Monte Carlo methods for light propagation in biological tissues

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

Monte Carlo methods for light propagation in biological tissues

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

[1]Inria, BIGS, Villers-lès-Nancy, F-54600
[2]Université de Lorraine, Institut Élie Cartan de Lorraine, UMR 7502, Vandœuvre-lès-Nancy, F-54506
[3]Université de Lorraine, CRAN, UMR 7039, 9, avenue de la forêt de Haye, Vandœuvre-lès-Nancy Cedex, 54516
[4]Université de Fribourg, Département de Mathématiques, chemin du Musée 23, CH-1700 Fribourg

October 15, 2014

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

The results presented in this article have initially been motivated by several research projects grounded on Photodynamic therapy (PDT), which 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. See [29, 30] for a review on PDT.

The project that initiated this work focuses on an innovative application : the interstitial PDT for the treatment of high-grade brain tumors [6, 5]. This strategy requires the installation of optical fibers to deliver light directly into the tumor tissue to be treated, while nanoparticles are used to carry the photosensitizer into the cancer cells.

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, mathematical and physical models must be developed to better control and understand PDT responses. In this new challenge, the two main questions to which these models should answer are:

1. What is the optimal shape, position and number of light sources 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 [10]. 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 [8, 9, 32] 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 [28], firstly described by S. A. Prahl in [21] and S. A. Prahl et al. in [22]. 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 [18, 13, 4, 20]), 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. We first derive a rigorous probabilistic representation of the solution to

ERT in homogeneous tissues, which will help us to propose an alternative MC method to Wang's algorithm [28]. Then we also propose a variance reduction method.

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. 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 rate 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 [28]. 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 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

Let $D = \mathbb{R}^3$ be the set of positions in the biological homogeneous tissue and $\mathbb{S}^2$ be the unit sphere in $\mathbb{R}^3$. Let us denote the optical parameters of the tissue by $\mu_s > 0$ for the *scattering coefficient*, $\mu_a > 0$ for the *simple absorption coefficient* and $\mu = \mu_s + \mu_a$ for the *total absorption coefficient* (or *attenuation coefficient*). Moreover, let us denote by $L_e(x,\omega)$ the *emitted light* from x in direction ω and by $L(x,\omega)$ the *quantity of light* at x in the direction ω. Then the equation of radiative transfer takes the following form (see e.g. [21, 3]):

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

where $L_i(x,\omega)$ is the incident volume emittance and $T : L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R}) \to L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$ is the linear operator defined on the Banach space of essentially bounded real-valued functions $\ell : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$, given by

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

with f the so-called *bidirectional scattering distribution function* and σ the uniform probability measure on the unit sphere $\mathbb{S}^2$. The incident volume emittance L_i is also defined by applying a linear operator T_i to L_e:

$$L_i = T_i L_e, \quad \text{with} \quad T_i\ell(x,\omega) = \int_0^{+\infty} \ell(x - r\omega, \omega) \exp(-\mu r) dr. \tag{2.3}$$

In the following, we will denote the *albedo coefficient* by $\rho := \frac{\mu_s}{\mu} < 1$. Moreover, since we consider an homogeneous biological tissue, the scattering function is given by the so-called *Henyey-Greenstein function*, see [15], that is

$$f(\omega, x, \hat{\omega}) = f_{HG}(\omega, \hat{\omega}) = \frac{1 - g^2}{(1 + g^2 - 2g\langle \omega, \hat{\omega} \rangle)^{3/2}}, \quad \omega, \hat{\omega} \in \mathbb{S}^2,\, \forall x \in D, \tag{2.4}$$

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

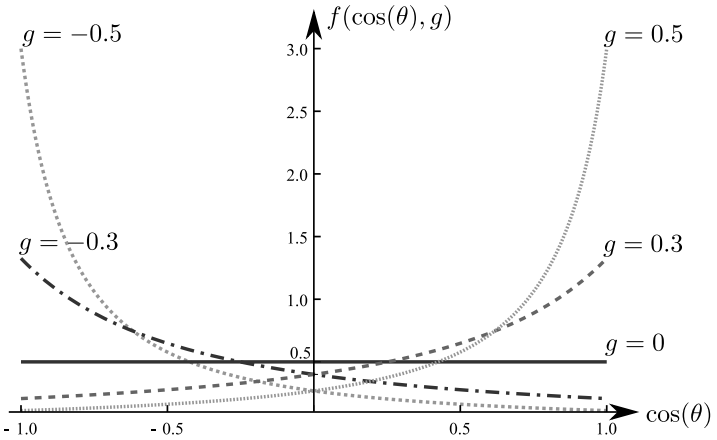

Figure 1: Density function of $\cos(\theta) = \langle \omega, \hat{\omega} \rangle$ when $\hat{\omega}$ is scattered according to the Henyey-Greenstein function $f_{HG}(\omega, \cdot)$ for several values of the anisotropy parameter g.

anisotropic parameter g, the more likely the large values of $\cos(\theta)$ and the less the scattering of the ray (see Fig. 1).

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 it in Neumann series. This method is based on the next classical and general result.

Theorem 1 ([31] 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 for any $x_0 \in B$, the equation $x = Ax + x_0$ admits a unique solution given by*

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

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. *The operator T defined in* (2.2)*, with f given by* (2.4)*, satisfies $\|T\| = \rho < 1$, where we recall that we have set $\rho := \frac{\mu_s}{\mu} < 1$.*

Proof. Let $\ell \in L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$. 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 $L_e \in L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$, then the radiative transfer equation* (2.1) *with a phase function given by* (2.4) *admits a unique solution L in $L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$. Moreover, L can be decomposed as*

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

$$[T^n \circ T_i]\, \ell(x, \omega_0) = \mu_s^n \int_{\mathbb{R}_+^{n+1}} dr_0 \cdots dr_n \, \exp\left(-\mu \sum_{j=0}^{n} r_j\right) \int_{(\mathbb{S}^2)^n} d\sigma^{\otimes n}(\omega_1, \ldots, \omega_n) \prod_{j=0}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) \ell\left(x - \sum_{k=0}^{n} \omega_k r_k, \omega_n\right). \quad (2.5)$$

Proof. Assume that $L_e \in L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$. It is readily checked from the definition (2.3) of L_i that we also have $L_i \in L^\infty(\mathbb{R}^3 \times \mathbb{S}^2; \mathbb{R})$. Indeed,

$$\|L_i\|_\infty \leq \frac{\|L_e\|_\infty}{\mu} < +\infty.$$

Hence, Theorem 1 and Lemma 2 provide the existence and uniqueness of the solution, as well as its expansion in Neumann series. Formula (2.5) is then found by induction. □

Our next step is now to recast representation (2.5) into a probabilistic formula.

2.3 Probabilistic representation

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

$$\mathcal{A} = \bigcup_{n=0}^{\infty} \mathcal{M}_n, \quad \text{with} \quad \mathcal{M}_n = \mathbb{R}_+^{n+1} \times \left(\mathbb{S}^2\right)^{n+1}. \quad (2.6)$$

We denote by $(\mathbf{r}, \boldsymbol{\omega})$ a generic element of $\mathcal{A}$ and by $(\mathbf{r}_n, \boldsymbol{\omega}_n)$ a generic element of $\mathcal{M}_n$ for $n \in \mathbb{N}$ with $\mathbf{r}_n = (r_0, \ldots, r_n)$ and $\boldsymbol{\omega}_n = (\omega_0, \ldots, \omega_n)$. If $(\mathbf{r}, \boldsymbol{\omega}) \in \mathcal{A}$, we set

$$|\mathbf{r}| = \sum_{n=1}^{\infty} n\, \mathbf{1}_{\mathcal{M}_n}(\mathbf{r}, \boldsymbol{\omega}) \quad (2.7)$$

and call it *size* or *length* of the path. For $n \in \mathbb{N}$, let

$$G_x^{(n)}(\mathbf{r}_n, \boldsymbol{\omega}_n) = L_e\left(x - \sum_{k=0}^{n} \omega_k r_k, \omega_n\right) \quad (2.8)$$

be defined on $\mathcal{M}_n$, and let

$$G_x(\mathbf{r}, \boldsymbol{\omega}) = \sum_{n=0}^{\infty} G_x^{(n)}(\mathbf{r}_n, \boldsymbol{\omega}_n)\, \mathbf{1}_{\mathcal{M}_n}(\mathbf{r}, \boldsymbol{\omega}) \quad (2.9)$$

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

$$\nu(F) = \sum_{n=0}^{\infty} (1-\rho)\rho^n \nu_n\left(F \cap \mathcal{M}_n\right), \quad (2.10)$$

where we recall that $\rho = \frac{\mu_s}{\mu}$ and where ν_n is the probability measure on $\mathcal{M}_n$ defined by

$$\nu_n(d\mathbf{r}_n, d\boldsymbol{\omega}_n) = \mu^{n+1}\, e^{-\mu \sum_{j=0}^{n} r_j} \prod_{j=0}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) d\mathbf{r}_n\, \sigma^{\otimes(n+1)}(d\boldsymbol{\omega}_n) \quad (2.11)$$

with $\prod_{j=0}^{-1} a_j = 1$ by convention. Before we express L as an expectation involving G_x and Y, let us state some properties of $Y = (\mathbf{R}, \boldsymbol{W})$.

