Monte Carlo methods for light propagation in biological tissues

Laura Vinckenbosch, Samy Tindel, Magalie Thomassin, Céline Lacaux, Obara Tiphaine

Monte Carlo methods for light propagation in biological tissues

Laura Vinckenbosch[1], Samy Tindel[1,2,3], Magalie Thomassin[4,5], Céline Lacaux[1,2,3], and Tiphaine Obara[4,5]

[1]Inria, BIGS, Villers-lès-Nancy, F-54600, France
[2]Université de Lorraine, Institut Élie Cartan de Lorraine, UMR 7502, Vandœuvre-lès-Nancy, F-54506, France
[3]CNRS, Institut Élie Cartan de Lorraine, UMR 7502, Vandœuvre-lès-Nancy, F-54506, France
[4]Université de Lorraine, CRAN, UMR 7039, 9, avenue de la forêt de Haye, Vandœuvre-lès-Nancy Cedex, 54516, France
[5]CNRS, CRAN, UMR 7039, France

November 21, 2013

Contents

Abstract

Light propagation in turbid media is driven by the equation of radiative transfer. We give a formal probabilistic representation of its solution in the framework of biological tissues and we implement algorithms based on Monte Carlo methods in order to estimate the quantity of light that is received by an homogeneous tissue when emitted by an optic fiber. A variance reduction method is studied and implemented, as well as a Markov chain Monte Carlo method based on the Metropolis-Hastings algorithm. The resulting estimating methods are then compared to the so-called Wang-Prahl (or Wang) method. Finally, the formal representation allows to derive a non-linear optimization algorithm close to Levenberg-Marquardt that is used for the estimation of the scattering and absorption coefficients of the tissue from measurements.

Keywords. Light propagation, equation of radiative transfer, Markov chain Monte Carlo methods, parameter estimation.

2010 Mathematics Subject Classification. Primary : 78M31; Secondary : 65C40, 60J20.

1 Introduction

Photodynamic therapy (PDT) is a type of phototherapy used for treating several diseases such as acne, bacterial infection, viruses and some cancers. The aim of this treatment is to kill pathological cells with a photosensitive drug that is absorbed by the target cells and that is then activated by light. For appropriate wavelength and power, the light beam makes the photosensitizer produce singlet oxygen at high doses and induces the apoptosis and necrosis of the malignant cells. This technique is also used in some framework to localize lesions.

Since the late 1970s, there have been a large development of this kind of therapy for cancers following the discovery of the hematoporphyrin derivative as an effective tumor-localizing photosensitizer. Restricted, at the beginning, to skin because of its easy access, this localization and treatment tool spread rapidly to a wide range of lesions such as bladder, lung, brain or liver cancer. See [27, 28] for a review on PDT and [9] for a review on tumor localization by hematoporphyrin derivative fluorescence.

Due to the complexity of interactions between physical, chemical and biological aspects and due to the high cost and the poor reproducibility of the experiments, there is a real demand in good mathematical and physical models which might help to better control and understand PDT responses. The two main questions to which these models should answer are:

1. What is the optimal shape and position of the light source in order to optimize the damage on malignant cells?

2. Is there a way to identify the physical parameters of the tissue which drive the light propagation?

The light propagation phenomenon involves three processes: absorption, emission and scattering that are described by the so-called equation of radiative transfer (ERT), see [7]. In general, this equation does not admit any explicit solution, and its study relies on methods of approximation. One of them is its approximation by the diffusion equation and the use of finite elements methods to solve it numerically (see for example [2]). An other approach, which appeared in the 1970s, is the simulation of particle-transport with Monte Carlo (MC) method (see [5, 6, 30] and references therein). This method has been extended by several authors in order to deal with the special case of biological tissues and there is now a consensus in favor of the algorithm proposed by L. Wang and S. L. Jacques in [26] and initiated by S. A. Prahl in [19] and S. A. Prahl et al. in [20]. This method is based on a probabilistic interpretation of the trajectory of a photon. It is widely used and there exist now turnkey softwares based on this method. However, this method is time consuming in 3D and the associated softwares lie inside some kind of black boxes. Due to a slight lack of formalism, it is difficult to speed it up while controlling the estimation error, or to adapt it to inhomogenous tissues such as infiltrating gliomas. Finally, even though there exist several methods in order to estimate the optical parameters of the tissue (see for example [16, 11, 3, 18]), one still misses formal representations that answer to the questions of identifiability.

In the current work, we wish to give a new point of view on simulation issues for ERT, starting from the very beginning. Specifically, we address three problems:

(i) We first derive a rigorous probabilistic representation of the solution to the ERT in homogeneous tissues, which will help us to propose an alternative MC method to Wang's algorithm [26]. Then we also propose a variance reduction method.

(ii) Interestingly enough, our formulation of the problem also allows us to design quite easily a Markov chain Monte Carlo (MCMC) method based on Metropolis-Hastings algorithm. We have compared both MC and MCMC algorithms, and our simulation results show that the plain MC method is still superior in case of an homogeneous tissue. However, MCMC methods induce quick mutations, which paves the way to very promising algorithms in the inhomogenous case. We shall go back to this issue in a subsequent publication.

(iii) Finally we handle the inverse problem (of crucial importance for practitioners), consisting in estimating the optical coefficients of the tissue according to a series of measurements. Towards this aim, we derive a probabilistic representation of the variation of the fluence with respect to the absorption and scattering coefficients. This leads us to the implementation of a Levenberg-Marquardt type algorithm that gives an approximate solution to the inverse problem.

Our work should thus be seen as a complement to the standard algorithm described in [26]. Focusing on a rigorous formulation, it opens the way to a thorough analysis of convergence, generalizations to MCMC type methods and a mathematical formulation of the inverse problem.

The paper is organized as follows. We derive the probabilistic representation of the solution to the ERT in Section 2. In Sections 3 and 4, we describe the MC and MCMC algorithms which are compared to Wang's algorithm in Section 5. Finally, the sensitivity of the measures with respect to the optical parameters of the medium, as well as their estimation are treated in Section 6.

2 Probabilistic representation of the fluence rate

2.1 The radiative transfer equation

In the particular case of biological tissues the equation of radiative transfer takes the following form, see [19]. Let $D = \mathbb{R}^3$ be the set of positions in the tissue and $\mathbb{S}^2$ be the unit sphere in $\mathbb{R}^3$. The *quantity of light* at x in the direction ω is denoted by $L(x,\omega)$ and satisfies

$$L(x,\omega) = L_e(x,\omega) + TL(x,\omega), \qquad x \in D,\, \omega \in \mathbb{S}^2, \tag{2.1}$$

where $L_e(x,\omega)$ is the *emitted light* from x in direction ω and T is a linear operator defined by

$$TL(x,\omega) = \mu_s \int_{\mathbb{R}_+} dr\, \exp(-\mu r) \int_{\mathbb{S}^2} d\sigma(\hat{\omega})\, f(\omega, x-\omega r, \hat{\omega})\, L(x-\omega r, \hat{\omega}), \tag{2.2}$$

where $\mu_s > 0$ is the *scaterring coefficient*, μ the *total absorption coefficient* (or *attenuation coefficient*), f the so-called *bidirectional scattering distribution function* and σ the uniform probability measure on the unit sphere $\mathbb{S}^2$. The total absorption coefficient μ is the sum of the scaterring coefficient μ_s and the *simple absorption coefficient* $\mu_a > 0$. In the following, we shall always assume that the *albedo coefficient* $\rho := \frac{\mu_s}{\mu}$ satisfies $\rho < 1$.

In the particular case of homogeneous biological tissues, the scattering function is given by the so-called *Henyey-Greenstein function*, see [13], that is

$$f(\omega, x, \hat{\omega}) = f_{HG}(\omega, \hat{\omega}) = \frac{1-g^2}{(1+g^2-2g\langle\omega,\hat{\omega}\rangle)^{3/2}}, \quad \omega, \hat{\omega} \in \mathbb{S}^2,\, \forall x \in D, \tag{2.3}$$

where the constant $g \in [0,1)$ is the *anisotropy factor* of the medium. The function $\hat{\omega} \mapsto f_{HG}(\omega,\hat{\omega})$ is a bounded and infinitely differentiable probability density function on $\mathbb{S}^2$ with respect to the uniform probability σ. It only depends on the angle θ between ω and $\hat{\omega}$ since $\langle\omega,\hat{\omega}\rangle = \cos(\theta)$. The greater the anisotropic parameter g, the more likely the large values of $\cos(\theta)$ and the less the scattering of the ray (see Fig. 1).

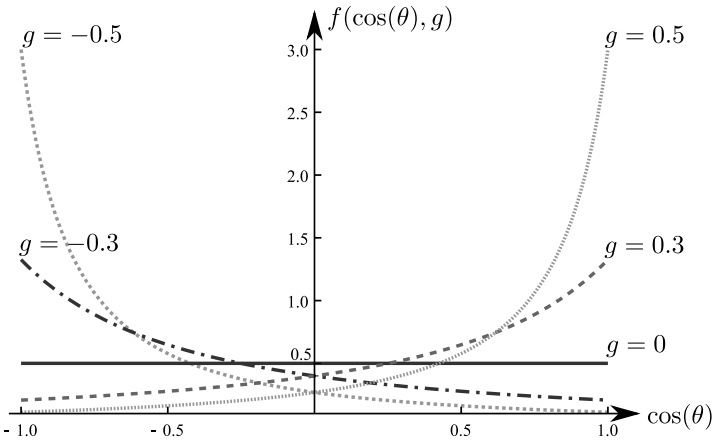

Figure 1: Density function of $\cos(\theta) = \langle \omega, \hat{\omega} \rangle$ when $\hat{\omega}$ is scattered according to the Henyen-Greenstein function $f_{HG}(\omega, \cdot)$ for several values of the anisotropy parameter g.

2.2 Neumann series expansion of the solution

In general, (2.1) admits no analytical solution and a classical way to express its solution is to expand the equation in Neumann series. This method is based on the next classical and general result.

Theorem 1 ([29] p.69)**.** *Let B be a Banach space equipped with a norm $\|\cdot\|$ and A a linear operator on B. If $\|A\| < 1$, then the Neumann series $\sum_{n=0}^{\infty} A^n$ converges, the operator $Id - A$ is invertible and the equation $x = Ax + x_0$ admits a unique solution given by*

$$x = (Id - A)^{-1} x_0 = \sum_{n=0}^{\infty} A^n x_0,$$

for any $x_0 \in B$.

In order to apply Theorem 1 in our context, let us now bound the norm of the operator T defined above by (2.2).

