DIFFUSIVE OPTICAL TOMOGRAPHY IN THE BAYESIAN FRAMEWORK[*]

KIT NEWTON[†], QIN LI[‡], AND ANDREW M. STUART[§]

Abstract. Many naturally-occuring models in the sciences are well-approximated by simplified models, using multiscale techniques. In such settings it is natural to ask about the relationship between inverse problems defined by the original problem and by the multiscale approximation. We develop an approach to this problem and exemplify it in the context of optical tomographic imaging.

Optical tomographic imaging is a technique for infering the properties of biological tissue via measurements of the incoming and outgoing light intensity; it may be used as a medical imaging methodology. Mathematically, light propagation is modeled by the radiative transfer equation (RTE), and optical tomography amounts to reconstructing the scattering and the absorption coefficients in the RTE from boundary measurements. We study this problem in the Bayesian framework, focussing on the strong scattering regime. In this regime the forward RTE is close to the diffusion equation (DE). We study the RTE in the asymptotic regime where the forward problem approaches the DE, and prove convergence of the inverse RTE to the inverse DE in both nonlinear and linear settings. Convergence is proved by studying the distance between the two posterior distributions using the Hellinger metric, and using Kullback-Leibler divergence.

1. Introduction.

1.1. Background. Optical imaging is one of many medical imaging techniques that uses light to probe tissue structure [29, 3]. Near infrared light is sent into biological tissue, and the outgoing photon current at the surface of the tissue is then measured. Using these measurements, it is possible to infer properties of the tissue. While traditional imaging methods such as X-ray imaging provide good reconstructions of the tissue's properties, they are typically more expensive and more invasive than optical imaging. Optical imaging can be used for brain, breast, and joint imaging, as well as monitoring blood oxygenation [19, 20, 27].

To study optical imaging mathematically, one may use the radiative transfer equation (RTE). The forward RTE describes the dynamics of photons in materials with given optical properties. We denote the distribution of particles at location x with velocity v by $f(x,v)$, where $x \in \Omega \subset \mathbb{R}^d$, $d = 2, 3$, and $v \in \mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$. The model enforces particle motion with constant unit speed, and the velocity affects the problem only through the direction of travel of the particle. The optical properties are characterized by two parameters, the scattering coefficient and the absorption coefficient. The scattering coefficient, denoted $k(x, v, v')$, determined by the probability of a photon, currently moving in direction v at position x, scattering off a particle in the material and changing direction to v'. The total absorption coefficient, denoted by $\alpha(x, v)$, accounts for photon absorbtion into the material where they are lost. With this notation established, the RTE is, in its most general form,

$$(1.1) \qquad v \cdot \nabla f(x,v) = \int_{\mathbb{S}^{d-1}} k(x, v, v') f(x, v') \mathrm{d}v' - \alpha(x, v) f(x, v).$$

Here the gradient operation is with respect to x, as are related contractions of the gradient to a divergence. The forward problem is to determine the particle distribution function f, given the optical properties of the medium as characterized by k and α. Optical imaging amounts to solving a related inverse problem: the map from incoming data (light intensity injected into the tissue) to the measured outgoing data (light intensity collected outside the tissue) is termed the albedo operator, and the absorption and scattering coefficients in RTE are reconstructed from knowledge of the albedo operator. There are a number of theoretical results concerning the inverse RTE, primarily focussed on the setting in which the entire albedo map is known: it was shown in [11] that the medium is uniquely recoverable in dimension $d = 3$, and then in [33] that the reconstruction is Lipschitz stable. For further literature surveys see the reviews in [4, 5].

Another model for photon dynamics is the diffusion equation (DE). The diffusion equation typically governs lower energy photons than the RTE, leading to a larger scattering effect and less absorption. Let $\rho(x)$ denote the light intensity at location x where, as before, $x \in \Omega \subset \mathbb{R}^d$, and let $a(x)$ denote the diffusion

†Department of Mathematics, University of Wisconsin-Madison (kcnewton@math.wisc.edu, http://kitcnewton.github.io).

‡Department of Mathematics, University of Wisconsin-Madison (qinli@math.wisc.edu).

§Computing + Mathematical Sciences, California Institute of Technology (astuart@caltech.edu).

coefficient. Then the DE is

$$-\nabla \cdot (a(x)\nabla\rho(x)) = 0.$$

In this setup, the map from the Dirichlet data (light intensity injected into the tissue) to the Neumann data (light propagating out) is termed the Dirichlet-to-Neumann (DtN) map, and is used to reconstruct the diffusion coefficient $a(x)$. Using the DtN map to reconstruct the medium for the elliptic equation is the famous Calderón problem, widely studied from a theoretical perspective. Two foundational papers are [32], where the uniqueness was shown, and [1] in which logarithmic stability of the inversion was demonstrated. The review [8] contains further citations to literature in this area.

It is natural to examine the relation between the two forward models, and to understand, from both a physical as well as a mathematial perspective, why they give distinct stability performances in the inverse problem. When $a(x)$ and $(k(x, v, v'), \alpha(x, v))$ satisfy certain relations, the two forward models are asymptotically "close" when the laser beam is composed of low-energy photons. In the forward setting, physically, high-energy photons experience little scattering before leaving the domain, while in comparison, low-energy photons are scattered frequently by the tissue before being released and measured at the boundary. As a consequence, high-energy photons present a crisp resolution, and the images from low-energy photons are rather blurred. Mathematically, the RTE is taken as the correct forward model, and we can use the Knudsen number to present the number of times that an average photon scatters. In the low-energy regime, the number of times the photon scatters increases, the Knudsen number shrinks to zero, and the RTE converges to the DE in the forward setting. Correspondingly, the inverse RTE is expected to converge to the inverse DE: meaning the information carried in the albedo operator is almost the same as that in the DtN map, and the reconstruction should also converge. This has been numerically observed in [2, 17, 10], and proved rigorously in [21].

The literature refered to thus far focusses on settings in which the entire albedo or DtN map is known, and this leads to a deep mathematical theory. However it is arguably far from the practical setting in which partial and noisy information about these mappings is all that is available. The Bayesian formulation of the inverse problem is useful in this setting as it allows for incorporation of prior information, partial obeservation and noise level in a natural fashion. The practicality of this approach was demonstrated in the monograph [18], and recent work has led to a mathematical framework [22, 23, 31] suitable for well-posedness analyses [15] and computations which blend state-of-the-art computational PDE and computatioal statistical approaches [9, 13]. In the Bayesian approach to the solution of the inverse problem, all quantities are viewed as random variables, and the solution is the probability distribution of the unknown quantity conditioned on the data [15]. Bayes' theorem allows determination of this conditional distribution (the posterior) from the prior distribution on the unknown and from the likelihood, the distribution of the data conditioned on fixing the unknown. Our work is focussed on understanding the relationship between the two inverse problems in this Bayesian setting.

1.2. Our Contribution. The goal of this paper is to connect the two inverse problems in optical imaging, and specifically to prove convergence of the inverse RTE to the inverse DE in the Bayesian framework. Multiscale techniques provide the desired estimates on the forward problem and we show how these may be transfered to the Bayesian inverse problems. To this end we make the following contributions:

- we provide multiscale-based error estimates which relate the solution of the forward problems for the RTE and DE – see Theorem 2.2 for which we provide a formal asymptotic justification in the main body of the paper, and a rigorous proof in the appendix;
- we compare the two posterior distributions for the RTE and DE, measuring distance between them in the Kullback-Leibler divergence and the Hellinger distance in the optically thick regime (zero limit of the Knudsen number) when the scattering coefficient is large – see Theorems 3.2 and 3.3;
- we extend the convergence result linking posterior distributions to a setting in which the albedo operator's dependence on the medium is approximated by linearization – see Theorem 4.3 and Corollary 4.2.

The approach we adopt will apply to other Bayesian inverse problems whose forward models are linked through multiscale analyses. The paper is organized as follows. In section 2 we provide the mathematical setting for our work, including discussion of the Bayesian formulation of inverse problems, and the diffusion limit of the radiative transfer equation. In section 3 we estimate the distance between the Bayesian solution

of the RTE and DE inverse problems, and in section 4 we address the same question in the linearized setting. We conclude in section 5.

2. The Setting. In this section we establish the mathematical framework within which all our results are derived. In subsection 2.1 we describe Bayesian inverse problems in general. We then discuss the setting of linear inverse problems with Gaussian priors and Gaussian additive noise, in which the posterior is also Gaussian; and we discuss linearization of the forward operator to obtain an approximate Gaussian posterior. Subsection 2.2 describes the forward problems from the RTE and for the DE, providing error estimates linking their solutions in the small Knudsen number regime. In subsection 2.3 we formulate the Bayesian inverse problem for the RTE and DE. Subsection 2.4 is denoted to linearization of the forward operator for RTE and DE, and hence forms the basis for defining an approximate Gaussian posterior distributions for their respective inverse problems.

2.1. Bayesian Inversion. Consider the inverse problem of finding σ from y where

$$(2.1) \qquad y = \mathcal{G}(\sigma) + \eta,$$

$\mathcal{G}$ is a known forward map that takes parameter to the data space, and η a noise pollution. In the Bayesian formulation of inversion y, σ and η are viewed as random variables, linked by equation (2.1), and assumption that the prior distribution on σ, denoted by μ_0, and the distribution of the noise η, denoted by μ_{error}, are known. The objective is to find the conditional distribution on σ given y (denoted $\sigma|y$.) In this paper we will assume that η is independent of σ *a priori* and denote the distribution of y given σ, which is then a translation by $\mathcal{G}(\sigma)$ of $\mu_{\text{error}}(\eta)$, by $\mu^\sigma(y)$. We will concentrate on the commonly occuring case in which η is in a function space and the data y is finite dimensional; then $\mu_{\text{error}}(\eta)$ may be identified with its Lebesgue density, whilst μ_0 and μ^y are measures on a separable Banach space.

If we denote by μ^y the posterior distribution on σ given observation y then Bayes' theorem gives

$$(2.2) \qquad \mu^y(d\sigma) = \tfrac{1}{Z}\mu^\sigma(y)\mu_0(d\sigma),$$

where

$$Z = \int_X \mu^\sigma(y)\mathrm{d}\mu_0(d\sigma)$$

and X is a subset of a separable Banach space which contains the support of μ_0; then Z normalizes μ^y to a probability density. If we make the additional assumption that both μ_0 and $\mu_{\text{error}}(\eta)$ are Gaussian and finite dimensional so that $\mu_0 = \mathcal{N}(m_0, \mathcal{C}_{\text{prior}})$ and $\mu_{\text{error}}(\eta) = N(0, \mathcal{C}_{\text{prior}})$ then we may write a formula for the Lebesgue density of the posterior:

$$(2.3) \qquad \mu^y(\sigma) = \tfrac{1}{Z}\exp\left(-(\sigma - m_0)^\top \mathcal{C}_{\text{prior}}^{-1}(\sigma - m_0) - (y - \mathcal{G}(\sigma))^\top \mathcal{C}_{\text{error}}^{-1}(y - \mathcal{G}(\sigma))\right).$$

We note that analogous formulae are also available in the infinite dimensional case; see [31, 15] and the references therein.