Proposition 4. *Let $n \in \mathbb{N}$ and let π be a permutation of $\{0, \dots, n\}$. Let us recall that ν defined by (2.10) is the distribution of $Y = (\mathbf{R}, \boldsymbol{W})$. Then, conditionally on the event $\{Y \in \mathcal{M}_n\}$, the distribution of the variable $(R_0, \dots, R_n, W_0, \dots, W_n)$ is the probability measure ν_n defined by (2.11), which satisfies*

$$\nu_n(dr_0, \dots, dr_n, d\omega_0, \dots, d\omega_n) = \nu_n(dr_{\pi(0)}, \dots, dr_{\pi(n)}, d\omega_n, d\omega_{n-1}, \dots, d\omega_0). \tag{2.12}$$

In other words, on the event $\{Y \in \mathcal{M}_n\}$, the random variables $(R_{\pi(0)}, \dots, R_{\pi(n)}, W_n, W_{n-1}, \dots, W_0)$ and $(R_0, \dots, R_n, W_0, \dots, W_n)$ have the same distribution ν_n.

Proof. By definition of $\nu = \mathcal{L}(Y)$ and since $(\mathcal{M}_p)_{p\in\mathbb{N}}$ is a collection of pairwise disjoint sets, it is straightforward that ν_n is the distribution of the variable $(R_0, \dots, R_n, W_0, \dots, W_n)$ on the set $\{Y \in \mathcal{M}_n\}$. Let us now emphasize that the phase function is symmetric, i.e. $f_{HG}(\omega, \cdot) = f_{HG}(\cdot, \omega)$. Then replacing $f_{HG}(\omega_j, \omega_{j+1})$ by $f_{HG}(\omega_{j+1}, \omega_j)$ and using the invariance of the definition of ν_n by permutations of the variables $(r_0, \dots, r_n)$, we obtain Equation (2.12). □

Corollary 5. *For all $n \in \mathbb{N}$, conditionally on the events $\{Y \in \mathcal{M}_n\}$ and $\{Y \in \bigcup_{p\geq n} \mathcal{M}_p\}$, for any $j = 0, \dots, n$, the marginal distribution γ_{W_j} of the direction W_j is the uniform probability σ on $\mathbb{S}^2$. In particular, W_0 is uniformly distributed on $\mathbb{S}^2$.*

Proof. Let $n, j \in \mathbb{N}$ such that $j \leq n$. By proposition 4, on the event $\{Y \in \mathcal{M}_n\}$, the probability measure ν_n defined by (2.11) is the distribution of $(\mathbf{R}_n, \boldsymbol{W}_n)$. Then, integrating the law ν_n with respect to all variables except w_j and using that $f_{HG}(\omega, \cdot) = f_{HG}(\cdot, \omega)$ is a density function on $\mathbb{S}^2$, one obtains that on the event $\{Y \in \mathcal{M}_n\}$, W_j is uniformly distributed on $\mathbb{S}^2$. Since this also holds replacing n by any $p \geq n$ and since $\mathcal{M}_p$, $p \geq n$, are pairwise disjoint sets, this implies that on the event $\{Y \in \bigcup_{p\geq n} \mathcal{M}_n\}$, W_j is also uniformly distributed on $\mathbb{S}^2$. □

Proposition 6. *The series $\sum_{n=0}^{\infty} T^n L_i$ can be expressed as*

$$L(x, \omega) = \sum_{n=0}^{\infty} T^n L_i(x, \omega) = \frac{1}{\mu_a} \mathbb{E}\left[G_x(Y) \mid W_0 = \omega\right]. \tag{2.13}$$

Proof. We shall relate $[T_n \circ T_i] L_e$ with the measure ν_n defined above, which will be sufficient for our purposes. To this aim, consider $\hat{\omega}_0 \in \mathbb{S}^2$, and write a somehow more cumbersome version of formula (2.5) with $\ell = L_e$:

$$[T^n \circ T_i]\, L_e(x, \hat{\omega}_0) = \mu_s^n \int_{\mathbb{R}_+^{n+1}} dr_0 \cdots dr_n \, \exp\left(-\mu \sum_{j=0}^{n} r_j\right) \int_{(\mathbb{S}^2)^n} d\sigma^{\otimes n}(\omega_1, \dots, \omega_n)$$
$$f_{HG}\left(\hat{\omega}_0, \omega_1\right) \prod_{j=1}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) \, L_e\left(x - \hat{\omega}_0 r_0 - \sum_{k=0}^{n} \omega_k r_k, \omega_n\right).$$

Noting that $\mu_s^n = \frac{(1-\rho)\,\rho^n\,\mu^{n+1}}{\mu_a}$, we thus get the following identity:

$$[T^n \circ T_i]\, L_e(x, \hat{\omega}_0) = \frac{(1-\rho)\,\rho^n\,\mu^{n+1}}{\mu_a} \int_{\mathbb{R}_+^{n+1}} dr_0 \cdots dr_n \, \exp\left(-\mu \sum_{j=0}^{n} r_j\right) \int_{(\mathbb{S}^2)^{n+1}} d\hat{\sigma}_{n,\hat{\omega}_0}(\omega_0, \dots, \omega_n)$$
$$\prod_{j=0}^{n-1} f_{HG}\left(\omega_j, \omega_{j+1}\right) \, L_e\left(x - \sum_{k=0}^{n} \omega_k r_k, \omega_n\right),$$

where the measure $\hat{\sigma}_{n,\hat{\omega}_0}$ on $\mathbb{S}^{n+1}$ is given by $\hat{\sigma}_{n,\hat{\omega}_0}(d\omega_0, \dots, d\omega_n) = \delta_{\hat{\omega}_0}(d\omega_0) \otimes \sigma^{\otimes n}(d\omega_1, \dots, d\omega_n)$.

Finally set $\varphi_n(\mathbf{r}_n, \boldsymbol{\omega}_n) = L_e(x - \sum_{k=0}^{n} \omega_k r_k, \omega_n)$. Taking into account the identity above and (2.11), we easily get

$$[T^n \circ T_i]\, L_e(x, \hat{\omega}_0) = \frac{(1-\rho)\,\rho^n}{\mu_a} \, \nu_n\left(\varphi_n \middle| \, \omega_0 = \hat{\omega}_0\right),$$

from which our claim is straightforward. □

Remark 7. In the following, we will call the random variable $Y = (\mathbf{R}, \boldsymbol{W})$ 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 (2.13) 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 8. *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*

$$L_e(x, \omega) = c\, \mathbf{1}_{V_0 \times C^{2\alpha}}(x, \omega) := \begin{cases} c, & \text{if } (x, \omega) \in V_0 \times C^{2\alpha}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.14}$$

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 8 in mind, we are interested in estimating the *fluence rate* at the center of the voxels V_k, $k \neq 0$, that is the mean light intensity averaged in all directions

$$L(x_k) := \int_{\mathbb{S}^2} L(x_k, \omega_0)\, \sigma(d\omega_0). \tag{2.15}$$

This quantity admits a nice probabilistic representation.