Lemma 2. *Let $L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$ be the Banach space of essentially bounded real-valued functions defined on $\mathbb{S}^2 \times \mathbb{R}^3$. The operator $T : L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R}) \to L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$ defined in (2.2), with f given by (2.3), satisfies $\|T\| = \rho < 1$, where we recall that we have set $\rho := \frac{\mu_s}{\mu} < 1$.*

Proof. Let $\ell \in L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$. We have

$$\begin{aligned} |T\ell(x,\omega)| &\leqslant \mu_s \int_{\mathbb{R}_+} dr\, \exp(-\mu r) \int_{\mathbb{S}^2} d\sigma(\hat{\omega})\, |f_{HG}(\omega, \hat{\omega})\, \ell(x - \omega r, \hat{\omega})| \\ &\leqslant \mu_s \|\ell\|_\infty \int_{\mathbb{R}_+} dr\, \exp(-\mu r) = \frac{\mu_s}{\mu} \|\ell\|_\infty, \end{aligned}$$

since f_{HG} is a density function on $\mathbb{S}^2$. Thus, $\|T\| \leqslant \frac{\mu_s}{\mu}$ and since $T\mathbf{1} \equiv \frac{\mu_s}{\mu}$, we obtain $\|T\| = \frac{\mu_s}{\mu}$ and the proof is complete. □

As a corollary of the previous considerations, we are able to derive an analytic expansion for the solution to equation (2.1):

Corollary 3. *If $L_e \in L^\infty((\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$, then the radiative transfert equation (2.1) with a phase function given by (2.3) admits a unique solution L in $L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$. Moreover, L can be decomposed*

as $L = \sum_{n=0}^{\infty} T^n L_e$*, where* $T^0 \equiv Id$ *and for* $n \geq 1$*,* T^n *is the linear operator on* $L^\infty(\mathbb{S}^2 \times \mathbb{R}^3; \mathbb{R})$ *defined by*

$$T^n \ell(x, \omega_0) = \mu_s^n \int_{\mathbb{R}_+^n} dr_1 \cdots dr_n \, \exp\left(-\mu \sum_{j=1}^{n} r_j\right) \int_{\mathbb{S}^{2n}} d\sigma^{\otimes n}(\omega_1, \ldots, \omega_n) \prod_{j=0}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) \ell\left(x - \sum_{k=0}^{n-1} \omega_k r_{k+1}, \omega_n\right). \quad (2.4)$$

Proof. Theorem 1 and Lemma 2 provide the existence and uniqueness of the solution, as well as its form. Formula (2.4) is then found by induction. □

Our next step is now to recast representation (2.4) into a probabilistic formula.

2.3 Probabilistic representation

The Neumann expansion of T enables us to express $\sum_{n=0}^{\infty} T^n L_e$ as an expectation. To this aim, let us introduce some notation. Let us define

$$\mathcal{A} = \bigcup_{n=1}^{\infty} \mathbb{S}^2 \times \mathcal{M}_n, \quad \text{with} \quad \mathcal{M}_n = \mathbb{R}_+^n \times \left(\mathbb{S}^2\right)^n. \quad (2.5)$$

We denote by $(\omega_0, \mathbf{r}, \boldsymbol{\omega})$ a generic element of $\mathcal{A}$ and by $(\omega_0, \mathbf{r}_n, \boldsymbol{\omega}_n)$ a generic element of $\mathbb{S}^2 \times \mathcal{M}_n$ for $n \geq 1$ with $\boldsymbol{r}_n = (r_1, \ldots, r_n)$ and $\boldsymbol{\omega}_n = (\omega_1, \ldots, \omega_n)$. If $(\omega_0, \mathbf{r}, \boldsymbol{\omega}) \in \mathcal{A}$, we set

$$|\mathbf{r}| = \sum_{n=1}^{\infty} n \, \mathbf{1}_{\mathbb{S}^2 \times \mathcal{M}_n}(\omega_0, \mathbf{r}, \boldsymbol{\omega}) \quad (2.6)$$

and call it *size* or *length* of the path. For $n \geqslant 1$, let

$$G_x^{(n)}(\omega_0, \mathbf{r}_n, \boldsymbol{\omega}_n) = L_e\left(x - \sum_{k=0}^{n-1} \omega_k r_{k+1}, \omega_n\right) \quad (2.7)$$

be defined on $\mathbb{S}^2 \times \mathcal{M}_n$ and let

$$G_x(\omega_0, \mathbf{r}, \boldsymbol{\omega}) = \sum_{n=1}^{\infty} G_x^{(n)}(\omega_0, \mathbf{r}, \boldsymbol{\omega}) \, \mathbf{1}_{\mathcal{M}_n}(\mathbf{r}, \boldsymbol{\omega}) \quad (2.8)$$

be a function on $\mathcal{A}$. Let $Y = (W_0, \mathbf{R}, \boldsymbol{W})$ be a $\mathcal{A}$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, whose law ν is given by

$$\nu(F) = \sum_{n=1}^{\infty} (1-\rho)\rho^{n-1} \nu_n\left(F \cap \left(\mathbb{S}^2 \times \mathcal{M}_n\right)\right), \quad (2.9)$$

with ν_n the probability measure on $\mathbb{S}^2 \times \mathcal{M}_n$ defined by

$$\nu_n(d\omega_0, d\mathbf{r}_n, d\boldsymbol{\omega}_n) = \mu^n \, e^{-\mu \sum_{j=1}^{n} r_j} \prod_{j=0}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) \gamma_{W_0}(d\omega_0) d\mathbf{r}_n \, \sigma^{\otimes n}(d\boldsymbol{\omega}_n), \quad (2.10)$$

where γ_{W_0} is any probability measure on $\mathbb{S}^2$. With these notations in hand, we have the following representation formula for the solution of (2.1).

Proposition 4. *The series* $\sum_{n=0}^{\infty} T^n L_e$ *can be expressed as*

$$L(x,\omega) = \sum_{n=0}^{\infty} T^n L_e(x,\omega) = L_e(x,\omega) + \frac{\rho}{1-\rho}\mathbb{E}\left[G_x(Y) \mid W_0 = \omega\right]. \tag{2.11}$$

Proof. The proof is straightforward starting from (2.4). □

Remark 5. In the following, we will call the random variable $Y = (W_0, \mathbf{R}, \boldsymbol{W})$ a *ray*. Notice that it does not correspond exactly to a ray of light in the physics sense, since Y has a finite length (though random) and since a given realization of Y does not carry the information due to light absorption. Also notice that Y owns a complete probabilistic description which allows to exactly simulate it (see Proposition 10 for simulation considerations).

2.4 Model for light propagation

Observe that our formula (2.11) induces a Monte Carlo procedure to estimate $L(x,\omega)$ for each $(x,\omega) \in \mathbb{R}^3 \times \mathbb{S}^2$ based on the simulation of independent copies of Y. Nevertheless this procedure is time consuming. Indeed, assuming that the light is only emitted by an optical fiber, many realizations y of Y lead to a null contribution in the estimation of $L(x,\omega)$. Our aim is now to accelerate our simulation by means of a coupling between random variables corresponding to different (x,ω). Towards this aim, we now focus on an averaged model for light propagation.

Let thus $V \subset \mathbb{R}^3$ be a cube whose center coincides with the origin. We discretize it into a partition of K smaller cubes (voxels in the image processing terminology) $\{V_k, k = 0, \dots, K-1\}$, whose volume equals h^3, $h \in \mathbb{R}_+$ and such that the origin is the center of V_0. Let us denote by x_k the center of the voxel V_k. We work under the following simplified assumption for the form of the light source:

Hypothesis 6. *We assume that the only emission of light in the domain* V *comes from the optical fiber. Let* $C^{2\alpha} \subset \mathbb{S}^2$ *denote the cone with opening angle* 2α*, whose summit is placed at the origin and whose axis follows* $-\vec{e}_3$*. The light source is defined by* $S = \{(x,\omega) : x \in V_0,\, \omega \in C^{2\alpha}\}$*. We assume that the emission of light satisfies*

$$L_e(x,\omega) = c\,\mathbf{1}_{V_0 \times C^{2\alpha}}(x,\omega) := \begin{cases} c, & \text{if } (x,\omega) \in V_0 \times C^{2\alpha}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.12}$$

where $c > 0$ *is a given constant.*

This model remains close to reality and it is possible to refine it by weighting the light directions of the source in order to stick better to the shape of the fiber. With Hypothesis 6 in mind, we are interested in estimating the *fluence* at the center of the voxels V_k, $k \neq 0$, that is the mean light intensity averaged in all directions with respect to a certain probability measure γ_{W_0}:

$$L(x_k) := \int_{\mathbb{S}^2} L(x_k, \omega_0)\, \gamma_{W_0}(d\omega_0). \tag{2.13}$$

This quantity admits a nice probabilistic representation.

Proposition 7. *Let* $k \in \{0, \dots, K-1\}$ *and let* $Y = (W_0, \boldsymbol{R}, \boldsymbol{W})$ *be a random variable with distribution* ν *defined by* (2.9)*. Then the fluence* $L(x_k)$ *at the center* x_k *of the voxel* V_k*, which is defined by* (2.13)*, can also be expressed as*

$$L(x_k) = L_e(x_k) + \frac{\rho\, c}{1-\rho}\,\mathbb{P}\left(x_k - \sum_{i=0}^{|\boldsymbol{R}|-1} R_{i+1} W_i \in V_0,\, W_{|\boldsymbol{R}|} \in C^{2\alpha}\right) \tag{2.14}$$

where we recall that the length $|\boldsymbol{R}|$ *of the ray* Y *is defined by* (2.6)*.*

Proof. Noting that γ_{W_0} is the distribution of W_0 and invoking Proposition 4 we get

$$L(x_k) = L_e(x_k) + \frac{\rho}{1-\rho}\,\mathbb{E}\left[G_{x_k}(Y)\right]$$

where G_{x_k} is defined by (2.8). Then using equations (2.7) and (2.6), we get

$$G_{x_k}(Y) = L_e\left(x_k - \sum_{i=0}^{|\boldsymbol{R}|-1} R_{i+1}W_i,\, W_{|\boldsymbol{R}|}\right).$$

Now applying Hypothesis 6, the random variable $G_{x_k}(Y)$ can be expressed as

$$G_{x_k}(Y) = c\,\mathbf{1}_{\left\{x_k - \sum_{i=0}^{|\boldsymbol{R}|-1} R_{i+1}W_i \in V_0,\, W_{|\boldsymbol{R}|} \in C^{2\alpha}\right\}},$$

which finishes our proof. □

Let us, from now on, consider the particular case where W_0 is distributed uniformly on the unit sphere, that is $\gamma_{W_0}(d\omega_0) \equiv \sigma(d\omega_0)$. In this particular (but very natural) case, the following property will be the basis of our coupling method:

Proposition 8. *If $\gamma_{W_0}(d\omega_0) \equiv \sigma(d\omega_0)$, then for all $n \in \mathbb{N}^*$, the probability measure ν_n defined in* (2.10) *satisfies*

$$\nu_n(d\omega_0, d\mathbf{r}_n, d\omega_1, \ldots, d\omega_n) = \nu_n(d\omega_n, d\mathbf{r}_{\pi_n}, d\omega_{n-1}, \ldots, d\omega_0),$$

where π_n is any permutation of $(r_1, \ldots, r_n)$.

Proof. Notice first that in the definition of ν_n in (2.10), $(r_1, \ldots, r_n)$ are mutually independent, so the vector $(r_{\pi_n(1)}, \ldots, r_{\pi_n(n)})$ has the same law as $(r_1, \ldots, r_n)$. Moreover, the phase function is symmetric: $f_{HG}(\omega, \cdot) = f_{HG}(\cdot, \omega)$. Replacing $\gamma_{W_0}(d\omega_0)$ by $\sigma(d\omega_0)$ in (2.10) and replacing the terms $f_{HG}(\omega_j, \omega_{j+1})$ by $f_{HG}(\omega_{j+1}, \omega_j)$, we obtain immediately the desired property. □

Corollary 9. *Assume that $\gamma_{W_0}(d\omega_0) \equiv \sigma(d\omega_0)$ and consider $Y = (W_0, \boldsymbol{R}, \boldsymbol{W})$ a random variable with distribution ν defined by* (2.9). *Then for all $n \in \mathbb{N}^*$, on the event $\left\{Y \in \mathbb{S}^2 \times \mathcal{M}_n\right\}$, for any $j = 1, \ldots, n$, the marginal distribution γ_{W_j} of any direction W_j is the uniform probability on $\mathbb{S}^2$.*

Proof. This is a direct consequence of the previous proposition. To see this, it suffices to integrate the law ν_n defined in (2.10) with respect to all variables except for W_j and use the fact that $f_{HG}(\omega, \cdot) = f_{HG}(\cdot, \omega)$ is a density function on $\mathbb{S}^2$. □

Now, instead of seeing Y as a ray starting at x_k which possibly hits the light source $V_0 \times C^{2\alpha}$, we can imagine that it starts at the center of the light source and it possibly hits the voxel V_k in any direction.