In optical tomography, one has two fundamental models for describing light propagation: the radiative transfer equation (RTE), and the diffusion equation (DE). We will denote the solution of the respective inverse problems by $\mu_{\text{DE}}^y(\sigma)$ and $\mu_{\text{RTE}}^y(\sigma)$. This paper is primarily concerned with showing that these two distributions are close in the small Knudsen number regime, and quantifying the difference. There are multiple ways to quantify the distance between two probability distributions μ and μ'. We will use the Kullback-Leibler (KL) divergence and the Hellinger distance. If μ has density with respect to μ' and μ has support in X defined as above, then the KL divergence is given by

$$\mathrm{d}_{\text{KL}}(\mu, \mu') = \int_X \log \frac{d\mu}{d\mu'}(\sigma)\mathrm{d}\mu(d\sigma);$$

if μ and μ' have density with respect to common reference measure λ, with support in X defined as above, then the Hellinger distance is given by

$$d_{\text{Hell}}(\mu, \mu')^2 = \frac{1}{2}\int_X \left(\sqrt{\frac{d\mu}{d\lambda}} - \sqrt{\frac{d\mu'}{d\lambda}}\right)^2 \mathrm{d}\lambda.$$

These formulae have interpretations in the infinite dimensional setting; see the appendix of [15]. The KL divergence has an information theoretic interpretation which makes it attractive. However, the Hellinger metric is particularly useful because, for square integrable test functions, it translates directly into bounds of differences of expectations of test functions, see Lemma 7.14 in [15]. The square root of the KL divergence bounds the Hellinger metric, but often sharper bounds on differences of expectations of test functions are obtained by studying the Hellinger distance directly. KL divergence was used to quantify the error incurred when approximating posterior distributions in [26] in finite dimensions, and the Hellinger metric was used in [12] in the infinite dimensional setting required in this paper.

In some contexts the unknown σ is naturally a positive random variable and so we seek instead u where $\sigma = \exp(u)$. If we define $\mathcal{F} = \mathcal{G} \circ \exp(\cdot)$ then the inverse problem (2.1) becomes

$$(2.4) \qquad\qquad y = \mathcal{F}(u) + \eta.$$

Often we have an approximate solution u_0 to (2.4) and it is natural to seek a solution which deviates slightly from this. In this situation we write $u = u_0 + v$ and linearize (2.4) to obtain

$$y \approx \mathcal{F}(u_0) + D\mathcal{F}(u_0)v + \eta.$$

This suggests studying the linear inverse problem

$$(2.5) \qquad\qquad z = Gv + \eta$$

where $z = y - \mathcal{F}(u_0)$ and $G = D\mathcal{F}(u_0)$. If we put Gaussian prior $\mathcal{N}(m_{\text{prior}}, \mathcal{C}_{\text{prior}})$ on v then the posterior on $v|z$ is also Gaussian $\mathcal{N}(m_{\text{post}}, \mathcal{C}_{\text{post}})$ determined by

$$(2.6) \qquad \mathcal{C}_{\text{post}}^{-1} = \mathcal{C}_{\text{prior}}^{-1} + G^T \mathcal{C}_{\text{error}}^{-1} G, \quad \text{and} \quad m_{\text{post}} = m_{\text{prior}} + \mathcal{C}_{\text{post}} G^T \mathcal{C}_{\text{error}}^{-1} \left(z - G m_{\text{prior}} \right).$$

These formulae can also be interpreted in the infinite dimensional setting; see [24] and further citations in [31, 15, 28].

When Bayesian inversion is based on a nonlinear forward model, characterization of the resulting non-Gaussian posterior distribution can be quite complicated, requiring MCMC or SMC techniques [30]. One possible approach to deal with this is to perform the linearization described above and work with Gaussian priors and posterior distributions, leading to closed form solutions. These can be augmented with constraints by means of rejection sampling based on independent sampling from the Gaussian posterior.

2.2. Diffusion Limit Of The RTE. We consider the RTE (1.1) in the setting where the the absorption coefficient α and the scattering coefficient $k(x, v, v')$ are set to

$$\alpha(x, v) = k(x, v, v') = \epsilon^{-1} \sigma(x),$$

where ϵ is the Knudsen number. The thickness of the material physically corresponds to the number of times a photon scatters between being injected in a medium and escaping. The physical quantity is termed the Knudsen number, which stands for the ratio of mean free path and the domain length. The mean free path is the average distance a particle travels before being scattered. When the Knudsen number is small, photons, on average, scatter many times before they are emitted, and the material is thus regarded as optically thick.

In equation (1.1) dv denotes normalized unit measure, meaning

$$\langle 1 \rangle_v = \int_{\mathbb{S}^{d-1}} 1 dv = 1,$$

where we have used the notation $\langle \cdot \rangle_v$ to denote normalized integration over v. Thus equation (1.1) may be written as

$$v \cdot \nabla f = \frac{1}{\epsilon} \sigma \mathcal{L} f,$$

where the collision operator is

$$(2.7) \qquad\qquad \mathcal{L}f = \int_{\mathbb{S}^{d-1}} f(x, v') dv' - f = \langle f \rangle_v - f.$$

To ensure a unique solution we impose an incoming boundary condition, the analogue of a Dirichlet boundary condition for equations lacking velocity dependence. To this end define

$$\Gamma_\pm = \{(x,v) : x \in \partial\Omega, \pm v \cdot n_x > 0\}$$

which denotes the collection of coordinates on the boundary $x \in \partial\Omega$ on which the velocity v points in/out of the domain, respectively where $\pm v \cdot n_x > 0$. Here n_x is the normal vector at point x pointing out of Ω. The incoming boundary condition is imposed on Γ_-. We also define, for any fixed $y \in \partial\Omega$,

$$\Gamma_\pm(y) = \{(x,v) : x = y, \pm v \cdot n_y > 0\},$$

and set

$$\Gamma = \Gamma_+ \cup \Gamma_-, \quad \text{and} \quad \Gamma(y) = \Gamma_+(y) \cup \Gamma_-(y).$$

For a unique solution to (1.1), boundary conditions must be imposed on Γ_- as follows:

$$f|_{\Gamma_-} = \phi(x,v).$$

Combining the foregoing considerations we obtain

(2.8)
$$\begin{cases} v \cdot \nabla f = \frac{1}{\epsilon}\sigma\mathcal{L}f, & (x,v) \in \Omega \times \mathbb{S}^{d-1}, \\ f|_{\Gamma_-} = \phi(x,v) \end{cases},$$

with $\mathcal{L}$ as defined in (2.7). The domain Ω has a smooth C^1 boundary $\partial\Omega$. In the small ϵ regime, it was conjectured in [7] and then proved in [6] that the equation is asymptotically close to the following diffusion equation:

(2.9)
$$\begin{cases} -\nabla \cdot \left(\frac{1}{\sigma}\nabla\rho\right) = 0, & x \in \Omega \subset \mathbb{R}^d, \\ \rho|_{\partial\Omega} = \xi(x) \end{cases},$$

We make this convergence explicit under the following assumptions.

ASSUMPTION 2.1. *The functions σ, ϕ and ξ characterizing the medium and the boundary conditions are smooth functions, bounded in the following sense:*
- *the admissible medium is bounded, meaning there is a constant C_1 so that:*

$$\max\{\|\sigma\|_{L_\infty(\Omega)}, \|\sigma^{-1}\|_{L_\infty(\Omega)}, \|\nabla\left(\sigma^{-1}\right)\|_{L_\infty(\Omega)}\} < C_1;$$

- *and the boundary conditions are smooth and bounded, meaning:*

$$\max\{\|\xi\|_{L_\infty(\partial\Omega)}, \|\phi\|_{L_\infty(\Gamma)}\} < C_1.$$

We also term the set of admissible media:

(2.10)
$$\mathcal{A} = \{\sigma \in C^3(\Omega) : \max\{\|\sigma\|_{L_\infty(\Omega)}, \|\sigma^{-1}\|_{L_\infty(\Omega)}, \|\nabla\left(\sigma^{-1}\right)\|_{L_\infty(\Omega)}\} < C_1\}.$$

Here C^3 is the collection of third-order differentiable function set.

With this assumption, we first have the uniform boundedness of the Neumann data over $\mathcal{A}$.

PROPOSITION 2.1 ([16]). *Suppose ρ solves (2.9) with medium σ and the smooth boundary condition ξ satisfying Assumption 2.1, then there is a constant C that only depends on C_1 and Ω, so that*

(2.11)
$$\sup_{\sigma \in \mathcal{A}} \|\frac{1}{\sigma}\partial_n\rho\|_\infty < C.$$

Note that we assume only that the medium is smooth enough so the Neumann data is bounded. The regularity of the medium could certainly be relaxed but we do not pursue that direction in this paper. The key point here is to have the uniform bound over the set $\mathcal{A}$.

THEOREM 2.2. *Suppose $f(x, v)$ satisfies equation (2.8) with smooth boundary conditions and $\rho(x)$ solves (2.9). Then, as $\epsilon \to 0$, $f(x, v) \to \rho(x)$, assuming suitable compatibility relationships between the boundary data ϕ and ξ of the two equations. In particular, with compatible boundary conditions at different orders, one approximates f through different forms:*

- *if $\phi(x, v) = \xi(x)$:*

$$\|f - \rho\|_{L_\infty(\Omega \times \mathbb{S}^{d-1})} < C_{\mathcal{A}} \epsilon \, ;$$

- *if $\phi(x, v) = \xi(x) - \epsilon \frac{1}{\sigma(x)} v(x) \cdot \nabla \xi(x)$:*

$$\|f - \rho + \frac{\epsilon}{\sigma} v \cdot \nabla \rho\|_{L_\infty(\Omega \times \mathbb{S}^{d-1})} < C_{\mathcal{A}} \epsilon^2 \, .$$

Here the constant $C_{\mathcal{A}}$ depends on C_1, the upper bound in Assumption 2.1 for the admissible set.

We leave the rigorous proof to the appendix, and present here the formal perturbation expansion derivation; the latter is useful in building intuition.