Proposition 9. *Let $k \in \{0, \dots, K-1\}$ and let $Y = (\boldsymbol{R}, \boldsymbol{W})$ be a random variable with distribution ν defined by* (2.10)*. Then, the fluence rate $L(x_k)$ at the center x_k of the voxel V_k, which is defined by* (2.15)*, can also be expressed as*

$$L(x_k) = \frac{c}{\mu_a} \mathbb{P}\left(x_k - \sum_{j=0}^{|\boldsymbol{R}|} R_j W_j \in V_0,\, W_{|\boldsymbol{R}|} \in C^{2\alpha} \right) \tag{2.16}$$

where we recall that the length $|\boldsymbol{R}|$ of the ray Y is defined by (2.7)*.*

Proof. Invoking Proposition 6 and by definition of L, we get

$$L(x_k) = \frac{1}{\mu_a} \int_{\mathbb{S}^2} \mathbb{E}\left[G_{x_k}(Y) | W_0 = \omega_0 \right] \sigma(d\omega_0)$$

where G_{x_k} is defined by (2.9). Since by Corollary 5, σ is the distribution of W_0, the previous equation can be written as

$$L(x_k) = \frac{1}{\mu_a} \mathbb{E}\left[G_{x_k}(Y) \right].$$

Then using equations (2.8) and (2.7), we get

$$G_{x_k}(Y) = L_e\left(x_k - \sum_{j=0}^{|\boldsymbol{R}|} R_j W_j,\, W_{|\boldsymbol{R}|}\right).$$

Now applying Hypothesis 8, the random variable $G_{x_k}(Y)$ can be expressed as

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

which finishes our proof. □

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. This possibility stems from the invariance of ν stated in Proposition 4, and is exploited in the next result.

Proposition 10. *For any $0 \leqslant k \leqslant K-1$ we have :*

$$L(x_k) = \frac{c(1-\cos\alpha)}{2\mu_a}\,\mathbb{P}\left(S_N \in V_k\right), \tag{2.17}$$

where $N \sim \mathrm{NB}(1,\rho)$ is a negative binomial random variable with parameter $(1,\rho)$, that is a random variable whose law is given by $\mathbb{P}(N=n) = (1-\rho)\rho^n$ for all $n \in \mathbb{N}$, and where for $n \in \mathbb{N}$

$$S_n = \sum_{i=0}^{n} R_i W_i$$

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

- *$(R_i)_{i\geq 0}$ is a sequence of independent identically distributed (i.i.d.) exponentially random variables of parameter μ (i.e 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 $(W_0,\ldots,W_{i-1}) = (\omega_0,\ldots,\omega_{i-1})$ is $f_{HG}(\omega_{i-1},\omega_i)\sigma(d\omega_i)$.*
- *N, $(R_i)_{i\geqslant 0}$ and $(W_i)_{i\geqslant 0}$ are independent.*

Proof. Let $k \in \{0,\ldots,K-1\}$. 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 $V_0 + x_k = V_k$. Therefore, we can rewrite (2.16) as

$$L(x_k) = \frac{c}{\mu_a}\mathbb{P}\left(\sum_{j=0}^{|\boldsymbol{R}|} R_j W_j \in V_k, W_{|\boldsymbol{R}|} \in C^{2\alpha}\right), \tag{2.18}$$

where the distribution of $Y = (\boldsymbol{R}, \boldsymbol{W})$ is the probability measure ν defined by (2.10). Then,

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

where the distribution of $(R_0,\ldots,R_n,W_0,\ldots,W_n)$ is the probability ν_n defined by (2.11). Therefore, applying Proposition 4, we get

$$L(x_k) = \frac{c}{\mu_a}\sum_{n=0}^{+\infty}(1-\rho)\rho^n\mathbb{P}\left(\sum_{j=0}^{n} R_j W_j \in V_k, W_0 \in C^{2\alpha}\right).$$

By definition of ν_n, we easily see that

$$\mathbb{P}\left(\sum_{i=0}^{n} R_i W_i \in V_k,\, W_0 \in C^{2\alpha}\right) = \sigma(C^{2\alpha})\mathbb{P}\left(\sum_{i=0}^{n} R'_i W'_i \in V_k\right),$$

where $(R'_0, \dots, R'_n, W'_0, \dots, W'_n)$ is a random variable with distribution ν'_n, defined by replacing in (2.11) the uniform probability $\sigma(d\omega_0)$ on the sphere $\mathbb{S}^2$ by the uniform probability $\mathbf{1}_{\{\omega'_0 \in C^{2\alpha}\}}\sigma(d\omega'_0)/\sigma(C^{2\alpha})$ on the cone $C^{2\alpha}$. By definition of ν'_n, this means that

$$\mathbb{P}\left(\sum_{i=0}^{n} R_i W_i \in V_k,\, W_0 \in C^{2\alpha}\right) = \sigma(C^{2\alpha})\mathbb{P}(S_n \in V_k)$$

where $S = (S_p)_{p\geq 0}$ is the random walk defined in the statement of the proposition. Thanks to the fact that N is a $\mathrm{NB}(1,\rho)$ random variable independent of the random walk S, we easily get relation (2.17). □

Remark 11. The random variables N, R_i and W_0 in Proposition 10 are simulated in a straightforward way. 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 (2.4). To recover the cartesian coordinates of the directions, we inductively apply appropriate changes of frame. The corresponding formulas can be found in [21, p. 37].

Remark 12. 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 (2.17) for all $k \in \{0, \dots, K-1\}$ 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 negative binomial time. This classically means that $L(x_k)$ can be approximated by MC methods. We first derive the expression of the approximate fluence rate by means of the basic MC method and then describe the variance reduction method that we implemented.

Proposition 13. *Let us consider a random walk $S = (S_n)_{n\geq 0}$ and a negative binomial 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 $k = 0, \dots, K-1$,*

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

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 12, the family $(S^i, N_i)_{1\leq i\leq M}$ enables to estimate the fluence rate $L(x_k)$ for all $k = 0, \ldots, K-1$ at once.

The reader should be aware of the fact that the quantity ρ is large in general in biological tissues, which means that the size of the ray, given by $N \sim \mathrm{NB}(1, \rho)$, 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 (2.18), 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 14 (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 hold:

- $(W_0^j)_{1\leq j\leq M_{\mathrm{rot}}}$ 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 negative binomial random variable with parameter $(1, \rho)$.
- $\left(S_n^i\right)_{n\geq 0}$, $1 \leq i \leq M$, 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 $\left(S_n^i\right)_{n\geq 0}$ after a rotation corresponding to the j-th initial direction $W_0 = W_0^j$.
Then, for $k \in \{0, \ldots, K-1\}$, the MC-SOME estimate of $L(x_k)$ is defined by

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

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

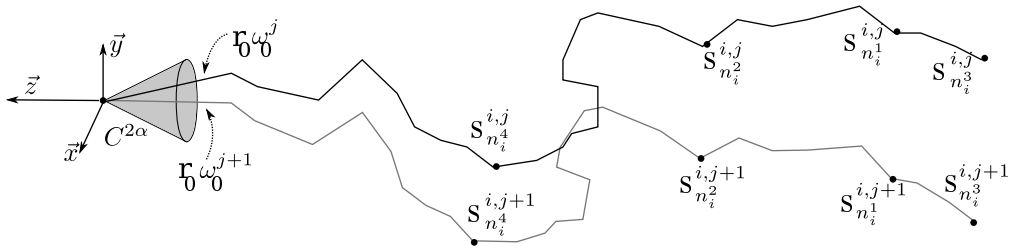

Figure 2: Description of MC-SOME method. In this example, the grey path is a rotation of the black one with respect to its initial direction ω_0^{j+1} and $M = 4$.

Remark 15. Choosing only some points $(S_{N^\ell}, 1 \leqslant \ell \leqslant M_{\text{points}})$ of the path instead of all $(S_i, 1 \leqslant i \leqslant n)$ provides a more efficient estimation. Indeed, it would take a lot of time to run over all voxels V_k in order to evaluate the indicator functions $\mathbf{1}_{\{S_i \in V_k\}}$. Moreover, the information brought by close points is in a sense redundant. Choosing the points according to a negative binomial law maintain the estimator unbiased, while speeding up the estimation.