Proposition 10. *Assume that $\gamma_{W_0}(d\omega_0) \equiv \sigma(d\omega_0)$. Then, for any $1 \leqslant k \leqslant K-1$*

$$L(x_k) = \frac{\rho c(1-\cos\alpha)}{2(1-\rho)}\,\mathbb{P}\left(S_N \in V_k\right), \tag{2.15}$$

where S_N is a random variable that can be exactly simulated in the following way: consider a geometric random variable N is a with parameter $1-\rho$ and for all $n \geqslant 1$, set

$$S_n = \sum_{i=1}^{n} R_i W_i,$$

with $(R_i)_{i\geq 1}$ and $(W_i)_{i\geq 0}$ satisfying the following assertions:

- *$(R_i)_{i\geq 1}$ is a sequence of independent identically distributed (i.i.d.) exponentially random variables of parameter μ such that $\mathbb{E}(R_i) = \mu^{-1}$.*

- *W_0 is uniformly distributed on the cone $C^{2\alpha}$.*
- *for any $i \geq 1$, the conditional distribution of W_i given $(W_0, \dots, W_{i-1}) = (\omega_0, \dots, \omega_i)$ is $f_{HG}(\omega_{i-1}, \omega_i)\sigma(d\omega_i)$.*
- *N, $(R_i)_{i\geqslant 1}$ and $(W_i)_{i\geqslant 0}$ are independent.*

Proof. Let $k \in \{1, \dots, K-1\}$. Notice that, if $V_k + x$ denotes the translation of the voxel V_k by the vector $x \in \mathbb{R}^3$, then it is clear that $V_0 + x_k = V_k$. Furthermore, since $k \neq 0$, $L_e(x_k) = 0$. Therefore, we can rewrite (2.14) as

$$L(x_k) = \frac{\rho c}{1-\rho}\mathbb{P}\left(\sum_{i=0}^{|\boldsymbol{R}|-1} R_{i+1}W_i \in V_k, W_{|\boldsymbol{R}|} \in C^{2\alpha}\right), \tag{2.16}$$

where the distribution of $Y = (W_0, \boldsymbol{R}, \boldsymbol{W})$ is the probability measure ν defined by (2.9) with $\gamma_{W_0}(d\omega_0) \equiv \sigma(d\omega_0)$. Then,

$$L(x_k) = \frac{\rho c}{1-\rho}\sum_{n=1}^{+\infty}(1-\rho)\rho^{n-1}\mathbb{P}\left(\sum_{i=0}^{n-1} R_{i+1}W_i \in V_k, W_n \in C^{2\alpha}\right)$$

where the distribution of $(W_0, R_1, \dots, R_n, W_1, \dots, W_n)$ is the probability ν_n defined by (2.10). Therefore, applying Proposition 8, we get

$$L(x_k) = \frac{\rho c}{1-\rho}\sum_{n=1}^{+\infty}(1-\rho)\rho^{n-1}\mathbb{P}\left(\sum_{i=1}^{n} R_iW_i \in V_k, W_0 \in C^{2\alpha}\right).$$

By definition of ν_n, we easily see that

$$\mathbb{P}\left(\sum_{i=1}^{n} R_iW_i \in V_k,\, W_0 \in C^{2\alpha}\right) = \sigma(C^{2\alpha})\mathbb{P}\left(S_n \in V_k\right). \tag{2.17}$$

This allows to replace, in the description of the random walk $S = (S_n)_{n\geqslant 1}$, the random variable $W_0 \sim \mathcal{U}(\mathbb{S}^2)$ by a random variable W_0' uniformly distributed on the cone $C^{2\alpha}$. Since N is a geometric random variable of parameter $1-\rho$ independent with the random walk S, this leads to equation (2.15).

Finally, the simulation of the sequence $(W_i)_{i\geqslant 1}$ is obtained as follows. The direction W_i of the i-th step of the random walk is sampled relatively to the direction W_{i-1}. We sample the spherical angles (Θ_i, Φ_i) *between* the two directions according to the Henyey-Greenstein phase function. Namely, $\cos(\Theta_i) = \langle W_{i-1}, W_i\rangle$ owns the following invertible cumulative distribution function

$$F^{-1}(y) = \frac{1}{2g}\left(1 + g^2 - \left(\frac{1-g^2}{1-g+2gy}\right)^2\right), \qquad y \in [0,1]$$

and Φ_i, the azimuth angle of W_i in the frame linked to W_{i-1}, is uniformly distributed on $[0, 2\pi]$, see (2.3). To recover the cartesian coordinates of the directions, we inductively apply appropriate changes of frame. The corresponding formulas can be found in [19, p. 37]. □

Remark 11. Notice that S_N does not depend on the voxel x_k under consideration. This permits to use a single sample of realizations of this random variable in order to estimate the right hand side of (2.15) for all $k \in \{1, \dots, K-1\}$ simultaneously. We call this improvement coupling, meaning that the random variables related to the Monte Carlo evaluations at different voxels are completely correlated.

3 Monte Carlo approach with variance reduction

In the last section, we have derived a probabilistic representation of $L(x_k)$ for every voxel V_k by means of the arrival position of a random walk $(S_n)_{n\geqslant 1}$ stopped at a geometric time. This classically means that $L(x_k)$ can be approximated by MC methods. We first derive the expression of the approximate fluence by means of the basic MC method and then describe the variance reduction method that we implemented.

Proposition 12. *Let us consider a random walk $S = (S_n)_{n\geq 1}$ and a geometric random variable N as defined in Proposition 10. Let $(S^i, N_i)_{1\leq i\leq M}$ be M independent copies of (S, N). Then, for $k = 1, \dots, K-1$,*

$$\widehat{L}_{\mathrm{MC}}(x_k) := \frac{\rho c\,(1-\cos\alpha)}{2(1-\rho)M} \sum_{i=1}^{M} \mathbf{1}_{\left\{S^i_{N_i}\in V_k\right\}} \tag{3.1}$$

is an unbiased and strongly consistent estimator of $L(x_k)$.

Proof. This statement follows simply from the discussion of the previous section and the law of large numbers (LNN). □

In addition to this proposition, let us highlight the fact that central limit theorem provides confidence intervals for the estimators $\widehat{L}_{\mathrm{MC}}(x_k)$. Furthermore, owing to Remark 11, the family $(S^i, N_i)_{1\leq i\leq M}$ enables to estimate the fluence $L(x_k)$ for all $k = 1, \dots, K-1$ at once.

The reader should be aware of the fact that the quantity $1-\rho$ is small in general in biological tissues, which means that the size of the ray will often be large. Typical values of the parameters are provided in Table 1. Therefore, sampling a ray is relatively time consuming and it is necessary to improve the basic Monte Carlo algorithm in order to reduce the variance of the estimates.

	μ_s	μ_a	μ	ρ	g
healthy tissue	280 cm^{-1}	0.57 cm^{-1}	280.57 cm^{-1}	0.998	0.9
tumor	73 cm^{-1}	1.39 cm^{-1}	74.39 cm^{-1}	0.981	0.9

Table 1: Values given by [1] for the optical parameters in the rat brain for a wavelength $\lambda = 632$ nm (red light).

Furthermore, because of the formulation (2.16), only the last point of each whole ray is used in the estimation. It is however possible to take into account more points of the rays and still have unbiased estimators. Finally, the angular symmetry of the problem, allows us to replicate observed rays by applying rotation. We took these two considerations into account and named the resulting method **Monte Carlo with some points (MC-SOME)**. The idea is to firstly draw some random walks which share the same initial direction and to pick a given number of points of each walk. Then, we apply rotations to that set of points with respect to different initial directions. We finally count the number of points in each voxel. This artificially increase the size of the samples and thus reduce the variance of our estimation of $L(x_k)$. Specifically, the resulting estimator is given by :

Definition 13 (MC-SOME). Let $M, M_{\mathrm{points}}, M_{\mathrm{rot}} \in \mathbb{N}^*$ be the parameters of the method. Let us assume that the following assertions holds:

- $(W_0^j)_{1\leq j\leq M_{\mathrm{rot}}}$ are i.i.d. copies of $W_0 \sim \mathcal{U}(C^{2\alpha})$.
- N_i^ℓ, $1 \leqslant i \leqslant M$, $1 \leqslant \ell \leqslant M_{\mathrm{points}}$, are i.i.d. copies of a geometric random variable with parameter $1-\rho$.
- $\left(S_n^i\right)_{n\geq 1}$, $1 \leq i \leq M$, are i.i.d. copies of the random walk S defined in Proposition 10, all sharing the same initial direction W_0^1.
- The sequences $\left(N_i^\ell\right)_{i,\ell}$ and $\left(W_0^j, S_n^i\right)_{i,j}$ are independent.

Let $S^{i,j}_{N_i^\ell}$ denotes the N_i^ℓ-th point of the i-th random walk $\left(S_n^i\right)_{n\geq 1}$ after a rotation corresponding to the j-th initial direction $W_0 = W_0^j$.

Then, for $k \in \{1, \dots, K-1\}$, the MC-SOME estimate of $L(x_k)$ is defined by

$$\widehat{L}_{\text{MC-SOME}}(x_k) = \frac{\rho c\,(1-\cos\alpha)}{2(1-\rho)M_{\text{rot}}M_{\text{points}}M} \sum_{i=1}^{M} \sum_{\ell=1}^{M_{\text{points}}} \sum_{j=1}^{M_{\text{rot}}} \mathbf{1}_{\left\{S_{N_i^\ell}^{i,j} \in V_k\right\}}. \tag{3.2}$$

This estimator is unbiased and strongly consistent. Its construction is illustrated in Figure 2 and it will be compared to Wang's algorithm estimator in Section 5.

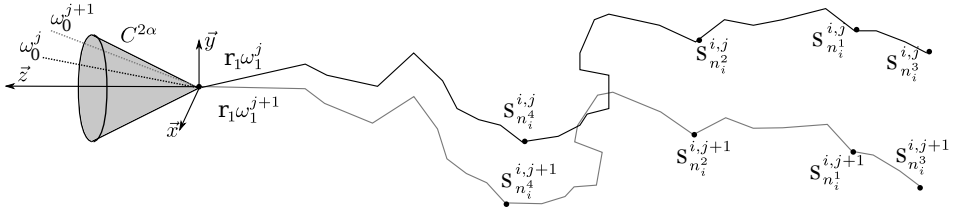

Figure 2: Description of MC-SOME method. In this example, the grey path is a rotation of the black one with respect to its initial direction ω_0^{j+1} and $M = 4$.

4 A Metropolis-Hastings algorithm for light propagation

Inspired by results in computer graphics, see [14, 23, 25], we implemented a Metropolis Hastings algorithm which is a *Markov chain Monte Carlo method* (MCMC) by random walk [12, 21]. We shall first discuss general principles and then practical implementation issues.

For simplicity reasons, by slightly abusing the notations, we identify the stopped random walk $S = (S_i)_{1 \leqslant i \leqslant N}$ of Proposition 10 with a ray and consider that S is a $\mathcal{A}$-valued random variable distributed according to ν in (2.9), with $\gamma_{W_0} = \mathcal{U}(C^{2\alpha})$ in (2.10). A realization of the walk S stopped at $N = n$ will be indifferently referred to by $(S_1, \dots, S_n)$, by $(\omega_0, r_1, \dots, r_n, \omega_1, \dots, \omega_n)$ or by $(\theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_n, \theta_n, \varphi_n)$ where (θ_0, φ_0) are the spherical coordinates of ω_0 and for $1 \leqslant i \leqslant n$, $\cos(\theta_i) = \langle \omega_{i-1}, \omega_i \rangle$ and φ_i is the azimuth angle of ω_i in the frame linked to ω_{i-1}.

