

Article

# Diffusion Equation-Assisted Markov Chain Monte Carlo Methods for the Inverse Radiative Transfer Equation

Qin Li \*  and Kit Newton

Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53705, USA;  
k.cole.newton@gmail.com

\* Correspondence: qinli@math.wisc.edu

Received: 18 December 2018; Accepted: 8 March 2019; Published: 18 March 2019



**Abstract:** Optical tomography is the process of reconstructing the optical properties of biological tissue using measurements of incoming and outgoing light intensity at the tissue boundary. Mathematically, light propagation is modeled by the radiative transfer equation (RTE), and optical tomography amounts to reconstructing the scattering coefficient in the RTE using the boundary measurements. In the strong scattering regime, the RTE is asymptotically equivalent to the diffusion equation (DE), and the inverse problem becomes reconstructing the diffusion coefficient using Dirichlet and Neumann data on the boundary. We study this problem in the Bayesian framework, meaning that we examine the posterior distribution of the scattering coefficient after the measurements have been taken. However, sampling from this distribution is computationally expensive, since to evaluate each Markov Chain Monte Carlo (MCMC) sample, one needs to run the RTE solvers multiple times. We therefore propose the DE-assisted two-level MCMC technique, in which bad samples are filtered out using DE solvers that are significantly cheaper than RTE solvers. This allows us to make sampling from the RTE posterior distribution computationally feasible.

**Keywords:** multi-level MCMC; radiative transfer equation; inverse problems; diffusion limit

## 1. Introduction

Optical tomography is a medical imaging technique in which near infrared light is sent into biological tissue, and the reflected and transmitted outgoing light at the surface of the tissue is measured [1–4]. Using information about the incoming and outgoing light, one can determine the properties of the tissue. The use of low-energy light makes optical tomography cheaper and less invasive than traditional methods such as X-ray imaging, although the reconstruction of the properties of the tissue can be more difficult than when high-energy photons are used. Optical imaging can be used to study and monitor many different kinds of tissue, including brain, breast, and joint imaging, as well as monitoring blood oxygenation [5–7].

The process of collecting information about the incoming and outgoing light and using it to reconstruct the tissue's properties is an inverse problem. There are two associated forward models that both map the incoming data (light intensity sent into the tissue) to the outgoing data (light intensity measured coming off the tissue). The first forward model corresponds to the radiative transfer equation (RTE) and is called the albedo operator. Using the albedo operator, the inverse problem may be solved to obtain the scattering coefficient in the RTE. The second forward model corresponds to the diffusion equation (DE) and is called the Dirichlet-to-Neumann (DtN) map. Using the DtN map, one may similarly solve the inverse problem and reconstruct the diffusion coefficient in the DE. Typically, RTE is used for high-energy photons and DE is used for low-energy photons. The energy of

photons determines the “mean-free-path” of free transport, and then further determines the strength of scattering. Mathematically, it can be characterized by the Knudsen number, and one can show that in the small Knudsen number regime, the two equations are equivalent, as will be made clearer in Section 3. It was shown in [8] that the coefficients of the RTE are uniquely recoverable in 3D when the entire albedo operator is known, and it was shown in [9] that this reconstruction is Lipschitz stable. On the other hand, using the DtN map to reconstruct the corresponding coefficients of the DE is the famously ill-posed Calderón problem. The uniqueness of the reconstruction has been presented in [10], and the logarithmically ill-posed nature of the problem has been proved in [11].

These two forward models are related, as we will describe. The RTE describes light propagating through a material with some optical properties, here taken to be the scattering coefficient of the material. Let  $f(x, v)$  denote the distribution of particles at location  $x$  with direction  $v$ , where  $x \in \Omega \subset \mathbb{R}^d$ , and  $v \in \mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$ , where  $d$  is the dimension of the problem. In other words, all particles are taken to move with constant unit speed. In the later parts of the paper, we use “velocity” and “direction” interchangeably when no ambiguity is present. For simplicity, we work with a version of the RTE in which the scattering coefficient  $\sigma(x)$  depends only on position. Then the RTE is

$$v \cdot \nabla f(x, v) = \sigma(x) \mathcal{L}f(x, v),$$

where  $\mathcal{L}$  is an integral operator describing photons colliding with the media and scattering off. Its explicit formulation is in Section 3.

The diffusion equation, on the other hand, is a simplified model from the RTE, accurate for low-energy photons in the high-scattering, low-absorption regime. Since the scattering is very strong, the distribution achieves equilibrium in the velocity domain, and the light intensity, also known as fluence, becomes a function only on physical space. Suppose that  $\rho(x)$  is the light intensity at position  $x$ , and  $a(x)$  is the diffusion coefficient (corresponding to  $\frac{1}{\sigma(x)}$  from the RTE, as will be shown in Theorem 2). Then the DE is

$$C \nabla \cdot (a(x) \nabla \rho(x)) = 0,$$

where  $C$  is a constant depending on dimension. In this case, the map from the Dirichlet data (light intensity or fluence injected into the tissue) to the Neumann data (light propagating out) is used to reconstruct the diffusion coefficient  $a(x)$ . This map is known as the Dirichlet-to-Neumann (DtN) map.

There is a relationship between the two forward models RTE and DE. One would like to understand, physically and mathematically, why the stability of each model is different in the inverse setting. It turns out that if  $a(x)$  and  $\sigma(x)$  satisfy certain relations, the RTE and DE are asymptotically close in the near infrared case. This is made precise below in Theorem 2. Physically, in the forward setting, high-energy photons experience little scattering before exiting the domain, whereas low-energy photons scatter frequently in the tissue before they exit. As a result, the reconstruction of the tissue using high-energy photons is generally crisp, whereas low-energy photons provide more blurred images. We use the “Knudsen number”  $\epsilon$  to quantify how much scattering a photon experiences in the material. In the low-energy regime, scattering increases, the Knudsen number shrinks to zero, and the RTE converges to the DE in the forward setting. On the inverse side, the inverse problem for the RTE converges to the inverse problem for the DE, meaning that the information carried in the albedo operator is almost the same as that in the DtN map, and therefore the reconstruction of the tissue properties converges in this limit as well. This has been numerically observed in [12,13], and further proved rigorously in [14,15].

A Bayesian solution to the inverse problem is seen as the posterior distribution for the quantity of interest (QoI), in our case the scattering coefficient in the RTE and the diffusion coefficient in the DE. Bayesian methods are particularly useful for inverse problems, because noise in the measurement as well as prior information about the QoI is taken into account naturally [16]. If we observe some noisy

data in order to obtain information about the QoI, the solution of the inverse problem is the probability distribution of the QoI given this data, known as the posterior distribution [17]. Bayes' theorem allows us to determine this distribution given a guess for the probability distribution of the QoI (the prior distribution) and the probability distribution for the data given the QoI (the likelihood function, obtainable from the forward model). As such, there is a sharp distinction between this probabilistic view and other deterministic tools, and the Bayesian formulation gives one the ability to “regularize” an inverse problem with some prior knowledge.

While we theoretically have the posterior distribution in hand, accessing concrete information about it such as its mean and variance can be challenging. One way to obtain this information is by creating a list of samples from the distribution by some means, such as the Markov chain Monte Carlo (MCMC) method. Even this, however, is not always straightforward. One sample from the RTE represents an entirely new configuration of the media, meaning that one must re-compute the forward albedo map for each MCMC sample. This is especially expensive because the RTE is over phase space: if the domain is three-dimensional, then  $f$  is supported on  $(x, v) \in \Omega \times \mathbb{S}^{d-1}$ , a five-dimensional space. The same process applies for the DE as well, but since the DE is only over physical space, these computations are much faster. Knowing that the RTE converges to the DE in the high scattering regime, one wishes to combine the two and speed up the computation.

There is a balance to be struck here, and the goal of this paper is to find a way to sample from the posterior distribution for  $\sigma$  in a way that combines the DE and RTE. To that end, we employ the two-level MCMC method [18], also known as the two-stage method. The two-level method uses two models, which we will call the low-resolution model and the high-resolution model. The high-resolution model gives rise to the desired target distribution—for us, the posterior distribution for the RTE. The low-resolution model gives rise to a distribution that approximates the target distribution, and is ideally fast to compute. It is used to filter out poor draws so we need not waste time evaluating the target distribution on them. We know the DE is fast to compute, and that when scattering is sufficiently strong, the posterior based on the DE is a good approximation to the posterior based on the RTE, so we use the posterior based on the DE as the low-resolution model. Then, we can use the DE to reduce the number of times we must solve the RTE by rejecting bad draws. This method combines the inverse problem for the DE and the inverse problem for the RTE to create a faster method for sampling from the inverse problem for the RTE in the diffusion limit.

There are similar methods to approach related problems, such as the multilevel Monte Carlo method for path simulation [19] or the multilevel Monte Carlo method for parametric integration [20]. These methods also combine low-resolution models with a high-resolution model. The algorithm our method relies on, introduced in [18], is similar to the methods proposed in [21,22]. In this view, our algorithm is a two-stage method that increases the number of MCMC iterations for a given computational cost. It uses a two-stage delayed acceptance, in which the candidate sample  $y$  has to be accepted with a low-resolution model before it is passed on to be either accepted or rejected by the high-resolution model. However, typically for these methods, a coarse-grid approximation is used for the low-resolution model [23]. No matter how coarse the grids are for the RTE, however, they are still on phase space and the dimensionality issue is still left open. In our method, however, a completely different model is used: the inverse diffusion equation is used as a low-resolution model for the inverse radiative transfer equation.

This paper is organized as follows. In Section 2, we provide some necessary background, including a discussion of Bayesian methods for inverse problems, as well as an introduction to Markov chain Monte Carlo methods. In Section 3, we discuss the diffusion limit of the radiative transfer equation in the forward setting, and go on to discuss the inverse setting convergence of the posterior distributions. In Section 4, we discuss the DE-assisted two-level MCMC method, and prove its convergence. We also discuss the dependence of the second-level acceptance rate on  $\epsilon$ , the Knudsen number. In Section 5, we present our numerical evidence. In particular, we show the convergence of the two forward models, the convergence of the posterior distribution functions

quantified using the Hellinger distance, and the improvement of the DE-assisted two-level MCMC over the standard MCMC.

## 2. Background

In this section, we present preliminaries for studying the inverse problem for the RTE in the diffusion limit. In particular, we will first review basic concepts of the Bayesian formulation, and then review the previous algorithms on which our algorithm is based.

These results will be used to study the convergence of the posterior distribution with the RTE or the DE as the forward model, which will be used to develop our numerical method that combines the two posterior distributions.

### 2.1. Bayesian Formulation for Inverse Problems

Bayesian inference is a technique for estimating the distribution of some quantity of interest when some measurements are available. It is based on Bayes' theorem. A given physical problem may be denoted as

$$b = \mathcal{G}(\sigma) + \eta, \quad (1)$$

where  $\mathcal{G}$  is the forward map that takes a parameter  $\sigma$  to the measurement  $b$ , with an added noise  $\eta$ . The forward problem is to find  $b = \mathcal{G}(\sigma)$  for a given  $\sigma$ , while most problems in practice are naturally inverse problems, meaning that one conducts experiments and obtains data  $b$ , and tries to use it to reconstruct the quantity of interest  $\sigma$ .

