

A Monte Carlo Method for 3D Radiative Transfer Equations with Multifractional Singular Kernels

Christophe Gomez* Olivier Pinaud†

October 10, 2022

Abstract

We propose in this work a Monte Carlo method for three dimensional scalar radiative transfer equations with non-integrable, space-dependent scattering kernels. Such kernels typically account for long-range statistical features, and arise for instance in the context of wave propagation in turbulent atmosphere, geophysics, and medical imaging in the peaked-forward regime. In contrast to the classical case where the scattering cross section is integrable, which results in a non-zero mean free time, the latter here vanishes. This creates numerical difficulties as standard Monte Carlo methods based on a naive regularization exhibit large jump intensities and an increased computational cost. We propose a method inspired by the finance literature based on a small jumps - large jumps decomposition, allowing us to treat the small jumps efficiently and reduce the computational burden. We demonstrate the performance of the approach with numerical simulations and provide a complete error analysis. The multifractional terminology refers to the fact that the high frequency contribution of the scattering operator is a fractional Laplace-Beltrami operator on the unit sphere with space-dependent index.

Key words. radiative transfer, singular scattering kernels, Monte Carlo method, wave propagation, random media, long-range correlations.

1 Introduction

Radiative transfer models have been used for more than a century to describe wave energy propagation through complex/random media [23, 9], as well as neutron transport [30, 40], heat transfer [43], and are still an active area of research in astrophysics, geophysics, and optical tomography [29, 33, 34, 35] for instance. In this work, we propose a new Monte Carlo (MC) method to simulate the following radiative transfer equation (RTE)

$$\begin{cases} \partial_t u + \hat{k} \cdot \nabla_x u = \mathcal{Q}u, \\ u(t=0, x, \hat{k}) = u_0(x, \hat{k}), \end{cases} \quad (t, x, \hat{k}) \in (0, \infty) \times \mathbb{R}^3 \times \mathbb{S}^2, \quad (1)$$

where \mathbb{S}^2 denotes the unit sphere in \mathbb{R}^3 , and u is the wave energy density in the context of wave propagation or a particle distribution function in the context of neutronics. The scattering operator \mathcal{Q} has the standard form

$$(\mathcal{Q}u)(x, \hat{k}) = \lambda(x) \int_{\mathbb{S}^2} \Phi(x, |\hat{p} - \hat{k}|) (u(x, \hat{p}) - u(x, \hat{k})) \sigma(d\hat{p}), \quad (2)$$

for $\sigma(d\hat{p})$ the surface measure on \mathbb{S}^2 , Φ the scattering kernel, and $\lambda > 0$ a function modeling the support of the scattering process. Regions where $\lambda(x) = 0$ are homogeneous and u undergoes free transport. MC methods have long been used for the resolution of (1), see e.g. [27, 40]. The originality and difficulty in our work lies in the fact that we consider situations where the mean free time t_0 associated with \mathcal{Q} vanishes in the scattering regions, that is

$$\frac{1}{t_0(x)} = \lambda(x) \int_{\mathbb{S}^2} \Phi(x, |\hat{k} - \hat{p}|) \sigma(d\hat{p}) = +\infty, \quad \text{where } \lambda(x) > 0, \quad (3)$$

and as a consequence the standard MC representations of u do not apply. Such a scenario arises for instance in the context of highly peaked-forward light scattering in biological tissues and in turbulent atmosphere, or more

*Aix-Marseille Univ, CNRS, I2M, christophe.gomez@univ-amu.fr

†Department of Mathematics, Colorado State University, Fort Collins CO, pinaud@math.colostate.edu

generally in the context of wave propagation in random media with long-range correlations that we describe below. In this paper we write Φ as

$$\Phi(x, |\hat{p} - \hat{k}|) := \frac{a(|\hat{p} - \hat{k}|)}{|\hat{p} - \hat{k}|^{2+\alpha(x)}} = \frac{1}{2^{1+\alpha(x)/2}} \rho(x, \hat{k} \cdot \hat{p}), \quad \text{with} \quad \rho(x, s) := \frac{a(\sqrt{2(1-s)})}{(1-s)^{1+\alpha(x)/2}} \quad s \in [-1, 1]. \quad (4)$$

Above, $\alpha : \mathbb{R}^3 \rightarrow [0, 2)$ accounts for the slow variations of scattering across the ambient space, and a is a smooth bounded function characterizing some statistical properties of the medium and such that $a(0) > 0$. Practical examples are given further. A direct calculation shows that (3) holds when $\alpha \in [0, 2)$. Also, the integral in (2) has to be understood in the principal value sense when $\alpha \in [1, 2)$, see [18]. The multifractional terminology that we use is motivated by the fact that the unbounded operator \mathcal{Q} can be expressed as a (multi)-fractional Laplace-Beltrami operator $(-\Delta_{\mathbb{S}^2})^{\alpha(x)/2}$ on the unit sphere up to a bounded operator w.r.t. the \hat{k} variable [17, 18].

We would like to emphasize that we focus in this work on kernels of the form (4) for simplicity of the exposition, and that our method applies, after proper decomposition (see [18]), to more general kernels that behave like (4) at the singularity.