Sketch Proof: Perturbation Expansion. We will use the standard asymptotic expansion technique in ϵ away from the boundary. Set

$$f_{\text{in}} = f_0 + \epsilon f_1 + \epsilon^2 f_2 + \cdots .$$

Plugging the expansion into (2.8), we obtain

$$v \cdot \nabla f_0 + \epsilon v \cdot \nabla f_1 + \epsilon^2 v \cdot \nabla f_2 + \cdots = \frac{1}{\epsilon} \sigma \mathcal{L}[f_0 + \epsilon f_1 + \epsilon^2 f_2 + \cdots] .$$

Multiplying by ϵ and equating in powers of ϵ gives

$$\begin{aligned} \epsilon^0 : \quad & \mathcal{L}[f_0] = 0 \, , \\ \epsilon^1 : \quad & v \cdot \nabla f_0 = \sigma \mathcal{L}[f_1] \, , \\ \epsilon^2 : \quad & v \cdot \nabla f_1 = \sigma \mathcal{L}[f_2] \, . \end{aligned}$$

The zeroth order expansion indicates that f_0 is in the null space of $\mathcal{L}$. From equation (2.7) we deduce that f_0 must be velocity independent, and thus we write $f_0(x, v) = \rho(x)$. With this expression, considering the equation at $\mathcal{O}(\epsilon)$, we have

$$f_1 = \mathcal{L}^{-1} \left[\frac{1}{\sigma} v \cdot \nabla \rho \right] \quad \Rightarrow \quad f_1 = -\frac{1}{\sigma} v \cdot \nabla \rho \, .$$

Here we have used the fact that $\mathcal{L}$ is one-to-one on the domain $(\text{Null} \, \mathcal{L})^\perp$ and that $v \cdot \nabla \rho \perp (\text{Null} \, \mathcal{L})^\perp$, since v integrates to zero on the unit sphere. To close the system we consider the equation at $\mathcal{O}(\epsilon^2)$, substituting $f_0 = \rho$ and $f_1 = -\frac{1}{\sigma} v \cdot \nabla \rho$ to obtain

$$-v \cdot \nabla \left(\frac{1}{\sigma} v \cdot \nabla \rho \right) = \sigma \mathcal{L}[f_2] .$$

Integrating the equation on both sides with respect to v and taking into account the fact that $\int_{\mathbb{S}^{d-1}} \mathcal{L}[f_2] dv = 0$, we have, using the summation convention,

$$\begin{aligned} 0 &= -\int_{\mathbb{S}^{d-1}} v \cdot \nabla \left(\frac{1}{\sigma} v \cdot \nabla \rho \right) dv = -\int_{\mathbb{S}^{d-1}} v_i v_j \partial_i \left(\frac{1}{\sigma} \partial_j \rho \right) dv = -C_d \partial_i \left(\frac{1}{\sigma} \partial_i \rho \right) \\ &= -C_d \nabla \cdot \left(\frac{1}{\sigma} \nabla \rho \right) . \end{aligned}$$

implying that

$$-\nabla \cdot \left(\frac{1}{\sigma} \nabla \rho \right) = 0 \, .$$

Here we have used that

$$\text{(2.12)} \qquad \int_{\mathbb{S}^{d-1}} v_i v_j \mathrm{d}v = C_d \delta_{ij}, \quad C_d := \int_{\mathbb{S}^{d-1}} v_i^2 \mathrm{d}v.$$

Note that C_d depends on dimension. For example, in $\mathbb{S}^2$, $C_d = 1/3$. Thus, we have shown that the radiative transfer equation in the diffusion limit becomes the diffusion equation, which concludes the sketch proof of the theorem. $\qquad\qquad \square$

Remark 2.2. In the preceding formal derivation we have ignored boundary conditions. In practice, unless these are chosen carefully, there will be a mismatch between the DE and the small ϵ solution of the RTE near the boundary. The boundary conditions stated in the theorem give different levels of consistency between the two equations, and hence lead to differing error estimates. See the proof in the appendix for details. When the boundary conditions are incompatible the analysis is considerably more subtle – see [7, 34, 25] for details.

2.3. Inverse Problems for the RTE and DE. We now define Bayesian inverse problems for the RTE and DE, relating to partial and noisy observations of the albedo and DtN operators respectively. The first ingredient is definition of the forward map $\mathcal{G}$, which we now do for the RTE and DE equations. We conclude this subsection with a discussion of the prior distribution, which we choose in common between the RTE and DE settings.

In optical tomography, high energy light with a known intensity is injected into the material, and detectors are placed on the tissue boundary to collect the light current emitted. For the RTE, the albedo operator is defined by $\mathcal{H}^{\mathrm{RTE}}$ which is a $\sigma-$dependent linear transformation of boundary data ϕ into the measurement space, defined by

$$\text{(2.13)} \qquad \mathcal{H}^{\mathrm{RTE}}(\sigma)\phi = h^{\mathrm{RTE}}.$$

where

$$\text{(2.14)} \qquad h^{\mathrm{RTE}}(x) = -\frac{1}{C_d \epsilon} \int_{\Gamma(x)} v \cdot n f(x, v) \mathrm{d}v,$$

and f satisfies (2.8). It is important to note that, whilst $\mathcal{H}^{\mathrm{RTE}}$ is linear in its action on ϕ, it depends nonlinearly on the unknown medium σ. The inverse problem of reconstructing σ from measurements of h^{RTE} is thus a nonlinear inverse problem.

In practice, finitely many smooth incoming data ϕ_k are injected and finitely many measurements are made at the boundary for each ϕ_k; we assume that these measurements may be expressed as linear functionals l_j of h^{RTE}. We thus define the forward map to be inverted by

$$\text{(2.15)} \qquad \mathcal{G}^{\mathrm{RTE}}(\sigma)_{j,k} = l_j(\mathcal{H}^{\mathrm{RTE}}(\sigma)\phi_k),$$

where $(j, k) \in \{1, \cdots, J\} \otimes \{1, \cdots, K\}$. We assume additive Gaussian noise η to obtain the compact representation of the inverse problem

$$\text{(2.16)} \qquad \mathsf{y} = \mathcal{G}^{\mathrm{RTE}}(\sigma) + \eta,$$

where $\eta \in \mathbb{R}^{JK}$ is drawn from a Gaussian distribution which we assume to have the form

$$\text{(2.17)} \qquad \eta \sim \mathcal{N}(0, \gamma^2 \mathbb{I}),$$

meaning

$$\text{(2.18)} \qquad \mathsf{y} \,|\, \sigma \sim \mathcal{N}(\mathcal{G}^{\mathrm{RTE}}(\sigma), \gamma^2 \mathbb{I}).$$

For the DE model the situation is analogous. The map that takes the Dirichlet data to the Neumann outflow is termed the DtN map and is defined by

$$\text{(2.19)} \qquad \mathcal{H}^{\mathrm{DE}}(\sigma)\phi = h^{\mathrm{DE}},$$

where

$$(2.20) \qquad h^{\mathrm{DE}}(x) = \frac{1}{\sigma}\frac{\partial \rho}{\partial n}(x)$$

and ρ satisfies (2.9). In practice, finitely many incoming data ξ_k are injected and finitely many linear functionals l_j of h^{DE} are measured, noisily, leading to an inverse problem of the form

$$(2.21) \qquad \mathsf{y} = \mathcal{G}^{\mathrm{DE}}(\sigma) + \eta\,,$$

where $\eta \in \mathbb{R}^{JK}$ denotes observational noise and where the forward map is defined by

$$(2.22) \qquad \mathcal{G}^{\mathrm{DE}}(\sigma)_{j,k} = l_j(\mathcal{H}^{\mathrm{DE}}(\sigma)\phi_k),$$

where $(j,k) \in \{1,\cdots,J\} \otimes \{1,\cdots,K\}$. For simplicity we assume the same noise model (2.17) for η.

Together with (2.17) and the assumption that η and σ are *a priori* independent, (2.16) and (2.21) define the likelihood for a Bayesian formulation of the inverse problem of determining σ from y from RTE and DE respectively. We now define the prior on σ, which we will choose in common between the two inverse problems. To this end recall the set (2.10) and define prior distribution $\mu_0(d\sigma)$ to be a probability measure supported on $\mathcal{A}$:

ASSUMPTION 2.3. *The prior measure μ_0 is supported on an infinite dimensional separable Banach space, and the support is contained in the admissible set $\mathcal{A}$ given by (2.10):*

$$\int_{\mathcal{A}} 1\, d\mu_0(\sigma) = 1\,.$$

In our case $\mathcal{A}$ is a subset of C^3. If we further relax the regularity assumptions, to let $\mathcal{A}$ be a subset of $W^{1,\infty}$, for example, then $W^{1,\infty}$ is not separable. But it is possible to construct useful measures with support in $W^{1,\infty}$ which are separable, for example through the closure of sets of random series expansions; see [15] for a related example in L^∞.

Bayes' theorem (2.2) for both models is then given by

$$(2.23) \qquad \mu^{\mathsf{y}}_{\mathrm{RTE}}(d\sigma) = \frac{1}{Z^{\mathrm{RTE}}}\mu^\sigma_{\mathrm{RTE}}(\mathsf{y})\mu_0(d\sigma)\,, \quad \text{and} \quad \mu^{\mathsf{y}}_{\mathrm{DE}}(\sigma) = \frac{1}{Z^{\mathrm{DE}}}\mu^\sigma_{\mathrm{DE}}(\mathsf{y})\mu_0(d\sigma)\,,$$

where

$$(2.24) \qquad \mu^\sigma_{\mathrm{RTE}}(\mathsf{y}) = \exp\left(-\frac{1}{2\gamma^2}\|\mathsf{y} - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|_2^2\right)\,, \quad \text{and} \quad \mu^\sigma_{\mathrm{DE}}(\mathsf{y}) = \exp\left(-\frac{1}{2\gamma^2}\|\mathsf{y} - \mathcal{G}^{\mathrm{DE}}(\sigma)\|_2^2\right)\,.$$

The functions $\mathcal{G}^{\mathrm{RTE}}$ and $\mathcal{G}^{\mathrm{DE}}$ are here both viewed as mappings from $\mathcal{A}$ into $\mathbb{R}^{JK}$. The normalization factors are given by

$$(2.25) \qquad Z^{\mathrm{RTE}} = \int_{\mathcal{A}} \mu^\sigma_{\mathrm{RTE}}(\mathsf{y})\mathrm{d}\mu_0(\sigma)\,, \quad \text{and} \quad Z^{\mathrm{DE}} = \int_{\mathcal{A}} \mu^\sigma_{\mathrm{DE}}(\mathsf{y})\mathrm{d}\mu_0(\sigma)\,.$$

Note also that the likelihoods $\mu^\sigma_{\mathrm{RTE}}(\mathsf{y})$ and $\mu^\sigma_{\mathrm{DE}}(\mathsf{y})$ are, for fixed σ, proportional to densities on $\mathbb{R}^{JK}$; hence we write them as functions of y. On the other hand $\mu^{\mathsf{y}}_{\mathrm{RTE}}(d\sigma), \mu^{\mathsf{y}}_{\mathrm{DE}}(d\sigma)$ and $\mu_0(d\sigma)$ are measures with support in $\mathcal{A}$, a subset of an infinite dimensional separable Banach space.

Theorem 2.2 shows that given compatible ϕ and ξ, h^{RTE} and h^{DE} are close when the Knudsen number ϵ is small, so that $\mathcal{G}^{\mathrm{RTE}}$ and $\mathcal{G}^{\mathrm{DE}}$ are close for every fixed σ, when ϵ is small. In section 3 we use these facts to demonstrate the convergence of $\mu^\sigma_{\mathrm{RTE}}$ to μ^σ_{DE} as $\epsilon \to 0$.