The Bayesian formulation is a method for retrieving information about  $\sigma$  using  $b$ . It requires knowledge of the following two probability distribution functions  $\mu_0$  and  $\mu_{\text{error}}$  ahead of time,

- without knowledge of the measurement  $b$ , a priori  $\sigma$  obeys a certain law,

$$\sigma \sim \mu_0(\sigma),$$

- and the noise is distributed as:

$$\eta = b - \mathcal{G}(\sigma) \sim \mu_{\text{error}}(\eta).$$

In many cases, both distribution functions can be assumed to be normal, i.e., to have a Gaussian-type density function. Suppose that  $\mu_0$  is a Gaussian distribution concentrated at  $m_0$  with covariance  $\mathcal{C}_0$  and the noise is concentrated at zero with covariance  $\mathcal{C}_{\text{error}}$ . If  $\sigma$  is finite dimensional then we may express the prior  $\mu_0$  and likelihood  $\mu^\sigma$  as follows:

$$\begin{cases} \mu_0(\sigma) = \mathcal{N}(m_0, \mathcal{C}_0) \propto \exp\left(-\frac{1}{2}(\sigma - m_0)^\top \mathcal{C}_0^{-1}(\sigma - m_0)\right) \\ \mu^\sigma(b) = \mu_{\text{error}}(\eta) \propto \exp\left(-\frac{1}{2}(b - \mathcal{G}(\sigma))^\top \mathcal{C}_{\text{error}}^{-1}(b - \mathcal{G}(\sigma))\right) \end{cases}.$$

Analogous formulae are also available in the infinite dimensional setting; see [16,17] and the references therein.

With this, the posterior distribution of  $\sigma$ , under the condition that  $b$  is obtained in experiment, is given by Bayes' theorem,

$$\begin{aligned} \mu^b(\sigma) &= \frac{1}{Z} \mu_0(\sigma) \mu^\sigma(b) \\ &= \frac{1}{Z} \exp\left(-\frac{1}{2}(\sigma - m_0)^\top \mathcal{C}_{\text{prior}}^{-1}(\sigma - m_0) - (b - \mathcal{G}(\sigma))^\top \mathcal{C}_{\text{error}}^{-1}(b - \mathcal{G}(\sigma))\right), \end{aligned} \quad (2)$$

where

$$Z = \int \mu^\sigma(y) d\mu_0(\sigma)$$

is the normalization factor.

In optical tomography, there are two fundamental models for describing light propagation: the radiative transfer equation (RTE), and the diffusion equation (DE). They rely on the albedo operator and the DtN map for the reconstruction. In the optically thick case (i.e., when photons scatter frequently), the two models are asymptotically close in some sense,

$$\mathcal{G}_{\text{RTE}}(\sigma) \approx \mathcal{G}_{\text{DE}}(\sigma).$$

Correspondingly, the posterior distributions given by the two models are close to each other [24], meaning:

$$\mu_{\text{DE}}^b(\sigma) \approx \mu_{\text{RTE}}^b(\sigma).$$

To be more precise, there are multiple ways to quantify the distance between two distribution functions. Typically, this distance is quantified by the Kullback–Leibler (KL) divergence or the Hellinger distance [25]. For any two distributions  $\mu$  and  $\mu'$  supported on the same space, they are defined by the following,

- KL divergence

$$d_{\text{KL}}(\mu, \mu') = \int \log \frac{\mu}{\mu'} d\mu,$$

- Hellinger distance

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

where  $\lambda$  is any pre-chosen distribution function. The Hellinger distance is invariant to the choice of  $\lambda$ . Again, these formulae have interpretations in the infinite dimensional setting; see the Appendix of [17]. In Theorem 3, we quantify the similarity between  $\mu_{\text{DE}}^b(\sigma)$  and  $\mu_{\text{RTE}}^b(\sigma)$  and find that the two distributions are  $\epsilon$  apart in the diffusion limit. Therefore, since  $\mu_{\text{DE}}^b$  is close to  $\mu_{\text{RTE}}^b$  in the optically thick case, the main task in this paper is to use  $\mu_{\text{DE}}^b$  to sample from  $\mu_{\text{RTE}}^b$ .

### 2.2. The Markov Chain Monte Carlo Method

We first present the standard Metropolis–Hastings (MH) algorithm. Given a probability density  $\mu$  called the target distribution and defined on  $X$ , the MH algorithm constructs a Markov chain on  $X$  that is stationary with respect to  $\mu$ . The elements of the Markov chain are then regarded as samples from the distribution  $\mu$ . More specifically, the MH algorithm starts with an initial guess  $x_0$ , and draws new samples according to a proposal distribution  $q$ . By adjusting the acceptance rate using the target distribution  $\mu$ , the MH algorithm accepts or rejects the draws so that the accepted samples form an empirical distribution that resembles the target distribution  $\mu$ . We now present the MH algorithm, shown in Algorithm 1.

---

#### Algorithm 1 Metropolis Hastings

---

1. Given  $x_k$ , draw  $y \sim q(\cdot, x_k)$ .
2. Let

$$\alpha(x_k, y) := \min \left\{ 1, \frac{q(y, x_k)\mu(y)}{q(x_k, y)\mu(x_k)} \right\}.$$

3. With probability  $\alpha$ , **accept**  $y$  and set  $x_{k+1} = y$ . Otherwise, set  $x_{k+1} = x_k$ .
-

The transition kernel  $P$  of this algorithm is

$$P(x_k, A) = P\{x_{k+1} \in A | x_0, \dots, x_k\}.$$

In order to demonstrate the convergence of MCMC in Theorem 1, it is necessary to examine the transition kernel, which is the probability that the next draw  $x_{k+1}$  is in the set  $A$  given the previous elements of the chain  $x_0, \dots, x_k$ . It may be written

$$P(x, y) = p(x, y) + \delta_y(x)r(x).$$

where the off-diagonal density of the kernel is

$$p(x, y) = \begin{cases} q(x, y)\alpha(x, y) & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}, \quad (3)$$

and the probability that the process remains at  $x$  is

$$r(x) = 1 - \int p(x, y)dy.$$

The goal is to show that the target distribution  $\mu$  is an invariant distribution of the Markov chain  $\{x_k\}_{k \geq 0}$  in the sense that all measurable sets  $A$  satisfy

$$\mu(A) = \int \mu(x)P(x, A)dx. \quad (4)$$

This means that elements of the Markov chain generated by Algorithm 1 give a good representation of the distribution  $\mu$ . To show that  $\mu$  is the invariant distribution, one first needs  $p(x, y)$  to satisfy the detailed balance lemma.

**Lemma 1.** *The off-diagonal density  $p(x, y)$  satisfies the following equation known as “detailed balance”,*

$$\mu(x)p(x, y) = \mu(y)p(y, x). \quad (5)$$

This lemma allows us to show the following theorem.

**Theorem 1.** *The target distribution  $\mu$  is an invariant distribution of the Markov chain  $x_n$  with transition kernel  $P$ , i.e.,*

$$\mu(A) = \int P(x, A)\mu(x)dx. \quad (6)$$

For our problem, each new proposal  $y$  is a new configuration of the media  $\sigma(x)$ . Thus, to compute  $\alpha(x_k, y)$ , one must evaluate  $\mu(y)$  and thus re-compute the forward map, which requires solving the RTE many times. As this is computationally prohibitive, we instead use the two-level MCMC method, described in the next section. The cost comparison is discussed further in Section 5.

### 2.3. Two-Level MCMC Method

The two-level MCMC method is a method to increase the efficiency of sampling from the target distribution  $\mu$ . It requires two distributions: a target distribution  $\mu$ , and a second distribution  $\mu'$ . Here, the target distribution  $\mu$  calls for the evaluation of a “high-resolution” model, but if it can be approximated in some sense by  $\mu'$  which only calls for the evaluation of some “low-resolution” model, then  $\mu'$  can serve as a good filter to reject poor draws in the MCMC sampling. More specifically, the algorithm goes through two levels of evaluating a proposed sample. First, the sample is evaluated using the low-resolution model and is either accepted or rejected. If it is accepted, the algorithm goes



on to evaluate the sample using the high-resolution model. This “pre-acceptance” stage filters out poor draws, allowing one to make more courageous proposals and not waste time evaluating the forward model on them. We present the two-level scheme below in Algorithm 2.

---

**Algorithm 2** Two-level Metropolis Hastings
 

---

1. Given  $x_k$ , draw  $y \sim q(\cdot, x_k)$
2. Let

$$\alpha(x_k, y) := \min \left\{ 1, \frac{q(y, x_k)\mu'(y)}{q(x_k, y)\mu'(x_k)} \right\}.$$

3. With probability  $\alpha$ , **pre-accept**  $y$  and continue to 4. Otherwise, set  $x_{k+1} = x_k$  and start over.
4. The second-level proposal is now  $y$ , effectively drawn from

$$q_2(y, x_k) = \alpha(x_k, y)q(y, x_k) + \delta_{x_k}(y) \left( 1 - \int \alpha(x_k, y)q(y, x_k)dy \right).$$

Then set

$$\beta(x_k, y) := \min \left\{ 1, \frac{q_2(y, x_k)\mu(y)}{q_2(x_k, y)\mu(x_k)} \right\}.$$

With probability  $\beta(x_k, y)$ , **accept** and set  $x_{k+1} = y$ . Otherwise, set  $x_{k+1} = x_k$ .

---

Similarly to the MCMC method (Algorithm 1), in the two-level MCMC method, the draw of  $x_{k+1}$  merely depends on the evaluation of  $x_k$ , and the previous draws  $x_1, \dots, x_{k-1}$  are irrelevant. The transition kernel that brings  $x_k$  to  $x_{k+1}$  is denoted by  $P_2$ , which we will discuss in more detail in Section 4.1.

In this paper, the desired high-resolution model will be the posterior distribution for the RTE. The low-resolution model will be the posterior distribution for the DE. Evaluating  $\mu'(y)$  only involves computing the DE, which is much faster since the DE is supported on the physical domain. Once the DE accepts the proposal  $y$ , it passes to the inverse problem for the RTE's posterior distribution  $\mu(y)$ . Thus, we compute the RTE forward model fewer times overall, which saves time.

In Section 4, we discuss the convergence of this method and the dependence of the second-level acceptance rate  $\beta$  on the Knudsen number  $\epsilon$ , our limit parameter.

### 3. Diffusion Limit

In this section, we examine the diffusion limit of our problem in the forward and inverse case. We first study the convergence of the radiative transfer equation to the diffusion equation. Then, we explain how the inverse problems are solved using the forward map. Next, we discuss the convergence of the forward map for the radiative transfer equation to that of the diffusion equation. Finally, we discuss the convergence of the posterior distribution for the radiative transfer equation to the posterior for the diffusion equation.