4.1 General principle

For a given $\omega_0 \in C^{2\alpha}$ and for $1 \leqslant k \leqslant K-1$, we are willing to estimate the conditional probability $\mathbb{P}\left(S_N \in V_k \mid W_0 = \omega_0\right)$ by generating a Markov chain whose steady-state measure is the conditional distribution $\nu|_{W_0=\omega_0}$ and by applying LNN for ergodic Markov chains. We then combined this estimation with the classical LNN sampling the initial direction W_0 in $C^{2\alpha}$ to obtain an estimate of $L(x_k)$ viewed as in (2.15).

An overview of the MCMC dynamics in this context is the following: Let $\omega_0 \in C^{2\alpha}$ be a fixed initial direction. The Markov chain starts at time $t = 1$ in the state $S(1) \in \mathcal{A}$ with $W_0 = \omega_0$. At each time $t \in \mathbb{N}^*$, a move (mutation) is proposed from the current state $S(t)$ to the state $S'(t)$ according to a *proposal density* $q(r, s)$ and such that the initial direction of $S'(t)$ is still ω_0. The chain then jumps to $S'(t)$ with *acceptance probability* $\alpha(S(t), S'(t))$ or stay in $S(t)$ with probability $1 - \alpha(S(t), S'(t))$. This is described in pseudo-code in Algorithm 1. The MCMC simulation generates a Markov chain $\{S(t); t \geqslant 1\}$ on the space of rays $\mathcal{A}$ whose steady-state measure is the desired distribution $\nu|_{W_0=\omega_0}$ (see [24, §2.3.1]).

Algorithm 1 Metropolis-Hastings algorithm for light propagation

Initialization:
draw ω_0 uniformly on $\mathcal{C}^{2\alpha}$,
draw $S(1)$ according to $\nu|_{W_0=\omega_0}$
for $t = 1$ to $T-1$ **do**
  $S'(t) \sim q(S(t), \cdot)$
  $\alpha(S(t), S'(t)) \leftarrow \min\left\{1, \dfrac{\nu|_{W_0=\omega_0}(S'(t))\ q(S'(t), S(t))}{\nu|_{W_0=\omega_0}(S(t))\ q(S(t), S'(t))}\right\}$
  if Rand()$< \alpha(S(t), S'(t))$ **then**
    $S(t+1) \leftarrow S'(t)$
  else
    $S(t+1) \leftarrow S(t)$
  end if
end for

If the ray $S(t) = (S_1(t), \dots, S_{N_t}(t))$ denotes the position of the chain at a given time t, then, for $1 \leqslant k \leqslant K-1$, then, almost surely,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\left\{S_{N_t}(t)\in V_k\right\}} = \mathbb{P}\left(S_N \in V_k \mid W_0 = \omega_0\right), \tag{4.1}$$

on condition that the chain $\{S(t); t \geqslant 1\}$ is Harris positive with respect to $\nu|_{W_0=\omega_0}$. Indeed, this statement relies on LLN for Harris recurrent ergodic chains, see [17, Theorem 17.0.1].

We can then sample the law of W_0 to recover an estimate of $\mathbb{P}(S_N \in V_k)$. Let $(\omega_0^1, \dots, \omega_0^{M_{\rm rot}})$, $M_{\rm rot} \in \mathbb{N}^*$ be a sample of i.i.d. initial directions drawn according to $\mathcal{U}\left(C^{2\alpha}\right)$ and for $i = 1, \dots, M_{\rm rot}$, $t = 1, \dots, T$, let $S^{(i)}(t)$ denote the rotation of the random walk $S(t)$ with respect to the initial direction ω_0^i, see Fig. 2. For $k \in \{1, \dots, K-1\}$, our Metropolis-Hastings estimator of $L(x_k)$ is defined by

$$\widehat{L}_{\rm MH}(x_k) = \frac{\rho c\,(1-\cos\alpha)}{2(1-\rho) M_{\rm rot}\, T} \sum_{i=1}^{M_{\rm rot}} \sum_{t=1}^{T} \mathbf{1}_{\left\{S^{(i)}_{N_t}(t)\in V_k\right\}}. \tag{4.2}$$

This estimator is strongly consistent on condition that the proposal density q of Algorithm 1 provides a Harris recurrent chain.

4.2 Mutation strategies

One of the delicate issues in the implementation of MCMC methods is the choice of a convenient proposal q. In our case, we have tested a mixture of mutations of three types for the Metropolis-Hastings algorithm. In order to describe them, let us introduce some notation: first, we have made use of a *perturbed phase function*. Namely, let $\epsilon \in [-1, 1]$. Then the perturbed phase function is defined by a function $f_{HG}^{\epsilon g}$, with ϵg as anisotropic coefficient instead of g. We shall also define some length change by two coprime integers $1 \leqslant j < J$, which denote respectively the small and the big size of a length change.

Remark 14. We have chosen two coprime numbers j and J in order to produce paths of arbitrary length with our successive length changes.

Definition 15 (Mutation rules). Let us assume that, at time $t \in \mathbb{N}^*$, the current ray is given by $S(t) = (\theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t})$. Our proposition for the next move from $S(t)$ to $S(t+1)$ is the following.

(i) With probability $\frac{1}{2}$, the mutation is a *rotation*:

- Choose an index i uniformly over $\{1, \dots, n_t\}$.
- Draw a new angle $\theta_i^{\rm new}$ according to $f_{HG}^{\epsilon g}$.

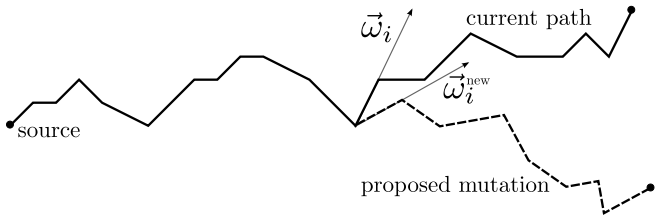

Figure 3: Metropolis-Hastings algorithm. Example of a mutation by rotation.

- Draw a new angle φ_i^{new} uniformly on $[0, 2\pi]$.

If $n_t > 1$, then

- The proposed path is $S'(t) = (\theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_i, \theta_i^{\text{new}}, \varphi_i^{\text{new}}, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t})$, see Fig. 3.

If $n_t = 1$, then

- Draw a new edge length r_1^{new} according to $\mathcal{E}(\mu)$.
- The proposed path is $S'(t) = (\theta_0, \varphi_0, r_1^{\text{new}}, \theta_1^{\text{new}}, \varphi_1^{\text{new}})$.

(ii) With probability $\frac{1}{2}$, the mutation is of type *deletion-addition*: Let $\Delta(n_t)$ be a random function of the size of the current ray, defined by

$$\Delta(n) \sim \begin{cases} \mathcal{U}\left(\{-J, -j, j, J\}\right), & \text{if } n \geqslant J+1, \\ \mathcal{U}\left(\{-j, j, J\}\right), & \text{if } j+1 \leqslant n \leqslant J, \\ \mathcal{U}\left(\{j, J\}\right), & \text{if } 1 \leqslant n \leqslant j. \end{cases}$$

Then:

- If $\Delta(n_t) < 0$, we delete $|\Delta(n_t)|$ edges of $S(t)$ and the proposed path is

$$S'(t) = \left(\theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_{n_t - \Delta(n_t)}, \theta_{n_t - \Delta(n_t)}, \varphi_{n_t - \Delta(n_t)}\right).$$

- If $\Delta(n_t) > 0$, then add $\Delta(n_t)$ edges at the end of $S(t)$ such that
  - $\left(r_{n_t+1}^{\text{new}}, \dots, r_{n_t+\Delta(n_t)}^{\text{new}}\right)$ are i.i.d. according to $\mathcal{E}(\mu)$.
  - $\left(\theta_{n_t+1}^{\text{new}}, \dots, \theta_{n_t+\Delta(n_t)}^{\text{new}}\right)$ are i.i.d. according to $f_{HG}^{\epsilon g}$.
  - $\left(\varphi_{n_t+1}^{\text{new}}, \dots, \varphi_{n_t+\Delta(n_t)}^{\text{new}}\right)$ are i.i.d. uniformly on $[0, 2\pi]$.

  then

$$S'(t) = (\theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t}, r_{n_t+1}^{\text{new}}, \theta_{n_t+1}^{\text{new}}, \varphi_{n_t+1}^{\text{new}}, \dots, r_{n_t+\Delta(n_t)}^{\text{new}}, \theta_{n_t+\Delta(n_t)}^{\text{new}}, \varphi_{n_t+\Delta(n_t)}^{\text{new}}).$$

The proposal density $q(S, \cdot)$ of this mutation rule is easy to compute. For $m \in \mathbb{N}^*$, let

$$\zeta(m) = \begin{cases} \frac{1}{2}\,\frac{1}{4} = \frac{1}{8}, & \text{if } m \geqslant J+1, \\ \frac{1}{2}\,\frac{1}{3} = \frac{1}{6}, & \text{if } j+1 \leqslant m \leqslant J, \\ \frac{1}{2}\,\frac{1}{2} = \frac{1}{4}, & \text{if } 1 \leqslant m \leqslant j. \end{cases}$$

Assume that S' is a mutation of S, denote by n' and n their respective length and set $\Delta = n' - n$. We have

- if $\Delta = 0$ and $n > 1$, then $q(S, S') = \frac{1}{2}\,\frac{1}{n}\,\frac{1}{2\pi}\,f_{HG}^{\epsilon g}(\theta'_i)$;
- if $\Delta = 0$ and $n = 1$, then $q(S, S') = \frac{1}{2}\,e^{-\mu r'_1}\,\frac{1}{2\pi}\,f_{HG}^{\epsilon g}(\theta'_i)$;
- if $\Delta < 0$, then $q(S, S') = \zeta(n)$;
- if $\Delta > 0$, then $q(S, S') = \zeta(n) e^{-\mu \sum_{i=1}^{\Delta} r'_{n'+i}} \left(\frac{\mu}{2\pi}\right)^{\Delta} \prod_{i=1}^{\Delta} f_{HG}^{\epsilon g}\left(\theta'_{n'+i}\right)$,

where r'_i and θ'_i denote respectively the i-th edge length and angle of S'. From these formulas, it is straightforward to recover the acceptance probability.

The idea behind this mixture of mutations is to find a compromise between large jumping size of the Markov chain which implies a lot of "burnt" samples, and smaller jumps which provide more correlated samples, hence a worse convergence. The rotations lead to a good exploration of the domain at low cost, whereas the addition-deletion mutations ensure the visit of the whole state space $\mathcal{A}$ with $W_0 = \omega_0$. The use of a perturbed phase function decreases the acceptance probability of the mutations and thus, increases the number of samples needed in order to converge to the invariant measure. But, it allows a better exploration of the domain and this why the parameter ϵ, as well as the sizes j, J, need to be adapted on a case by case basis. Finally, we can prove that, with this rules of mutations, Algorithm 1 produces a Markov chain that satisfies the LLN. This guarantees the convergence of the estimator defined in (4.2).