2.4. Linearized Albedo Operator And DtN Map. We derive linearized versions of the albedo operator and the DtN map by assuming that the unknown medium σ is close to a known background σ_0. In order to enforce positivity, we assume $\sigma = e^u$, define u_0 so that $\sigma_0 = e^{u_0}$, and find equations satisfied by the perturbation $w = u - u_0$. The corresponding inverse problem amounts to reconstructing w using the

measurements and some known information computed using the background medium σ_0. We express the admissible set $\mathcal{A}$ given in (2.10) on the log-scale and write

$$(2.26) \qquad \mathcal{A}_u = \{u \in C^3(\Omega) : \max\{\|e^u\|_{L_\infty(\Omega)}, \|e^{-u}\|_{L_\infty(\Omega)}, \|\nabla\left(e^{-u}\right)\|_{L_\infty(\Omega)}\} < C_1\}.$$

This implies

$$(2.27) \qquad \sup_{u \in \mathcal{A}_u} \|u\|_{L_\infty(\Omega)} < C_2 := \log C_1.$$

To start, we recall equation (2.8),

$$(2.28) \qquad \begin{cases} v \cdot \nabla f = \frac{1}{\epsilon} e^u \mathcal{L} f \\ f|_{\Gamma_-} = \phi(x,v) \end{cases}.$$

We assume that there is a background scattering coefficient $u_0 \in \mathcal{A}_u$, and that $w(x) \in C^3(\Omega)$ is a small fluctuation of u around the background u_0: $w = u - u_0$. Then

$$(2.29) \qquad \|w(x)\|_{L_\infty(\Omega)} = \|u(x) - u_0(x)\|_{L_\infty(\Omega)} \ll \|u\|_{L_\infty(\Omega)}.$$

We define a new function f_{lin} which solves the RTE with the background scattering coefficient and the same boundary condition,

$$(2.30) \qquad \begin{cases} v \cdot \nabla f_{\mathrm{lin}} = \frac{1}{\epsilon} e^{u_0} \mathcal{L} f_{\mathrm{lin}} \\ f_{\mathrm{lin}}|_{\Gamma_-} = \phi(x,v) \end{cases}.$$

The difference between f and f_{lin}, termed $\mathsf{f} = f - f_{\mathrm{lin}}$, then satisfies, neglecting terms of $\mathcal{O}(w^2)$, the following error equation:

$$(2.31) \qquad \begin{cases} v \cdot \nabla \mathsf{f} = \frac{1}{\epsilon} e^{u_0} \mathcal{L} \mathsf{f} + \frac{1}{\epsilon} e^{u_0} w \mathcal{L} f_{\mathrm{lin}} \\ \mathsf{f}|_{\Gamma_-} = 0 \end{cases}.$$

To extract boundary data from (2.31), we define the adjoint equation, with a delta function on the boundary,

$$(2.32) \qquad \begin{cases} -v \cdot \nabla g = \frac{1}{\epsilon} e^{u_0} \mathcal{L} g \\ g|_{\Gamma_+} = \delta_y(x) \end{cases}.$$

Here we have used the fact that $\mathcal{L}$ is self-adjoint, and for the adjoint equation, the incoming boundary condition should be imposed on Γ_+. We have also imposed a delta function concentrated at $y \in \partial\Omega$. Multiplying (2.31) by g and (2.32) by f and subtracting the two obtained equations, integrated over x and v, we obtain, upon using Green's identity

$$\int_{\Gamma_+(y) \cup \Gamma_-(y)} (v \cdot n) \mathsf{f} g \mathrm{d}x \mathrm{d}v = \frac{1}{\epsilon} \int_\Omega e^{u_0} w \int_{\mathbb{S}^{d-1}} g \mathcal{L} f_{\mathrm{lin}} \mathrm{d}v \mathrm{d}x.$$

Noting that $\mathsf{f}|_{\Gamma_-} = 0$ and $g|_{\Gamma_+} = \delta_y$, we may further simplify the left hand side, obtaining

$$(2.33) \qquad \int_{\Gamma_+(y)} v \cdot n_y \mathsf{f}(y,v) \mathrm{d}v = \frac{1}{\epsilon} \int_\Omega e^{u_0(x)} w(x) \int_{\mathbb{S}^{d-1}} g(x,v) \mathcal{L} f_{\mathrm{lin}}(x,v) \mathrm{d}v \mathrm{d}x.$$

As in the nonlinear case, we have finitely many measurements and experiments conducted. In the K experiments, we use incoming data ϕ_k, and for each experiment we measure data using the measurement-operator l_j:

$$(2.34) \qquad \{\phi_1, \cdots, \phi_K\}, \quad \{l_1, \cdots, l_J\}.$$

Letting g_j denote the solution to (2.32) with $\delta_{x_j}(x)$ on the boundary, and f_k, and $f_{\mathrm{lin},k}$ denote the solutions to (2.28) and (2.30) with ϕ_k as boundary data, we define

$$(2.35) \qquad \gamma_{jk}^{\mathrm{RTE}}(x) = -\frac{e^{u_0}}{C_d \epsilon^2} \int_{\mathbb{S}^{d-1}} g_j(x,v) \mathcal{L} f_{\mathrm{lin},k}(x,v)\mathrm{d}v,$$

and define

$$(2.36) \qquad G_{jk}^{\mathrm{RTE}}(w) := \langle \gamma_{jk}^{\mathrm{RTE}}(x), w \rangle_x,$$

with $\langle \cdot, \cdot \rangle_x$ denoting the inner-product defined by integration over Ω in the x variable alone. We note that, by (2.33),

$$(2.37) \qquad G_{jk}^{\mathrm{RTE}}(w) \approx \mathcal{G}_{jk}^{\mathrm{RTE}}(\log(u)) - \mathcal{G}_{jk}^{\mathrm{RTE}}(\log(u_0))$$

because

$$
\begin{aligned}
\mathcal{G}_{jk}^{\mathrm{RTE}}(\log(u)) - \mathcal{G}_{jk}^{\mathrm{RTE}}(\log(u_0)) &= l_j(\mathcal{H}^{\mathrm{RTE}}(\log(u))(\phi_k)) - l_j(\mathcal{H}^{\mathrm{RTE}}(\log(u_0))(\phi_k)) \\
&= -\frac{1}{C_d \epsilon} \int_{\Gamma_+(x_j)} v \cdot n_{x_j} f_k(x_j,v)\mathrm{d}v + \frac{1}{C_d \epsilon} \int_{\Gamma_+(x_j)} v \cdot n_{x_j} f_{\mathrm{lin},k}(x_j,v)\mathrm{d}v \\
&\approx -\frac{1}{C_d \epsilon} \int_{\Gamma_+(x_j)} v \cdot n_{x_j} \mathsf{f}_k(x_j,v)\mathrm{d}v \\
&= G_{jk}^{\mathrm{RTE}}(w).
\end{aligned}
$$

Here $\mathsf{f}_k = f_k - f_{\mathrm{lin},k}$.

Deriving the linearized DtN map for the diffusion equation is similar. For ease of notation we start with the form of the diffusion equation as in (2.9), where the scattering coefficient is shown in the denominator. For positivity, we use $\sigma = e^u$, and $\sigma_0 = e^{u_0}$ as before. We now derive an equation for $\tau = \rho - \rho_{lin}$, first noting that

$$
\begin{cases}
-\nabla \cdot \left(\frac{1}{e^u}\nabla\rho\right) = -\nabla \cdot \left(\frac{1}{e^{u_0}e^w}\nabla(\rho_{\mathrm{lin}} + \tau)\right) = 0 \\
\rho\big|_{\partial\Omega} = \xi(x)
\end{cases},
$$

where ρ_{lin} solves:

$$(2.38) \qquad
\begin{cases}
-\nabla \cdot \left(\frac{1}{e^{u_0}}\nabla\rho_{\mathrm{lin}}\right) = 0 \\
\rho_{\mathrm{lin}}\big|_{\partial\Omega} = \xi(x)
\end{cases}.
$$

Subtracting the two equations and neglecting higher order terms, we have

$$(2.39) \qquad
\begin{cases}
-\nabla \cdot \left(\frac{1}{e^{u_0}}\nabla\tau\right) = -\nabla \cdot \left(\frac{w}{e^{u_0}}\nabla\rho_{\mathrm{lin}}\right) \\
\tau\big|_{\partial\Omega} = 0
\end{cases}
$$

with $w = u - u_0$ as in the RTE case.

We also define ρ_g that solves the adjoint equation:

$$(2.40) \qquad
\begin{cases}
-\nabla \cdot \left(\frac{1}{e^{u_0}}\nabla\rho_g\right) = 0 \\
\rho_g\big|_{\partial\Omega} = \delta_y
\end{cases}.
$$

Multiplying (2.39) by ρ_g, (2.40) by τ and integrating over Ω, we obtain:

$$(2.41) \quad \int_\Omega \frac{w}{e^{u_0}} \nabla\rho_{\mathrm{lin}} \cdot \nabla\rho_g \mathrm{d}x = -\int_{\partial\Omega} \frac{w}{e^{u_0}}\frac{\partial\rho_{\mathrm{lin}}}{\partial n}\rho_g \mathrm{d}x + \int_{\partial\Omega} \frac{1}{e^{u_0}}\frac{\partial\tau}{\partial n}\rho_g \mathrm{d}x = -\frac{w(y)}{e^{u_0(y)}}\frac{\partial\rho_{\mathrm{lin}}(y)}{\partial n_y} + \frac{1}{e^{u_0(y)}}\frac{\partial\tau(y)}{\partial n_y}.$$

Similarly to the nonlinear case, we conduct finitely many experiments and make finitely many measurements as in (2.34). Defining

$$(2.42) \qquad \gamma_{jk}^{\mathrm{DE}}(x) = \frac{1}{e^{u_0}} \nabla \rho_{\mathrm{lin},k} \cdot \nabla \rho_{g,j} \,,$$

where $\rho_{g,j}$ satisfies (2.40) with δ_{y_j} boundary condition and $\rho_{\mathrm{lin},k}$ satisfies (2.38) with ξ_k with as the boundary condition. Using the approximation that τ satisfies (2.39), we write (2.41) as

$$(2.43) \qquad \langle \gamma_{jk}^{\mathrm{DE}}(x) , w \rangle_x = \mathcal{G}_{jk}^{\mathrm{DE}}(\log(u)) - \mathcal{G}_{jk}^{\mathrm{DE}} \log((u_0)) = G_{jk}^{\mathrm{DE}}(w) \,,$$

where again G_{jk}^{DE} is the linearized approximation and we have used the estimate

$$
\begin{aligned}
\mathcal{G}_{jk}^{\mathrm{DE}}(\log(u)) - \mathcal{G}_{jk}^{\mathrm{DE}}(\log(u_0)) &= \frac{1}{e^u} \frac{\partial \rho_k}{\partial n}(x_j) - \frac{1}{e^{u_0}} \frac{\partial \rho_{\mathrm{lin},k}}{\partial n}(x_j) \\
&= \frac{1}{e^w e^{u_0}} \frac{\partial (\tau_k + \rho_{\mathrm{lin},k})}{\partial n}(x_j) - \frac{1}{e^{u_0}} \frac{\partial \rho_{\mathrm{lin},k}}{\partial n}(x_j) \\
&\approx \frac{1}{e^{u_0}} \frac{\partial \tau_k}{\partial n}(x_j) - \frac{w}{e^{u_0}} \frac{\partial \rho_{\mathrm{lin},k}}{\partial n}(x_j) \,.
\end{aligned}
$$

Here $\tau_k = \rho_k - \rho_{lin,k}$. This defines the (linear) action of G_{jk}^{DE} on w. Notice that the linearized albedo operator (2.36) and the linearized DtN map (2.43) have the same format: they are both Fredholm first type integrals, determined by the $\gamma_{jk}^{\mathrm{RTE}}$ and $\gamma_{jk}^{\mathrm{DE}}$ respectively defined in (2.35) and (2.42). To show the convergence of the two problems amounts to showing, in the small ϵ regime, the convergence of the two forward maps $\gamma_{jk}^{\mathrm{RTE}} \sim \gamma_{jk}^{\mathrm{DE}}$ for all j and k, and the convergence of the data $G_{jk}^{\mathrm{RTE}}(w)$ to $G_{jk}^{\mathrm{DE}}(w)$ for reasonably small w.