#### 3.1. Diffusion Limit of the Radiative Transfer Equation

The optical “thickness” of the material physically corresponds to the number of times a photon scatters between entering a medium and escaping. The physical quantity is termed the Knudsen number, which stands for the ratio of mean free path to the domain length. The mean free path is the average distance a photon 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 this regime, the two mathematical models for light propagation carry the same information, namely, the radiative transfer equation and the diffusion equation are asymptotically converging.

The radiative transfer equation takes a statistical mechanics viewpoint, and describes the distribution of photons on the phase domain. Let  $f(x, v)$  denote the number of photons at position

$x \in \Omega \subset \mathbb{R}^d$ , a bounded domain, moving in direction  $v \in \mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$  (i.e., the speed is normalized to be 1). This distribution satisfies the RTE,

$$v \cdot \nabla f(x, v) = \sigma(x)\mathcal{L}f(x, v), \tag{7}$$

where the collision operator is

$$\mathcal{L}f = \int f(x, v')dv' - f = \langle f \rangle_v - f. \tag{8}$$

In the equation, the term  $v \cdot \nabla f$  on the left shows that the photons move with direction  $v$ , and the term on the right shows that the photons colliding with the media and being scattered. We have used the notation  $\langle \cdot \rangle_v$  to denote normalized integration over  $v$ , and  $dv$  is the normalized unit measure, meaning

$$\int_{\mathbb{S}^{d-1}} 1dv = 1.$$

The equation has a unique solution when it is equipped with the incoming boundary condition, which is the analogue of a Dirichlet boundary condition for equations lacking velocity space. Let

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

denote the collection of coordinates on the boundary  $x \in \partial\Omega$ , so that the velocity  $v$  points in/out of the domain:  $\pm v \cdot n_x > 0$ . Here,  $n_x$  is the normal vector at  $x$  pointing out of  $\Omega$ . The incoming boundary condition is imposed on  $\Gamma_-$ , whereas  $\Gamma_+$  represents the particles going out of  $\Omega$ . For a unique solution to (7), boundary conditions must be imposed on  $\Gamma_-$ ,

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

**Remark 1.** In fact, a more general model of the radiative transfer equation is

$$v \cdot \nabla_x f(x, v) = \int k(x, v, v')f(x, v')dv' - \sigma_a(x, v)f(x, v).$$

It concerns the case when the scattering coefficient, now seen as  $k(x, v, v')$ , may depend on the changing velocity of the particles during a collision, and also includes an absorption coefficient  $\sigma_a(x, v)$  representing the photons being absorbed into the material and lost. A standard  $k$  which takes into account other kinds of scattering is given by the Henyey–Greenstein model. The absorption coefficient can be taken to be zero if scattering is sufficiently high, and this is the case we focus on.

The equation is asymptotically equivalent to the diffusion equation in the optically thick regime when the Knudsen number is small. We denote the Knudsen number by  $\epsilon$  and rescale the problem by setting  $\sigma \rightarrow \sigma/\epsilon$ . Then, as the Knudsen number becomes small, the scattering effect dominates. Equation (7) may then be written as

$$\begin{cases} v \cdot \nabla f = \frac{1}{\epsilon}\sigma\mathcal{L}f, & (x, v) \in \Omega \times \mathbb{S}^{d-1} \\ f|_{\Gamma_-} = \phi(x, v) \end{cases}. \tag{9}$$

In the small  $\epsilon$  regime, it was conjectured in [26] and then proved in [27,28] that the equation is asymptotically equivalent to the diffusion equation. One can make the convergence explicit under the following assumptions.



**Assumption 1.** Both the media and the boundary conditions are bounded.

- the admissible media is bounded, meaning that there is a constant  $C_*$  so that:

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

- and the boundary conditions are bounded, meaning:

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

We also term the set of admissible media:

$$\mathcal{A} = \{\sigma \in W_{1,\infty}(\Omega) : \max\{\|\sigma\|_{L_\infty(\Omega)}, \|\sigma^{-1}\|_{L_\infty(\Omega)}, \|\nabla(\sigma^{-1})\|_{L_\infty(\Omega)}\} < C_*\}. \tag{10}$$

With the assumption, we have the following theorem.

**Theorem 2.** Suppose that  $f(x, v)$  satisfies Equation (9), then as  $\epsilon \rightarrow 0$ ,  $f(x, v) \rightarrow \rho(x)$ , which satisfies

$$\begin{cases} C_d \nabla \cdot \left(\frac{1}{\sigma} \nabla \rho\right) = 0, & x \in \Omega \subset \mathbb{R}^d \\ \rho|_{\partial\Omega} = \xi(x) \end{cases}, \tag{11}$$

where  $\xi(x)$  is defined by  $\phi(x)$ .  $C_d$  is a constant depending on the dimension  $d$  and could be dropped out of Equation (11). In particular, with compatible boundary conditions at different orders, one approximates  $f$  through different forms:

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

$$\|f - \rho\|_{L_\infty(\Omega)} < C_A \epsilon.$$

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

$$\|f - \rho + \epsilon \sigma^{-1} v \cdot \nabla \rho\|_{L_\infty(\Omega)} < C_A \epsilon^2.$$

Here, the constant  $C_A$  depends on  $C_*$ , the upper bound in Assumption 1 for the admissible set.

The proof, which we omit here, relies on asymptotic expansion away from the boundary.

### 3.2. Convergence in the Inverse Setting

We examine the convergence in the inverse setting in this section. To describe light propagation in a given tissue in  $\Omega$ , there are two models, the radiative transfer equation that gives a statistical description, and the diffusion equation that characterizes the macroscopic behavior. The two models are asymptotically equivalent, as discussed in the previous section.

In optical tomography, light with a known intensity is injected into the material, and detectors are placed on the tissue boundary to collect the light current emitted from the material. For the RTE, the mapping from the incoming data to the outgoing data is termed the albedo operator, defined by  $\mathcal{H}^{\text{RTE}}$ ,

$$\mathcal{H}^{\text{RTE}}(\sigma) : \phi(x) \rightarrow h^{\text{RTE}}(x) = \frac{1}{C_d \epsilon} \int v \cdot n f|_{\Gamma_+} dv, \tag{12}$$

where  $f$  satisfies (9). It may also be written:

$$h^{\text{RTE}} = \mathcal{H}^{\text{RTE}}(\sigma)\phi. \tag{13}$$

In practice, finitely many incoming data  $\phi_k$  are injected and finitely many measurements are taken at the boundary location  $l_j$  per experiment. We define the map, determined by the to-be-reconstructed  $\sigma$ , from the known incoming data to the measured outgoing data as:

$$b_{j,k} = l_j(h^{\text{RTE}}) + \eta_{j,k} = l_j(\mathcal{H}^{\text{RTE}}(\sigma)\phi_k) + \eta_{j,k}, \quad (j,k) \in \{1, \dots, J\} \otimes \{1, \dots, K\}, \quad (14)$$

or in a compact form:

$$\vec{b} = \mathcal{G}^{\text{RTE}}(\sigma) + \vec{\eta}, \quad (15)$$

where  $\vec{\eta}$  is a vector of  $JK$  length that contains the noise in the measurements. Clearly, the  $JK$  length vector  $\vec{b}$  is the result of a forward map  $\mathcal{G}$  acting on the quantity of interest  $\sigma$ , with a small perturbation due to the noise. We assume that the noise is distributed as a Gaussian,

$$\vec{\eta} \sim \mathcal{N}(0, \gamma^2 \mathbb{I}) \propto \exp\left(-\frac{1}{2\gamma^2} \sigma^2\right), \quad (16)$$

meaning:

$$\vec{b}|\sigma \sim \mathcal{N}(\mathcal{G}^{\text{RTE}}(\sigma), \gamma^2 \mathbb{I}) \propto \exp\left(-\frac{1}{2\gamma^2} \left(b - \mathcal{G}^{\text{RTE}}(\sigma)\right)^\top \left(b - \mathcal{G}^{\text{RTE}}(\sigma)\right)\right). \quad (17)$$

According to Bayes' theorem, one then has:

$$\mu_{\text{RTE}}^{\vec{b}}(\sigma) = \frac{1}{Z_{\text{RTE}}} \mu_{\text{RTE}}^\sigma(\vec{b}) \mu_0(\sigma). \quad (18)$$

For the DE model, we consider the forward map to be the map that takes the Dirichlet data to the Neumann outflow. It is termed the DtN map:

$$\mathcal{H}^{\text{DE}}(\sigma) : \quad \xi(x) \rightarrow h^{\text{DE}}(x) = \frac{1}{\sigma} \frac{\partial \rho}{\partial n}, \quad (19)$$

where  $\rho$  satisfies (11). Another way to write it is:

$$h^{\text{DE}} = \mathcal{H}^{\text{DE}}(\sigma)\xi. \quad (20)$$

Again, in practice, finitely many incoming data  $\xi_k$  are injected and finitely many measurements are taken at the boundary  $l_j$  per experiment. We define the map, determined by the to-be-reconstructed media  $\sigma$ , from  $\xi_k$  to the measured data as:

$$b_{j,k} = l_j(h^{\text{DE}}) + \eta_{j,k} = l_j(\mathcal{H}^{\text{DE}}(\sigma)\phi_k) + \eta_{j,k}, \quad (j,k) \in \{1, \dots, J\} \otimes \{1, \dots, K\}, \quad (21)$$

or:

$$\vec{b} = \mathcal{G}^{\text{DE}}(\sigma) + \vec{\eta}, \quad (22)$$

where  $\vec{\eta}$  is the same pollution in the measurement. Again, the vector  $\vec{b}$  is the result of the forward map for the diffusion equation acting on the to-be-reconstructed  $\sigma$ , with the perturbation from the noise. Then, similarly, we have

$$\mu_{\text{DE}}^{\vec{b}}(\sigma) = \frac{1}{Z_{\text{DE}}} \mu_{\text{DE}}^\sigma(\vec{b}) \mu_0(\sigma). \quad (23)$$

It is proved in [24] that the two forward maps converge, namely:

**Proposition 1.** Under Assumption 1, the forward maps  $\mathcal{G}^{\text{RTE}}$  and  $\mathcal{G}^{\text{DE}}$  satisfy

$$\|\mathcal{G}^{\text{RTE}}(\sigma) - \mathcal{G}^{\text{DE}}(\sigma)\|_\infty = \mathcal{O}(\epsilon),$$

for all  $\sigma \in \mathcal{A}$ .

Using this convergence, it was also proved that the posterior distributions are close in the diffusion limit, as in the following theorem.

**Theorem 3.** Under Assumption 1, the Hellinger distance between the two posterior distribution is bounded by  $\epsilon$  in the optically thick regime when  $\epsilon \rightarrow 0$ , namely,

$$d_{\text{Hell}}(\mu_{\text{RTE}}^{\vec{b}}, \mu_{\text{DE}}^{\vec{b}}) = \mathcal{O}(\epsilon).$$

Similarly, the Kullback–Leibler divergence between the posterior distribution for the RTE and the posterior distribution for the DE is also  $\mathcal{O}(\epsilon)$ ,

$$d_{\text{KL}}(\mu_{\text{RTE}}^{\vec{b}}, \mu_{\text{DE}}^{\vec{b}}) = \mathcal{O}(\epsilon).$$

The proof largely depends on the Lipschitz continuity of the Gaussian form in the likelihood function, and the convergence result in Proposition 1. We omit the details and refer interested readers to [24].