Proposition 16. *If the chain $(S(t))_{t\in\mathbb{N}^*}$ is obtained by Algorithm 1 with the mutation rule given in Definition 15, then it is Harris positive with respect to the measure $\nu|_{W_0=\omega_0}$ and the estimator $\widehat{L}_{\mathrm{MH}}(x_k)$ defined in (4.2) is strongly consistent for all $1 \leqslant k \leqslant K-1$.*

Proof. The fact that $\nu|_{W_0=\omega_0}$ is an invariant measure of $(S(t))_{t\in\mathbb{N}^*}$ is an inherent property of Metropolis-Hastings algorithm ([22, 24]). The Harris recurrence is then obtained by checking that the chain is irreducible with respect to $\nu|_{W_0=\omega_0}$, see [24, Corollary 2]. Let $\tau_A = \inf\{t \in \mathbb{N}^* : S(t) \in A\}$ denote the hitting time of any $A \subset \mathcal{A}$ such that $\nu|_{W_0=\omega_0}(A) > 0$. We must demonstrate that $(S(t))_{t\in\mathbb{N}^*}$ is irreducible with respect to $\nu|_{W_0=\omega_0}$, that is,

$$\mathbb{P}_s(\tau_A < +\infty) > 0, \qquad \text{for all } s \in \mathcal{A}, \tag{4.3}$$

where $\mathbb{P}_s(S(1) = s) = 1$. Furthermore, notice that it is sufficient to check this property for subsets A of the type

$$A = \{\omega_0\} \times \prod_{i=1}^{n} (I_i \times S_i), \tag{4.4}$$

where $n \in \mathbb{N}^*$ and where for all $1 \leqslant i \leqslant n$, the sets $I_i \subset \mathbb{R}_+$ and $S_i \subset \mathbb{S}^2$ are all sets of positive Lebesgue measure.

In order to prove relation (4.3) for sets of the form (4.4), consider the conditional measure $\nu_n|_{W_0=\omega_0}$ using (2.10). This measure is equivalent to the Lebesgue measure on $\mathcal{M}_n$ and so $\nu|_{W_0=\omega_0}(A) > 0$. Now, notice that by the Markov property, if τ_A and $\tau_{I_1\times S_1}$ denote respectively the time for the chain to be in A, resp. in $\{\omega_0\} \times (I_1 \times S_1)$, then we have

$$\mathbb{P}_s(\tau_A < +\infty) \geqslant \mathbb{P}_s(\tau_{I_1\times S_1} < +\infty)\,\mathbb{P}_{I_1\times S_1}(\tau_A < +\infty).$$

Now we can lower bound the right hand side of this relation in the following way:
(i) We have that $\mathbb{P}_s(\tau_{I_1\times S_1} < +\infty)$ is greater than the probability of deleting all the edges of s except (r_1, ω_1) and of modifying this edge so that $(r'_1, \omega'_1) \in I_1 \times S_1$. This probability is strictly positive, as well as its acceptance. Indeed, we use here the fact that j and J are coprime (through Bezout's lemma) plus elementary relations for uniform distributions to assert that the probability of deleting all the edges is strictly positive. The positivity of acceptance is due to some absolute continuity properties of q.
(ii) The same kind of argument works in order to estimate $\mathbb{P}_{I_1\times S_1}(\tau_A < +\infty)$ from below. Namely, this quantity is greater than the probability to construct directly a ray $s \in A$, which is itself strictly

positive. Indeed, since j and J are mutually prime, it is possible to construct a ray of any desired length. Moreover, at each step, the probability of adding an edge $(r_i, \omega_i) \in I_i \times S_i$, as well as its acceptance are always strictly positive.
We have thus obtained that $\mathbb{P}_s\left(\tau_{I_1\times S_1} < +\infty\right)\mathbb{P}_{I_1\times S_1}\left(\tau_A < +\infty\right) > 0$, which concludes the proof. □

Remark 17. The process $(N_t)_{t\geqslant 1}$ that gives the length of the ray $S(t)$ at time t behaves like a birth-death process with time inhomogeneous rates. If there exists an invariant measure of the process $(N_t)_{t\geqslant 1}$, then it must coincide with the geometric distribution of parameter $1-\rho$ that drives the length of a path $S \sim \nu$. This provides an easy criterium in order to check that the chain has already mixed, for example with a chi-squared test on the empirical distribution of $(N_t)_{t\geqslant 1}$.

5 Simulation and comparison of the methods

In this section, we compare the estimates of the fluence $L(x_k)$ provided by three methods: Monte Carlo with Wang-Prahl algorithm (denoted by WANG, see [19, 26]), MC-SOME (see (3.2)) and the Metropolis-Hastings (MH) (see (4.2)) with the mutation rules given in Definition 15. We tested the methods in different settings. Here, we present results in a framework corresponding to a healthy homogeneous rat brain tissue. We chose to follow [1] for the values of the optical parameters (see Table 1). Other values for rat or human brain can be found in [4, 8, 15]. The volume of the cube V equals 8 cm^3, that is $V = [-1, 1]^3$. It is discretized into $K = 50^3$ voxels so that the volume of each voxel is $(0.04)^3$ cm^3. The half-opening angle of the optical fiber was set to $\alpha = \frac{\pi}{10}$ and the constant c in (2.14) was set $c = 1$.

We chose simulation parameters for the three methods so that they need the same amount of computational time. Those are

WANG: $M = 5000$ photons trajectory.

MC-SOME: $M = 40000$ rays, $M_{\text{points}} = 50$ points chosen in each ray and $M_{\text{rot}} = 30$ rotations with respect to the initial direction.

MH $j = 10$, $J = 21$, $\epsilon = 0.9$, $T = 2 \cdot 10^5$ steps of the chain and $M_{\text{rot}} = 30$ rotations with respect to the initial direction,

In Fig. 4, we picture a contour plot of the estimates on the plane $x = 0$ of each methods. The shape of the contour lines are similar. However, we can notice that MH appears more noisy and that its halo is more spread-out than those of WANG and MC-SOME.

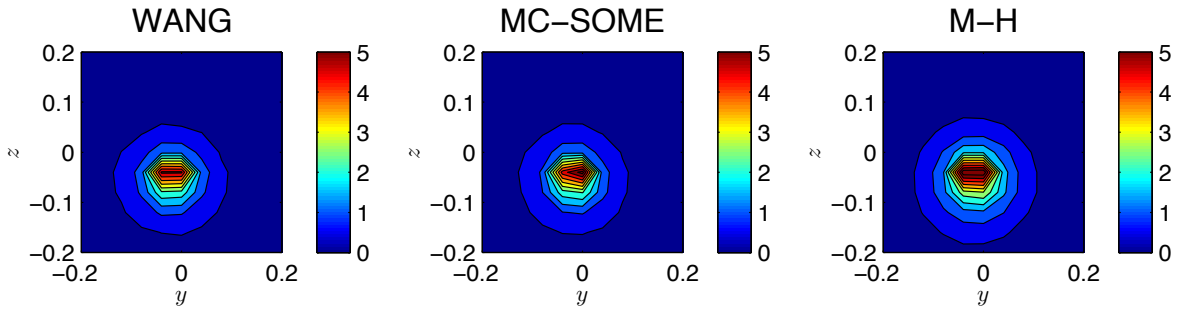

Figure 4: Contour plots of the estimates of the fluence rate in the plane $x = 0$ for WANG, MC-SOME and MH. Notice that we zoomed into the volume V.

In Fig. 6, we compare the estimates along several lines $(\ell_i)_{i=1,\dots,6}$ of voxels parallel to the y-axis and passes through the points $(0, 0, -0.04)$, $(0, 0, -0.08)$, $(0, 0, -0.12)$, $(0, 0, -0.4)$, $(0, 0, -0.48)$ and $(0, 0, -0.6)$ respectively (see Fig. 5). Close to the light source, the three methods give similar estimates.

Further below the light source, we notice that MC-SOME gives much smoother estimates than the two other algorithms. Moreover, it seems that this running of MH undervalued the fluence far from the light source. Perhaps because it had not converged yet.

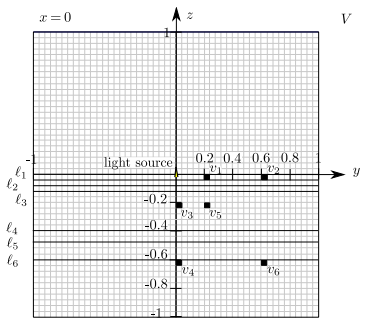

Figure 5: Choice of six particular voxels and position of the lines $(\ell_i)_{i=1,\dots,6}$ in the cube V.

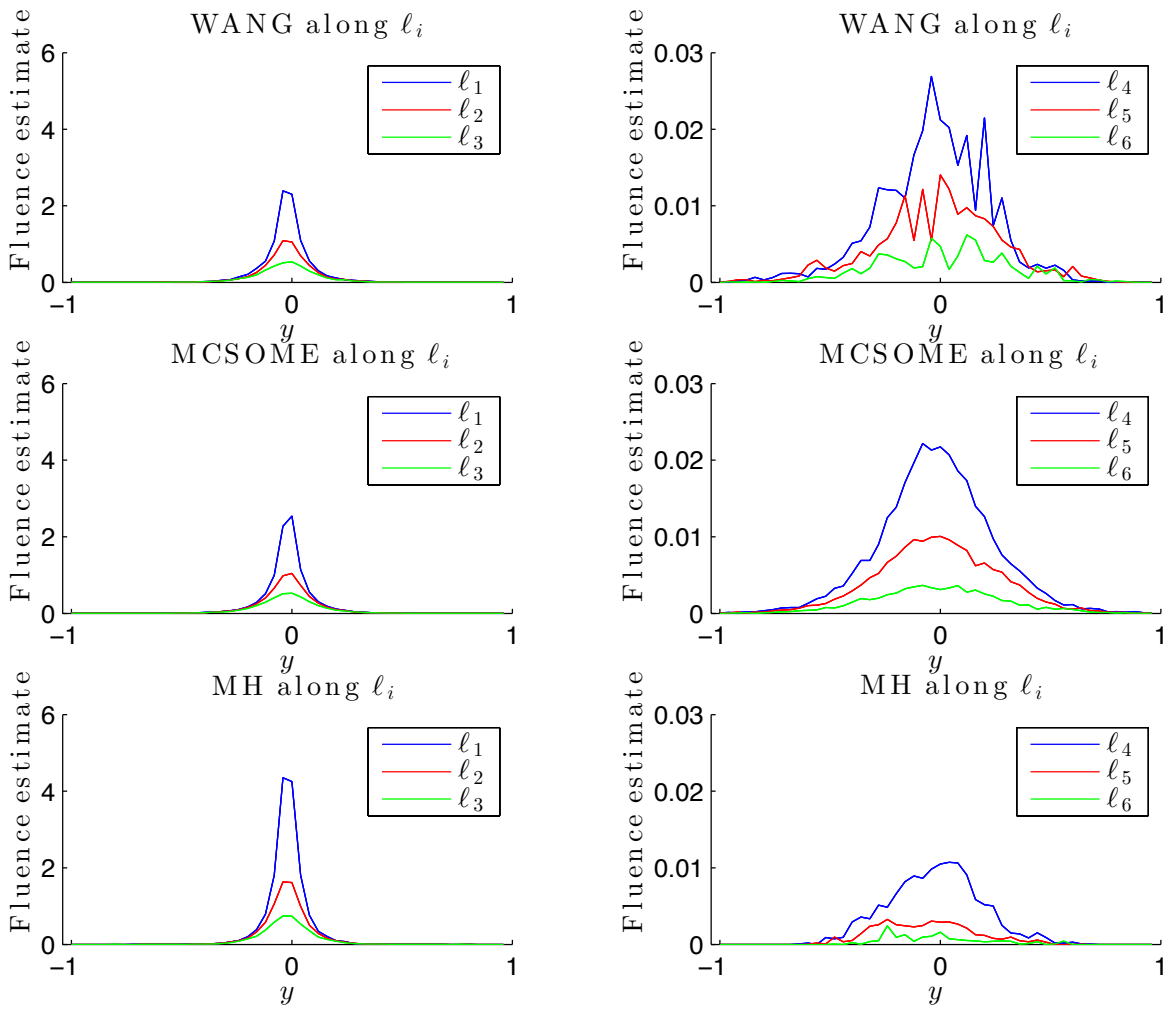

Figure 6: Estimates of the fluence rate along the lines $(\ell_i)_{i=1,\dots,6}$ with WANG, MC-SOME and MH.