3. Nonlinear Inverse Problems. In this section we analyze the distance between the posterior distributions of the nonlinear inverse problems defined by the RTE and DE, namely $\mu_{\mathrm{RTE}}^{\mathsf{y}}(\sigma)$ and $\mu_{\mathrm{DE}}^{\mathsf{y}}(\sigma)$ respectively. We consider the setting in which the Knudsen number ϵ is small. We show that the two distributions converge in the Kullback-Leibler divergence and the Hellinger distance as $\epsilon \to 0$. The three subsections concern, in turn, the following convergence results as $\epsilon \to 0$:

1. convergence of the forward map $\mathcal{G}^{\mathrm{RTE}}(\sigma)$ to $\mathcal{G}^{\mathrm{DE}}(\sigma)$ for a fixed list of (ϕ_k, l_j);
2. convergence of the KL divergence between $\mu_{\mathrm{RTE}}^{\mathsf{y}}(\sigma)$ and $\mu_{\mathrm{DE}}^{\mathsf{y}}(\sigma)$ to zero;
3. convergence of the Hellinger metric $\mu_{\mathrm{RTE}}^{\mathsf{y}}(\sigma)$ and $\mu_{\mathrm{DE}}^{\mathsf{y}}(\sigma)$ to zero.

Before these three pieces of analysis, recall that the forward measurement operators for the RTE and DE are defined in (2.15) and (2.22) respectively and that Bayes' theorem (2.2) delivers the formulae for the posterior distributions in (2.23)–(2.25).

3.1. Convergence Of The Forward Map. For simplicity we assume that l_j is the linear funtional that takes corresponding to evaluation at point $x_j \in \partial\Omega$:

$$(3.1) \qquad l_j(f) = f(x_j), \quad \text{where} \quad x_j \in \partial\Omega \,.$$

Other linear functionals can be handled with similar analysis. Then

$$(3.2) \qquad \mathcal{G}_{jk}^{\mathrm{RTE}}(\sigma) = l_j(\mathcal{H}^{\mathrm{RTE}}(\sigma)\phi_k) = -\frac{1}{C_d \epsilon} \int_{\Gamma(x_j)} (v \cdot n) f(x_j, v) \mathrm{d}v \,,$$

and

$$(3.3) \qquad \mathcal{G}_{jk}^{\mathrm{DE}}(\sigma) = l_j(\mathcal{H}^{\mathrm{DE}}(\sigma)\xi_k) = \frac{1}{\sigma(x_j)} \frac{\partial \rho}{\partial n_{x_j}}(x_j) \,.$$

where f and ρ are the solutions to the RTE and the DE with ϕ_k and ξ_k as incoming conditions, respectively.

We now have the following proposition:

PROPOSITION 3.1. *Assume that $\phi_k(x,v) = \xi(x) - \epsilon \frac{1}{\sigma(x)} v(x) \cdot \nabla \xi_k(x)$. Then, under Assumption 2.1, the forward maps $\mathcal{G}^{RTE}$ and $\mathcal{G}^{DE}$ satisfy*

$$(3.4) \qquad \sup_{\sigma \in \mathcal{A}} \|\mathcal{G}^{RTE}(\sigma) - \mathcal{G}^{DE}(\sigma)\|_\infty \le \frac{C_{\mathcal{A}}}{C_d} \epsilon \,.$$

Furthermore, there is a constant C that only depends on C_1 and Ω so that

$$(3.5) \qquad \max\left\{ \sup_{\sigma\in\mathcal{A}} \|\mathcal{G}^{RTE}(\sigma)\|_\infty , \sup_{\sigma\in\mathcal{A}} \|\mathcal{G}^{DE}(\sigma)\|_\infty \right\} \leq C .$$

Proof. To show the first item it is enough to prove that for every j and k,

$$|\mathcal{G}^{\mathrm{RTE}}_{jk}(\sigma) - \mathcal{G}^{\mathrm{DE}}_{jk}(\sigma)| \leq \frac{C_{\mathcal{A}}}{C_d}\epsilon .$$

Note that, for any y, $\mathbb{S}^{d-1}\backslash\Gamma(y)$ is the set on which $(v \cdot n) = 0$. Hence, employing (2.12), (3.2) and (3.3), and defining

$$r = f - \rho + \frac{\epsilon}{\sigma} v \cdot \nabla\rho,$$

we then have

$$(3.6) \qquad \begin{aligned} |\mathcal{G}^{\mathrm{RTE}}_{jk}(\sigma) - \mathcal{G}^{\mathrm{DE}}_{jk}(\sigma)| &= \left| \frac{1}{\sigma}\frac{\partial\rho}{\partial n}(x_j) + \frac{1}{C_d\epsilon}\int_{\Gamma(x_j)} v \cdot n f \mathrm{d}v \right| \\ &= \left| \frac{1}{\sigma}\frac{\partial\rho}{\partial n}(x_j) + \frac{1}{C_d\epsilon}\int_{\mathbb{S}^{d-1}} (v\cdot n)(\rho - \frac{\epsilon}{\sigma}v\cdot\nabla\rho + r)\mathrm{d}v \right| \\ &= \left| \frac{1}{\sigma}\frac{\partial\rho}{\partial n}(x_j) + \int_{\mathbb{S}^{d-1}}\frac{1}{C_d}\left[\frac{-1}{\sigma}(v\cdot n)(v\cdot\nabla\rho) + \frac{r}{\epsilon}\right]\mathrm{d}v \right| \\ &\leq \left| \frac{1}{\sigma}\frac{\partial\rho}{\partial n}(x_j) - \frac{1}{\sigma}\frac{\partial\rho}{\partial n}(x_j) \right| + \frac{C_{\mathcal{A}}}{C_d}\epsilon = \frac{C_{\mathcal{A}}}{C_d}\epsilon . \end{aligned}$$

Here we used Theorem 2.2 which states

$$\|r\|_{L_\infty(\Omega\times\mathbb{S}^{d-1})} = \left\|f - \left(\rho - \frac{\epsilon}{\sigma}v\cdot\nabla\rho\right)\right\|_{L_\infty(\Omega\times\mathbb{S}^{d-1})} \leq C_{\mathcal{A}}\epsilon^2 ,$$

which concludes the proof of equation (3.4). Equation (3.5) is a direct consequence of Proposition 2.1 and by combining the conclusion of Proposition 2.1 with equation (3.4). $\qquad\square$

3.2. Convergence In Kullback-Leibler Divergence. We use the convergence of the forward map to show the convergence in the posterior distribution using the Kullback-Leibler divergence.

THEOREM 3.2. *Let the assumptions of Proposition 3.1, together with Assumption 2.3, hold. Then*

$$\mathrm{d}_{KL}(\mu^y_{RTE}, \mu^y_{DE}) \leq \mathcal{O}(\epsilon) .$$

Proof. We first note that, over the set $\mathcal{A}$ which contains the support of the common prior measure μ_0, the likelihoods $\mu^\sigma_{\mathrm{RTE}}(y)$ and $\mu^\sigma_{\mathrm{DE}}(y)$ are bounded uniformly from above and below. Hence the measures μ^y_{RTE} and μ^y_{DE} are mutually absolutely continuous and have densities with respect to one another. In particular we may define

$$(3.7) \qquad \mathrm{d}_{\mathrm{KL}}(\mu^y_{\mathrm{RTE}}, \mu^y_{\mathrm{DE}}) = \int_{\mathcal{A}}\left(\log\frac{d\mu^y_{\mathrm{RTE}}}{d\mu^y_{\mathrm{DE}}}(\sigma)\right)\mathrm{d}\mu^y_{\mathrm{DE}}(\sigma) ,$$

where $\sigma \in \mathcal{A}$. Clearly μ^y_{DE} has no ϵ dependence, and so it suffices to show that $\log\frac{d\mu^y_{\mathrm{RTE}}}{d\mu^y_{\mathrm{DE}}}$ is $\mathcal{O}(\epsilon)$, uniformly on $\mathcal{A}$. Using (2.23)–(2.25), we find

$$\log\frac{d\mu^y_{\mathrm{RTE}}}{d\mu^y_{\mathrm{DE}}}(\sigma) = \log\left(\frac{\mu_0(\sigma)\mu^\sigma_{\mathrm{RTE}}(y)}{Z^{\mathrm{RTE}}}\frac{Z^{\mathrm{DE}}}{\mu_0(\sigma)\mu^\sigma_{\mathrm{DE}}(y)}\right) = \log\frac{Z^{\mathrm{DE}}}{Z^{\mathrm{RTE}}} + \log\frac{\mu^\sigma_{\mathrm{RTE}}(y)}{\mu^\sigma_{\mathrm{DE}}(y)} .$$

We will show that both terms are $\mathcal{O}(\epsilon)$. Recalling (2.24) we see that

$$\begin{aligned} \left|\mu^\sigma_{\mathrm{RTE}}(y) - \mu^\sigma_{\mathrm{DE}}(y)\right| &= \left| \exp\left(-\frac{\|y - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|^2}{2\gamma^2}\right) - \exp\left(-\frac{\|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|^2}{2\gamma^2}\right) \right| \\ &\leq c\left| \|y - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|^2 - \|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|^2 \right| , \end{aligned}$$

where $c < \infty$ is the Lipschitz constant for $\exp(-|x|/2\gamma^2)$. Now note that

$$\|y - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|^2 - \|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|^2 = -\left(2y - \mathcal{G}^{\mathrm{RTE}}(\sigma) - \mathcal{G}^{\mathrm{DE}}(\sigma)\right)^{\top}\left(\mathcal{G}^{\mathrm{RTE}}(\sigma) - \mathcal{G}^{\mathrm{DE}}(\sigma)\right),$$

and according to Proposition 3.1,

$$(3.8) \qquad \sup_{\sigma \in \mathcal{A}} \|c\left(2y - \mathcal{G}^{\mathrm{RTE}}(\sigma) - \mathcal{G}^{\mathrm{DE}}(\sigma)\right)\|_{\infty} < \infty,$$

we deduce that

$$\sup_{\sigma \in \mathcal{A}} \left|\mu^{\sigma}_{\mathrm{RTE}}(y) - \mu^{\sigma}_{\mathrm{DE}}(y)\right| = \mathcal{O}(\epsilon).$$

Using the definition of the normalization factor and noting that $\int d\mu_0(\mathcal{A}) = 1$, we also have

$$\left|Z^{\mathrm{RTE}} - Z^{\mathrm{DE}}\right| \leq \int_{\mathcal{A}} \left|\mu^{\sigma}_{\mathrm{RTE}}(y) - \mu^{\sigma}_{\mathrm{DE}}(y)\right| d\mu_0(\sigma) = \mathcal{O}(\epsilon).$$

Noting that Z^{DE} and $\mu^{\sigma}_{\mathrm{DE}}(y)$ are bounded from below uniformly with respect to $\sigma \in \mathcal{A}$, we deduce from the two preceding displays that

$$\sup_{\sigma \in \mathcal{A}} \left| \log \frac{Z^{\mathrm{DE}}}{Z^{\mathrm{RTE}}} + \log \frac{\mu^{\sigma}_{\mathrm{RTE}}(y)}{\mu^{\sigma}_{\mathrm{DE}}(y)} \right| = \mathcal{O}(\epsilon)$$

which completes the proof. □

3.3. Convergence In Hellinger Metric. Convergence in the Hellinger metric has a very similar proof to that used in KL divergence.