Using Theorem 3, when  $\epsilon$  is sufficiently small, one may use the diffusion equation posterior to approximate the radiative transfer equation posterior. This approximation allows us to speed up the MCMC computation by setting  $\mu_{\text{DE}}^{\vec{b}} = \mu'$  as the low-resolution model in the first level. This filters out bad draws, passing better draws to  $\mu_{\text{RTE}}^{\vec{b}} = \mu$  on the second level.

#### 4. Algorithm

In this section, we discuss the DE-assisted two-level MCMC method and its convergence for our case, the inverse problem for the RTE in the diffusion limit. We present a result about the second-level acceptance rate of two-level MCMC and its dependence on  $\epsilon$  and a result about the computational cost of our algorithm compared to the one-level MCMC method.

##### 4.1. DE-Assisted Two-Level MCMC Method

In this section, we discuss our algorithm, the DE-assisted two-level MCMC method. It is shown below in Algorithm 3.

---

#### Algorithm 3 DE-assisted two-level Metropolis Hastings

---

1. Given  $x_k$ , draw  $y \sim q(\cdot, x_k)$
2. Let

$$\alpha(x_k, y) := \min \left\{ 1, \frac{q(y, x_k) \mu_{\text{DE}}^{\vec{b}}(y)}{q(x_k, y) \mu_{\text{DE}}^{\vec{b}}(x_k)} \right\}.$$

3. With probability  $\alpha$ , **pre-accept**  $y$  and continue to 4. Otherwise, set  $x_{k+1} = x_k$  and start over.
4. The second-level proposal is now  $y$ , effectively drawn from

$$q_2(y, x_k) = \alpha(x_k, y)q(y, x_k) + \delta_{x_k}(y) \left( 1 - \int \alpha(x_k, y)q(y, x_k)dy \right).$$

Then set

$$\beta(x_k, y) := \min \left\{ 1, \frac{q_2(y, x_k) \mu_{\text{RTE}}^{\vec{b}}(y)}{q_2(x_k, y) \mu_{\text{RTE}}^{\vec{b}}(x_k)} \right\}.$$

With probability  $\beta(x_k, y)$ , **accept** and set  $x_{k+1} = y$ . Otherwise, set  $x_{k+1} = x_k$ .

---

The transition kernel  $P_2$  may be written

$$P_2(x, y) = p_2(x, y) + r_2(x)\delta_x(y),$$

where

$$p_2(x, y) = \begin{cases} q_2(x, y)\beta(x, y) & x \neq y \\ 0 & x = y \end{cases}$$

is called the second-level off-diagonal density, and  $r_2(x) = 1 - \int p_2(x, y)dy$ . As before, to demonstrate the convergence of MCMC, we will examine the transition kernel, which is the probability that the next draw  $x_{k+1}$  is in the set  $A$  given the previous elements of the chain  $x_0, \dots, x_k$ .

For the two-level MCMC method, one desires that the high-resolution target distribution  $\mu_{\text{RTE}}^{\vec{b}}$  is an invariant distribution of the Markov chain  $\{x_k\}_{k \geq 0}$ . By definition, this is true if for all measurable sets  $A$ ,

$$\mu_{\text{RTE}}^{\vec{b}}(A) = \int \mu_{\text{RTE}}^{\vec{b}}(x)P(x, A)dx.$$

This means that elements of the Markov chain generated by Algorithm 3 give a good representation of the distribution  $\mu_{\text{RTE}}^{\vec{b}}$ . To show that  $\mu_{\text{RTE}}^{\vec{b}}$  is the invariant distribution, we first need the following two lemmas.

**Lemma 2.** *The second-level acceptance rate may be written*

$$\beta(x_k, y) = \min \left\{ 1, \frac{q_2(y, x_k)\mu_{\text{RTE}}^{\vec{b}}(y)}{q_2(x_k, y)\mu_{\text{RTE}}^{\vec{b}}(x_k)} \right\} = \min \left\{ 1, \frac{\mu_{\text{DE}}^{\vec{b}}(x_k)\mu_{\text{RTE}}^{\vec{b}}(y)}{\mu_{\text{DE}}^{\vec{b}}(y)\mu_{\text{RTE}}^{\vec{b}}(x_k)} \right\},$$

where  $q_2$  is defined as in Algorithm 2.

**Proof.** From Lemma 3, for  $x_k \neq y$ , we have

$$\begin{aligned} \frac{q_2(y, x_k)}{q_2(x_k, y)} &= \frac{\alpha(y, x)q(y, x)}{\alpha(x, y)q(x, y)} \\ &= \frac{\mu_{\text{DE}}^{\vec{b}}(x)}{\mu_{\text{DE}}^{\vec{b}}(y)}. \end{aligned}$$

Plugging this in to our definition of  $\beta$ , we find

$$\beta(x_k, y) = \min \left\{ 1, \frac{\mu_{\text{DE}}^{\vec{b}}(x_k)\mu_{\text{RTE}}^{\vec{b}}(y)}{\mu_{\text{DE}}^{\vec{b}}(y)\mu_{\text{RTE}}^{\vec{b}}(x_k)} \right\}. \quad (24)$$

□

The importance of this lemma is that it equates  $\beta$ , which depends on  $q_2$ , to a form that is computable. Note that  $q_2$  has a complicated integral form and thus is numerically challenging to compute. This form of the second-stage acceptance rate  $\beta$  has been calculated in [18]. Later, we will discuss the dependence of  $\beta$  on  $\epsilon$  through the inverse problem for the RTE's posterior distribution and show that when the RTE and DE are close,  $\beta$  is close to one.

We next discuss a property of  $\beta$  that we will need to show the convergence of the two-level MCMC method.

**Lemma 3.** *The second-level off-diagonal density satisfies the detailed balance equation*

$$p_2(x, y)\mu_{\text{RTE}}^{\vec{b}}(x) = p_2(y, x)\mu_{\text{RTE}}^{\vec{b}}(y).$$

**Proof.** Considering the form of  $\beta$  in Equation (24), when  $\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y) < \mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)$ ,

$$\beta(x, y) = \frac{\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y)}{\mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)}, \quad \beta(y, x) = 1.$$

Similarly, when  $\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y) > \mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)$ ,

$$\beta(x, y) = 1, \quad \beta(y, x) = \frac{\mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)}{\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y)},$$

so dividing them gives

$$\frac{\beta(x, y)}{\beta(y, x)} = \frac{\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y)}{\mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)} = \frac{\mu_{RTE}^{\vec{b}}(y)q_2(y, x)}{\mu_{RTE}^{\vec{b}}(x)q_2(x, y)},$$

using Lemma 1. Simplifying gives

$$\mu_{RTE}^{\vec{b}}(x)q_2(x, y)\beta(x, y) = \mu_{RTE}^{\vec{b}}(y)q_2(y, x)\beta(y, x).$$

□

This lemma gives rise to the following theorem.

**Theorem 4.** The second-level distribution  $\mu_{RTE}^{\vec{b}}$  is the invariant distribution of the Markov chain  $\{x_n\}$  with second-level transition kernel  $P_2(x, A)$ , i.e.,

$$\mu_{RTE}^{\vec{b}}(A) = \int p_2(x, A)\mu_{RTE}^{\vec{b}}(x)dx, \tag{25}$$

where the transition kernel  $p_2(x, y)$  is defined as

$$P_2(x, y) = p_2(x, y) + r_2(x)\delta_x(y).$$

**Proof.** Consider

$$\begin{aligned} \int P_2(x, A)\mu_{RTE}^{\vec{b}}(x)dx &= \int p_2(x, y)\mu_{RTE}^{\vec{b}}(x)dx + \int r_2(x)\delta_x(A)\mu_{RTE}^{\vec{b}}(x)dx \\ &= \int_A \int p_2(y, x)\mu_{RTE}^{\vec{b}}(y)dx dy + \int_A r_2(x)\mu_{RTE}^{\vec{b}}(x)dx, \end{aligned}$$

using Lemma 3 and using the delta function to perform the integration over  $x$  in the second term. Then, from the definition of  $p_2$ ,

$$\begin{aligned} \int P_2(x, A)\mu_{RTE}^{\vec{b}}(x)dx &= \int_A (1 - r_2(y))\mu_{RTE}^{\vec{b}}(y)dy + \int_A r_2(x)\mu_{RTE}^{\vec{b}}(x)dx \\ &= \int_A \mu_{RTE}^{\vec{b}}(y)dy \\ &= \mu_{RTE}^{\vec{b}}(A). \end{aligned}$$

□

This theorem demonstrates that the high-resolution target distribution, for us  $\mu_{\text{RTE}}^{\vec{b}}$ , is the invariant distribution of  $P_2$ . Thus, the two-level MCMC method gives a list of elements  $\{x_k\}$  that can be regarded as samples from  $\mu_{\text{RTE}}^{\vec{b}}$ .

To generate samples from the posterior based on the RTE, using the one-level MCMC method, for each new proposal, the forward map must be computed. To do this, one injects  $K$  boundary data  $\phi_k$  and computes  $J$  outgoing data  $l_j$ , meaning that the RTE is solved  $K$  times and each time the solution is evaluated at  $J$  locations. As this is computationally expensive, one looks for a cheaper way to sample from the distribution. From Theorem 3, we have that in the diffusion limit, the posterior based on the RTE is close to the posterior based on the DE. The diffusion equation is significantly faster to solve because it is only over physical space. Therefore, in the strong scattering regime, we can save time on the computation by using the DE-assisted two-level MCMC method, in which the DE is used as a surrogate model to filter out bad draws. This saves computational effort spent on evaluating bad samples using the high-resolution model. The diffusion equation posterior can be used to reject bad samples, so that we have to evaluate the RTE posterior fewer times overall.

#### 4.2. Properties of DE-Assisted Two-Level MCMC

In this section, we first present our result concerning the acceptance rate of the DE-assisted MCMC method and its dependence on  $\epsilon$ , and then present the computational cost of our method compared to the one-level MCMC method.

##### 4.2.1. Acceptance Rate of DE-Assisted Two-Level MCMC

There are many ways to improve MCMC. Different sampling methods such as the Gibbs sampler or independence samplers can make the algorithm more efficient. There are also delayed acceptance/rejection methods, and adaptive methods in which the low-resolution model is improved at each stage using results from the high-resolution model [23]. However, this is not the focus of the current paper. Our method is a delayed acceptance method. It relies on the two-level MCMC algorithm and the diffusion limit to improve the efficiency of sampling from the posterior distribution for the RTE.