The RTE can be derived from high frequency wave propagation in random media, see e.g. [38]. In such a context, the velocity field $c(x)$ has the form

$$\frac{1}{c^2(x)} = \frac{1}{c_0^2} \left(1 + \sqrt{\eta} V_0 \left(x, \frac{x}{\eta} \right) \right) \quad x \in \mathbb{R}^3, \quad \eta \ll 1,$$

where c_0 is the background velocity (that we set to one in the sequel for simplicity), V_0 is a mean zero random field modeling fluctuations around the background, and η is the correlation length of the random medium, assumed to be small after proper rescaling. The first variable in V_0 represents the slow variations of the random perturbations, while the second one corresponds to their high frequency oscillations. The latter are responsible for the strong interaction between the wave and the medium over sufficient distances. The scattering kernel Φ is related to the correlation function of V_0 , and assuming V_0 is stationary (in the statistical sense) with respect to the fast variable, a kernel of the form (4) can be obtained from random fields such that

$$\mathbb{E}[V_0(x, x')V_0(y, y')] = \sqrt{\lambda(x)\lambda(y)} \int_{\mathbb{R}^3} \frac{a(|p|)}{|p|^{1+\frac{\alpha(x)+\alpha(y)}{2}}} e^{ip \cdot (x' - y')} dp, \quad (5)$$

with α ranging from 0 to 2. Denoting by $R(x)$ the expectation in (5) with $y = x$, $y' = x' + x/\eta$, one can show that R behaves like $|x|^{\alpha(x)-2}$ for $|x| \gg 1$, and is therefore not integrable. This is how random fields with long-range correlations are defined, as opposed to random fields with short-range correlations that exhibit an integrable correlation function.

In Figure 1, we provide examples of such 2D random fields. The top-left picture represents a random medium with short-range correlations (with a standard Gaussian covariance kernel), while the top-right picture illustrates a random medium with long-range correlations with $\alpha \equiv 1$. Because of the singularity at $p = 0$, one can observe significantly larger statistical patterns than in the short-range case. In the bottom two pictures, we highlight the roles of λ and α : λ characterizes scattering regions, and α defines the correlation structure. In the inner circle of the bottom-left picture we have $\alpha \equiv 0.1$, which tends to create shorter range fluctuations than in the outside where $\alpha \equiv 1$. In the bottom-right picture, we have a three-layer model for α in which the inner band exhibits smaller statistical patterns than the outer ones. This type of model is used for modeling non-Kolmogorov atmospheric turbulences, while standard atmospheric turbulence is modeled with the so-called Kolmogorov power spectrum

$$\Phi(|k|) \propto \frac{a(|k|)}{|k|^{11/3}},$$

for $|k|$ in the inertial range of turbulence. This corresponds to the case $\alpha = 5/3$. Experimental observations of atmospheric turbulence statistics have though been reported to not obey to the 11/3 power law and can vary with altitude [3, 41, 44]. Models have been derived for instance (see [26] for a review) by considering three ranges (0-2km, 2-8km, and above 8km) with distinct power laws (see Figure 16 for an illustration).

In the context of biological tissues, the following the Gegenbauer scattering kernel ρ_G and Henyey-Greenstein (HG) kernel ρ_{HG} are commonly used in the peaked-forward regime [21, 37]:

$$\rho_G(x, s) := \frac{\alpha g (1 + g^2 - 2g s)^{-1-\alpha/2}}{2\pi((1-g)^{-\alpha} - (1+g)^{-\alpha})}, \quad \rho_{HG}(x, s) := \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g s)^{3/2}}. \quad (6)$$

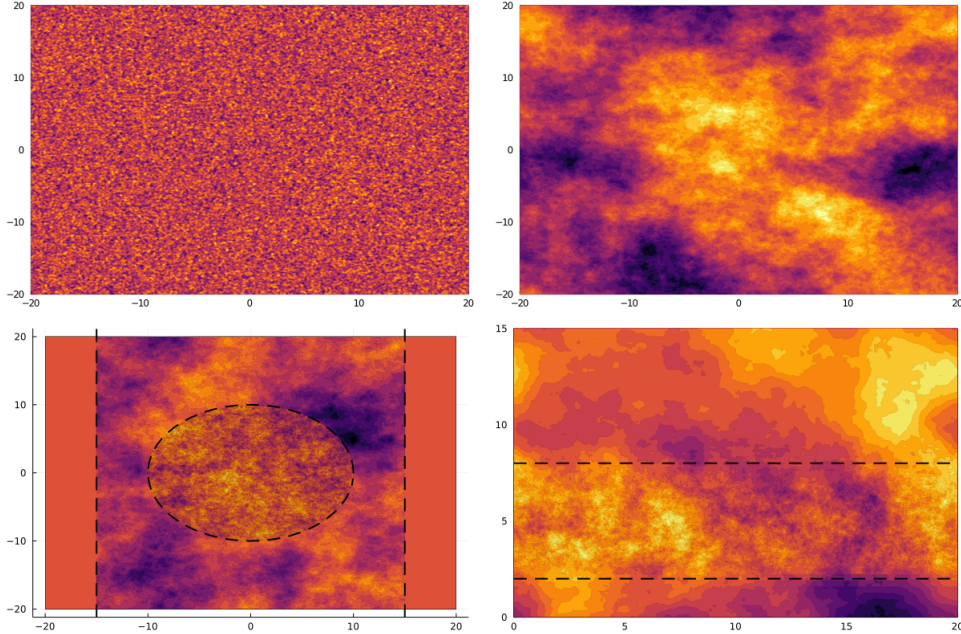


Figure 1: Realizations of Gaussian random fields. The upper-left picture represents a field with short-range correlations, while the upper-right picture depicts a field with long-range correlations with $\lambda \equiv 1$ and $\alpha \equiv 1$. In the lower-left picture, we have $\lambda = \mathbf{1}_{\{|x_1| < 15\}}$, and $\alpha = 0.1 \cdot \mathbf{1}_{\{|x| \leq 10\}} + 1 \cdot \mathbf{1}_{\{10 < |x|\}}$. In the lower-right, we have $\lambda \equiv 1$ and $\alpha(x_2) = 5/3 \cdot \mathbf{1}_{\{x_2 \leq 2\}} + 0.5 \cdot \mathbf{1}_{\{2 < x_2 \leq 8\}} + 1.9 \cdot \mathbf{1}_{\{8 < x_2\}}$.

