we get

$$
\widetilde{\mathcal{K}}_j = \left(\sum_{m=1}^{j-1} \mathcal{K}_m \sum_{i_1+\ldots+i_m=j} K_{i_1} \otimes \ldots \otimes K_{i_m} \right) (I - \mathcal{K}_1 K_1 \otimes \ldots \otimes \mathcal{K}_1 K_1).
$$

Since $\eta - \widetilde{\eta} = \eta - \mathcal{K}_1 K_1 \eta - \widetilde{\mathcal{K}}_2 \eta \otimes \eta + \ldots$, we get

$$
\|\eta - \widetilde{\eta}\|_{H^\mathfrak{a}} \leqslant \|\eta - \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}} + \sum_{j=2}^\infty \|\widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta\|_{H^\mathfrak{a}}
$$

using the triangle inequality. Next, consider $\|\widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta\|_{H^\mathfrak{a}}$, for $j \geqslant 2$. We first get

$$\|\widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta\|_{H^\mathfrak{a}} \leqslant \sum_{m=1}^{j-1} \|\mathcal{K}_m\|_{\mathfrak{b}\mathfrak{a}} \sum_{i_1 + \ldots + i_m = j} \|K_{i_1}\|_{H^\mathfrak{a} \to H^\mathfrak{b}} \ldots \|K_{i_m}\|_{H^\mathfrak{a} \to H^\mathfrak{b}}$$
$$\cdot \|\eta \otimes \ldots \otimes \eta - \mathcal{K}_1 K_1 \eta \otimes \ldots \otimes \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}}.$$

Letting $\psi = \eta - \mathcal{K}_1 K_1 \eta$ and using [31, equation (29)] gives

$$\|\eta \otimes \ldots \otimes \eta - \mathcal{K}_1 K_1 \eta \otimes \ldots \otimes \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}} \leqslant j \mathcal{M}^{j-1} \|\psi\|_{H^\mathfrak{a}}.$$

Our proof is distinct from the literature after this point. From the proof of lemma 8,

$$\sum_{m=1}^{j-1} \|\mathcal{K}_m\|_{\mathfrak{b}\mathfrak{a}} \sum_{i_1 + \ldots + i_m = j} \|K_{i_1}\|_{H^\mathfrak{a} \to H^\mathfrak{b}} \ldots \|K_{i_m}\|_{H^\mathfrak{a} \to H^\mathfrak{b}} \leqslant C(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^j$$

since the left-hand side of this equation is the same as the right-hand side of equation (A.1), except with the factor of $\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}^j$ removed. This means, for $j \geqslant 2$, we have

$$\|\widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta\|_{H^\mathfrak{a}} \leqslant C\|\psi\|_{H^\mathfrak{a}} j \mathcal{M}^{j-1} (\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^j,$$

which gives

$$\|\eta - \widetilde{\eta}\|_{H^\mathfrak{a}} \leqslant \|\psi\|_{H^\mathfrak{a}} + C\|\psi\|_{H^\mathfrak{a}} \sum_{j=2}^{\infty} j \mathcal{M}^{j-1} (\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^j$$
$$\leqslant C^* \|\psi\|_{H^\mathfrak{a}} \sum_{j=1}^{\infty} j \mathcal{M}^{j-1} (\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})^j, \tag{A.3}$$

and this series converges since $(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})\mathcal{M} < 1$. Computing this sum gives

$$\|\eta - \widetilde{\eta}\|_{H^\mathfrak{a}} \leqslant \frac{C^*(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})\|\eta - \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}}}{(1 - (\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})\mathcal{M})^2} = C_{\mathfrak{a}\mathfrak{b}} \|\eta - \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}}. \tag{A.4}$$

We can split the approximation error into two terms with the triangle inequality:

$$\left\| \eta - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \ldots \otimes \phi \right\|_{H^\mathfrak{a}} \leqslant \left\| \eta - \sum_{j} \widetilde{\mathcal{K}}_j \eta \otimes \ldots \otimes \eta \right\|_{H^\mathfrak{a}} + \left\| \widetilde{\eta} - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \ldots \otimes \phi \right\|_{H^\mathfrak{a}}.$$