We emphasize that the DE-assisted two-level MCMC algorithm in Algorithm 3 has two acceptance rates,  $\alpha$  and  $\beta$ . The first-level acceptance rate  $\alpha$  depends only on the posterior based on the DE, so it can have no  $\epsilon$  dependence. The second-level acceptance rate  $\beta$ , depends on  $\epsilon$  through the posterior based on the RTE. As  $\epsilon \rightarrow 0$ , the posterior based on the RTE becomes closer to the posterior based on the DE, so we expect that the samples that pass the selection criterion have a high chance of being accepted in the second step as well. We quantify this in the following proposition.

**Proposition 2.** *In the diffusion limit, the acceptance rate  $\beta$  is high:*

$$|\beta - 1| = \mathcal{O}(\epsilon^{1-\alpha}), \quad (26)$$

as long as the proposal is reasonably close to the measured data,

$$\|\mathcal{G}_{\text{DE}}(y) - \vec{b}\|_2 \leq \gamma\sqrt{-\alpha \ln \epsilon},$$

where  $\gamma$  is the variance of the noise and  $\alpha$  is any constant between 0 and 1.

**Proof.** Suppose  $\|\mathcal{G}_{\text{DE}}(y) - \vec{b}\|_2 \leq \gamma\sqrt{-\alpha \ln \epsilon}$ . Then, the likelihood function for the inverse diffusion equation is

$$\mu_{\text{DE}}^{\sigma}(y) = \exp\left(-\frac{1}{\gamma^2}\|\mathcal{G}^{\text{DE}}(y) - \vec{b}\|_2^2\right) > \exp\left(-\frac{1}{\gamma^2}(\gamma^2(-\alpha \ln \epsilon))\right).$$

In other words,

$$\mu_{DE}^\sigma(y) > \epsilon^\alpha.$$

Since the likelihood function is greater than  $\epsilon$  and the prior has no  $\epsilon$  dependence, the posterior distribution will also be greater than  $\epsilon$ ,

$$\mu_{DE}^{\vec{b}}(y) = \mu_0(\sigma)\mu_{DE}^\sigma(y) = \mathcal{O}(\epsilon^\alpha). \tag{27}$$

Considering the form of  $\beta$  in Equation (24), we have

$$\beta = \min \left\{ 1, \frac{\mu_{DE}^{\vec{b}}(x)\mu_{RTE}^{\vec{b}}(y)}{\mu_{DE}^{\vec{b}}(y)\mu_{RTE}^{\vec{b}}(x)} \right\} = \min \left\{ 1, \frac{\mu_{DE}^{\vec{b}}(x)(\mu_{DE}^{\vec{b}}(y) + C(y)\epsilon)}{\mu_{DE}^{\vec{b}}(y)(\mu_{DE}^{\vec{b}}(x) + C(x)\epsilon)} \right\}, \tag{28}$$

using Theorem 3. Considering

$$\begin{aligned} \frac{\mu_{DE}^{\vec{b}}(x)(\mu_{DE}^{\vec{b}}(y) + C(y)\epsilon)}{\mu_{DE}^{\vec{b}}(y)(\mu_{DE}^{\vec{b}}(x) + C(x)\epsilon)} &= \frac{\mu_{DE}^{\vec{b}}(x)\mu_{DE}^{\vec{b}}(y) + \mu_{DE}^{\vec{b}}(x)C(y)\epsilon}{\mu_{DE}^{\vec{b}}(y)\mu_{DE}^{\vec{b}}(x) + \mu_{DE}^{\vec{b}}(y)C(x)\epsilon} \\ &= \frac{1 + C(y)\epsilon/\mu_{DE}^{\vec{b}}(y)}{1 + C(x)\epsilon/\mu_{DE}^{\vec{b}}(x)} \\ &= \left( 1 + \frac{C(y)\epsilon}{\mu_{DE}^{\vec{b}}(y)} \right) \left( 1 - \frac{C(x)\epsilon}{\mu_{DE}^{\vec{b}}(x)} + \dots \right) \\ &= 1 + \mathcal{O}(\epsilon^{1-\alpha}), \end{aligned}$$

where the Taylor expansion is possible due to Equation (27). Therefore, for small  $\epsilon$ ,

$$|\beta - 1| = \mathcal{O}(\epsilon^{1-\alpha}).$$

□

This proposition demonstrates that as  $\epsilon \rightarrow 0$  and the posterior based on the RTE becomes closer to the posterior based on the DE, more and more samples are accepted at the second level.

#### 4.2.2. Computational Cost Comparison

In this section, we discuss the computational cost of our method compared to the one-level MCMC method.

**Proposition 3.** Let  $Cost_1$  denote the cost of obtaining  $k$  accepted samples using the one-level MCMC method, and  $Cost_2$  denote the cost of obtaining  $k$  accepted samples using the DE-assisted two-level MCMC method. Let  $r_1$  denote the acceptance rate of the one-level MCMC method. Let  $\alpha$  denote the first-level acceptance rate of the DE-assisted two-level MCMC method, and  $\beta$  denote the second-level acceptance rate of the method.

- $Cost_1$  and  $Cost_2$  are

$$\begin{aligned} Cost_1 &= \frac{k}{r_1} C_R \\ Cost_2 &= \frac{k}{\alpha\beta} C_D + \frac{k}{\beta} C_R, \end{aligned}$$

where  $C_D = \mathcal{O}(N_x^{cd})$  is the cost of solving the diffusion equation with  $N_x$  grid points in  $d$  dimensions, where  $c$  is the cost of the linear algebra solver used. Similarly,  $C_R = \mathcal{O}(N_x^{cd}N_v^{c(d-1)})$  is the cost of solving



the radiative transfer equation with the same parameters, and  $N_v$  is the number of grid points in the velocity domain.

- Considering  $C_R \gg C_D$ , we have

$$\frac{\text{Cost}_2}{\text{Cost}_1} \approx \frac{r_1}{\beta}.$$

Thus, the cost saving of our method comes from  $\beta > r_1$ .

**Proof.** First, we examine the cost of evaluating the posterior distribution based on the diffusion equation compared to the posterior based on the radiative transfer equation. The cost of evaluating each posterior distribution is directly proportional to the cost of computing each solution to the equation, so we examine instead the cost of computing a solution to the diffusion equation and the radiative transfer equation. Considering  $C_D = \mathcal{O}(N_x^{cd})$ , and  $C_R = \mathcal{O}(N_x^{cd} N_v^{c(d-1)})$ , we see that the diffusion equation is significantly cheaper to compute than the radiative transfer equation, depending on the number of grid points in the velocity domain.

Next, we compute the number of MCMC iterations required to obtain  $k$  accepted samples for each method. For the one-level MCMC method, we simply have

$$N_1 r_1 = k,$$

where  $r_1$  is the acceptance rate and  $N_1$  is the total number of iterations considered. For the two-level method,

$$N_2 r_2 = k,$$

where  $N_2$  is the total number of iterations, and  $r_2$  is the overall acceptance rate. To be more specific, a single sample must go through both levels of evaluation in order to be accepted. Supposing that the acceptance rate at the first level is  $\alpha$ , and the acceptance rate at the second level is  $\beta$ , then, the overall acceptance rate is  $r_2 = \alpha\beta$ , which gives us

$$N_2 \alpha \beta = k.$$

Using the above results, we can compute the cost of obtaining  $k$  accepted samples using the one-level MCMC method,

$$\text{Cost}_1 = N_1 C_R = \frac{k}{r_1} C_R.$$

Similarly, the cost of obtaining  $k$  accepted samples using the two-level method is

$$\text{Cost}_2 = N_2 C_D + \frac{k}{\beta} C_R = \frac{k}{\alpha\beta} C_D + \frac{k}{\beta} C_R. \quad (29)$$

Considering  $C_D \ll C_R$ , the term containing  $C_D$  may be dropped out of Equation (29). Then we have

$$\frac{\text{Cost}_2}{\text{Cost}_1} = \frac{r_1}{\beta}.$$

Then for  $\beta \gg r_1$ , the cost of obtaining  $k$  samples using the DE-assisted two-level MCMC is much less than the cost of obtaining  $k$  samples using the one-level MCMC.  $\square$

In practice,  $\beta$  and  $r_1$  depend on the sampling strategy used, the initial values of the MCMC parameters, and the step size, so we give only a theoretical asymptotic estimate of the cost comparison.

### 5. Numerics

We summarize our numerical evidence in this section. We will first show the convergence in the forward setting, and demonstrate that the solution to the RTE indeed converges to that of the DE. We then demonstrate, with two different media configurations, the convergence of the posterior distribution using the Metropolis–Hastings MCMC method. The 2-level MCMC result will also be shown.

Throughout this section, the DE is computed using the standard finite element method, and the RTE solver is a preconditioned GMRES-based method with upwinding in the physical domain, designed in [29].

#### 5.1. Forward Model Convergence

We first review the convergence in the forward setup. As discussed in Theorem 2, in the zero limit of  $\epsilon$ , the solution to the RTE becomes the solution to the DE, which drives the convergence of the albedo operator to the DtN map. We show this numerically below.

We first use a pseudo-2D example, with the  $y$ -direction assumed to be homogeneous. The RTE is then a degenerate case given by

$$\cos \theta \partial_x f = \sigma(x) \mathcal{L} f .$$

Here, the media is set as

$$\sigma(x) = 1 + 9\chi_{[0.05,0.15]}(x) + 19\chi_{[0.35,0.45]}(x) + 29\chi_{[0.75,0.85]}(x) ,$$

where  $\chi_{[a,b]}$  is the characteristic function. Figure 1 contains a plot of the solution  $\rho$  as a function of  $x$  for the given  $\sigma$  for the RTE with  $\epsilon = 1, 2^{-3}$ , and  $2^{-6}$ , as well as the solution for the DE. Numerically, we set  $dx = 0.05$  and  $dv = 2\pi/16$ . As  $\epsilon$  becomes smaller and smaller, the density  $\rho$  of the radiative transfer equation becomes closer and closer to the solution of the diffusion equation. We plot the error on the right panel of Figure 1 on a log-log scale as a function of  $\epsilon$ , measured in  $L_2(dx)$ . The values of the error are shown in Table 1.

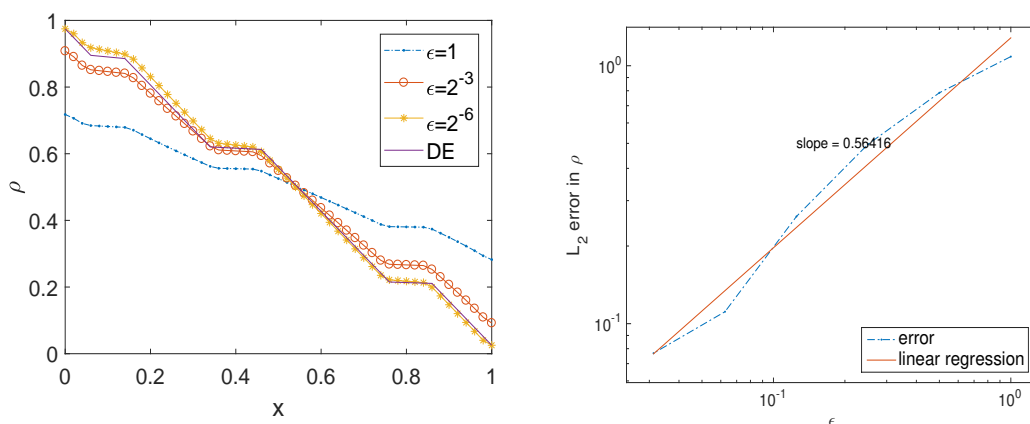


Figure 1. The solution  $\rho$  as a function of  $x$  for different  $\epsilon$ , and the error in the solution  $\rho$  as a function of  $\epsilon$ .