Let us conclude this section by studying the accuracy of the methods by mean of 50 independent replicates of these estimates. In Fig. 7, the boxplots compare the dispersion of the 50 estimates of each method in six voxels $(v_i)_{i=1,\dots,6}$ such that (see Fig. 5)

$$\begin{aligned} &(0, 0.2, 0) \in v_1, && (0, 0.6, 0) \in v_2, && (0, 0, -0.2) \in v_3, \\ &(0, 0, -0.6) \in v_4, && (0, 0.2, -0.2) \in v_5, && (0, 0.6, -0.6) \in v_6. \end{aligned} \tag{5.1}$$

On one hand, we see that MC-SOME is much more consistent than WANG, because of the variance reduction that we use in MC-SOME. On the other hand, MH is comparable to WANG concerning the dispersion of the estimates. This is quantified in Table 2 where we provide the mean of the estimates and their mean square error in each of the 6 voxels.

	Mean			Mean Square Error		
	Wang	MC-SOME	M-H	Wang	MC-SOME	M-H
v_1	0.1902	0.1855	0.1957	$0.4699 \cdot 10^{-3}$	$0.5151 \cdot 10^{-4}$	$0.6915 \cdot 10^{-3}$
v_2	0.0040	0.0037	0.0014	$0.0023 \cdot 10^{-3}$	$0.0034 \cdot 10^{-4}$	$0.0028 \cdot 10^{-3}$
v_3	0.1751	0.1722	0.1807	$0.2822 \cdot 10^{-3}$	$0.3160 \cdot 10^{-4}$	$0.5232 \cdot 10^{-3}$
v_4	0.0037	0.0034	0.0022	$0.0020 \cdot 10^{-3}$	$0.0008 \cdot 10^{-4}$	$0.0037 \cdot 10^{-3}$
v_5	0.0774	0.0758	0.0707	$0.1129 \cdot 10^{-3}$	$0.0658 \cdot 10^{-4}$	$0.1926 \cdot 10^{-3}$
v_6	0.0004	0.0004	0.0003	$0.0001 \cdot 10^{-3}$	$0.0001 \cdot 10^{-4}$	$0.0001 \cdot 10^{-3}$

Table 2: Mean and mean square error of the 50 estimates of the fluence rate in 6 voxels for WANG, MC-SOME and MH.

6 Inverse problem and sensitivity

Good estimates for the optical coefficients of the tissue under consideration are of considerable practical importance for biologists. This will be numerically solved thanks to our probabilistic representation (2.14), and we first proceed to a sensibility analysis with respect to the parameters g, μ_s and μ_a in order to get an intuition on the meaning of our estimations.

6.1 Sensitivity of the measurements

As a preliminary step towards a good resolution of the inverse problem, we first observe how the measurements vary with respect to the optical parameters. To this aim, we built a database of simulations for different values of g, μ_s and μ_a and then compared the estimated fluence. The estimates are computed by resorting to MC-SOME, which is the best performing method among the three we have implemented according to Section 5.

Our experiments are developed in the following way: we choose a reference simulation obtained for the reference parameters (g^*, μ_a^*, μ_s^*) and pick some voxels centered at $(x_{k_i})_{i=1,\dots,n}$ and in which we simulate the fluence

$$m_i = \widehat{L}(x_{k_i}; g^*, \mu_a^*, \mu_s^*), \qquad i = 1, \dots, n,$$

where we recall that $L(x_{k_i})$ is defined by (2.13) with $\gamma_{W_0} = \sigma$ and where we stress the dependence on the optical coefficients by writing $\widehat{L}(x_{k_i}; g^*, \mu_a^*, \mu_s^*) \equiv \widehat{L}(x_{k_i})$. Now for each possible triplet of parameters (g, μ_a, μ_s), we compute the normalized quadratic error (or *evaluation error*)

$$J(\mu_a, \mu_s, g) = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\widehat{L}(x_{k_i}; g, \mu_a, \mu_s) - m_i}{m_i} \right)^2. \tag{6.1}$$

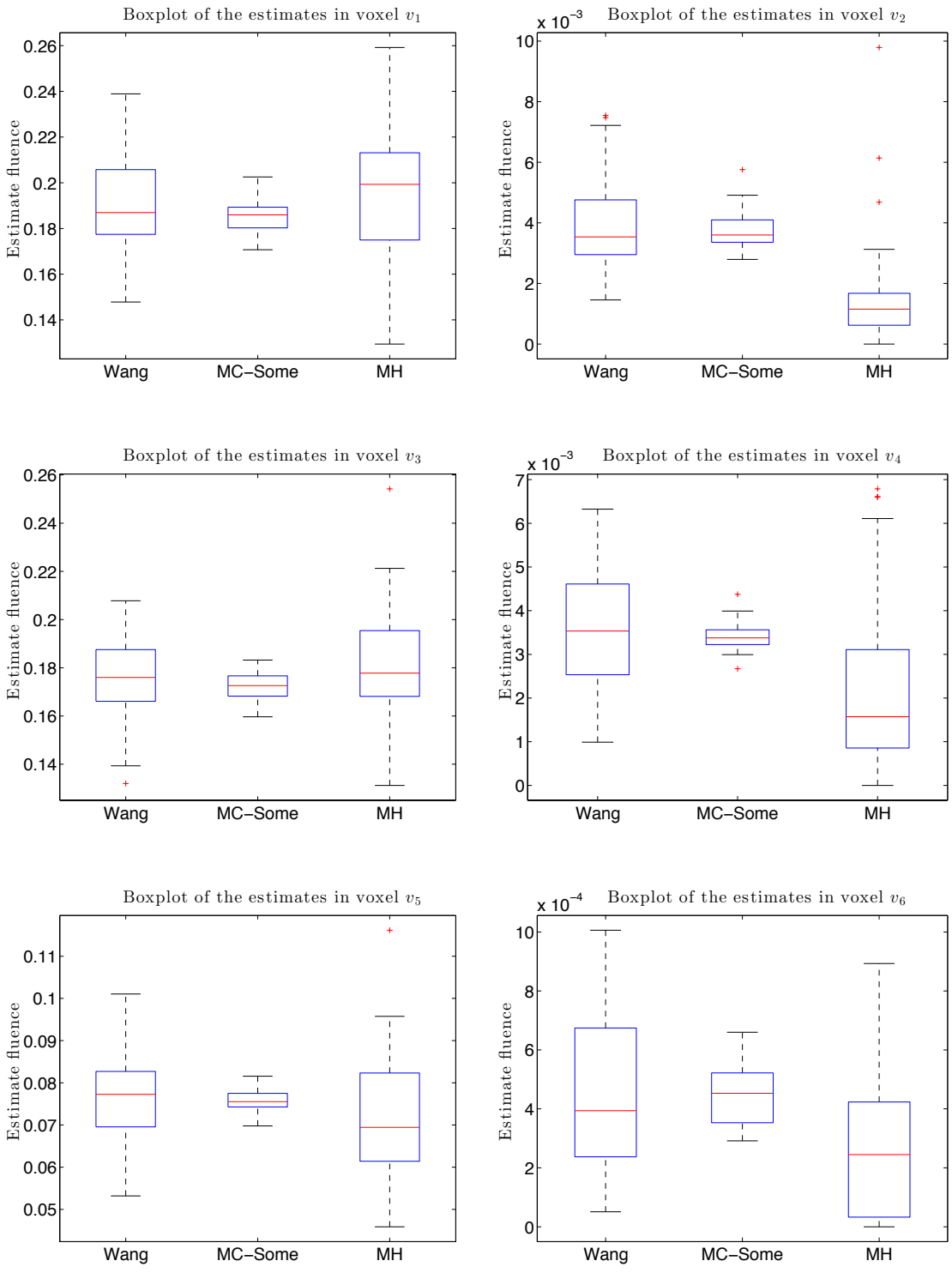

Figure 7: Boxplots of 50 estimates of the fluence for WANG, MC-SOME and MH in the voxels $(v_i)_{i=1,\dots,6}$.

For the dataset of simulations, we use the same settings as in Section 5 ($|V| = 8$ cm^3, the volume of a voxel is $(0.04)^3$ cm^3 and $\alpha = \frac{\pi}{10}$). The variable parameters are: g, μ_a and μ_s. Their values are given in Table 3. This choice is motivated by [1, 4, 8, 15]. The anisotropy parameter g does not vary a lot between tissue type (healthy or tumorous) and it is often even hidden in a reduction of the scattering coefficient $\mu_s' = \mu_s(1-g)$. For this reason, we chose only three values in a small range of common values. Concerning the other parameters, we chose five values in intervals covering values corresponding to healthy and tumorous brain tissues according to [1, 15].

g	0.85	0.90	0.95		
μ_a in cm^{-1}	0.5	0.75	1	1.25	1.5
μ_s in cm^{-1}	75	90	105	120	135

Table 3: Values of the optical parameters for the study of sensitivity.

Figures 8 to 11 give different representation of the variation of the error $J(\mu_a, \mu_s, g)$ with respect to the optical parameters. The real values are $(\mu_a^*, \mu_s^*, g^*) = (0.75, 105, 0.9)$ and we set $n = 3$, $x_{k_1} \in v_2$, $x_{k_2} \in v_4$, $x_{k_3} \in v_6$ respectively (see Fig. 5). We see that the sensitivity in the parameters μ_s and g is very low compared to the sensitivity in μ_a. In Fig. 11, we see that a wrong value of μ_a has strong effects on the error function and that it becomes then almost impossible to see any tendency for the anisotropy parameter g. Notice also that an undervaluation of μ_a is worse than an overvaluation in terms of the error.

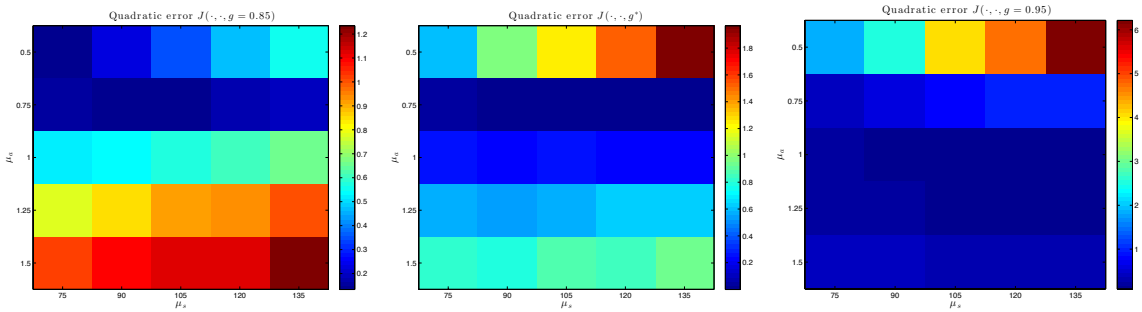

Figure 8: Colormap of the quadratic error $(\mu_a, \mu_s) \mapsto J(\mu_a, \mu_s, g)$ for three values of g, where μ_s is displayed on the x axis and μ_a on the y axis.