4 A Metropolis-Hastings algorithm for light propagation

Inspired by results in computer graphics, see [16, 25, 27], we implemented a Metropolis Hastings algorithm which is a *Markov chain Monte Carlo method* (MCMC) by random walk [14, 23]. 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 $S = (S_n)_{0 \leqslant n \leqslant N}$ of Proposition 10 with the ray $(R_0, \dots, R_N, W_0, \dots, W_N)$ which defines it. The law of this ray, which will still be denoted by ν, is given by replacing in (2.10) the uniform measure $\sigma(d\omega_0)$ on the sphere by the uniform measure on the cone $C^{2\alpha}$.

A realization of the walk S stopped at $N = n$ will be indifferently referred to as $(S_0, \dots, S_n)$, as $(r_0, \dots, r_n, \omega_0, \dots, \omega_n)$ or as $(r_0, \theta_0, \varphi_0, \dots, r_n, \theta_n, \varphi_n)$ 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 $0 \leqslant k \leqslant K-1$, we are willing to estimate the conditional probability $\mathbb{P}(S_N \in V_k \mid W_0 = \omega_0)$ 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 (2.17).

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* $q(S(t), \cdot)$ 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 stays 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 [26, §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 $S(t) = (S_0(t), \dots, S_{N_t}(t))$ denotes the position of the chain at a given time t, then, for $0 \leqslant k \leqslant K-1$,

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

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 [19, 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 $M_{\text{rot}} \in \mathbb{N}^*$ and let $(\omega_0^1, \dots, \omega_0^{M_{\text{rot}}})$ be a sample of i.i.d. initial directions drawn according to $\mathcal{U}\left(C^{2\alpha}\right)$. For $i = 1, \dots, M_{\text{rot}}$ and $t = 1, \dots, T$, 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 $k \in \{0, \dots, K-1\}$, our Metropolis-Hastings estimator of $L(x_k)$ is defined by

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

This estimator is strongly consistent on condition that the proposal density $q(\cdot,\cdot)$ of Algorithm 1 provides a Harris recurrent chain.

4.2 Mutation strategy

One of the delicate issues in the implementation of MCMC methods is the choice of a convenient proposal density q. In our case, we have tested a mixture of mutations of two types for the Metropolis-Hastings algorithm: *rotation-translation* and *deletion-addition*. In order to describe them, let us introduce some notations. The method uses a *perturbed phase function* $f_{HG}^{\varepsilon g}$, $\varepsilon \in [-1, 1]$ which has an anisotropic coefficient εg instead of g. It also uses two coprime integers $1 \leqslant j < J$, which denote respectively the small and the big size of a length change.

Remark 16. The need for these two numbers to be coprime will be made clearer in the proof of Proposition 19, as it will ensure that the mutations produce paths of arbitrary length.

Definition 17 (Mutation rule). Let us assume that, at time $t \in \mathbb{N}^*$, the current ray is given by $S(t) = (r_0, \theta_0, \varphi_0, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t})$. 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 of type *deletion-addition*. Draw $\Delta(n_t)$ according to the following law that depends on the size of the current ray

$$\Delta(n) \sim \begin{cases} \mathcal{U}\left(\{-J, -j, j, J\}\right), & \text{if } n \geqslant J, \\ \mathcal{U}\left(\{-j, j, J\}\right), & \text{if } j \leqslant n < J, \\ \mathcal{U}\left(\{j, J\}\right), & \text{if } 0 \leqslant n < j. \end{cases}$$

- If $\Delta(n_t) < 0$, then delete the last $|\Delta(n_t)|$ edges of $S(t)$. The proposed path is
$$S'(t) = \left(r_0, \theta_0, \varphi_0, \dots, r_{n_t - \Delta(n_t)}, \theta_{n_t - \Delta(n_t)}, \varphi_{n_t - \Delta(n_t)}\right).$$
- If $\Delta(n_t) > 0$, then add $\Delta(n_t)$ new edges at the end of $S(t)$:
  - Draw $\left(r^{\text{new}}_{n_t+1}, \dots, r^{\text{new}}_{n_t+\Delta(n_t)}\right)$ i.i.d. according to $\mathcal{E}(\mu)$.
  - Draw $\left(\theta^{\text{new}}_{n_t+1}, \dots, \theta^{\text{new}}_{n_t+\Delta(n_t)}\right)$ i.i.d. according to $f^{\varepsilon g}_{HG}$.
  - Draw $\left(\varphi^{\text{new}}_{n_t+1}, \dots, \varphi^{\text{new}}_{n_t+\Delta(n_t)}\right)$ i.i.d. uniformly on $[0, 2\pi]$.

  The proposed path is
$$S'(t) = \left(r_0, \theta_0, \varphi_0, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t}, r^{\text{new}}_{n_t+1}, \theta^{\text{new}}_{n_t+1}, \varphi^{\text{new}}_{n_t+1}, \dots, r^{\text{new}}_{n_t+\Delta(n_t)}, \theta^{\text{new}}_{n_t+\Delta(n_t)}, \varphi^{\text{new}}_{n_t+\Delta(n_t)}\right).$$

(ii) With probability $\frac{1}{2}$, the mutation is of type *rotation-translation.* Choose an index i uniformly over $\{0, \dots, n_t\}$.

- If $i \neq 0$, then make a rotation of the path at the i-th edge:
  - Draw a new angle θ_i^{new} according to $f^{\varepsilon g}_{HG}$.
  - Draw a new angle φ_i^{new} uniformly on $[0, 2\pi]$.

  The proposed path is (see Fig. 3)
$$S'(t) = \left(r_0, \theta_0, \varphi_0, \dots, r_i, \theta_i^{\text{new}}, \varphi_i^{\text{new}}, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t}\right).$$
- If $i = 0$, then translate from the initial edge:
  - Draw a new edge length r_0^{new} according to $\mathcal{E}(\mu)$.

  The proposed path is
$$S'(t) = \left(r_0^{\text{new}}, \theta_0, \varphi_0, r_1, \theta_1, \varphi_1, \dots, r_{n_t}, \theta_{n_t}, \varphi_{n_t}\right).$$

Remark 18. The initial direction is fixed, thus mutations of (θ_0, φ_0) are forbidden. However, the length r_0 has to change from time to time. This is ensured by the translations and this is why the initial edge need to be considered separately in the mutation rule.

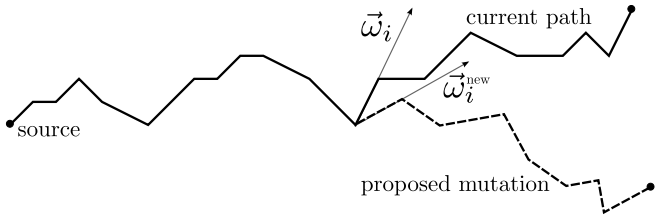

Figure 3: Metropolis-Hastings algorithm. Example of a mutation by rotation.

Let us now compute the proposal density $q(S, \cdot)$ of this mutation rule. For $m \in \mathbb{N}^*$, let
$$\zeta(m) = \begin{cases} \frac{1}{4}, & \text{if } m \geqslant J, \\ \frac{1}{3}, & \text{if } j \leqslant m < J, \\ \frac{1}{2}, & \text{if } 0 \leqslant m < j. \end{cases}$$

Assume that S' is a mutation of S, denote by i the first index where there is a difference between them, denote by n' and n their respective length and set $\Delta = n' - n$. We have

- if $\Delta = 0$ and $i \geqslant 1$, then $q(S, S') = \frac{1}{2}\,\frac{1}{n+1}\,\frac{1}{2\pi}\,f_{HG}^{\varepsilon g}(\theta'_i)$;
- if $\Delta = 0$ and $i = 0$, then $q(S, S') = \frac{1}{2}\,\frac{1}{n+1}\,\mu e^{-\mu r'_0}$;
- if $\Delta < 0$, then $q(S, S') = \frac{1}{2}\,\zeta(n)$;
- if $\Delta > 0$, then $q(S, S') = \frac{1}{2}\,\zeta(n) e^{-\mu \sum_{k=1}^{\Delta} r'_{n+k}} \left(\frac{\mu}{2\pi}\right)^{\Delta} \prod_{k=1}^{\Delta} f_{HG}^{\varepsilon g}\left(\theta'_{n+k}\right)$.