Table 1. The values of the error in the solution  $\rho$  as a function of  $\epsilon$ .

$\epsilon$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$
error	0.2191	0.2004	0.1737	0.1418	0.1130	0.0990

We furthermore plot the solution  $f$  as a function on the phase space. As shown in Theorem 2, in the diffusion limit,  $f$  loses its velocity dependence, and the solution becomes flat in  $v$ . This is seen numerically as well in Figure 2, which shows  $f(x, v)$  for different  $\epsilon$ .

Similar convergence is numerically evaluated in two dimensions where  $\Omega = [0, 1]^2$  and  $\sigma$  is chosen to be

$$\sigma(x) = 1 + h\chi(|x - c| < r),$$

where  $x = (x, y) \in [0, 1]^2$ ,  $h = 10$  and  $r = 0.4$ .  $\chi$  is the characteristic function, and the center  $c = (0.5, 0.5)$  is given. It is hard to visualize a 3D object, and we only plot the difference in  $\rho$  between the radiative transfer and diffusion solutions for different  $\epsilon$  in Figure 3. We observe that as  $\epsilon \rightarrow 0$ , the error decreases, as plotted in Figure 4.

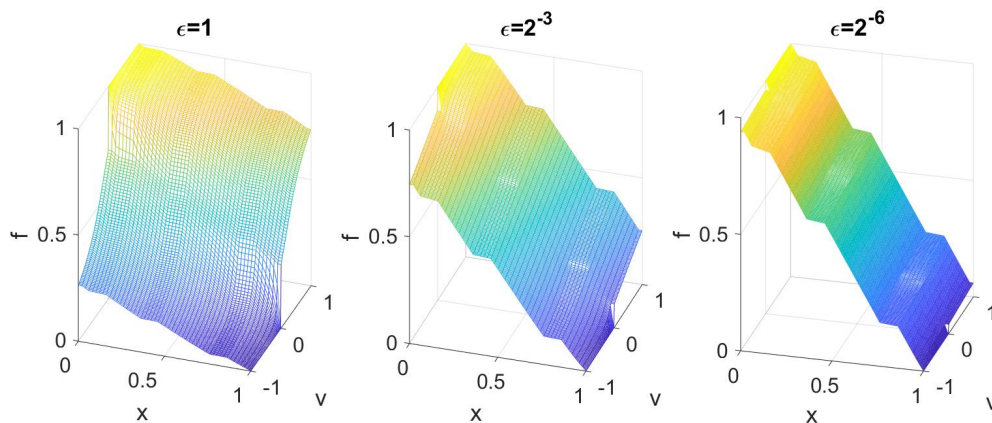


Figure 2. The solution as a function of  $x$  and  $v$  for different  $\epsilon$ .

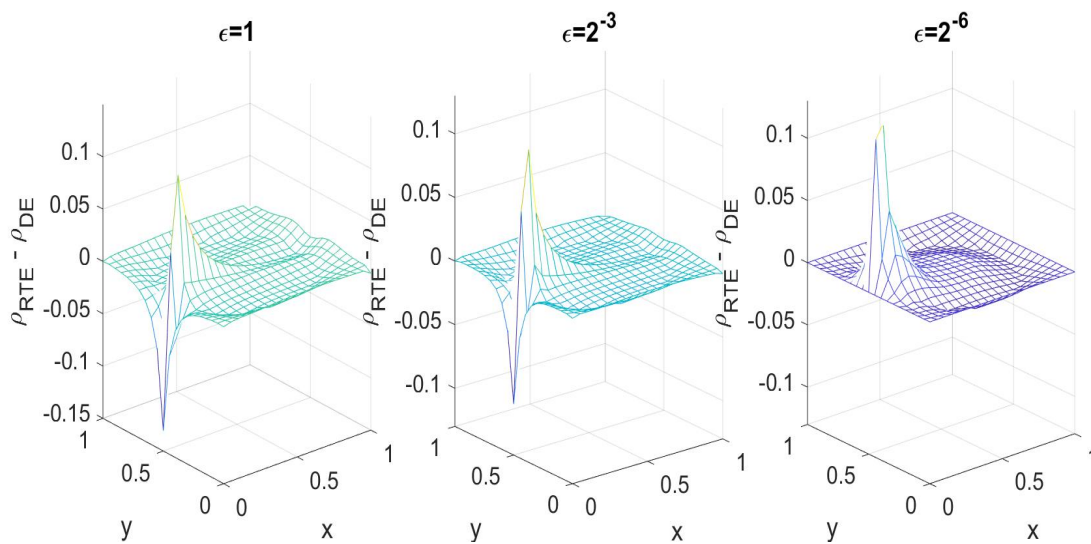


Figure 3. Difference between the RTE and DE solutions in 2D for different  $\epsilon$ .

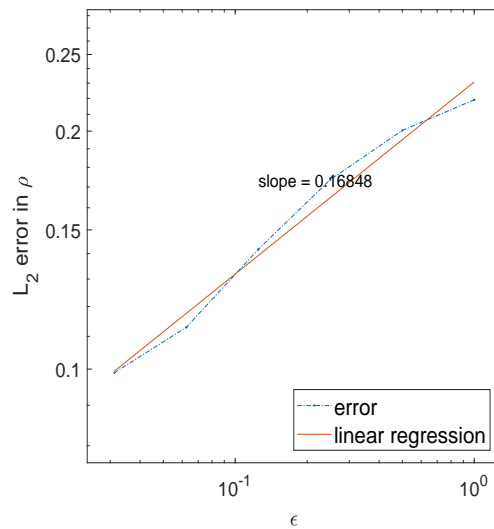


Figure 4.  $L_2$  norm in the error for the 2D solution  $\rho$  as a function of  $\epsilon$ .

5.2. Reconstruction Example 1

In this example, we test the inverse problem in  $\Omega = [0, 1]^2 \subset \mathbb{R}^2$  with the true media

$$\sigma(x) = 1 + h\chi(|x - c| < r),$$

where  $r = 0.4$  and  $h = 10$ , and  $c = (0.5, 0.5)$ . It represents a circle with radius  $r$  and height  $h$  in the middle of the unit square. Since the media configuration is uniquely determined by the two parameters  $r$  and  $h$ , the reconstructed posterior distribution then is a distribution function on the  $(r, h)$  domain. The convergence and the accuracy of the method is independent of dimension, but we need to use an example that can be visualized to the readers. Besides, in most application problems the bio-tissues to be reconstructed have two or three components that have very different optical properties, and the model used here is a realistic representation of the physical scenario. To compute the forward RTE, we use the GMRES method. We set  $dx = 0.05$  and  $dv = 2\pi/16$ . The light sources are placed on the left boundary ( $x = 0$ ) for all discrete grids in  $y$ -space. This means 20 experiments are being looked at. For all these experiments, the sensors are located at all grid points on the boundaries. For each MCMC sample  $(r, h)$ , to evaluate  $\mu_{\text{RTE}}^{\hat{b}}(r, h)$ , 20 forward RTE solvers are run with the media configuration  $(r, h)$  with the boundary condition being Kronecker delta functions at each light source.

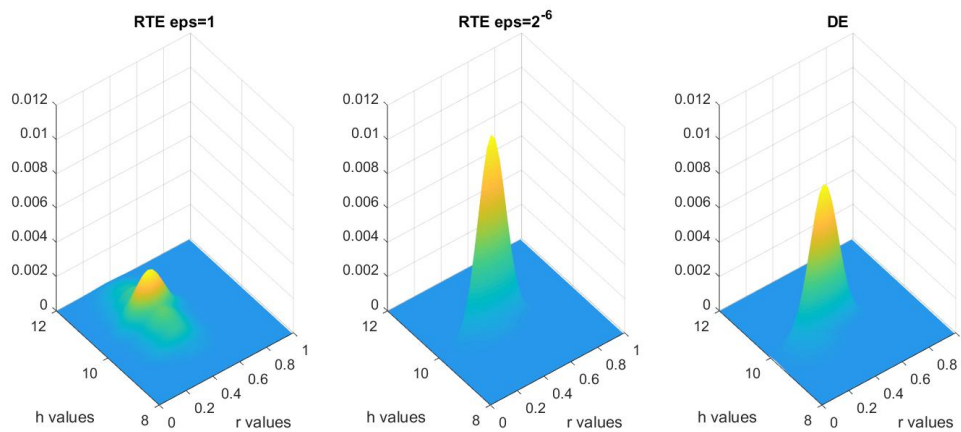
We briefly mention that to visualize the data, we do apply the kernel density estimator that smooths out the solution with a Gaussian kernel.

We first consider the solution to this problem in the one-level MCMC framework. The prior distribution is taken to be uniform in  $r$  from 0 to 0.5, and uniform in  $h$  from 8 to 12, and the variance of the noise is taken to be  $10^{-4}$ . With merely  $10^3$  MCMC steps, one can already distinguish the distribution functions. As seen in Figure 5, the posterior distribution using the RTE model with  $\epsilon = 1$  is much more smeared out than the one with  $\epsilon = 2^{-6}$ , and the posterior distribution using the RTE model with  $\epsilon = 2^{-6}$  is significantly closer to that using the DE model.

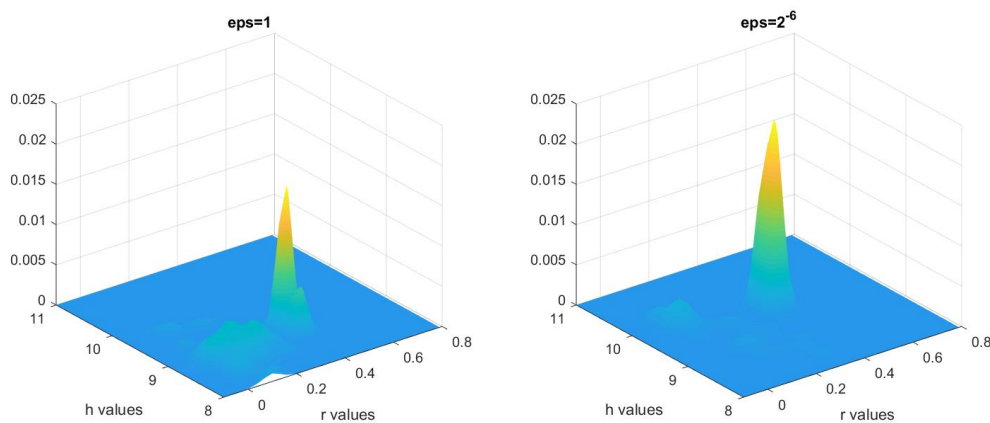
To quantize the difference between these three distributions, we compute the Hellinger distances, as documented in Table 2, which numerically confirms Theorem 3.