THEOREM 3.3. *Let the assumptions of Proposition 3.1, together with Assumption 2.3, hold. Then*

$$d_{Hell}(\mu^{y}_{RTE}, \mu^{y}_{DE}) \leq \mathcal{O}(\epsilon).$$

Proof. We first recall the definition of the Hellinger distance between two distributions in section 2.1, using $\lambda = \mu_0$ as the reference measure:

$$d_{\mathrm{Hell}}(\mu^{y}_{\mathrm{RTE}}, \mu^{y}_{\mathrm{DE}})^2 = \frac{1}{2} \int_{\mathcal{A}} \left(\sqrt{\frac{d\mu^{y}_{\mathrm{RTE}}}{d\mu_0}(\sigma)} - \sqrt{\frac{d\mu^{y}_{\mathrm{DE}}}{d\mu_0}(\sigma)} \right)^2 \mu_0(d\sigma).$$

Following [31] we obtain

$$(3.9)$$

$$d_{\mathrm{Hell}}(\mu^{y}_{\mathrm{RTE}}, \mu^{y}_{\mathrm{DE}})^2 = \frac{1}{2} \int_{\mathcal{A}} \left[\frac{1}{\sqrt{Z^{\mathrm{RTE}}}} \exp\left(\frac{-1}{2\gamma^2}\|y - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|_2^2\right) - \frac{1}{\sqrt{Z^{\mathrm{DE}}}} \exp\left(\frac{-1}{2\gamma^2}\|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|_2^2\right) \right]^2 d\mu_0$$

$$\leq I_1 + I_2,$$

where

$$I_1 = \frac{1}{Z^{\mathrm{RTE}}} \int_{\mathcal{A}} \left[\exp\left(-\frac{1}{2\gamma^2}\|y - \mathcal{G}^{\mathrm{RTE}}(\sigma)\|^2\right) - \exp\left(-\frac{1}{2\gamma^2}\|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|^2\right) \right]^2 d\mu_0(\sigma),$$

and

$$I_2 = \left|(Z^{\mathrm{RTE}})^{-1/2} - (Z^{\mathrm{DE}})^{-1/2}\right|_2^2 \int_{\mathcal{A}} \exp\left(-\frac{1}{2\gamma^2}\|y - \mathcal{G}^{\mathrm{DE}}(\sigma)\|_2^2\right) d\mu_0(\sigma).$$

With the same argument, we have

$$(3.10) \qquad I_1 \leq \frac{c}{Z^{\mathrm{RTE}}} \int_{\mathcal{A}} \|\mathcal{G}^{\mathrm{RTE}} - \mathcal{G}^{\mathrm{DE}}\|_2^2 \|\mathcal{G}^{\mathrm{RTE}} + \mathcal{G}^{\mathrm{DE}} - 2y\|_2^2 d\mu_0 = \mathcal{O}(\epsilon^2),$$

where we have used

$$\|\mathcal{G}^{\mathrm{RTE}} - \mathcal{G}^{\mathrm{DE}}\|_{\infty} < C_{\mathcal{A}}\epsilon/C_d,$$

and the Lipschitz argument as in (3.8). To deal with I_2, we notice that

$$(3.11) \qquad I_2 \le \max\left\{(Z^{\text{RTE}})^{-3}, (Z^{\text{DE}})^{-3}\right\} |Z^{\text{RTE}} - Z^{\text{DE}}|^2 \int_{\mathcal{A}} \exp\left(-\frac{1}{2}\|\mathsf{y} - \mathcal{G}^{\text{DE}}(\sigma)\|_2^2\right) d\mu_0(\sigma),$$

using the fact that

$$|(Z^{\text{RTE}})^{-1/2} - (Z^{\text{DE}})^{-1/2}|^2 \le \max\{(Z^{\text{RTE}})^{-3}, (Z^{\text{DE}})^{-3}\}|Z^{\text{RTE}} - Z^{\text{DE}}|^2.$$

According to the definition of $Z^{\text{RTE,DE}}$, we have

$$
\begin{aligned}
|Z^{\text{RTE}} - Z^{\text{DE}}| &\le \int_{\mathcal{A}} \left| \exp\left(-\frac{1}{\gamma^2}\|\mathsf{y} - \mathcal{G}(\sigma)^{\text{RTE}}\|_2^2\right) - \exp\left(-\frac{1}{\gamma^2}\|\mathsf{y} - \mathcal{G}(\sigma)^{\text{DE}}\|_2^2\right) \right| d\mu_0(\sigma) \\
&\le c \int_{\mathcal{A}} \left| \|\mathsf{y} - \mathcal{G}^{\text{RTE}}(\sigma)\|_2^2 - \|\mathsf{y} - \mathcal{G}^{\text{DE}}(\sigma)\|_2^2 \right| d\mu_0(\sigma) \\
&\le c \int_{\mathcal{A}} \|\mathcal{G}^{\text{RTE}} - \mathcal{G}^{\text{DE}}\|_2 \|\mathcal{G}^{\text{RTE}} + \mathcal{G}^{\text{DE}} - 2\mathsf{y}\|_2 d\mu_0(\sigma) \\
&= C_{\mathcal{A}}\epsilon/C_d,
\end{aligned}
$$

where we used (3.8). Plugging these back in (3.11), we have

$$I_2 = \mathcal{O}(\epsilon^2).$$

Together with the boundedness of I_1 and the inequality (3.9), we conclude

$$d_{\text{Hell}}(\mu_{\text{RTE}}^{\mathsf{y}}, \mu_{\text{DE}}^{\mathsf{y}}) = \mathcal{O}(\epsilon). \qquad \square$$

4. Linearized Inverse Problems.

In this section we study approximations of the two Bayesian inverse problems in the linearized setting. We show asymptotic closeness of the posterior distributions in the small Knudsen number regime $\epsilon \ll 1$. Equations (2.35)–(2.37) and (2.43) give rise to the following approximate inverse problems:

$$(4.1) \qquad \mathsf{y}_{\text{lin}}^{\text{RTE}} = G^{\text{RTE}}(w) + \eta, \quad \text{and} \quad \mathsf{y}_{\text{lin}}^{\text{DE}} = G^{\text{DE}}(w) + \eta,$$

where

$$\mathsf{y}_{\text{lin}}^{\text{RTE}} = \mathsf{y} - \mathcal{G}^{\text{RTE}}(\log(u_0)), \quad \text{and} \quad \mathsf{y}_{\text{lin}}^{\text{DE}} = \mathsf{y} - \mathcal{G}^{\text{DE}}(\log(u_0)).$$

is a vector of length JK and can be regarded as the linearized data. It can be obtained by subtracting y, the collected measurements with $\{\phi_k, k = 1, \ldots, K\}$ being the input data and $\{l_j, j = 1, \ldots, J\}$ being the pointwise evaluation operator, as defined in (3.1), and $\mathcal{G}^{\text{RTE,DE}}(\log(u_0))$, the background data that is precomputed using (2.30) or (2.38) with the same input and measurement operator, and the background medium u_0.

Assuming $\eta \sim \mathcal{N}(0, \gamma^2 \mathbb{I})$ as always, now we have the likelihood functions:

$$\nu_{\text{RTE}}^w(\mathsf{y}^{\text{RTE}}) = \exp\left(-\frac{1}{2\gamma^2}\|\mathsf{y}_{\text{lin}}^{\text{RTE}} - G^{\text{RTE}}(w)\|_2^2\right) \quad \text{and} \quad \nu_{\text{DE}}^w(\mathsf{y}^{\text{DE}}) = \exp\left(-\frac{1}{2\gamma^2}\|\mathsf{y}_{\text{lin}}^{\text{DE}} - G^{\text{DE}}(w)\|_2^2\right)$$

The two models use the same prior distribution function $\nu_0(w)$, satisfying

$$\int_{\mathcal{C}^3(\Omega)} 1 d\nu_0 = 1.$$

The posterior distributions are then

$$(4.2) \qquad \nu_{\text{RTE}}^{\mathsf{y}}(dw) = \frac{1}{Z^{\text{RTE}}} \nu_{\text{RTE}}^w(\mathsf{y}^{\text{RTE}}) \nu_0(dw), \quad \text{and} \quad \nu_{\text{DE}}^{\mathsf{y}}(w) = \frac{1}{Z^{\text{DE}}} \nu_{\text{DE}}^w(\mathsf{y}^{\text{DE}}) \nu_0(dw),$$

where Z^{RTE} and Z^{DE} are the normalization factors.

4.1. Convergence Of Linearized Forward Map. We first show the convergence of the interpreters $\gamma^{\mathrm{RTE,DE}}$, which will allow us to show the convergence of the two forward maps.

PROPOSITION 4.1. *Assume $u_0 \in \mathcal{A}_u$, then for ϵ sufficiently small, $\gamma^{RTE} \to \gamma^{DE}$. More specifically, for every j and k,*

$$(4.3) \qquad \|\gamma_{jk}^{RTE} - \gamma_{jk}^{DE}\|_{L_\infty(\Omega)} \le C\epsilon^2 \,.$$

Here the constant C only depends on $C_\mathcal{A}$ and C_1.

Proof. Recall the definition of γ_{jk} in (2.35)

$$\gamma_{jk}^{\mathrm{RTE}}(x) = -\frac{e^{u_0}}{C_d \epsilon^2} \int_{\mathbb{S}^{d-1}} g_j(x,v) \mathcal{L} f_{\mathrm{lin},k}(x,v) \mathrm{d}v \,,$$

where g_j and $f_{\mathrm{lin},k}$ solve (2.32) and (2.30) with δ_{y_j} and ϕ_k as boundary conditions. We further recall Theorem 2.2, so that we have

$$\|g_j - \rho_{g_j} - \epsilon e^{-u_0} v \cdot \nabla \rho_{g_j}\|_{L_\infty(\Omega \times \mathbb{S}^{d-1})} < C_\mathcal{A}\epsilon^2 \,, \quad \text{and} \quad \|f_{\mathrm{lin},k} - \rho_{f_k} + \epsilon e^{-u_0} v \cdot \nabla \rho_{f_k}\|_{L_\infty(\Omega \times \mathbb{S}^{d-1})} < C_\mathcal{A}\epsilon^2 \,,$$

where $\rho_{g,j}$ and $\rho_{\mathrm{lin},k}$ solve

$$-\nabla \cdot (e^{-u_0} \nabla \rho) = 0 \,,$$

with boundary condition δ_{y_j} and ξ_k respectively. Recalling $\mathcal{L}\rho = 0$ for all ρ, and that $\int_{\mathbb{S}^{d-1}} v \mathrm{d}v = 0$, then

$$\|\gamma_{jk}^{\mathrm{RTE}}(x) - \gamma_{jk}^{\mathrm{DE}}(x)\|_{L_\infty(\Omega)} = \|\gamma_{jk}^{\mathrm{RTE}}(x) - e^{-u_0} \nabla \rho_{g,j} \cdot \nabla \rho_{\mathrm{lin},k}\|_{L_\infty(\Omega)} = \mathcal{O}(\epsilon^2) \,. \qquad \square$$

We conclude the proof since this holds for every j and k.

We emphasize that γ^{RTE} is uniquely determined by g and f_{lin} that solve (2.32) and (2.30), and that the two equations depend merely on u_0. So the convergence holds true as long as $u_0 \in \mathcal{A}_u$, and there is no requirement on w.