The parameter $g \in (-1, 1)$ is called the anisotropy factor, and ρ_{HG} is obtained by setting $\alpha \equiv 1$ in ρ_G . The case $g = 0$ corresponds to isotropic energy transfer over the unit sphere, $g < 0$ to dominant transfer in the backward direction, and $g > 0$ to forward energy transfer. The peaked forward regime is obtained in the limit $g \rightarrow 1$, for which

$$\frac{1}{(1-g)^\alpha} \rho_G(x, \hat{k} \cdot \hat{p}) \underset{g \rightarrow 1}{\sim} \frac{\alpha}{2\pi(2 - 2\hat{k} \cdot \hat{p})^{1+\alpha/2}} = \frac{\alpha}{2\pi|\hat{k} - \hat{p}|^{2+\alpha}}. \quad (7)$$

The case $\alpha \equiv 1$ for the HG kernel is widely used in photon scattering in biological tissues [11, 16, 22]. A typical realization of the corresponding random field in 2D as $g \rightarrow 1$ is depicted in the top-right panel of Figure 1.

There exist a variety of methods for the resolution of (1) that handle the singular nature of the HG kernel, see e.g. [14, 15, 24, 25, 28]. They are based on finite differences type discretizations, projections over appropriate bases w.r.t. the \hat{k} variable, and approximations of the kernel. Here we propose an alternative approach to handle singular scattering kernel (4) that is based on a MC method. The latter are popular choices for the simulation of the RTE when the kernel is smooth, see e.g. [27, 31, 32, 36, 40], essentially for their adaptability to a wide range of configurations and their simplicity of implementation. A downside is their slow convergence rate, and there is a vast literature on variance reduction techniques for acceleration. In this work, we focus on the design of an efficient MC method and postpone any variance reduction considerations to future works.

Our approach is based on an adaptation of a method proposed by Asmussen-Cohen-Rosiński [2, 10] (ACR) for the simulation of Lévy processes with infinite jump intensity. It relies on a *small jumps/large jumps* decomposition of the corresponding infinitesimal generator. The main idea is to approximate the generator of the *small-jump* part, which possesses the infinite intensity due to the singularity of kernel, by a Laplace-Beltrami operator (with respect to the angular variables) on the unit sphere \mathbb{S}^2 . This requires us to simulate paths of a jump-diffusion process over the unit sphere. For this purpose, we use the characterization of Brownian motion on the unit sphere given in [4] based on a standard stochastic differential equation (SDE) in \mathbb{R}^3 that is suitable for space-dependent kernels. This situation is hence more involved than the 2D case we investigated in [19] where the small jumps part can be approximated by Brownian motion on the unit circle for which analytical expressions are available. Note that, as shown in [19], neglecting small jumps altogether in order to use standard MC methods leads to large errors, and reducing those comes at significantly increased computational cost.

Denoting by $\hat{\mu}(u)$ the estimator produced by our MC method for some observable $\mu(u)$ built on the solution

u to (1), we provide an error estimate of the form

$$\mathbb{P}\left(|\hat{\mu}(u) - \mu(u)| > E_1 + E_2 + E_3\right) \ll 1$$

as a theoretical support of our method. Above, E_1 , E_2 , and E_3 are small terms characterizing the various approximation errors from the original model: the Laplace-Beltrami (i.e. small jumps) approximation, the discretization error of the diffusion process over the unit sphere, and the MC error. Note that the method we propose here applies directly to the stationary version of (1)

$$\hat{k} \cdot \nabla_x \mathbf{u} - \mathcal{Q}\mathbf{u} = u_0, \quad (x, \hat{k}) \in \mathbb{R}^3 \times \mathbb{S}^2,$$

with source term u_0 , through the relation

$$\mathbf{u}(x, \hat{k}) := \int_0^\infty u(t, x, \hat{k}) dt.$$

The paper is organized as follows. In Section 2, we introduce probabilistic representations for (1) and its approximation based on the ACR method. In Section 3, we describe our MC method, state the main theoretical result regarding the overall approximation error, and detail the simulation algorithms. Section 4 is dedicated to the validation of the method using semi-analytical solutions. Numerical illustrations are given in Section 5, where we investigate the role of the strength α of the singularity, both when constant or space-dependent in the case of non-Kolmogorov turbulence, and compare with solutions for the HG kernel. Section 6 is devoted to the proofs of our main results and we recall in an Appendix the stochastic collocation method.

The numerical simulations are performed using the Julia programming language (v1.6.5) on a NVIDIA Quadro RTX 6000 GPU driven by a 24 Intel Xeon Sliver 2.20GHz CPUs station. The codes have been implemented using the CUDA.jl library [7, 8].

Acknowledgment. OP acknowledges support from NSF grant DMS-2006416.

2 Probabilistic representations and approximation

2.1 Representation for (1)

The starting point is the following standard probabilistic interpretation to (1):

$$u(t, x, k) = \mathbb{E}_{x, \hat{k}}[u_0(D(t))] := \mathbb{E}[u_0(D(t)) | D(0) = (x, \hat{k})],$$

where $D = (X, K)$ is a Markov process on $\mathbb{R}^3 \times \mathbb{S}^2$ with infinitesimal generator

$$\mathcal{L}f(x, \hat{k}) := -\hat{k} \cdot \nabla_x f(x, \hat{k}) + \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{\mathbb{S}^2} \rho(x, \hat{p} \cdot \hat{k}) (f(x, \hat{p}) - f(x, \hat{k})) \sigma(d\hat{p}).$$

A path, or a realization, of the Markov process D is often referred to as a *particle trajectory*. The X component of D represents the position of a particle, and the component K its direction. The generator \mathcal{L} comprises two terms, the transport part describing free propagation of the particle, and the scattering operator (often referred to as the jump part in the probabilistic literature) describing the evolution of its direction. The jump component exhibits a non-integrable singularity leading to a infinite jump intensity and a vanishing mean free time as expressed in (3).