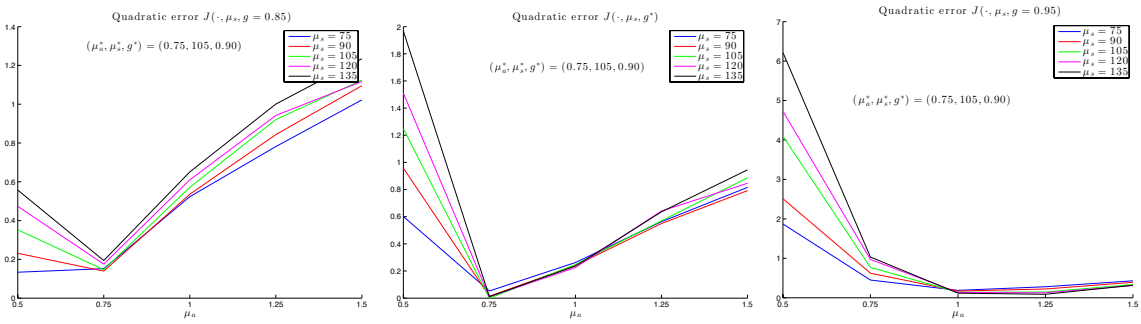

Figure 9: Quadratic error $\mu_a \mapsto J(\mu_a, \mu_s, g)$ for three values of g.

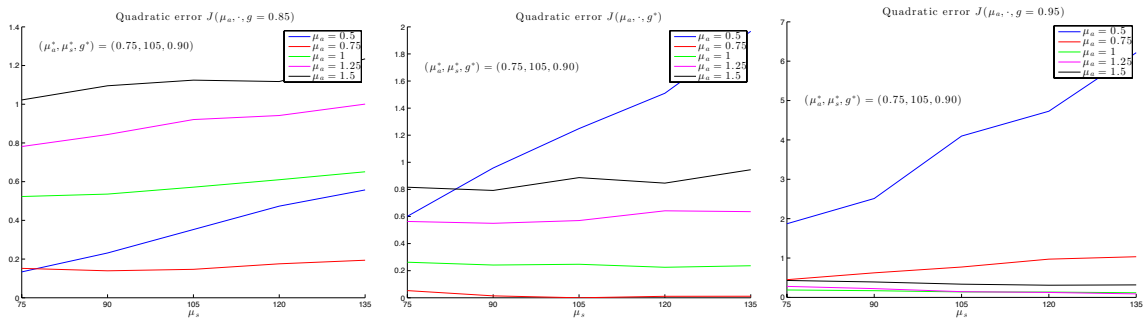

Figure 10: Quadratic error $\mu_s \mapsto J(\mu_a, \mu_s, g)$ for three values of g.

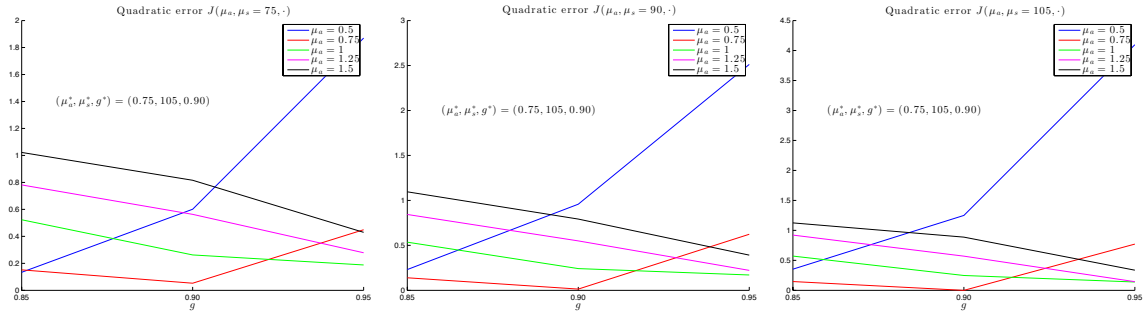

Figure 11: Quadratic error $g \mapsto J(\mu_a, \mu_s, g)$ for three values of μ_s.

6.2 Parameter estimation

This section is only devoted to an estimate of the parameters μ_a and μ_s. Indeed, we have seen in the last section that the sensitivity of the fluence with respect to the anisotropic parameter g is low. Moreover, our simulations do not show any monotonicity or tendency in the error for this parameter because, in our settings, the Monte Carlo error prevails over the evaluation error. In addition, for our purpose, the uncertainty about g is small in front of the uncertainty of the two other parameters (see [1]). We shall thus suppose in the sequel that g is known.

With these preliminary considerations in mind, our goal is to solve the following nonlinear least square minimization problem: Find (μ_s, μ_a) in order to minimize

$$J(\mu_s, \mu_a) = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{L(x_{k_i}; \mu_s, \mu_a) - m_i}{m_i} \right)^2, \tag{6.2}$$

where $(m_i)_{i=1,\dots,n}$ are the measurements in n different voxels whose center points are $(x_{k_i})_{i=1,\dots,n}$.

The optimization method that we use then in order to approximate the minimum of the *score* J is based on the Levenberg-Marquardt algorithm (see [10]). This gradient descent algorithm also involves the computation of the gradient, as well as the Hessian matrix of J. The algorithm is described in pseudo-code in Algorithm 2. In this description, for $k \geqslant 0$ we have set $H_k = \mathrm{Hess}(J)(\mu_s^k, \mu_a^k)$, $\mathrm{Diag}(H_k)$ is the diagonal matrix of H_k, λ_k is the *damping factor* which may be either constant or corrected at each step and $\tau_k \in \mathbb{R}_+$ controls the step size.

Let us highlight the fact that the form of the objective function in (6.2) allows to express the term on the right hand side of line 3 in Algorithm 2 explicitly as a function of the partial derivatives of L.

Algorithm 2 Gradient descent algorithm for the estimation of μ_a and μ_s

Input: measurements $(m_i)_{i=1,\dots,n}$, initial couple (μ_s^0, μ_a^0), precision $\epsilon > 0$.

$k \leftarrow 0$

while $J(\mu_s^k, \mu_a^k) > \epsilon$ **do**

$(\mu_s^{k+1}, \mu_a^{k+1}) \leftarrow (\mu_s^k, \mu_a^k) - \tau_k \left[H_k + \lambda_k \operatorname{Diag}(H_k)\right]^{-1} \nabla J(\mu_s^k, \mu_a^k)$

$k \leftarrow k + 1$

end while

Output: an approximation (μ_s^k, μ_a^k) of the real parameters (μ_s^*, μ_a^*)

Indeed, the gradient of J is given by

$$\nabla J(\cdot) = \sum_{i=1}^{n} \frac{L(x_{k_i};\cdot) - m_i}{m_i^2} \nabla L(x_{k_i};\cdot), \tag{6.3}$$

and its Hessian matrix is given by

$$\operatorname{Hess}(J)(\cdot) = \sum_{i=1}^{n} \left(\frac{L(x_{k_i};\cdot) - m_i}{m_i^2} \operatorname{Hess}(L)(x_i;\cdot) + \frac{1}{m_i^2} \nabla L(x_{k_i};\cdot) \nabla^t L(x_{k_i};\cdot) \right). \tag{6.4}$$

Moreover, as stated in the following proposition, the formal representation in Proposition 4 allow to also use the Monte Carlo method MC-SOME in order to estimate the first order and the second order partial derivatives of L which can be expressed similarly to (2.16).

Proposition 18. *The partial derivatives of $L(x_{k_i};\mu_s,\mu_a)$ can be expressed as the expectation of fully simulable random variables. Using the same notations as in* (2.15)*, they are given by*

$$\frac{\partial L}{\partial \mu_a}(x_{k_i};\mu_s,\mu_a) = -\frac{\mu_s}{\mu_a}\, \tilde{c}\, \mathbb{E}^{\mu_s,\mu_a} \left(\mathbf{1}_{\{S_N \in V_{k_i}\}} \sum_{k=1}^{N} r_k \right), \tag{6.5}$$

$$\frac{\partial L}{\partial \mu_s}(x_{k_i};\mu_s,\mu_a) = \frac{\mu_s}{\mu_a}\, \tilde{c}\, \mathbb{E}^{\mu_s,\mu_a} \left(\mathbf{1}_{\{S_N \in V_{k_i}\}} \left(\frac{N}{\mu_s} - \sum_{k=1}^{N} r_k \right) \right), \tag{6.6}$$

with $\tilde{c} = c\,\frac{1-\cos\alpha}{2}$ and $\frac{\mu_s}{\mu_a} = \frac{\rho}{1-\rho}$.

Proof. We start by differentiating term-by-term the Neumann series of Corollary 3. For $n \geqslant 1$, we have by definition of T^n in (2.4), that

$$\frac{\partial \left(T^{\mu_s,\mu_a}\right)^n}{\partial \mu_a} L_e(x,\omega_0) = \mu_s^n \int_{\mathbb{R}_+^n} dr_1 \cdots dr_n \left(-\sum_{j=1}^{n} r_j \right) \exp\left(-(\mu_s+\mu_a) \sum_{j=1}^{n} r_j \right) \int_{\mathbb{S}^{2n}} d\sigma^{\otimes n}(\omega_1,\dots,\omega_n) \prod_{j=0}^{n-1} f_{HG}\left(\omega_j,\omega_{j+1}\right) L_e\left(x - \sum_{k=0}^{n-1} \omega_k r_{k+1}, \omega_n \right).$$

Looking back at Section 2.3 and using the same notations, we deduce that

$$\sum_{n=0}^{\infty} \frac{\partial \left(T^{\mu_s,\mu_a}\right)^n}{\partial \mu_a} L_e(x,\omega_0) = -\int_{\mathcal{A}} d\nu(\omega_0,\boldsymbol{r},\boldsymbol{\omega})\, G_x(\omega_0,\boldsymbol{r},\boldsymbol{\omega}) \sum_{i=1}^{|\boldsymbol{r}|} r_i,$$

where we recall that $|\boldsymbol{r}|$ stands for the size of $\boldsymbol{r}$. Assuming that the left-hand side coincides with the partial derivative $\frac{\partial L}{\partial \mu_a}$, then (6.5) is found just like (2.15) and the same arguments provide (6.6), considering that

$$\frac{\partial \left(T^{\mu_s,\mu_a}\right)^n}{\partial \mu_s} = \frac{n}{\mu_s} \left(T^{\mu_s,\mu_a}\right)^n + \frac{\partial \left(T^{\mu_s,\mu_a}\right)^n}{\partial \mu_a}.$$

To conclude the proof, notice that the match between the partial derivatives and the term-by-term differentiation of the Neumann series is ensured by the fact that the operator $\left(T^{\mu_s,\mu_a}\right)^n$ is infinitely continuously differentiable for all n and by the uniform convergence of the corresponding sequences of truncated sums

$$s_m = \sum_{n=0}^{m} \frac{\partial \left(T^{\mu_s,\mu_a}\right)^n}{\partial \mu_s} L_e(x,\omega_0), \qquad m \geqslant 1.$$

□

Remark 19. Similar formula to (6.5) and (6.6) can be easily found for the second order derivatives $\frac{\partial^2 L}{\partial \mu_s^2}$ and $\frac{\partial^2 L}{\partial \mu_a^2}$.