COROLLARY 4.2. *For any fixed $u_0 \in \mathcal{A}_u$, assume $w \in C^3$, if ϵ significantly small, then $G^{RTE} \to G^{DE}$. More specifically,*

$$(4.4) \qquad \|G^{RTE} - G^{DE}\|_\infty \le C\epsilon^2 \|w\|_{L_2(\Omega)} \,,$$

where $G^{RTE,DE}$ are two vectors of length JK, and C only depends on $C_\mathcal{A}$, J and K.

Proof. According to the definition of $G^{\mathrm{RTE,DE}}$,

$$G_{jk}^{\mathrm{RTE}} - G_{jk}^{\mathrm{DE}} = \langle \gamma_{jk}^{\mathrm{RTE}} - \gamma_{jk}^{\mathrm{DE}}, w \rangle \le \|\gamma_{jk}^{\mathrm{RTE}} - \gamma_{jk}^{\mathrm{DE}}\|_{L_2(\Omega)} \|w\|_{L_2(\Omega)} \,.$$

We conclude using the result from Proposition 4.1, and that $\|\gamma_{jk}^{\mathrm{RTE}} - \gamma_{jk}^{\mathrm{DE}}\|_{L_2(\Omega)} \lesssim \|\gamma_{jk}^{\mathrm{RTE}} - \gamma_{jk}^{\mathrm{DE}}\|_{L_\infty(\Omega)}$ for:

$$\|G^{\mathrm{RTE}} - G^{\mathrm{DE}}\|_2 = \sqrt{\sum_{jk} |G_{jk}^{\mathrm{RTE}} - G_{jk}^{\mathrm{DE}}|^2} \le \sqrt{JK} C\epsilon^2 \|w\|_{L_2(\Omega)} \,. \qquad \square$$

4.2. Convergence In Hellinger Metric. The proof of the following result is a straightforward extension of Theorem 3.3 and hence we only sketch it.

THEOREM 4.3. *Consider the linearized setting with $u_0 \in \mathcal{A}_u$ and assume that ν_0 is a centred Guassian measure supported on the space $C^3(\Omega)$. Then the Hellinger distance between the posterior distribution ν_{RTE}^y and ν_{DE}^y is bounded by $\mathcal{O}(\epsilon)$:*

$$d_{Hell}(\nu_{RTE}^y, \nu_{DE}^y) \le \mathcal{O}(\epsilon^2) \,.$$

Sketch Proof. The primary difference of this theorem with Theorem 3.3 is that the data y is subtracted by $\mathcal{G}^{\mathrm{RTE}}(\log(u_0))$ and $\mathcal{G}^{\mathrm{DE}}(\log(u_0))$, and that the linear operator can be made explicit: $G^{\mathrm{RTE,DE}} = \langle \gamma^{\mathrm{RTE,DE}}, w \rangle$. As a result, the estimates for I_1 and I_2 change accordingly. The proof is rather similar to that for Theorem 3.3, so we omit the details and only estimate I_1 here:

$$I_1 \le \frac{c}{Z^{\mathrm{RTE}}} \int_{C^3(\Omega)} \|G^{\mathrm{RTE}}(w) - G^{\mathrm{DE}}(w)\|_2^2 \, \|G^{\mathrm{RTE}}(w) + G^{\mathrm{DE}}(w) - \mathsf{y}_{\mathrm{lin}}^{\mathrm{DE}} - \mathsf{y}_{\mathrm{lin}}^{\mathrm{RTE}}\|_2^2 \, \mathrm{d}\nu_0(dw)$$

$$\le C\epsilon^4 \int_{C^3(\Omega)} \|w\|_{L_2(\Omega)}^2 \big(1 + \|w\|_{L_2(\Omega)}^2\big) \mathrm{d}\nu_0 \,.$$

For the second inequality to hold true, we first use the conclusion from Corollary 4.2, and to bound the second term, we simply use:

$$G^{\mathrm{RTE}}(w) \le \langle \gamma^{\mathrm{DE}}, w \rangle + C\epsilon^2 \|w\|_{L_2(\Omega)} \,, \quad \mathcal{G}^{\mathrm{RTE}}(\log(u_0)) \le \mathcal{G}^{\mathrm{DE}}(\log(u_0)) + C\epsilon \,,$$

to obtain

$$\|G^{\mathrm{RTE}}(w) + G^{\mathrm{DE}}(w) - \mathsf{y}_{\mathrm{lin}}^{\mathrm{DE}} - \mathsf{y}_{\mathrm{lin}}^{\mathrm{RTE}}\|_2 \le \|2\langle \gamma^{\mathrm{DE}}, w \rangle - 2\mathcal{G}^{\mathrm{DE}}(\log(u_0)) - 2\mathsf{y}\|_2 + C\epsilon^2 \|w\|_{L_2(\Omega)} + C\epsilon$$

$$\le 2\|\mathcal{G}^{\mathrm{DE}}(\log(u_0)) + \mathsf{y}\|_2 + 2\|\langle \gamma^{\mathrm{DE}}, w \rangle\|_2 + C\epsilon^2 \|w\|_{L_2(\Omega)} + C\epsilon$$

$$\le C + C\|w\|_{L_2(\Omega)} \,.$$

Application of the Fernique theorem [14] shows that we have $\int_{C^3} \|w\|_{L_2(\Omega)}^p \mathrm{d}\nu_0$ bounded by a constant (independent of ϵ) for any p and that

$$I_1 \le C\epsilon^4 \,.$$

The estimate for I_2 is very similar, and therefore

$$d_{\mathrm{Hell}}(\nu_{\mathrm{RTE}}^{\mathsf{y}}, \nu_{\mathrm{DE}}^{\mathsf{y}})^2 \le I_1 + I_2 = \mathcal{O}(\epsilon^4) \,,$$

which leads to the conclusion of Theorem 4.3. □

Comparing the preceding theorem with Theorem 3.3, a very interesting phenomenon we immediately see is that the convergence in the linearized setting has a higher rate. This higher rate is a direct consequence of the convergence in γ in which the $\mathcal{O}(\epsilon)$ cancels due to the symmetry between the forward model and the adjoint.

4.3. Implications For Posterior Convergence. In the linear setup, if the prior distribution and the likelihood functions are both Gaussian functions, the posterior distribution is also Gaussian, and is thus completely determined by its mean and covariance, or the leading two moments. In our case, $u_0 \in \mathcal{A}_u$, and $w \in C^3$, and the prior is supported in $C^3(\Omega)$ for w. Thus distances between the posterior distributions computed using the RTE and DE can be estimated from distances between means and covariances. To this end, consider the following lemma.

LEMMA 4.4 (Lemma 7.14 from [15]). *Let ν and ν' be two probability measures on a separable Banach space X. Assume also that $f : X \to E$, where $(E, \|\cdot\|)$ is a separable Banach space, is measurable and has second moments with respect to both ν and ν'. Then*

$$\|\mathbb{E}^\nu f - \mathbb{E}^{\nu'} f\| \le 2 \left(\mathbb{E}^\nu \|f\|^2 + \mathbb{E}^{\nu'} \|f\|^2 \right)^{\frac{1}{2}} \mathrm{d}_{Hell}(\nu, \nu').$$

Furthermore, if E is a separable Hilbert space and $f : X \to E$ also has fourth moments, then

$$\|\mathbb{E}^\nu (f \otimes f) - \mathbb{E}^{\nu'} (f \otimes f)\| \le 2 \left(\mathbb{E}^\nu \|f\|^4 + \mathbb{E}^{\nu'} \|f\|^4 \right)^{\frac{1}{2}} \mathrm{d}_{Hell}(\nu, \nu'),$$

where the operator norm on E is employed.

When applied in our case, we obtain:

COROLLARY 4.5. *Let $m_{post}^{RTE,DE}$ and $\mathcal{C}_{post}^{RTE,DE}$ denote the mean function and the covariance operator computed from the posterior distribution of the radiative transfer and diffusion model in the linearized setting. Then*

$$\|m_{post}^{RTE} - m_{post}^{DE}\| \leq \mathcal{O}(\epsilon^2), \quad \|\mathcal{C}_{post}^{RTE} - \mathcal{C}_{post}^{DE}\| \leq \mathcal{O}(\epsilon^2).$$

Here the norm for the mean is the standard norm on $C^3(\Omega)$ and the norm for the covariance is the operator norm on $H^3(\Omega)$.

Proof. Let f as in the statement of Lemma 4.4 be the identity map, and the spaces X and E be $C^3(\Omega)$ equipped with L_2 norm, then $f(w) = w$. Let ν and ν' be the posterior distributions $\nu_{\mathrm{RTE}}^{\mathsf{y}}$ and $\nu_{\mathrm{DE}}^{\mathsf{y}}$ respectively. Then:

$$\|m_{\mathrm{post}}^{\mathrm{RTE}} - m_{\mathrm{post}}^{\mathrm{DE}}\|_{L_2(\Omega)} = \|\mathbb{E}^{\nu_{\mathrm{RTE}}^{\mathsf{y}}} w - \mathbb{E}^{\nu_{\mathrm{DE}}^{\mathsf{y}}} w\|_{L_2(\Omega)} \leq 2 \left(\mathbb{E}^{\nu_{\mathrm{RTE}}^{\mathsf{y}}} \|w\|_{L_2(\Omega)}^2 + \mathbb{E}^{\nu_{\mathrm{DE}}^{\mathsf{y}}} \|w\|_{L_2(\Omega)}^2 \right)^{\frac{1}{2}} \mathrm{d}_{\mathrm{Hell}}(\nu_{\mathrm{RTE}}^{\mathsf{y}}, \nu_{\mathrm{DE}}^{\mathsf{y}}).$$

Since

$$\mathbb{E}^{\nu_{\mathrm{RTE}}^{\mathsf{y}}} \|f\|_{L_2(\Omega)}^2 = \int_{C^3(\Omega)} \|w\|_{L_2(\Omega)}^2 \mathrm{d}\nu_{\mathrm{RTE}}^{\mathsf{y}} \lesssim \int_{C^3(\Omega)} \|w\|_{L_2(\Omega)}^2 \mathrm{d}\nu_0 < C$$

and that

$$\mathbb{E}^{\nu_{\mathrm{DE}}^{\mathsf{y}}} \|f\|^2 = \int_{\mathcal{A}} \|w\|_{L_2(\Omega)}^2 \mathrm{d}\nu_{\mathrm{DE}}^{\mathsf{y}} \lesssim \int_{C^3(\Omega)} \|w\|_{L_2(\Omega)}^2 \mathrm{d}\nu_0 < C,$$

where we have again used $\int_{\mathcal{C}^3(\Omega)} \|w\|_{L_2(\Omega)}^2 \mathrm{d}\nu_{\mathrm{RTE}}^{\mathsf{y}} < C$ using the Fernique theorem, and that $\mathrm{d}\nu_{\mathrm{RTE,DE}}^{\mathsf{y}} \lesssim \mathrm{d}\nu_0$ [15], we achieve the convergence of the mean function by applying Theorem 4.3. The same analysis is applied to analyze the covariance. □

5. Conclusion. In this paper, we study the inverse problem of diffuse optical tomography to reconstruct the scattering coefficient. Partial and noisy data is assumed, and hence a Bayesian formulation of inversion is natural. Two forward models are described, one employing the radiative transfer equation and the other employing the diffusion equation respectively. Multiscale analysis demonstrates that solutions of the two forward models are close in the optically thick (small Knudsen number) regime, and this allows us to quantify the convergence of the two Bayesian inverse problems. In particular, we show that $\mu_{\mathrm{RTE}}^{\mathsf{y}}$ and $\mu_{\mathrm{DE}}^{\mathsf{y}}$, the two posterior distribution functions, are $\mathcal{O}(\epsilon)$ away from each other in both Kullback-Leibler divergence sense, and the Hellinger sense, for both nonlinear and linear cases. Forward solution of the diffusion equation is computationally less burdensome than the radiative transfer equation, and the theory justifies using it within Bayesian inversion algorithms where multiple forward model evaluations are required. We have employed a setting in which compatible boundary conditions are used for the two forward models. It would also be of interest to study extensions of this. However the primary stumbling block here is the analysis of the forward problem itself. The approach developed in this paper will apply to other Bayesian inverse problems whose forward problems are close, and can be used to justify inversion algorithms which employ an averaged or (as in this case) homogenized approximate forward model, in order to speed-up computation.