Note that when λ and α are constant, it is shown in [18] that the solution u is unique and infinitely differentiable in all variables for $t > 0$ for any square integrable initial condition. When λ and α are infinitely differentiable with bounded derivatives at all orders, this result remains valid and we will assume throughout this work that u is smooth. The same applies to the function u_ε defined further in Proposition 2.1.

In order to adapt the ACR method, we introduce the following small *critical* region over which the singularity of the infinitesimal generator is not integrable:

$$S_\varepsilon^< = S_\varepsilon^<(\hat{k}) := \{\hat{p} \in \mathbb{S}^2 : 1 - \hat{p} \cdot \hat{k} < \varepsilon\} \quad \varepsilon \in (0, 1). \quad (8)$$

We can now decompose the jump part of the generator \mathcal{L} into two components

$$\begin{aligned} \mathcal{L}f(x, \hat{k}) &= -\hat{k} \cdot \nabla_x f(x, \hat{k}) + \mathcal{L}_\varepsilon^< f(x, \hat{k}) + \mathcal{L}_\varepsilon^> f(x, \hat{k}) \\ &:= -\hat{k} \cdot \nabla_x f(x, \hat{k}) + \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \left(\int_{S_\varepsilon^<} + \int_{S_\varepsilon^>} \right) \rho(x, \hat{p} \cdot \hat{k}) (f(x, \hat{p}) - f(x, \hat{k})) \sigma(d\hat{p}), \end{aligned}$$

where $S_{>}^\varepsilon = (S_{<}^\varepsilon)^c$ is the complementary set of the critical region over the unit sphere. The part of the scattering operator involving $S_{>}^\varepsilon$ (with no singularity) is the infinitesimal generator of a standard jump Markov process. Regarding $S_{<}^\varepsilon$ (with the singularity), the following result justifies the approximation of this singular part by a Laplace-Beltrami operator $\Delta_{\mathbb{S}^2}$ over the unit sphere \mathbb{S}^2 . We will use the notation $r'_\varepsilon = \sqrt{1 - (1 - \varepsilon)^2}/(2 - \varepsilon)$ in what follows, and set in the rest of the paper $0 < \varepsilon \leq \varepsilon_0 < 1$ and $0 \leq \alpha_m \leq \alpha(x) \leq \alpha_M < 2$.

Proposition 2.1 *Let u be the solution to (1) and u_ε be the solution to*

$$\begin{cases} \partial_t u_\varepsilon + \hat{k} \cdot \nabla_x u_\varepsilon = \sigma_\varepsilon^2(x) \Delta_{\mathbb{S}^2} u_\varepsilon + \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{S_{>}^\varepsilon} \rho(x, \hat{p} \cdot \hat{k}) (u_\varepsilon(\hat{p}) - u_\varepsilon(\hat{k})) \sigma(d\hat{p}), \\ u_\varepsilon(0, x, \hat{k}) = u_0(x, \hat{k}), \end{cases} \quad (9)$$

for $(t, x, \hat{k}) \in (0, \infty) \times \mathbb{R}^3 \times \mathbb{S}^2$, where

$$\sigma_\varepsilon^2(x) := \frac{2^{1-\alpha(x)} a(0) \pi \lambda(x)}{2 - \alpha(x)} r'_\varepsilon{}^{2-\alpha(x)}. \quad (10)$$

Assuming $a'(0) = 0$, for any $T > 0$, we have

$$\sup_{t \in [0, T]} \|u(t, \cdot, \cdot) - u_\varepsilon(t, \cdot, \cdot)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \leq \varepsilon^{2-\alpha_M/2} \sqrt{2T} E(u) \quad (11)$$

where $E(u)$ is defined in (29).

The proof of Proposition 2.1 is postponed to Section 6.1. The term $E(u)$ is independent of ε and depends on derivatives of u w.r.t. \hat{k} up to order 4. Note that the error is of order $\varepsilon^{1-\alpha_M/2}/(2-\alpha_M)$ when $a'(0) \neq 0$ yielding a less accurate approximation than for $a'(0) = 0$. This later assumption holds throughout the remaining of the paper. Based on (11), we then devise a MC method for (9) instead of (1). The advantage in using (9) is the fact that the angular diffusion term $\sigma_\varepsilon^2(x) \Delta_{\mathbb{S}^2}$ is the generator of a Markov process that can be easily simulated. Indeed, for W a standard 3D Brownian motion on \mathbb{R}^3 and \times the cross product in \mathbb{R}^3 , it is shown in [4] that the process B solving the SDE

$$dB = B \times dW - B dt, \quad B(0) \in \mathbb{S}^2,$$

has generator $\frac{1}{2} \Delta_{\mathbb{S}^2}$. A simple adaptation then gives the desired diffusion coefficient. Since the error is of order $\varepsilon^{2-\alpha_M/2}$, it is always smaller than ε , and can be adjusted to obtain a desired accuracy. Note also that $\sigma_\varepsilon(x)$ increases as $\alpha(x)$ gets to 2, and diffusion on the sphere eventually becomes the dominant dynamics.

2.2 Representation for (9)

We interpret (9) as the forward Kolmogorov equation of an appropriate Markov process, and as a consequence focus on *forward* MC methods, see e.g. [27] for terminology. Backward equations are simulated in a similar manner, and can be combined with forward methods for variance reduction techniques [6, 30, 40].