Using theorem 2, we finally obtain

$$\left\| \eta - \sum_{j=1}^{N} \mathcal{K}_j \phi \otimes \ldots \otimes \phi \right\|_{H^\mathfrak{a}} \leqslant C_{\mathfrak{a}\mathfrak{b}} \|\eta - \mathcal{K}_1 K_1 \eta\|_{H^\mathfrak{a}} + C \frac{[(\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}\|\phi\|_{H^\mathfrak{b}}]^{N+1}}{1 - (\mu_{\mathfrak{a}\mathfrak{b}} + \nu_{\mathfrak{a}\mathfrak{b}})\|\mathcal{K}_1\|_{\mathfrak{b}\mathfrak{a}}\|\phi\|_{H^\mathfrak{b}}}$$

where C is defined in equation (4.3). $\qquad \square$

ORCID iD

Yunan Yang ⓘD https://orcid.org/0000-0001-7238-7978

References

[1] Arbogast T and Bona J L 1999 *Methods of Applied Mathematics* vol 2008 (Austin, TX: Department of Mathematics, University of Texas)

[2] Arridge S, Moskow S and Schotland J C 2012 Inverse Born series for the Calderon problem *Inverse Problems* **28** 035003

[3] Bao G, Li P, Lin J and Triki F 2015 Inverse scattering problems with multi-frequencies *Inverse Problems* **31** 093001

[4] Barceló J A, Folch-Gabayet M, Luque T, Pérez-Esteva S and Vilela M C 2021 The Fourier extension operator of distributions in Sobolev spaces of the sphere and the Helmholtz equation *Proc. R. Soc. Edinburgh* A **151** 1768–89

[5] Barceló J A, Luque T and Pérez-Esteva S 2020 Characterization of Sobolev spaces on the sphere *J. Math. Anal. Appl.* **491** 124240

[6] Bardsley P and Vasquez F G 2014 Restarted inverse Born series for the Schrödinger problem with discrete internal measurements *Inverse Problems* **30** 045014

[7] Barnett A, Greengard L, Pataki A and Spivak M 2017 Rapid solution of the Cryo-EM reconstruction problem by frequency marching *SIAM J. Imaging Sci.* **10** 1170–95

[8] Brossier R, Operto Sephane and Virieux J 2010 Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics* **75** R37–46

[9] Bunks C, Saleck F M, Zaleski S and Chavent G 1995 Multiscale seismic waveform inversion *Geophysics* **60** 1457–73

[10] Calder J, Mansouri A and Yezzi A 2010 Image sharpening via Sobolev gradient flows *SIAM J. Imaging Sci.* **3** 981–1014

[11] Chung F J, Gilbert A C, Hoskins J G and Schotland J C 2017 Optical tomography on graphs *Inverse Problems* **33** 055016

[12] Colton D L and Kress R 1998 *Inverse Acoustic and Electromagnetic Scattering Theory* vol 93 (Berlin: Springer)

[13] Dunlop M M and Yang Y 2021 Stability of Gibbs posteriors from the Wasserstein loss for Bayesian full waveform inversion *SIAM/ASA J. Uncertain. Quantification* **9** 1499–526

[14] Engquist Born, Ren K and Yang Y 2020 The quadratic Wasserstein metric for inverse data matching *Inverse Problems* **36** 055001

[15] Engquist Born, Ren K and Yang Y 2022 A generalized weighted optimization method for computational learning and inversion *Int. Conf. on Learning Representations*

[16] Federer H 2014 *Geometric Measure Theory* (Berlin: Springer)

[17] Fichtner A, Trampert J, Cupillard P, Saygin E, Taymaz T, Capdeville Y and Villaseñor A 2013 Multiscale full waveform inversion *Geophys. J. Int.* **194** 534–56

[18] Hoskins J G and Schotland J C 2022 Analysis of the inverse Born series: an approach through geometric function theory *Inverse Problems* **38** 074001