The probabilistic representation of $L(x_{k_i};\mu_s,\mu_a)$ in (2.16) and its partial derivatives allows us to estimate the score $J(\mu_s,\mu_a)$, its gradient and its Hessian matrix by Monte Carlo methods. A sole sample $(y_1,\ldots,y_n) \in \mathcal{A}^n$ of n observations of the random ray Y can be used to estimate the expectations in L, $\nabla L(x_{k_i};\cdot)$ and $\mathrm{Hess}(L)(x_i;\cdot)$ at the same time. We denote these estimates by $\widehat{L}$, $\widehat{\nabla} L(x_{k_i};\cdot)$ and $\widehat{\mathrm{Hess}}(L)(x_i;\cdot)$ and the corresponding score by $\widehat{J}$. The updating rule at line 3 in Algorithm 2 becomes then

$$(\mu_s^{k+1},\mu_a^{k+1}) = (\mu_s^k,\mu_a^k) - \tau_k \left[\widehat{H}_k + \lambda_k \,\mathrm{Diag}(\widehat{H}_k)\right]^{-1} \widehat{\nabla} J(\mu_s^k,\mu_a^k). \tag{6.7}$$

The noise coming from the Monte Carlo estimation of the score, of its gradient and of its Hessian matrix during the run of the algorithm, makes a precise estimate of the real values of (μ_a^*,μ_s^*) difficult. Far from the real value, the eigenvalues of the Hessian matrix are very small and we have experienced that their sign can vary a lot because of the noise of the Monte Carlo estimation. Conversely, near the real value of the parameters, the estimation of the Hessian matrix is more robust and its eigenvalues are almost always both positive, which legitimates the quadratic approximation of Levenberg-Marquardt. For this reason, we implemented a hybrid algorithm between a classic steepest gradient descent and the Levenberg-Marquard descent. At each step, we test the sign of the eigenvalues of H_k. If they are both positive, one moves to the next point following (6.7), else one makes a move in the direction $-\nabla J(\mu_s^k,\mu_a^k)$. In order to test the algorithm, we ran it in the same settings as in the previous section. The damping parameter λ has been chosen as fixed and in order to prevent the descent to make too large steps, we forced the step size to be a given decreasing sequence ak^{-b} with $a>0$ and $0<b<1$, by setting

$$\tau_k = ak^{-b} \left\| \left[\widehat{H}_k + \lambda_k \,\mathrm{Diag}(\widehat{H}_k)\right]^{-1} \widehat{\nabla} J(\mu_s^k,\mu_a^k) \right\|^{-1}.$$

As we shall see, the descent goes rapidly in a neighborhood of the real couple. A satisfying approximate of μ_a^* comes up easily, whereas μ_s^* is more difficult to find. This is consistent with what we have seen in the study of sensitivity of the parameters in the previous section. In Figure 12, we can see a first example of the descent that we obtain with this algorithm. Notice the oscillations around the real value $\mu_a^*=1$, once we are close to it. Those descent zigzags near the real value of μ_a are also apparent in a second descent illustrated in Figure 13. The sequence τ_k we chose for the size step is proportional to $k^{-\frac{3}{4}}$ in both examples.

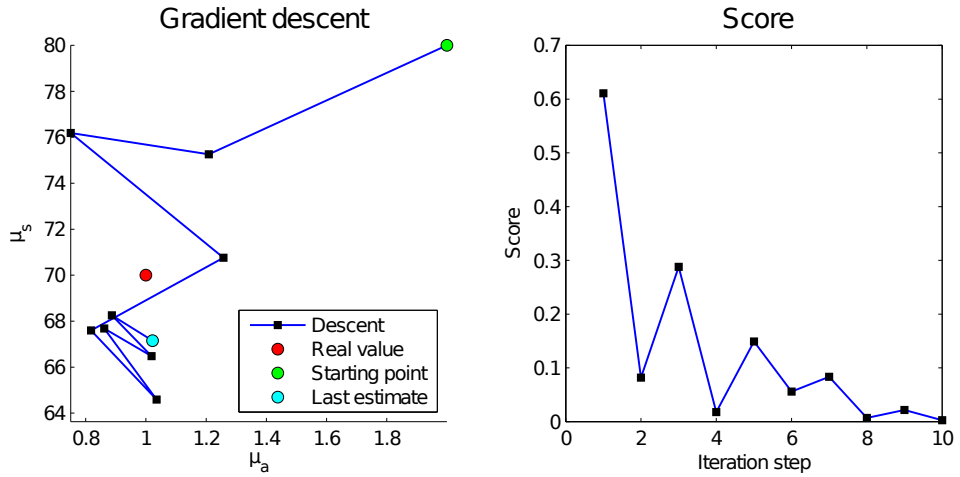

Figure 12: Parameter estimation with an adaptation of Levenberg-Marquardt descent algorithm. The real value of (μ_a^*, μ_s^*) is $(1, 70)$. The starting point is $(2, 80)$. For a precision $\epsilon = 2 \cdot 10^{-3}$, the algorithm stopped at $(1.02, 67.15)$ with a score equals to $J = 0.0018$.

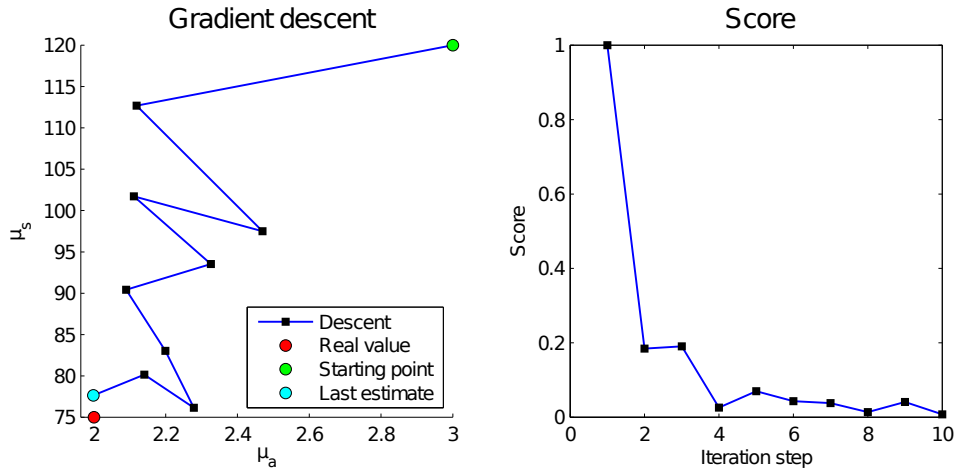

Figure 13: Parameter estimation with an adaptation of Levenberg-Marquardt descent algorithm. The real value of (μ_a^*, μ_s^*) is $(2, 75)$. The starting point is $(3, 120)$. For a precision $\epsilon = 5 \cdot 10^{-3}$, the algorithm stopped at $(1.998, 77.64)$ with a score equals to $J = 0.0041$.

References

[1] E. Angell-Petersen, H. Hirschberg, and S. J. Madsen. Determination of fluence rate and temperature distributions in the rat brain; implications for photodynamic therapy. *Journal of biomedical optics*, 12(1):014003–014003, 2007.

[2] S. Arridge, M. Schweiger, M. Hiraoka, and D. Delpy. A finite element approach for modeling photon transport in tissue. *Medical physics*, 20:299, 1993.

[3] P. R. Bargo, S. A. Prahl, T. T. Goodell, R. A. Sleven, G. Koval, G. Blair, and S. L. Jacques. In vivo determination of optical properties of normal and tumor tissue with white light reflectance and an empirical light transport model during endoscopy. *Journal of Biomedical Optics*, 10(3):034018–034018–15, 2005.

[4] F. Bevilacqua, D. Piguet, P. Marquet, J. D. Gross, B. J. Tromberg, and C. Depeursinge. In vivo local determination of tissue optical properties: Applications to human brain. *Applied optics*, 38(22):4939–4950, 1999.

[5] K. Bhan and J. Spanier. Condensed history Monte Carlo methods for photon transport problems. *Journal of computational physics*, 225(2):1673–1694, 2007.

[6] L. L. Carter and E. Cashwell. Particle-transport simulation with the Monte Carlo method. Technical report, Los Alamos Scientific Lab., N. Mex.(USA), 1975.

[7] S. Chandrasekhar. *Radiative transfer*. Dover Books on Physics Series. Dover Publications, Incorporated, 1960.

[8] W.-F. Cheong, S. A. Prahl, and A. J. Welch. A review of the optical properties of biological tissues. *Quantum Electronics, IEEE Journal of*, 26(12):2166–2185, 1990.

[9] T. J. Dougherty, C. J. Gomer, G. Jori, D. Kessel, M. Korbelik, J. Moan, and Q. Peng. Photodynamic therapy. *Journal of the National Cancer Institute*, 90(12):889–905, 1998.

[10] R. Fletcher. *Practical methods of optimization*. John Wiley & Sons, second edition, 2013.

[11] I. Fredriksson, M. Larsson, and T. Strömberg. Inverse Monte Carlo method in a multilayered tissue model for diffuse reflectance spectroscopy. *Journal of Biomedical Optics*, 17(4):047004–047004, 2012.

[12] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[13] L. G. Henyey and J. L. Greenstein. Diffuse radiation in the galaxy. *The Astrophysical Journal*, 93:70–83, 1941.

[14] H. W. Jensen, J. Arvo, P. Dutre, A. Keller, A. Owen, M. Pharr, and P. Shirley. Monte Carlo ray tracing. In *ACM SIGGRAPH*, 2003.

[15] M. Johns, C. Giller, D. German, and H. Liu. Determination of reduced scattering coefficient of biological tissue from a needle-like probe. *Optics express*, 13(13):4828–4842, 2005.

[16] H. Karlsson, I. Fredriksson, M. Larsson, and T. Strömberg. Inverse Monte Carlo for estimation of scattering and absorption in liquid optical phantoms. *Opt. Express*, 20(11):12233–12246, May 2012.

[17] S. P. Meyn, R. L. Tweedie, and P. W. Glynn. *Markov chains and stochastic stability*, volume 2. Cambridge University Press Cambridge, 2009.

[18] G. M. Palmer and N. Ramanujam. Monte Carlo-based inverse model for calculating tissue optical properties. Part I: Theory and validation on synthetic phantoms. *Applied Optics*, 45(5):1062–1071, Feb 2006.

[19] S. A. Prahl. *Light transport in tissue*. PhD thesis, University of Texas at Austin, 1988.

[20] S. A. Prahl, M. Keijzer, S. L. Jacques, and A. J. Welch. A Monte Carlo model of light propagation in tissue. *Dosimetry of laser radiation in medicine and biology*, 5:102–11, 1989.

[21] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[22] G. O. Roberts and J. S. Rosenthal. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, 16(4):2123–2139, 2006.

[23] P. Shirley, D. Edwards, and S. Boulos. Monte Carlo and quasi-Monte Carlo methods for computer graphics. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 167–177. Springer, 2008.

[24] L. Tierney. Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728, 1994.

[25] E. Veach. *Robust Monte Carlo methods for light transport simulation*. PhD thesis, Stanford University, 1997.

[26] L. Wang and S. L. Jacques. Monte Carlo modeling of light transport in multi-layered tissues in standard c. Technical report, The University of Texas, MD Anderson Cancer Center, Houston, 1992.

[27] B. C. Wilson and M. S. Patterson. The physics of photodynamic therapy. *Physics in medicine and biology*, 31(4):327, 1986.

[28] B. C. Wilson and M. S. Patterson. The physics, biophysics and technology of photodynamic therapy. *Physics in medicine and biology*, 53(9):R61, 2008.

[29] K. Yoshida. *Functional analysis*. Classics in mathematics / Springer. Springer London, Limited, 1980.

[30] C. Zhu and Q. Liu. Review of Monte Carlo modeling of light transport in tissues. *Journal of Biomedical Optics*, 18(5), 2013.