The Markov process we consider for this approach is defined by

$$D_\varepsilon(t) := \sum_{n \geq 0} \mathbf{1}_{[T_n, T_{n+1})}(t) \psi_n^z(t - T_n) \quad t \geq 0, \quad (12)$$

where (in the remaining of the paper we extensively make use of the notation $z = (x, \hat{k})$):

1. The flow $\psi_n^z = (X_n^x, K_n^{\hat{k}})$ is the unique strong solution to the SDE

$$\begin{cases} dX_n^x(t) &= K_n^{\hat{k}}(t) dt \\ dK_n^{\hat{k}}(t) &= \sqrt{2} \sigma_\varepsilon(X_n^x(t)) K_n^{\hat{k}}(t) \times dW_n(t) - 2 \sigma_\varepsilon^2(X_n^x(t)) K_n^{\hat{k}}(t) dt, \end{cases} \quad (13)$$

where $\psi_n^z(0) = z$, \times is the cross product in \mathbb{R}^3 , $(W_n)_n$ is a sequence of independent standard Brownian motions on \mathbb{R}^3 , and σ_ε is defined by (10).

2. The jump times $(T_n)_n$ are distributed according to

$$\mathbb{P}(T_{n+1} - T_n > t \mid D_\varepsilon(T_n) = z, (\psi_n^z(s))_{s \in [0, t]}) = \exp\left(-\int_0^t \Lambda_\varepsilon(\psi_n^z(s)) ds\right), \quad \forall n \geq 0,$$

with $T_0 = 0$, and for ρ given by (4),

$$\Lambda_\varepsilon(z) := \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{S_{>}^\varepsilon} \rho(x, \hat{p} \cdot \hat{k}) \sigma(d\hat{p}). \quad (14)$$

3. The jumps $(Z_n)_n$ describe a Markov chain with transition probability

$$\mathbb{P}(Z_{n+1} \in dy \otimes \sigma(d\hat{p}) | Z_n, T_{n+1} - T_n) = \Pi_\varepsilon(z_{n+1}, dz), \quad (15)$$

where $z_{n+1} := \psi_n^{Z_n}(T_{n+1} - T_n)$, and

$$\Pi_\varepsilon(z, dz) := \pi_\varepsilon(z, \hat{p}) \sigma(d\hat{p}) \delta_x(dy), \quad (16)$$

with density

$$\pi_\varepsilon(z, \hat{p}) := \frac{\rho(x, \hat{p} \cdot \hat{k})}{\int_{S_\varepsilon^2} \rho(x, \hat{p}' \cdot \hat{k}) d\sigma(\hat{p}')} \mathbf{1}_{S_\varepsilon^2}(\hat{p}), \quad z = (x, \hat{k}), \quad (17)$$

which is supported over S_ε^2 . The above Dirac mass $\delta_x(dy) := \delta(x-y)dy$ translates the fact that the jumps only hold w.r.t. the \hat{k} variable.

Let us note that the above family of standard Brownian motions $(W_n(t))_{t \in [0, T_{n+1} - T_n]}$ can be defined as

$$W_n(t) = W(t + T_n) - W(T_n), \quad t \in [0, T_{n+1} - T_n],$$

for any n , where W is a single standard Brownian motion on \mathbb{R}^3 . We have then the following probabilistic representation for the solution to (9).

Proposition 2.2 *The Markov process D_ε defined in (12) has for infinitesimal generator*

$$\mathcal{A}_\varepsilon g(z) := \hat{k} \cdot \nabla_x g(z) + \sigma_\varepsilon^2(x) \Delta_{\mathbb{S}^2} g(z) + \Lambda_\varepsilon(z) \int_{\mathbb{S}^2} \pi_\varepsilon(z, \hat{p}) (g(x, \hat{p}) - g(x, \hat{k})) \sigma(d\hat{p}),$$

and we have

$$\mathbb{P}_{\mu_0}(D_\varepsilon(t) \in dx \otimes \sigma(d\hat{k})) = \frac{1}{\bar{u}_0} u_\varepsilon(t, x, \hat{k}) dx \sigma(d\hat{k}), \quad (18)$$

where

$$\mu_0(dx, d\hat{k}) := \mathbb{P}(D_\varepsilon(0) \in dx \otimes \sigma(d\hat{k})) = \frac{u_0(x, \hat{k})}{\bar{u}_0} dx \sigma(d\hat{k}) \quad \text{with} \quad \bar{u}_0 := \int_{\mathbb{R}^3 \times \mathbb{S}^2} u_0(x, \hat{k}) dx \sigma(d\hat{k}). \quad (19)$$

The terminology *forward* comes from the fact that the particles are emitted at random points at time $t = 0$ (through μ_0) and propagate towards the observation position $z = (x, \hat{k})$. The proof of Proposition 2.2 is provided in Section 6.2. Let us illustrate two aspects of the representation (18). In order to obtain an estimation of $u_\varepsilon(t, x, \hat{k})$ at the point $z = (x, \hat{k})$, we calculate the probability

$$\frac{\bar{u}_0}{|\mathbf{B}(z, h)|} \mathbb{P}_{\mu_0}(D_\varepsilon(t) \in \mathbf{B}(z, h)) \simeq u_\varepsilon(t, x, \hat{k}),$$

where $\mathbf{B}(z, h) \subset \mathbb{R}^3 \times \mathbb{S}^2$ stands for the open ball centered at $z = (x, \hat{k})$ with radius $h \ll 1$. If we are only interested in e.g. the energy density at point x , we estimate

$$\frac{\bar{u}_0}{|B(x, h)|} \mathbb{P}_{\mu_0}(D_\varepsilon(t) \in B(x, h) \times \mathbb{S}^2) = \frac{\bar{u}_0}{|B(x, h)|} \mathbb{P}_{\mu_0}(D_{1,\varepsilon}(t) \in B(x, h)) \simeq \int_{\mathbb{S}^2} u_\varepsilon(t, x, \hat{k}) \sigma(d\hat{k}),$$

where $B(x, h) \subset \mathbb{R}^3$ stands for the open ball centered at x with radius $h \ll 1$, and $D_{1,\varepsilon}$ is the x component of D_ε .