We then study the two-level MCMC method with the same parameters. The Gaussian-kernel smoothed posterior distributions are shown in Figure 6.

We examined the acceptance rate  $\beta$ : the number of accepted particles on the second level versus the number of accepted particle on the first level. As suggested in Proposition 2, smaller  $\epsilon$  should give a higher acceptance rate, and this is truly seen in numerics. The results are shown in Table 3.



**Figure 5.** Multivariate kernel distribution representing the posterior distributions for DE, RTE with  $\epsilon = 1$  and  $\epsilon = 2^{-6}$  obtained using the one-level MCMC method.



**Figure 6.** Multivariate kernel distribution representing posterior distribution obtained from the two-level MCMC method for  $\epsilon = 1$  and  $\epsilon = 2^{-6}$ .

**Table 2.** Hellinger distance:  $d_{\text{Hell}}(\mu_{\text{DE}}^{\vec{b}}(\sigma), \mu_{\text{RTE}}^{\vec{b}}(\sigma))$  with  $\mu_{\text{RTE}}^{\vec{b}}(\sigma)$  computed with RTE as the forward solver using  $\epsilon = 1$ ,  $\epsilon = 2^{-3}$  and  $\epsilon = 2^{-6}$  respectively.

$\epsilon$	1	$2^{-3}$	$2^{-6}$
$d_{\text{Hell}}(\mu_{\text{DE}}^{\vec{b}}(\sigma), \mu_{\text{RTE}}^{\vec{b}}(\sigma))$	0.6418	0.5322	0.2219

**Table 3.** Acceptance rate  $\beta$  for Example 1.

$\epsilon$	1	$2^{-3}$	$2^{-6}$
acceptance rate	0.6931	0.8736	0.8939

We also expect for small  $\epsilon$ , distributions computed using the one-level MCMC should be similar to the ones computed using DE-assisted method. To compare them, we compute the Hellinger distance between the distribution based on the RTE computed using one-level MCMC for  $\epsilon = 2^{-6}$ , in Figure 5 and the distribution based on the RTE computed using the DE-assisted MCMC for  $\epsilon = 2^{-6}$ , shown in Figure 6. We obtain

$$d_{\text{Hell}}\left(\mu_{\text{RTE}}^{\vec{b}}(\sigma)_{\text{1-level}}, \mu_{\text{RTE}}^{\vec{b}}(\sigma)_{\text{DE-assisted}}\right) = 0.2289,$$

demonstrating that the distributions based on the methods are similar.

Finally, we report the computational cost savings of using the DE-assisted two-level MCMC method compared to the one-level MCMC method. The values of  $\beta$  and  $r_1$  are documented in Table 4. From the table, we see that using the DE-assisted two-level MCMC method saves us roughly 20% of the RTE evaluations, for each value of  $\epsilon$ .

**Table 4.** Computational cost comparison.

	$\epsilon = 1$	$\epsilon = 2^{-6}$
$\beta$	0.6931	0.8939
$r_1$	0.6000	0.7350
$r_1/\beta$	0.8656	0.8222

We emphasize that  $\beta$  and  $r_1$  highly depend on the sampling strategy used, the initial values of the MCMC parameters, the step size and how the prior distribution looks. If the prior distribution is centered far away from the maximum likelihood point, a large amount of sample points are needed to give a fair presentation of the posterior distribution function, and the behavior of the underlying equation would be crucial in computational saving.

### 5.3. Reconstruction Example 2

In this example, we take a more complicated media  $\sigma$ :

$$\sigma(x) = 1 + \sum_{i=1}^5 h_i \chi_i(|x - c_i| < r_i),$$

where  $x = (x, y) \in [0, 1]^2$  and  $c_i$  are given, and  $\chi_i$  are characteristic functions. The list of height  $h_i$  and radii  $r_i$  are parameters to be reconstructed, and thus we have ten unknown parameters. Numerically, we set  $dx = 0.05$ ,  $dv = 2\pi/16$ , and we place the light sources at the left boundary as in Example 1. Again, we take  $10^3$  MCMC steps. As before, to compute the forward RTE, we use the GMRES method. The light sources are placed on the left boundary ( $x = 0$ ) for all discrete grids in  $y$ -space. This means that 20 experiments are being looked at. For all these experiments, the sensors are located at all grid points on the boundaries. For each MCMC sample  $(r, h)$ , to evaluate  $\mu_{\text{RTE}}^{\bar{b}}(r, h)$ , 20 forward RTE solvers are run with the media configuration  $(r, h)$  with the boundary condition being Kronecker delta functions at each light source.

The true parameters are set as

$$\begin{aligned} (h_1, r_1) &= (5, 0.19), & (h_2, r_2) &= (1, 0.1), & (h_3, r_3) &= (7, 0.09), \\ (h_4, r_4) &= (4, 0.13), & (h_5, r_5) &= (10, 0.04), \end{aligned}$$

with the centers

$$\begin{aligned} c_1 &= (0.5, 0.5), & c_2 &= (0.2, 0.35), & c_3 &= (0.75, 0.2), \\ c_4 &= (0.8, 0.85), & c_5 &= (0.3, 0.8). \end{aligned}$$

The true  $\sigma$  is depicted in Figure 7.

To test the one-level MCMC result, we feed the algorithm an initial guess that could be 40% away from the true value:

$$h_i^{\text{guess}} = (0.6 + 0.4 \cdot \text{rand}) \cdot h_i^{\text{true}},$$

and similarly for each  $r_i$ . The prior is set as a uniform distribution, and the variance of the noise is again taken to be  $10^{-4}$ . In Figure 8, we plot the marginal posterior distribution of  $(r_1, r_4)$  of the DE, and RTE with  $\epsilon = 1$ , and  $\epsilon = 2^{-6}$ . The posterior distribution of RTE with small  $\epsilon$  is strikingly closer to

the DE model, while the RTE model with big  $\epsilon$  gives a very different result. In Table 5 we document the means and variances of the marginal distributions for  $r_1$  and  $r_4$ .

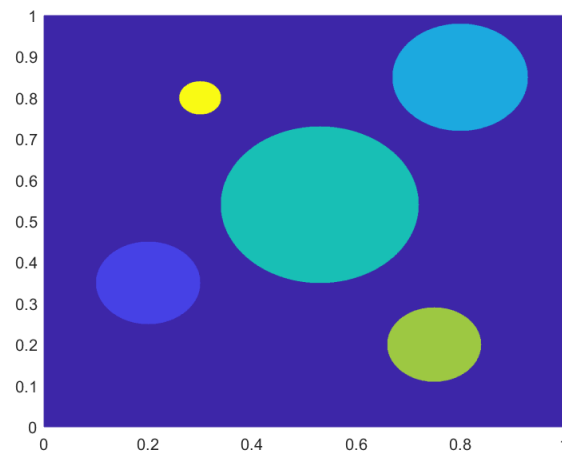


Figure 7. The true  $\sigma$  as a contour plot for Example 2.

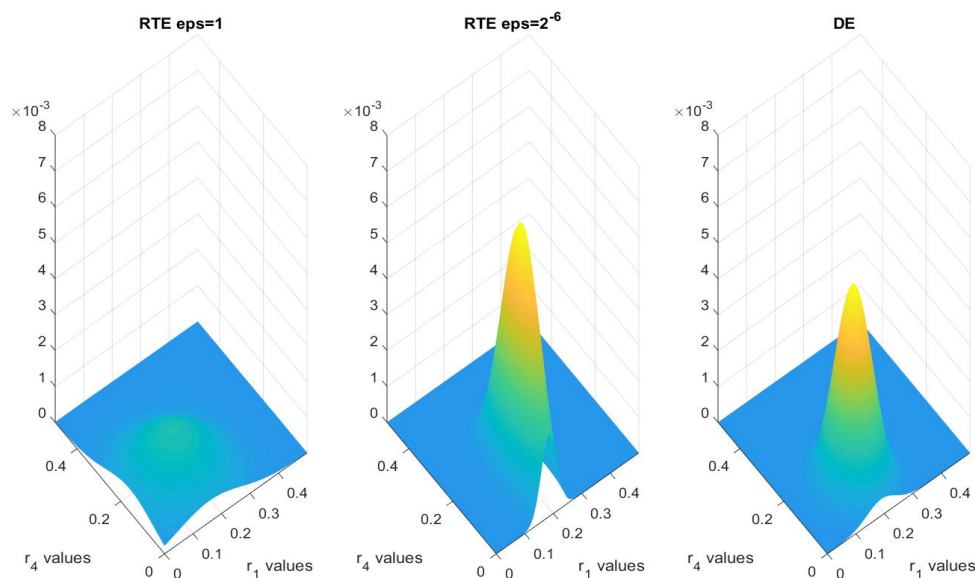


Figure 8. Multivariate kernel distribution for  $r_1$  and  $r_4$  for DE and RTE with  $\epsilon = 1$  and  $\epsilon = 2^{-6}$  using one-level MCMC.

Table 5. Means and variances of the marginal distributions for  $r_1$  and  $r_4$  for DE, RTE with  $\epsilon = 1$ , and RTE with  $\epsilon = 2^{-6}$  for two-level MCMC.

	Mean ( $r_1$ )	Mean ( $r_4$ )	Var ( $r_1$ )	Var ( $r_4$ )
RTE, $\epsilon = 1$	0.1483	0.1518	0.0468	0.0464
RTE, $\epsilon = 2^{-6}$	0.1856	0.1278	0.042	0.0076
DE	0.1895	0.1273	0.0078	0.0038

We again compute the Hellinger distance between the two marginal distributions for  $(r_1, r_4)$  of the posterior distributions with respect to  $\epsilon$ , and it is documented in Table 6. The results numerically confirm Theorem 3.



**Table 6.** Hellinger distance for the marginal distributions for  $r_1$  and  $r_4$ :  $d_{\text{Hell}}(\mu_{\text{DE}}^{\vec{b}}(\sigma), \mu_{\text{RTE}}^{\vec{b}}(\sigma))$  with  $\mu_{\text{RTE}}^{\vec{b}}(\sigma)$  computed with RTE as the forward solver using  $\epsilon = 1$ ,  $\epsilon = 2^{-3}$  and  $\epsilon = 2^{-6}$  respectively.

$\epsilon$	1	$2^{-3}$	$2^{-6}$
$d_{\text{Hell}}(\mu_{\text{DE}}^{\vec{b}}(\sigma), \mu_{\text{RTE}}^{\vec{b}}(\sigma))$	0.8693	0.6324	0.4440

As done in Example 1, we document the acceptance rate. As seen clearly in Table 7, the acceptance rate increases as  $\epsilon$  shrinks to zero.