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 rule of mutations, Algorithm 1 produces a Markov chain that satisfies the LLN. This guarantees the convergence of the estimator defined in (4.2).

Proposition 19. *If the chain $(S(t))_{t\in\mathbb{N}^*}$ is obtained by Algorithm 1 with the mutation rule given in Definition 17, 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 $0 \leqslant k \leqslant K-1$.*

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 ([24, 26]). The Harris recurrence is then obtained by checking that the chain is irreducible with respect to $\nu|_{W_0=\omega_0}$, see [26, 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,

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

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

$$A = \{\omega_0\} \times I_0 \times \prod_{i=1}^{n} \left(I_i \times E_i\right), \tag{4.4}$$

where $n \in \mathbb{N}^*$ and where for all $1 \leqslant i \leqslant n$, the sets $I_i \subset \mathbb{R}_+$ and $E_i \subset \mathbb{S}^2$ 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 (2.11). By assumptions on A, we see easily that $\nu|_{W_0=\omega_0}(A) > 0$. Now, notice that by the Markov property, if τ_A and $\tau_{\{\omega_0\}\times I_0}$ denote respectively the time for the chain to be in A, resp. in $\{\omega_0\} \times I_0$, then we have

$$\mathbb{P}_s\left(\tau_A < +\infty\right) \geqslant \mathbb{P}_s\left(\tau_{\{\omega_0\}\times I_0} < +\infty\right) \mathbb{P}_{\{\omega_0\}\times I_0}\left(\tau_A < +\infty\right), \qquad \text{for all } s \in \mathcal{A}.$$

Now we can lower bound the right hand side of this relation in the following way:
(i) We have that $\mathbb{P}_s\left(\tau_{\{\omega_0\}\times I_0} < +\infty\right)$ is greater than the probability of deleting all the edges of s except (r_0, ω_0) and of modifying its length so that $r'_0 \in I_0$. 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 absolute continuity properties of $q(s, \cdot)$.
(ii) The same kind of argument works in order to lower bound $\mathbb{P}_{\{\omega_0\}\times I_0}\left(\tau_A < +\infty\right)$. 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 $(r'_i, \omega'_i) \in I_i \times E_i$, as well as its acceptance are always strictly positive.
We have thus obtained that $\mathbb{P}_s\left(\tau_{\{\omega_0\}\times I_0} < +\infty\right)\mathbb{P}_{\{\omega_0\}\times I_0}\left(\tau_A < +\infty\right) > 0$, which concludes the proof. □

Remark 20. 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 negative binomial distribution $\mathrm{NB}(1,\rho)$ 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 rate $L(x_k)$ provided by three methods: Monte Carlo with Wang-Prahl algorithm (denoted by WANG, see [21, 28]), MC-SOME (see (3.2)) and the Metropolis-Hastings (MH) (see (4.2)) with the mutation rule given in Definition 17. 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 [7, 11, 17]. The volume of the cube V equals 8 cm^3, that is $V = [-1,1]^3$. It is discretized into voxels whose volume 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 (2.16) was set $c = 1$.

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

WANG: $M = 6000$ photons trajectory.

MC-SOME: $M = 30000$ rays, $M_{\text{points}} = 40$ points chosen in each ray and $M_{\text{rot}} = 30$ rotations with respect to the initial direction.

MH $j = 10$, $J = 21$, $\varepsilon = 0.9$, $T = 250000$ steps of the chain and $M_{\text{rot}} = 30$ rotations with respect to the initial direction.

In Fig. 4, we picture for each methods, a zoom of the contour plot of the estimates in the plane $x = -0.04$. In Fig. 6, we compare the estimates along several lines of voxels $(\ell_i)_{i=1,\dots,6}$ which are parallel to the y-axis and pass through the points $(0,0,-0.08)$, $(0,0,-0.12)$, $(0,0,-0.16)$, $(0,0,-0.4)$, $(0,0,-0.48)$ and $(0,0,-0.6)$ respectively (see Fig. 5). We notice that MC-SOME gives more consistant estimates than the two other algorithms whose estimates are more noisy. Moreover, it seems that the MH-estimates are not very symmetric. Perhaps because the algorithm had not converged yet.

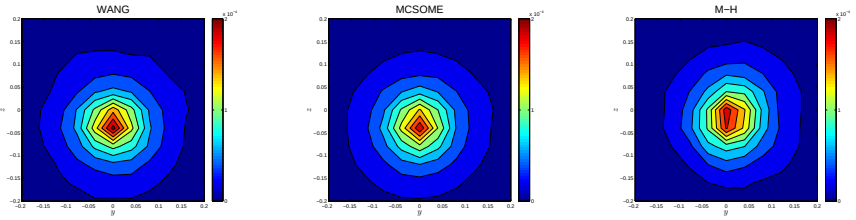

Figure 4: Contour plots of the fluence rate estimates in the plane $x = -0.04$ for WANG, MC-SOME and MH.

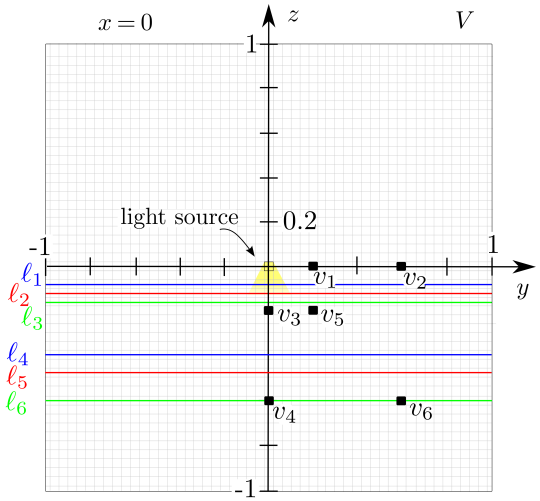

Figure 5: Choice of six particular voxels and position of the lines $(\ell_i)_{i=1,\dots,6}$ in the cube V.