3 Monte Carlo Method

Based on the previous probabilistic representation of (9), solving (9) requires the generation of random paths of the stochastic process D_ε . For any measurable bounded functions f , the convergence of the estimator

$$\mu_N(t, f) := \frac{1}{N} \sum_{j=1}^N f(D_\varepsilon^j(t)) \xrightarrow{N \rightarrow \infty} \int f(x, \hat{k}) u_\varepsilon(t, x, \hat{k}) dx \sigma(d\hat{k}) \quad \mathbb{P}_{\mu_0} - \text{almost surely,}$$

is guaranteed by the strong law of large numbers. Above $(D_\varepsilon^j)_j$ is a sample of D_ε . We detail next how to treat efficiently the diffusion and jump components of the process D_ε .

3.1 The jump part

Since the process D_ε is inhomogeneous, i.e. Λ_ε and Π_ε both depend on $z = (x, \hat{k})$, we use the so-called *thinning* method, also referred to as the *fictitious shocks* method [27]. It is based on an acceptance/rejection step and consists in simulating at first more jumps (or shocks) than necessary. In a second step, some of the jumps are rejected according to an appropriate probability distribution in order to recover the original dynamics. Assume $0 < \alpha_m \leq \alpha(x) \leq \alpha_M < 2$. A direct calculation shows that

$$\Lambda_\varepsilon(z) \leq \frac{2\pi \sup \lambda \sup a}{2^{1+\alpha(x)/2}} \int_{-1}^{1-\varepsilon} \frac{dt}{(1-t)^{1+\alpha(x)/2}} = \frac{4\pi \sup \lambda \sup a}{\alpha(x) 2^{1+\alpha(x)/2} \varepsilon^{\alpha(x)/2}} \leq \frac{2\pi \sup \lambda \sup a}{\alpha_m \varepsilon^{\alpha_M/2}} =: \bar{\Lambda}_\varepsilon. \quad (20)$$

The fictitious jump times are then drawn as

$$\bar{T}_n := \sum_{j=1}^n \xi_j, \quad \text{and} \quad \bar{T}_0 = 0,$$

where the $(\xi_j)_j$ are i.i.d. exponentially distributed random variables with parameter $\bar{\Lambda}_\varepsilon$.

The thinning method consists in the following acceptance/rejection step. At a jump time \bar{T}_n and current position z_n , we draw a jump z according to the probability distribution Π_ε given by (15). This jump is accepted with probability $p(z_n) = \Lambda_\varepsilon(z_n)/\bar{\Lambda}_\varepsilon$. Otherwise, the process D_ε continues to diffuse starting from z_n , and \bar{T}_n is not considered as a *true* jump time for D_ε . Practically, we can define the state as

$$\bar{Z}_n := z \mathbf{1}_{(U_n \leq p(z_n))} + z_n \mathbf{1}_{(U_n > p(z_n))}$$

at each fictitious jump times \bar{T}_n . Above, U_n is a random variable uniformly distributed over $[0, 1]$ and all the U_n 's are independent.

3.2 The diffusion part

The diffusion part between two jumps satisfies the linear SDE (13), and is simulated using the following Euler-Maruyama type scheme

$$(S_{n,m}) : \begin{cases} X_{n,m+1} &= X_{n,m} + h_{n,m} \hat{K}_{n,m} \\ K_{n,m+1} &= \hat{K}_{n,m} - 2h_{n,m} \sigma_\varepsilon^2(X_{n,m}) \hat{K}_{n,m} + \sqrt{2h_{n,m}} \sigma_\varepsilon(X_{n,m}) \hat{K}_{n,m} \times W_{n,m} \\ \hat{K}_{n,m+1} &= \frac{K_{n,m+1}}{|K_{n,m+1}|}, \end{cases} \quad (21)$$

where the $(W_{n,m})_{m,n}$ are i.i.d. mean-zero Gaussian random vectors with identity covariance matrix. Note that the above scheme does not conserve the Euclidean norm with respect to the angular variable, and as a consequence the evolution of $(K_{n,m})_{n,m}$ does not remain on the unit sphere over the iterations. This motivates the definition of $\hat{K}_{n,m}$. We have nevertheless $\mathbb{E}[|K_{n,m}^k|^2] = 1$, for all n and m , and Theorem 3.1 below guarantees that the distribution of $\hat{K}_{n,m}$ provides a converging approximation of the true statistics. The stepsizes $h_{n,m}$ are determined from a fixed stepsize h as follows. Since our convergence theorem further is stated at a fixed time T , we include $t = T$ in the discretization grid for simplicity. Let then N_T such that $\bar{T}_{N_T} \leq T < \bar{T}_{N_T+1}$. When $n \neq N_T$, let $dt_n = \bar{T}_{n+1} - \bar{T}_n$. With $m_n = [dt_n/h]$ ($[\cdot]$ the integer part), we set, for $n \neq N_T$,

$$h_{n,m} = \begin{cases} h & \text{for } m = 0, \dots, m_n - 1 \\ dt_n - m_n h & \text{for } m = m_n, \end{cases}$$

and if $m_n = 0$, we set $h_{n,0} = dt_n$. In the rest of the paper, the grid is denoted by $(T_{n,m})$, where $T_{n,m+1} = T_{n,m} + h_{n,m}$ for $T_{n,0} = \bar{T}_n$, $n \geq 0$ and $m = 0, \dots, m_n - 1$. When $n = N_T$, we divide the interval $[\bar{T}_{N_T}, \bar{T}_{N_T+1}]$ similarly into subintervals of length $h_{N_T,m}$ at most h (we suppose there are m_{N_T} of those) and such that $T_{N_T,m} = T$ for one m in $0, \dots, m_{N_T}$.