[19] Jacobs M, Léger F, Li W and Osher S 2019 Solving large-scale optimization problems with a convergence rate independent of grid size *SIAM J. Numer. Anal.* **57** 1100–23

[20] Jeong G, Hwang J and Min D-J 2017 Comparison of weighting techniques for acoustic full waveform inversion *J. Appl. Geophys.* **147** 16–27

[21] Kamei R, Pratt R G and Tsuji T 2014 Misfit functionals in Laplace–Fourier domain waveform inversion, with application to wide-angle ocean bottom seismograph data *Geophys. Prospect.* **62** 1054–74

[22] Kilgore K, Moskow S and Schotland J C 2012 Inverse Born series for scalar waves *J. Comput. Math.* **30** 601–14

[23] Kilgore K, Moskow S and Schotland J C 2017 Convergence of the Born and inverse Born series for electromagnetic scattering *Appl. Anal.* **96** 1737–48

[24] Leoni G 2017 *A First Course in Sobolev Spaces* (Providence, RI: American Mathematical Society)

[25] Machida M and Schotland J C 2015 Inverse Born series for the radiative transport equation *Inverse Problems* **31** 095009

[26] Mandelis A 2000 Diffusion waves and their uses *Phys. Today* **53** 29–34

[27] Mandelis A 2013 *Diffusion-Wave Fields: Mathematical Methods and Green Functions* (Berlin: Springer)

[28] Markel V A, O'Sullivan J A and Schotland J C 2003 Inverse problem in optical diffusion tomography. IV. Nonlinear inversion formulas *J. Opt. Soc. Am.* A **20** 903–12

[29] Moskow S and Schotland J C 2008 Convergence and stability of the inverse scattering series for diffuse waves *Inverse Problems* **24** 065005
[30] Moskow S and Schotland J C 2009 Numerical studies of the inverse Born series for diffuse waves *Inverse Problems* **25** 095007
[31] Moskow S and Schotland J C 2019 12 Inverse Born series *The Radon Transform, The First 100 Years and Beyond* (*Radon Series on Computational and Applied Mathematics vol 22*) (Berlin: De Gruyter) pp 273–95
[32] Necas J 2011 *Direct Methods in the Theory of Elliptic Equations* (Berlin: Springer)
[33] Neuberger J 2009 *Sobolev Gradients and Differential Equations* (Berlin: Springer)
[34] Nurbekyan L, Lei W and Yang Y 2022 Efficient natural gradient descent methods for large-scale optimization problems (arXiv:2202.06236)
[35] Oh J-W and Min D-J 2013 Weighting technique using backpropagated wavefields incited by deconvolved residuals for frequency-domain elastic full waveform inversion *Geophys. J. Int.* **194** 322–47
[36] Otto F and Villani Cedric 2000 Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality *J. Funct. Anal.* **173** 361–400
[37] Papadakis N, Peyré G and Oudet E 2014 Optimal transport with proximal splitting *SIAM J. Imaging Sci.* **7** 212–38
[38] Peyre Remi 2018 Comparison between W_2 distance and $\dot{H}^{-1}$ norm and localization of Wasserstein distance *ESAIM: Control, Optim. Calculus Variations* **24** 1489–501
[39] Rektorys K 2012 *Variational Methods in Mathematics, Science and Engineering* (Berlin: Springer)
[40] Ryan R A 2002 *Introduction to Tensor Products of Banach Spaces* vol 73 (Berlin: Springer)
[41] Sundaramoorthi G, Yezzi A and Mennucci A C 2007 Sobolev active contours *Int. J. Comput. Vis.* **73** 345–66
[42] Wieczorek M A and Meschede M 2018 Shtools: tools for working with spherical harmonics *Geochem. Geophys. Geosyst.* **19** 2574–92
[43] Yang Y, Engquist Born, Sun J and Hamfeldt B F 2018 Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion *Geophysics* **83** R43–R62
[44] Zhu B, Jingwei H, Lou Y and Yang Y 2021 Implicit regularization effects of the Sobolev norms in image processing (arXiv:2109.06255)