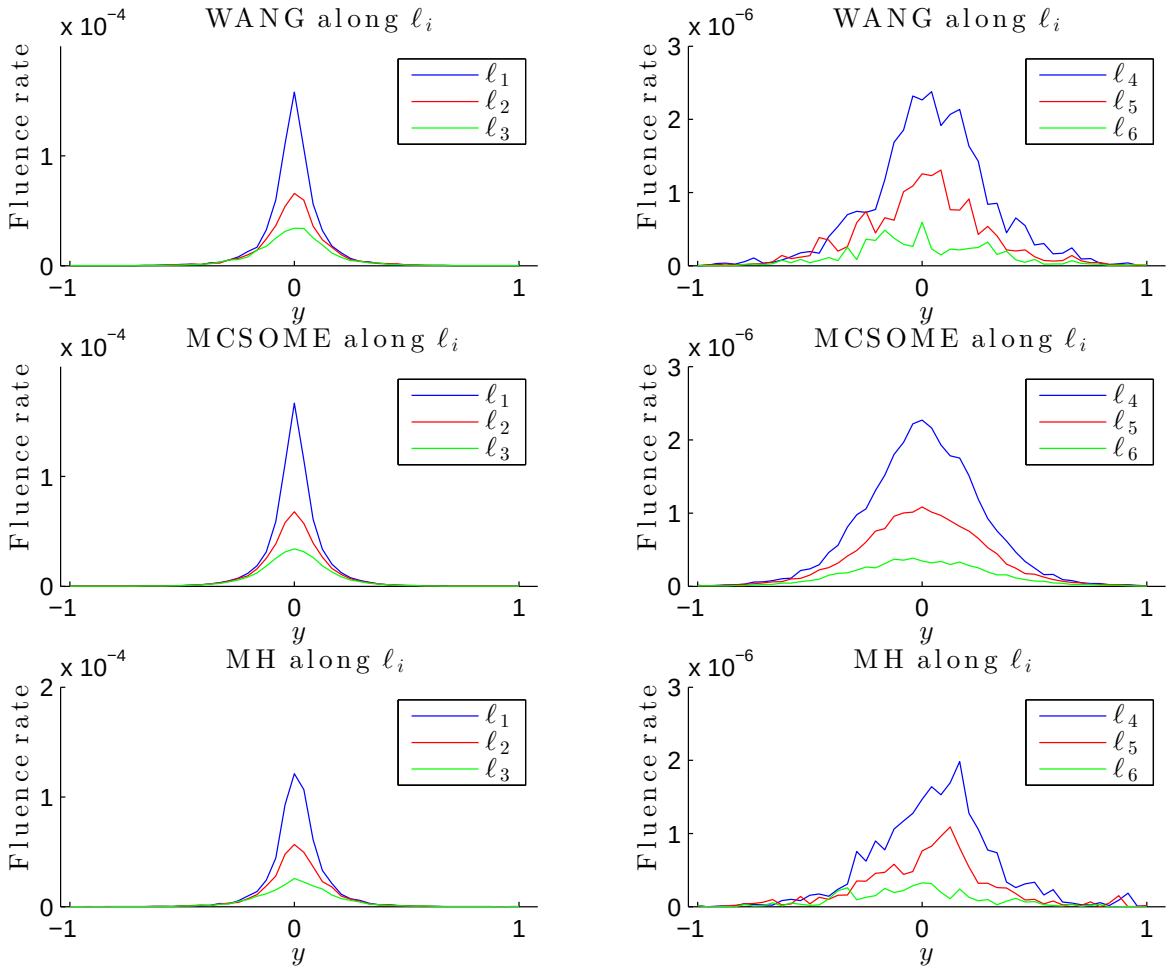

Figure 6: Fluence rate estimates along the lines $(\ell_i)_{i=1,\dots,6}$ with WANG, MC-SOME and MH.

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 gives spread estimates. This is due to the simulations for which the Markov chain has not yet converged. In Table 2, we have 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	$1.203 \cdot 10^{-5}$	$1.2366 \cdot 10^{-5}$	$1.3083 \cdot 10^{-5}$	$1.4484 \cdot 10^{-12}$	$3.9108 \cdot 10^{-13}$	$4.5158 \cdot 10^{-12}$
v_2	$2.4309 \cdot 10^{-7}$	$2.7177 \cdot 10^{-7}$	$1.6545 \cdot 10^{-7}$	$6.3788 \cdot 10^{-15}$	$2.5597 \cdot 10^{-15}$	$1.4411 \cdot 10^{-14}$
v_3	$2.0014 \cdot 10^{-5}$	$2.0033 \cdot 10^{-5}$	$1.8337 \cdot 10^{-5}$	$2.503 \cdot 10^{-12}$	$3.0408 \cdot 10^{-13}$	$1.6200 \cdot 10^{-11}$
v_4	$3.4947 \cdot 10^{-7}$	$3.5713 \cdot 10^{-7}$	$2.1497 \cdot 10^{-7}$	$1.4136 \cdot 10^{-14}$	$8.1737 \cdot 10^{-16}$	$1.5819 \cdot 10^{-14}$
v_5	$6.7519 \cdot 10^{-6}$	$6.6047 \cdot 10^{-6}$	$5.4786 \cdot 10^{-6}$	$8.0457 \cdot 10^{-13}$	$4.4781 \cdot 10^{-14}$	$1.1378 \cdot 10^{-12}$
v_6	$3.9952 \cdot 10^{-8}$	$4.217 \cdot 10^{-8}$	$1.8755 \cdot 10^{-8}$	$8.4058 \cdot 10^{-16}$	$6.6977 \cdot 10^{-17}$	$6.5413 \cdot 10^{-16}$

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

6 Inverse problem and sensitivity

For biologists, it is of considerable practical importance to have good estimates of the optical coefficients of the tissue they consider. One way to do this estimate is to compare simulated data with measurements of the fluence rate in the tissue and adjust the optical parameters of the simulation until obtaining values close to the measurements. Thanks to the probabilistic representation (2.16), this problem can be numerically solved as we shall see.

6.1 Sensitivity of the measurements

As a preliminary step towards a good resolution of the inverse problem, we first observe how fluence rate measurements vary with respect to the optical parameters g, μ_s and μ_a. To this aim, we built a small database of simulations for different values of the parameters and then compared the estimated fluence rate. 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.

First, we choose a reference simulation obtained with *reference parameters* (g^*, μ_a^*, μ_s^*). Then, we pick $n \geqslant 1$ voxels and consider their respective fluence rate estimates as *measurements*. That is, we choose n voxel centers $(x_{k_i})_{i=1,\dots,n}$ and define

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

where we recall that $L(x_{k_i})$ is defined by (2.15) with $\gamma_{W_0} = \mathbf{1}_{\{\omega_0' \in C^{2\alpha}\}} \sigma(d\omega_0')/\sigma(C^{2\alpha})$ 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

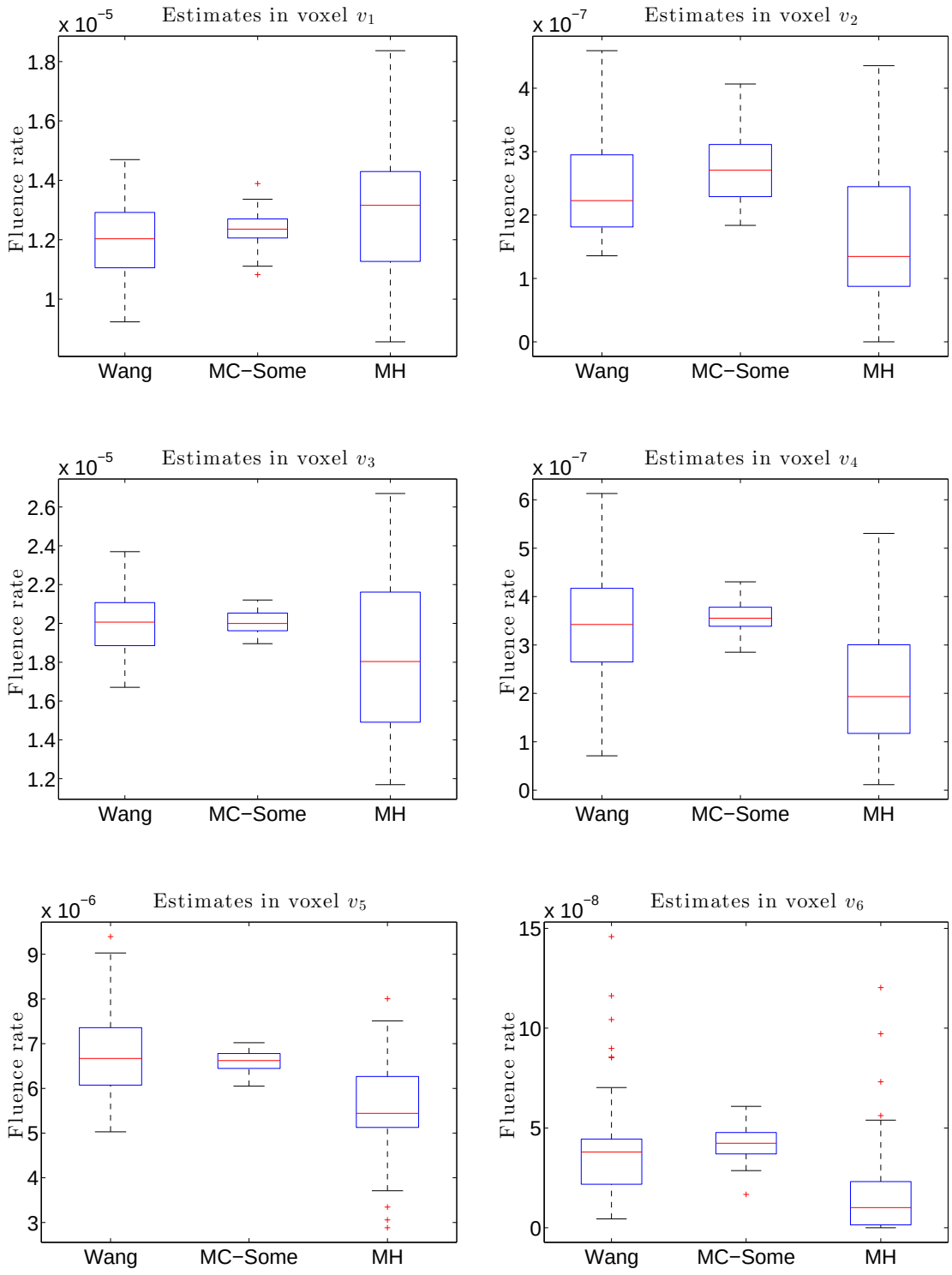

Figure 7: Boxplots of 50 fluence rate estimates in the voxels $(v_i)_{i=1,\dots,6}$ with WANG, MC-SOME and MH.

possible triplet of parameters (g, μ_a, μ_s), we compute the normalized quadratic error (or *evaluation error*)

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

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, 7, 11, 17]. 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, 17].