3.3 The overall discretized process and convergence

For any $t \geq 0$, the approximate version of the process D_ε , denoted $D_{h,\varepsilon}$, is defined by:

$$D_{h,\varepsilon}(t) = (X_{h,\varepsilon}(t), K_{h,\varepsilon}(t)) := \sum_{n=0}^{\infty} \sum_{m=0}^{m_n} \mathbf{1}_{[T_{n,m}, T_{n,m+h_{n,m}})}(t) Z_{n,m},$$

where

1. For any $m \in \{0, \dots, m_n\}$,

$$Z_{n,m+1} = S_{n,m}(Z_{n,m}),$$

where $Z_{n,0} = \bar{Z}_n$ for the $(\bar{Z}_n)_n$ defined below, and where $S_{n,m}(Z_{n,m}) = (X_{n,m+1}, \hat{K}_{n,m+1})$ is given by the scheme (21) with initial condition $Z_{n,m} = (X_{n,m}, \hat{K}_{n,m})$.

2. The sequence $(\bar{Z}_n)_n$, is defined by

$$\bar{Z}_{n+1} = z \mathbf{1}_{(U_n \leq p(Z_{n,m_n+1}))} + Z_{n,m_n+1} \mathbf{1}_{(U_n > p(Z_{n,m_n+1}))} \quad n \geq 0,$$

where z is drawn according to the probability measure $\Pi_\varepsilon(Z_{n,m_n+1}, dz)$ defined by (16).

Below, b is the backward solution to (1) with terminal condition $b(T, x, \hat{k}) = f(x, \hat{k})$, see (36). Our convergence result is then the following (we set h such that $4h \sup_x \sigma_\varepsilon^2(x) \leq 1$ to simplify some expressions):

Theorem 3.1 *Consider*

$$\mu_{N,h,\varepsilon}(t, f) = \frac{1}{N} \sum_{j=1}^N f(D_{h,\varepsilon}^j(t)), \quad \mu(t, f) = \int f(x, \hat{k}) u(t, x, \hat{k}) dx \sigma(d\hat{k}),$$

where $(D_{h,\varepsilon}^j)_j$ is a sample of $D_{h,\varepsilon}$. For any $T > 0$, $\eta > 0$ and any smooth bounded function f on $\mathbb{R}^3 \times \mathbb{S}^2$, we have

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(|\mu_{N,h,\varepsilon}(T, f) - \mu(T, f)| > \frac{\eta \Sigma_{h,\varepsilon}}{\sqrt{N}} + \varepsilon^{2-\alpha_M/2} F_0(u, b, f) + h F_1(b) \right) \leq \operatorname{erfc}(\eta/\sqrt{2}), \quad (22)$$

where

$$\Sigma_{h,\varepsilon} = \sqrt{\operatorname{Var}(f(D_{h,\varepsilon}(T)))} \leq \sup |f|.$$

The functions F_0 and F_1 are explicit and independent of ε and h , and are defined in the proof of the theorem in Section 6.3.

In (22), there are three terms that quantify the approximation error of our estimator $\mu_{N,h,\varepsilon}(t, f)$: one of order $\varepsilon^{2-\alpha_M/2}$ due to the approximation of u by u_ε (the smaller the α_M , i.e. the less singular the kernel is, the smaller the error), one of order h due to the numerical approximation of the diffusion over the unit sphere, and one due to the MC approximation with the standard $1/\sqrt{N}$ convergence rate. Note that the discretization error of the diffusion process is only of order h and not of order the standard \sqrt{h} . The reason is that we are only interested in the convergence of Monte Carlo estimators, allowing us to consider this discretization error in the weak sense [42]. The MC error is controlled by the standard deviation $\Sigma_{h,\varepsilon}$, and variance reduction techniques can be designed to reduce this term. The parameter h will be chosen proportionally to the shortest mean free time Λ_ε^{-1} . Theorem 3.1 is proved in Section 6.3.

3.4 Algorithms

We discuss in this section practical aspects of the method. Before stating the algorithm itself, let us emphasize that a key point is to sample efficiently the jumps from Π_ε given by (16).

Let us fix the current state of the process $D_{h,\varepsilon}$ at a point $z = (x, \hat{k})$. In spherical coordinates, π_ε defined in (17) is equivalent to a probability density function drawing a polar angle θ and an azimuthal angle φ . Here, the north pole of the spherical system is the current direction \hat{k} , and it is direct to see that the azimuthal angle φ is uniformly distributed over $(0, 2\pi)$. We denote this by $\varphi \sim \mathcal{U}(0, 2\pi)$. For the polar angle, a change of variables leads to $\theta = \arccos(1 - \chi)$, where χ has probability density function

$$f_\chi(\chi|x) := \frac{a(\sqrt{2\chi})}{C_\chi \chi^{1+\alpha(x)/2}} \mathbf{1}_{(\varepsilon, 2)}(\chi),$$

and C_χ is a normalizing constant. Therefore, to draw a jump according to (17) starting from \hat{k} , we compute

$$\hat{p} = R(\theta, \varphi, \hat{k}) := \cos(\theta) \hat{k} + \sin(\theta) \left(I_3 + \sin(\varphi) Q(\hat{k}) + (1 - \cos^2(\varphi)) Q^2(\hat{k}) \right) \hat{k}^\perp, \quad (23)$$

where \hat{k}^\perp is an orthonormal vector to \hat{k} , I_3 is the 3×3 identity matrix, and

$$Q(k) = \begin{pmatrix} 0 & -k_3 & k_2 \\ k_3 & 0 & -k_1 \\ -k_2 & k_1 & 0 \end{pmatrix}, \quad \text{where } k = (k_1, k_2, k_3). \quad (24)$$

The transformation R corresponds to a rotation from \hat{k} to \hat{p} with polar angle θ with respect to \hat{k} and azimuthal angle φ with respect to \hat{k}^\perp . Note that the choice of \hat{k}^\perp is not important since φ is uniformly distributed over $(0, 2\pi)$.