REFERENCES

[1] G. Alessandrini. Stable determination of conductivity by boundary measurements. *Appl. Anal.*, 27(1-3):153–172, 1988.
[2] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93, 1999.
[3] G. Bal. Inverse transport theory and applications. *Inverse Problems*, 25(5):053001, 48, 2009.
[4] G. Bal and A. Jollivet. Time-dependent angularly averaged inverse transport. *Inverse Problems*, 25(7):075010, 32, 2009.
[5] G. Bal, I Langmore, and F. Monard. Inverse transport with isotropic sources and angularly averaged measurements. *Inverse Probl. Imaging*, 2(1):23–42, 2008.
[6] C. Bardos, R. Santos, and R Sentis. Diffusion approximation and computation of the critical size. *Trans. Amer. Math. Soc.*, 284(2):617–649, 1984.
[7] A. Bensoussan, L. Lions, J, and G.C. Papanicolaou. Boundary layers and homogenization of transport processes. *Publ. Res. Inst. Math. Sci.*, 15(1):53–157, 1979.
[8] L. Borcea. Electrical impedance tomography. *Inverse problems*, 18(6):R99, 2002.

[9] T. Bui-Thanh, C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L.C. Wilcox. Extreme-scale uq for bayesian inverse problems governed by pdes. *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*, 2012.

[10] K. Chen, Q. Li, and L. Wang. Stability of stationary inverse transport equation in diffusion scaling. *Inverse Problems*, 34(2):025004, 26, 2018.

[11] M. Choulli and P. Stefanov. Inverse scattering and inverse boundary value problems for the linear Boltzmann equation. *Comm. Partial Differential Equations*, 21(5-6):763–785, 1996.

[12] S.L. Cotter, M. Dashti, and A.M. Stuart. Approximation of bayesian inverse problems for pdes. *SIAM Journal on Numerical Analysis*, 48(1):322–345, 2010.

[13] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. Mcmc methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pages 424–446, 2013.

[14] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152. Cambridge university press, 2014.

[15] M. Dashti and A.M. Stuart. The Bayesian approach to inverse problems. *Handbook of Uncertainty Quantification*, 2017.

[16] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1983.

[17] A.H. Hielscher, R. Alcouffe, and R Barbour. Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. *Phys. Med. Biol.*, 43(5):1285–1302, 1998.

[18] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[19] A.D. Klose, U. Netz, J. Beuthan, and A.H. Hielscher. Optical tomography using the time-independent equation of radiative transfer - part 1: forward model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 72, 2002.

[20] A.D. Klose, U. Netz, J. Beuthan, and A.H. Hielscher. Optical tomography using the time-independent equation of radiative transfer - part 2: inverse model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 72, 2002.

[21] R.-Y. Lai, Q. Li, and G. Uhlmann. Inverse problems for the stationary transport equation in the diffusion scaling. *arXiv: 1808.02071*, 2018.

[22] S. Lasanen. Non-gaussian statistical inverse problems. part i: Posterior distributions. *Inverse Problems & Imaging*, 6(2):215–266, 2012.

[23] S. Lasanen. Non-gaussian statistical inverse problems. part ii: Posterior convergence for approximated unknowns. *Inverse Problems & Imaging*, 6(2), 2012.

[24] M.S. Lehtinen, L. Paivarinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599, 1989.

[25] Q. Li, J. Lu, and W. Sun. Validity and regularization of classical half-space equations. *Journal of Statistical Physics*, 166(2):398–433, 2017.

[26] Y. Marzouk and D. Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 6:826–847, 2009.

[27] J.R. Mourant, J. Freyer, A.H. Hielscher, A. Eick, D. Shen, and T. Johnson. Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. *Appl. Opt.*, 37(16):3586–3593, 1998.

[28] F.J. Pinski, G. Simpson, A.M. Stuart, and H. Weber. Algorithms for Kullback-Leibler approximation of probability measures in infinite dimensions. *SIAM J. Sci. Comput.*, 37(6):A2733–A2757, 2015.

[29] K. Ren. Recent developments in numerical techniques for transport-based medical imaging methods. *Commun. Comput. Phys.*, 8(1):1–50, 2010.

[30] D. Sanz-Alonso, A.M. Stuart, and A. Taeb. Inverse problems and data assimilation. *arXiv preprint arXiv:1810.06191*, 2018.

[31] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[32] J. Sylvester and G. Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Ann. of Math. (2)*, 125(1):153–169, 1987.

[33] J.N. Wang. Stability estimates of an inverse problem for the stationary transport equation. *Ann. Inst. H. Poincaré Phys. Théor.*, 70(5):473–495, 1999.

[34] L. Wu and Y. Guo. Geometric correction for diffusive expansion of steady neutron transport equation. *Communications in Mathematical Physics*, 336(3):1473–1553, 2015.

6. Appendix. We give the rigorous proof here for Theorem 2.2. The two statements are proved in the same way, and for generality, we will only prove the second one, and the proof for the first statement, or even for higher order expansions, are easy extensions.

Proof. We repeat the RTE with a specially designed boundary condition,

$$\begin{cases} v \cdot \nabla f = \frac{\sigma}{\epsilon} \mathcal{L} f \\ f|_{\Gamma_-} = \xi(x) - \frac{\epsilon}{\sigma} v \cdot \nabla \rho(x) \end{cases},$$

where ρ satisfies

$$\nabla \cdot (\sigma^{-1} \nabla \rho) = 0 \quad x \in \Omega, \quad \text{with} \quad \rho|_{\partial\Omega} = \xi(x).$$

Now we decompose the solution to the RTE as

$$f = f_0 + \epsilon f_1 + \epsilon^2 f_2 + f_r \,,$$

where $f_0 = \rho$, $f_1 = -\frac{1}{\sigma} v \cdot \nabla \rho$, and $f_2 = \frac{1}{\sigma} \mathcal{L}^{-1} \left[(v \cdot \nabla) \frac{1}{\sigma} (v \cdot \nabla) \rho \right]$. Note that for the definition of f_2 to hold true, it is necessary that

$$(v \cdot \nabla) \frac{1}{\sigma} (v \cdot \nabla) \rho \in \text{Range} \, \mathcal{L} \,,$$

which in turn requires

$$\langle (v \cdot \nabla) \frac{1}{\sigma} (v \cdot \nabla) \rho \rangle_v = C_d \nabla \cdot (\frac{1}{\sigma} \nabla \rho) = 0 \,.$$

Since ρ is $\frac{1}{\sigma}$-harmonic with smooth boundary condition $\|\xi\|_{L_\infty(\partial\Omega)} < C_1$, then by the maximum principle [16],

$$\|\rho\|_{L_\infty(\Omega)} < C_1 \,, \quad \text{and} \quad \|\partial_i \rho\|_{L_\infty(\Omega)} < C_2 \quad \text{and} \quad \|\partial_{ij} \rho\|_{L_\infty(\Omega)} < C_3 \,.$$

Then since $\mathcal{L}^{-1}$ is a bounded operator on Null $\mathcal{L}^\perp$, we know that both f_1 and f_2 are bounded, meaning there is a constant C_4 that depends on C_1, C_2 and C_3:

$$\|f_1\|_\infty = \|\sigma^{-1} v \cdot \nabla \rho\|_{L_\infty(\Omega)} = \|\sigma^{-1}\|_{L_\infty(\Omega)} \|\partial_i \rho\|_{L_\infty(\Omega)} < C_4 \,,$$

and

$$\|f_2\|_\infty = \|\sigma^{-1} \mathcal{L}^{-1} \left[(v \cdot \nabla) \sigma^{-1} (v \cdot \nabla) \rho \right] \|_{L_\infty(\Omega)} \leq \|\sigma^{-1} (v \cdot \nabla) \sigma^{-1} (v \cdot \nabla) \rho\|_{L_\infty(\Omega)} < C_4 \,,$$

where we used the boundedness of $\|\sigma^{-1}\|_{L_\infty(\Omega)} < C_1$, $\|\nabla \left(\frac{1}{\sigma} \right)\|_{L_\infty(\Omega)} < C_1$, and the boundedness of the harmonic function and its derivatives.

Plugging it back into the equation, we have

$$v \cdot \nabla \left(\rho - \epsilon \frac{1}{\sigma} v \cdot \nabla \rho + \epsilon^2 f_2 + f_r \right) = \frac{\sigma}{\epsilon} \mathcal{L} \left(\rho - \epsilon \frac{1}{\sigma} v \cdot \nabla \rho + \frac{\epsilon^2}{\sigma} \mathcal{L}^{-1} \left[(v \cdot \nabla) \frac{1}{\sigma} (v \cdot \nabla) \rho \right] + f_r \right) \,.$$

Since ρ is a constant in v, and is thus in Null $\mathcal{L}$, then $\mathcal{L} \rho = 0$. Using the definition of $\mathcal{L}$, we also have $\mathcal{L}(v \cdot \nabla \rho) = -v \cdot \nabla \rho$. Now we cancel the terms and obtain the following equation for f_r,

$$v \cdot \nabla f_r = \frac{\sigma}{\epsilon} \mathcal{L} f_r - \epsilon^2 v \cdot \nabla f_2 \,.$$

It is immediate that f_r satisfies RTE and is equipped with a source term of $\mathcal{O}(\epsilon^2)$. The boundary condition for f_r is of the same order,

$$f_r|_{\Gamma_-} = -\epsilon^2 f_2|_{\Gamma_-} \,.$$

By maximum principle, the solution is bounded in L_∞ by the boundary condition and the source term, so we have

$$\|f_r\|_{L_\infty(\Omega)} \leq C_5 \|\epsilon^2 v \cdot \nabla f_2\|_{L_\infty(\Omega)} + \|\epsilon^2 f_2\|_{L_\infty(\Gamma)} = \mathcal{O}(\epsilon^2) \,,$$

where $C_5 \leq e^{C_1 l}$ and l is the longest radius of the domain. This leads to the fact that:

$$\|f - \rho + \epsilon v \cdot \nabla \rho\|_{L_\infty(\Omega)} = \|\epsilon^2 f_2 + f_r\|_{L_\infty(\Omega)} \mathcal{O}(\epsilon^2) \,.$$

We note again the constant merely depends on the boundedness of C_1, the upper bound of the infinite norm of $\frac{1}{\sigma}$, its derivative and ξ. $\qquad \square$