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 and 9 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. 8, 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.

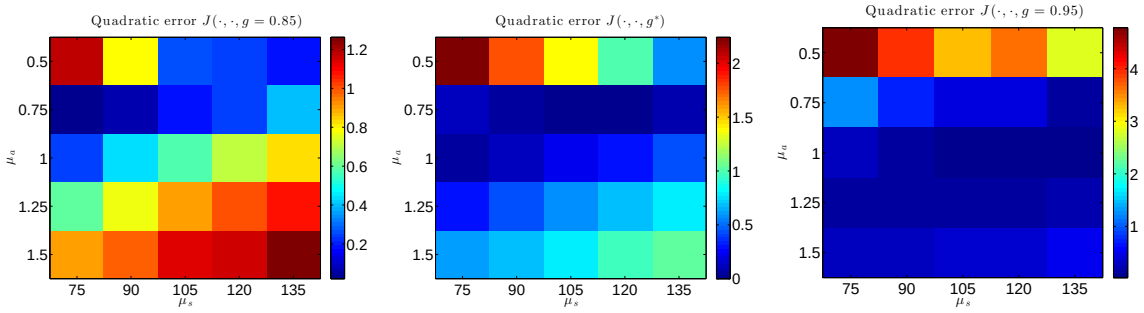

Figure 8: Colormap of the quadratic error $(\mu_a, \mu_s) \mapsto J(\mu_a, \mu_s, g)$ for three values of g, where μ_s is displayed on the x axis and μ_a on the y axis.

6.2 Parameters estimation

This section is devoted to the estimation of the parameters μ_a and μ_s only. Indeed, we have seen in the last section that the sensitivity of the fluence rate with respect to the anisotropic parameter g is low. Moreover, 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.

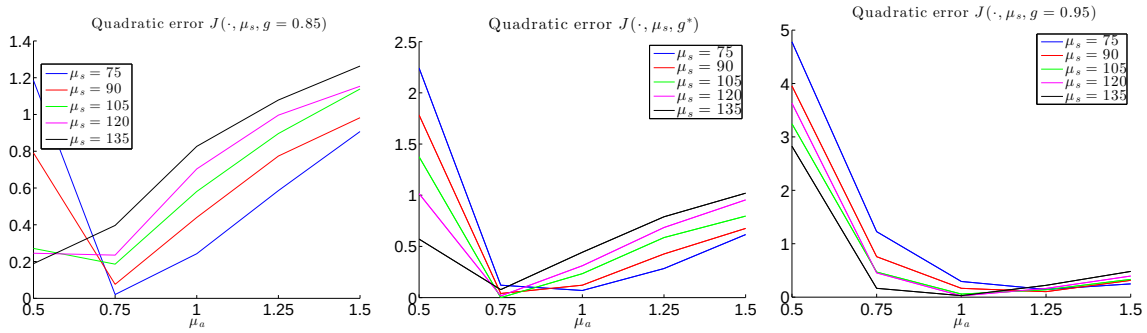

Figure 9: Quadratic error $\mu_a \mapsto J(\mu_a, \mu_s, g)$ for three values of g.

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 measurements in n different voxels centered at $(x_{k_i})_{i=1,\dots,n}$.

The optimization method that we implemented to solve this problem is based on the Levenberg-Marquardt algorithm (see [12]). This gradient descent algorithm involves the computation of the gradient, as well as the Hessian matrix of the *score function* J. It is described in pseudo-code in Algorithm 2. In this description, we have set $H_k = \mathrm{Hess}(J)(\mu_s^k, \mu_a^k)$ for $k \geqslant 0$. The term $\mathrm{Diag}(H_k)$ is the diagonal matrix of H_k, λ_k is the so-called *damping factor* which may be either constant or corrected at each step, and $\tau_k \in \mathbb{R}_+$ controls the step size of each iteration.

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 $\varepsilon > 0$.

1: $k \leftarrow 0$
2: **while** $J(\mu_s^k, \mu_a^k) > \varepsilon$ **do**
3: $\quad (\mu_s^{k+1}, \mu_a^{k+1}) \leftarrow (\mu_s^k, \mu_a^k) - \tau_k \left[H_k + \lambda_k \, \mathrm{Diag}(H_k) \right]^{-1} \nabla J(\mu_s^k, \mu_a^k)$
4: $\quad k \leftarrow k + 1$
5: **end while**

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

The simple 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. 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

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

Moreover, as stated in the following proposition, the formal representation in Proposition 6 allows 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 (2.18).

Proposition 21. *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* (2.17)*, they are given by*

$$\frac{\partial L}{\partial \mu_a}(x_{k_i};\mu_s,\mu_a) = -\frac{c(1-\cos(\alpha))}{2\mu_a}\mathbb{E}\left(\mathbf{1}_{\{S_N\in V_{k_i}\}}\sum_{j=0}^{N} R_j\right) \tag{6.5}$$

$$\text{and}\quad \frac{\partial L}{\partial \mu_s}(x_{k_i};\mu_s,\mu_a) = \frac{c(1-\cos(\alpha))}{2\mu_a}\mathbb{E}\left(\mathbf{1}_{\{S_N\in V_{k_i}\}}\left(\frac{N}{\mu_s}-\sum_{j=0}^{N} R_j\right)\right). \tag{6.6}$$

Proof. We start by differentiating term-by-term the Neumann series of Corollary 3. Let us first note that by definition of L_i we have:

$$\frac{\partial L_i}{\partial \mu_a}(x,\omega_0;\mu_s,\mu_a) = -\int_0^{+\infty} r\exp(-(\mu_s+\mu_a)r)L_e(x-rw,w)dr.$$

Moreover, for $n\geqslant 1$, by definition of $T^n\circ T_i$ (see (2.5)), we also have:

$$\frac{\partial\left[T^n\circ T_i\right]L_e}{\partial\mu_a}(x,\omega_0;\mu_s,\mu_a) = \mu_s^n\int_{\mathbb{R}_+^n} dr_0\cdots dr_n\left(-\sum_{j=0}^{n} r_j\right)\exp\left(-(\mu_s+\mu_a)\sum_{j=0}^{n} r_j\right)$$
$$\int_{(\mathbb{S}^2)^{(n+1)}} d\sigma^{\otimes(n+1)}(\omega_0,\ldots,\omega_n)\prod_{j=0}^{n-1} f_{HG}\left(\omega_j,\omega_{j+1}\right)\,L_e\left(x-\sum_{k=0}^{n} r_k\omega_k,\omega_n\right).$$

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

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

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 (2.17) and the same arguments provide (6.6), considering that for all $n\geq 0$

$$\frac{\partial\left[T^n\circ T_i\right]L_e}{\partial\mu_s} = \frac{n}{\mu_s}\left[T^n\circ T_i\right]L_e + \frac{\partial\left[T^n\circ T_i\right]L_e}{\partial\mu_a}.$$

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 $T^n\circ T_i$ is infinitely continuously differentiable for all n and by the uniform convergence of the corresponding sequences of truncated sums

$$s_m = \sum_{n=0}^{m}\frac{\partial[T^n\circ T_i]L_e}{\partial\mu_s}(x,\omega_0;\mu_s,\mu_a),\qquad m\geqslant 0.$$

□

Remark 22. Similar formula to (6.5) and (6.6) can be easily found for the second order derivatives $\frac{\partial^2 L}{\partial\mu_s^2}$, $\frac{\partial^2 L}{\partial\mu_s\partial\mu_a}$ and $\frac{\partial^2 L}{\partial\mu_a^2}$.