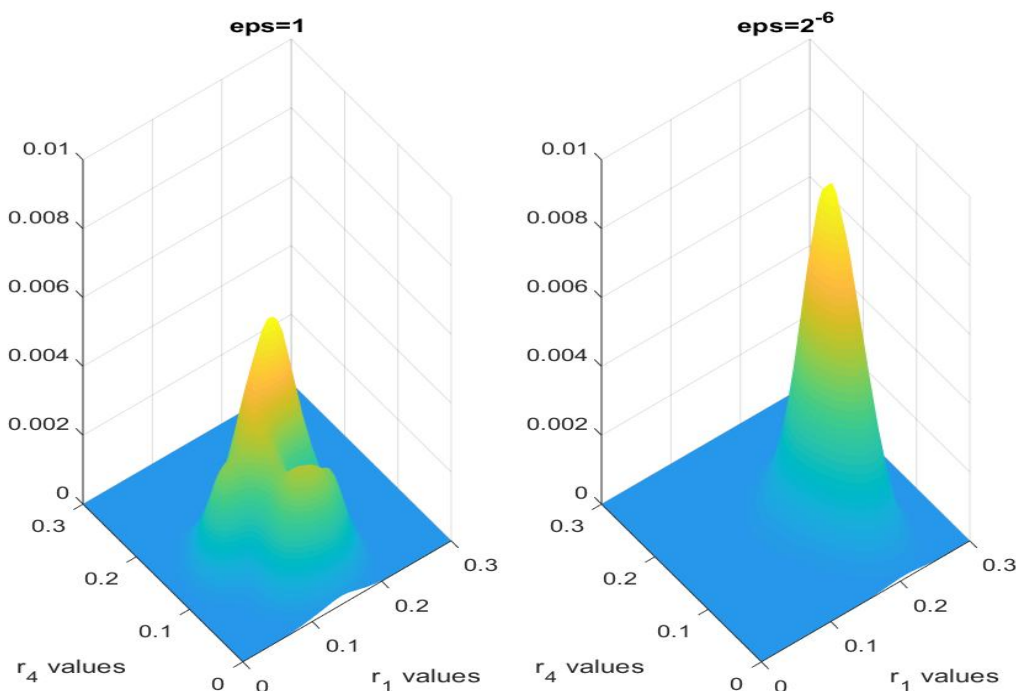
**Table 7.** Acceptance rate  $\beta$  for Example 2.

$\epsilon$	1	$2^{-3}$	$2^{-6}$
acceptance rate	0.7230	0.8969	0.9305

As before, we expect that the small  $\epsilon$  distributions computed using the one-level MCMC case and using our method should be similar. To compare them, we compute the Hellinger distance between the distribution based on the RTE computed using one-level MCMC for  $\epsilon = 2^{-6}$ , in Figure 8 and the distribution based on the RTE computed using the DE-assisted MCMC for  $\epsilon = 2^{-6}$ , shown in Figure 9. We obtain

$$d_{\text{Hell}}\left(\mu_{\text{RTE}}^{\vec{b}}(\sigma)_{\text{1-level}}, \mu_{\text{RTE}}^{\vec{b}}(\sigma)_{\text{DE-assisted}}\right) = 0.1357,$$

again demonstrating that the distributions based on the methods are similar.



**Figure 9.** Multivariate kernel distribution for  $r_1$  and  $r_4$  using two-level MCMC with  $\epsilon = 1$  and  $\epsilon = 2^{-6}$ .

Finally, we report the computational cost savings of using the DE-assisted two-level MCMC method compared to the one-level MCMC method for Example 2. The values of  $\beta$  and  $r_1$  for Example 2 are documented in Table 8. From the table, we see that using the DE-assisted two-level MCMC method saves us roughly 12% of the RTE evaluations, for both values of  $\epsilon$  that we considered.

**Table 8.** Computational cost comparison.

	$\epsilon = 1$	$\epsilon = 2^{-6}$
$\beta$	0.7230	0.9305
$r_1$	0.6450	0.8180
$r_1/\beta$	0.8921	0.8791

We again emphasize that  $\beta$  and  $r_1$  highly depend on the sampling strategy used, the initial values of the MCMC parameters, and the step size.

## 6. Conclusions

In this paper, we solve the inverse problem for the RTE using Bayesian inference, which gives us a posterior distribution for the quantity of interest. Accessing concrete information about this distribution using MCMC can be difficult, because the RTE is fairly expensive to solve. However, in the strong scattering regime with the Knudsen number  $\epsilon$  going to zero, the posterior distribution based on the RTE converges to the posterior distribution based on the DE. With this knowledge, we employ a two-level MCMC technique, in which the posterior based on the DE is used to create good samples for the posterior based on the RTE. In this way, we save time by rejecting poor samples using the posterior based on the DE distribution, which is fast to compute. This reduces the number of times we must solve the inverse problem for the RTE, which improves the efficiency of the computation. We also prove that the second-level acceptance rate of this method is close to one in the diffusion limit, meaning that samples accepted at the first level have a high chance of being accepted at the second level, if  $\epsilon$  is small.

**Author Contributions:** Conceptualization, Q.L. and K.N.; Methodology, Q.L. and K.N.; Software, Q.L. and K.N.; Validation, Q.L. and K.N.; Formal Analysis, Q.L. and K.N.; Investigation, Q.L. and K.N.; Resources, Q.L. and K.N.; Data Curation, Q.L. and K.N.; Writing-Original Draft Preparation, Q.L. and K.N.; Writing-Review & Editing, Q.L. and K.N.; Visualization, Q.L. and K.N.; Supervision, Q.L.; Project Administration, Q.L.; Funding Acquisition, Q.L.

**Funding:** This research was funded by NSF DMS 1619778, NSF DMS 1750488 and NSF TRIPDS 1740707.

**Acknowledgments:** The authors would like to thank Chi Zhang for her work running the MCMC code and help editing the draft.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arridge, S.R.; Schotland, J.C. Optical tomography: Forward and inverse problems. *Inverse Probl.* **2009**, *25*, 123010. [[CrossRef](#)]
2. Bal, G. Inverse transport theory and applications. *Inverse Probl.* **2009**, *25*, 053001. [[CrossRef](#)]
3. Egger, H.; Schlottbom, M. Numerical methods for parameter identification in stationary radiative transfer. *Comput. Optim. Appl.* **2015**, *62*, 67–83. [[CrossRef](#)]
4. Ren, K. Recent developments in numerical techniques for transport-based medical imaging methods. *Commun. Comput. Phys.* **2010**, *8*, 1–50. [[CrossRef](#)]
5. Klose, A.D.; Netz, U.; Beuthan, J.; Hielscher, A.H. Optical tomography using the time-independent equation of radiative transfer—Part 1: Forward model. *J. Quant. Spectrosc. Radiat. Transf.* **2002**, *72*, 691–713. [[CrossRef](#)]
6. Klose, A.D.; Netz, U.; Beuthan, J.; Hielscher, A.H. Optical tomography using the time-independent equation of radiative transfer—Part 2: Inverse model. *J. Quant. Spectrosc. Radiat. Transf.* **2002**, *72*, 715–732. [[CrossRef](#)]
7. Mourant, J.R.; Freyer, J.; Hielscher, A.H.; Eick, A.; Shen, D.; Johnson, T. Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics. *Appl. Opt.* **1998**, *37*, 3586–3593. [[CrossRef](#)] [[PubMed](#)]
8. Choulli, M.; Stefanov, P. Inverse scattering and inverse boundary value problems for the linear Boltzmann equation. *Commun. Partial Differ. Equ.* **1996**, *21*, 763–785. [[CrossRef](#)]
9. Wang, J.N. Stability estimates of an inverse problem for the stationary transport equation. *Annales de l'I.H.P. Physique Théorique* **1999**, *70*, 473–495.

10. Sylvester, J.; Uhlmann, G. A global uniqueness theorem for an inverse boundary value problem. *Ann. Math.* **1987**, *125*, 153–169. [[CrossRef](#)]
11. Alessandrini, G. Stable determination of conductivity by boundary measurements. *Appl. Anal.* **1988**, *27*, 153–172. [[CrossRef](#)]
12. Hielscher, A.H.; Alcouffe, R.; Barbour, R. Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. *Phys. Med. Biol.* **1988**, *43*, 1285–1302. [[CrossRef](#)]
13. Arridge, S.R. Optical tomography in medical imaging. *Inverse Probl.* **1999**, *15*, R41–R93. [[CrossRef](#)]
14. Chen, K.; Li, Q.; Wang, L. Stability of stationary inverse transport equation in diffusion scaling. *Inverse Probl.* **2018**, *34*, 025004. [[CrossRef](#)]
15. Lai, R.-Y.; Li, Q.; Uhlmann, G. Inverse problems for the stationary transport equation in the diffusion scaling. *arXiv* **2018**, arXiv:1808.02071.
16. Stuart, A.M. Inverse problems: A Bayesian perspective. *Acta Numer.* **2010**, *19*, 451–559. [[CrossRef](#)]
17. Dashti, M.; Stuart, A.M. The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification*; Springer International Publishing: Basel, Switzerland, 2017.
18. Bal, G.; Langmore, I.; Marzouk, Y. Bayesian inverse problems with Monte Carlo forward models. *Inverse Probl. Imaging* **2013**, *7*, 81–105.
19. Giles, M.B. Multi-level Monte Carlo path simulation. *Oper. Res.* **2008**, *56*, 607–617. [[CrossRef](#)]
20. Heinrich, S. Multilevel monte carlo methods. In *Large-Scale Scientific Computing*; Birkhäuser Basel: Basel, Switzerland, 2001.
21. Christen, J.A.; Fox, C. Markov chain monte carlo using an approximation. *J. Comput. Graph. Stat.* **2005**, *14*, 795–810. [[CrossRef](#)]
22. Fox, C.; Nicholls, G. Sampling conductivity images via MCMC. In *The Art and Science of Bayesian Image Analysis*; Leeds University Press: Leeds, UK, 1997; pp. 91–100.
23. Peherstorfer, B.; Willcox, K.; Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* **2018**, *60*, 550–591. [[CrossRef](#)]
24. Newton, K.; Li, Q.; Stuart, A.M. Diffuse optical tomography in the Bayesian framework. *arXiv* **2019**, arXiv:1902.10317.
25. Kullback, S. *Information Theory and Statistics*; Reprint of the second (1968) ed.; Dover Publications, Inc.: Mineola, NY, USA, 1997.
26. Bensoussan, A.; Lions, L.J.; Papanicolaou, G.C. Boundary layers and homogenization of transport processes. *Publ. Res. Inst. Math. Sci.* **1979**, *15*, 53–157. [[CrossRef](#)]
27. Bardos, C.; Santos, R.; Sentis, R. Diffusion approximation and computation of the critical size. *Trans. Am. Math. Soc.* **1984**, *284*, 617–649. [[CrossRef](#)]
28. Wu, L.; Guo, Y. Geometric correction for diffusive expansion of steady neutron transport equation. *Commun. Math. Phys.* **2015**, *336*, 1473–1553. [[CrossRef](#)]
29. Li, Q.; Wang, L. Implicit asymptotic preserving method for linear transport equations. *Commun. Comput. Phys.* **2017**, *22*, 157–181. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).