The probabilistic representation of $L(x_{k_i};\mu_s,\mu_a)$ in (2.18) 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

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

Implementation and discussion. The randomness coming from Monte Carlo estimation of the score J, of its gradient and of its Hessian matrix during the run of the algorithm, makes a precise estimation of the real values of (μ_a^*, μ_s^*) difficult. We observed that far from the real value μ_a^*, the eigenvalues of the Hessian matrix $\hat{H}_k$ are very small and that their sign can vary a lot because of the volatility of the estimates. Conversely, near the real value μ_a^*, the estimate $\hat{H}_k$ is more robust and its eigenvalues are almost always both positive, which legitimates the quadratic approximation of Levenberg-Marquardt algorithm. For these reasons, we implemented a *hybrid* algorithm which chooses between the Levenberg-Marquardt descent and the classic steepest gradient descent depending on 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 opposite direction of the gradient, $-\hat{\nabla} J$.

In Fig. 10 and 11, we can see two examples of descent of our algorithm. The settings are the following: *(1)* we choose $n = 3$ positions for the measurements: $x_{k_1} \in v_2$, $x_{k_2} \in v_4$, $x_{k_3} \in v_6$ (see Fig. 5) and the values of the measurements m_1, m_2 and m_3 are taken from the database of simulations described in Sec. 6.1 with the desired parameters, *(2)* the anisotropy factor is set to $g = 0.9$, *(3)* the damping parameter is constant $\lambda = 0.01$, *(4)* the precision parameter (see Algo. 2) is set to $\varepsilon = 0.005$, *(5)* the sequence $(\tau_k)_{k \geqslant 1}$ controlling the step size of each iteration is decreasing in k and depends on the score of the iteration, as well as on the sign of the eigenvalues of H_k.

In Fig. 10, the reference parameter are $(\mu_a^* = 1, \mu_s^* = 75)$. Notice the oscillations around μ_a^*. Those descent zigzags near the real value of μ_a are also apparent in the other descent in Fig. 11 for which $(\mu_a^* = 1, \mu_s^* = 105)$. They correspond to iteration where the descent is done according to the classic steepest descent. The iterations for which one moves more vertically correspond, as for them, to the case where the descent is done according to (6.7). As we can see, a satisfying estimate of μ_a^* comes up rapidly, whereas μ_s^* is more difficult to approach. This is due to the low sensitivity of J with respect to μ_s discussed in the previous section.

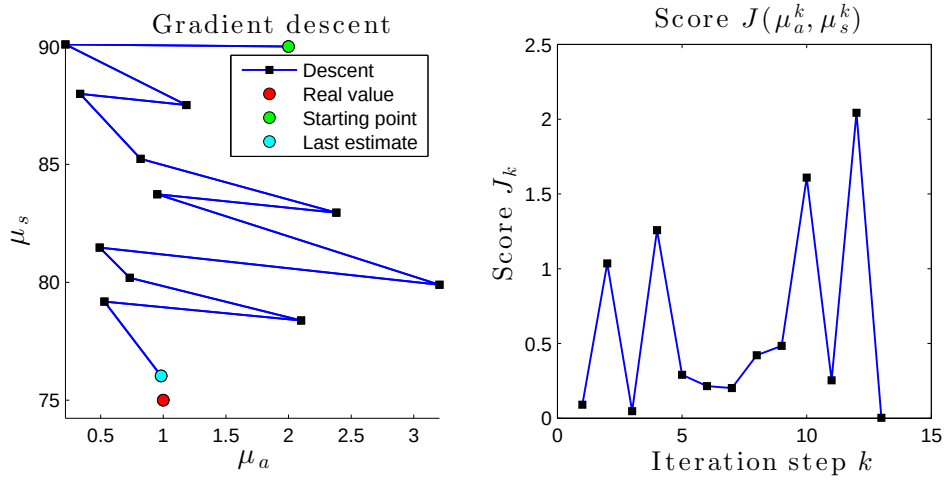

Figure 10: Parameters estimation with an adaptation of Levenberg-Marquardt descent algorithm. The real value of (μ_a^*, μ_s^*) is $(1, 75)$. The starting point is $(2, 90)$. The final estimate is $(1.07, 76.02)$ with a score equal to $J = 0.0035$.

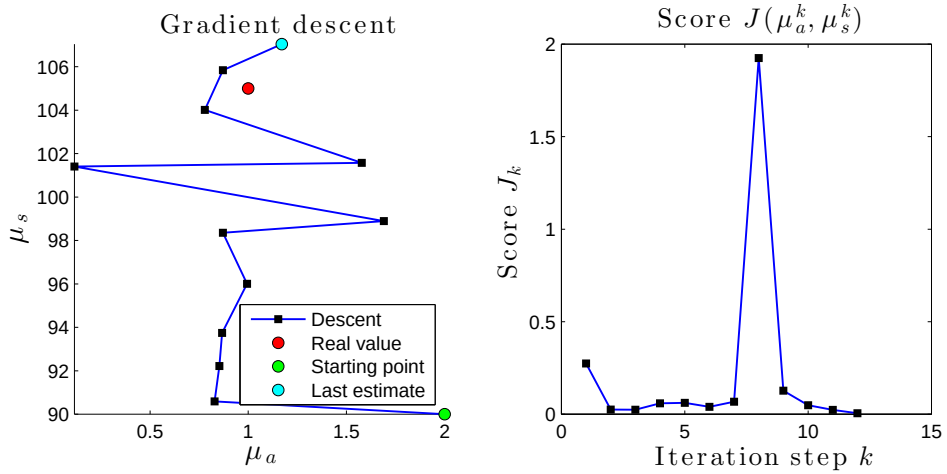

Figure 11: Parameters estimation with an adaptation of Levenberg-Marquardt descent algorithm. The real value of (μ_a^*, μ_s^*) is $(1, 105)$. The starting point is $(2, 90)$. The final estimate is $(1.17, 107.04)$ with a score equal to $J = 0.0048$.

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), 2007.

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

[3] J. Arvo. Transfer functions in global illumination. In *ACM SIGGRAPH*, 1993.

[4] 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), 2005.

[5] D. Bechet, S. R. Mordon, F. Guillemin, and M. A. Barberi-Heyob. Photodynamic therapy of malignant brain tumours: A complementary approach to conventional therapies. *Cancer Treatment Reviews*, 40(2), 2014.

[6] H. Benachour, A. Sève, T. Bastogne, C. Frochot, R. Vanderesse, J. Jasniewski, I. Miladi, C. Billotey, O. Tillement, F. Lux, and M. Barberi-Heyob. Multifunctional peptide-conjugated hybrid silica nanoparticles for photodynamic therapy and mri. *Theranostics*, 2(9), 2012.

[7] 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), 1999.

[8] K. Bhan and J. Spanier. Condensed history Monte Carlo methods for photon transport problems. *Journal of Computational Physics*, 225(2), 2007.

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

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

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

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

[13] 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), 2012.

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

[15] L. G. Henyey and J. L. Greenstein. Diffuse radiation in the galaxy. *The Astrophysical Journal*, 93, 1941.

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

[17] 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), 2005.

[18] H. Karlsson, I. Fredriksson, M. Larsson, and T. Strömberg. Inverse Monte Carlo for estimation of scattering and absorption in liquid optical phantoms. *Optics Express*, 20(11), 2012.

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

[20] 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), 2006.

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

[22] 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, 1989.

[23] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1, 2004.

[24] 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), 2006.

[25] 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*. Springer, 2008.

[26] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 1994.

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

[28] 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.

[29] B. C. Wilson and M. S. Patterson. The physics of photodynamic therapy. *Physics in Medicine and Biology*, 31(4), 1986.

[30] B. C. Wilson and M. S. Patterson. The physics, biophysics and technology of photodynamic therapy. *Physics in Medicine and Biology*, 53(9), 2008.

[31] K. Yoshida. *Functional analysis*. Classics in Mathematics, Springer. Springer London, Limited, 1980.

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