We notice that in the case of a constant function $a \equiv a_0$, one obtains a truncated Pareto distribution for χ . The corresponding cumulative distribution function can be exactly inverted giving then a direct simulation method. In this case, the cumulative distribution function is given by, for $\chi \in (\varepsilon, 2)$,

$$F_\chi(\chi|x) = \frac{a_0}{C_\chi} \int_\varepsilon^\chi \frac{dv}{v^{1+\alpha(x)/2}} = \frac{1 - (\varepsilon/\chi)^{\alpha(x)/2}}{1 - (\varepsilon/2)^{\alpha(x)/2}}.$$

The random variable χ can then be generated by

$$\chi = F_\chi^{-1}(U|x) = \varepsilon(1 - (1 - (\varepsilon/2)^{\alpha(x)/2})U)^{-2/\alpha(x)},$$

where U is a random variable uniformly distributed over $(0, 1)$ ($U \sim \mathcal{U}(0, 1)$). In the case of a non constant function a , the main features of the density $f_\chi(\cdot|x)$ are similar to those of the truncated Pareto distribution, and a stochastic collocation method can be considered to simulate $f_\chi(\cdot|x)$. This method is described in Appendix A in our context. It is based on the simulation of the above truncated Pareto distribution and proves to be very effective.

The algorithm used to simulate a trajectory of $D_{h,\varepsilon}$ can be summarized in the following two procedures. The first one corresponds to the simulation of the diffusion process between two (fictitious) jumps, and we use the notation

$$S(Z, W, h) = \begin{cases} X + hK \\ K - 2h\sigma_\varepsilon^2(X)K + \sqrt{2h}\sigma_\varepsilon(X)W \times K, \end{cases}$$

with $Z = (X, K)$. Below, $N(0, I_3)$ stands for the three dimensional multivariate Normal distribution with identity covariance matrix.

Algorithm 1: Diffusion

input : current state of the particle $z = (x, \hat{k})$, duration of the diffusion δt
output : state of the particle after the diffusion process
initialization: $n \leftarrow \lceil \delta t/h \rceil$ // number of iterations
 $Z \leftarrow z$ // initialization of the diffusion state
// Main loop of the diffusion
for $j \leftarrow 1$ **to** n **do**
 $W \sim N(0, I_3)$
 $Z \leftarrow S(Z, W, h)$
 $K \leftarrow K/|K|$
// Add a diffusion step with stepsize $h' \leq h$ to match the duration δt
 $h' \leftarrow \delta t - nh$
 $W \sim N(0, I_3)$
 $Z \leftarrow S(Z, W, h')$
 $K \leftarrow K/|K|$
return Z

The second procedure combines the diffusion step with the jump process. Below, we denote by $\mathcal{E}(\bar{\Lambda}_\varepsilon)$ the exponential distribution with parameter $\bar{\Lambda}_\varepsilon$ defined by (20).

Algorithm 2: TrajectorySimulation

```
input      : Duration  $T$  of the particle evolution
output    : state of the particle at time  $T$ 
initialization:  $Z \leftarrow (x, \hat{k}) \sim \mu_0$  // initialization of the particle state at random
 $t \leftarrow 0$  // temporary time variable
 $\delta t \sim \mathcal{E}(\bar{\Lambda}_\varepsilon)$  // first jump time
// main loop for the path evolution
while  $t + \delta t < T$  do
   $Z \leftarrow \text{Diffusion}(Z, \delta t)$ 
   $U \sim \mathcal{U}(0, 1)$ 
  if  $U \leq p(Z)$  then
    // the jump is accepted,  $Z$  is transformed
     $\chi \sim f_\chi(\cdot | x)$ 
     $\theta \leftarrow \arccos(1 - \chi)$ 
     $\varphi \sim \mathcal{U}(0, 2\pi)$ 
     $\hat{p} \leftarrow R(\theta, \varphi, \hat{k})$ 
     $Z \leftarrow (x, \hat{p})$ 
   $t \leftarrow t + \delta t$ 
   $\delta t \sim \mathcal{E}(\bar{\Lambda}_\varepsilon)$ 
// remaining diffusion step of duration  $T - t$ 
 $Z \leftarrow \text{Diffusion}(Z, T - t)$ 
return  $Z$ 
```

The rest of the paper is dedicated to numerical simulations and the proofs of our main results.

4 Validation

In this section, we first derive a semi-analytical solution to validate our method in the simplest situation where α, a, λ are constant functions. We then highlight the crucial role of the small jumps correction for computational efficiency.

4.1 Semi-analytical solution

We set $\lambda \equiv 1$ and the RTE (1) reads

$$\partial_t u + \hat{k} \cdot \nabla_x u = \mathcal{Q}u \quad (25)$$

with scattering kernel

$$\mathcal{Q}f(\hat{k}) = a \int_{\mathbb{S}^2} \frac{\sigma(d\hat{p})}{|\hat{p} - \hat{k}|^{2+\alpha}} (f(\hat{p}) - f(\hat{k})), \quad \hat{k} \in \mathbb{S}^2.$$

Using the the Funk-Hecke formula [39], this operator can be diagonalized in $L^2(\mathbb{S}^2)$ equipped with the inner product

$$\langle f, g \rangle_{L^2(\mathbb{S}^2)} = \int_{\mathbb{S}^2} f(\hat{p}) \overline{g(\hat{p})} \sigma(d\hat{p}) = \int_0^\pi \int_0^{2\pi} f(\theta, \varphi) \overline{g(\theta, \varphi)} \sin(\theta) d\theta d\varphi.$$

The eigenvalues are given by

$$\lambda_l = \frac{a\pi\Gamma(-\alpha/2)}{2^\alpha\Gamma(1+\alpha/2)} \left(\frac{\Gamma(l+1+\alpha/2)}{\Gamma(l+1-\alpha/2)} - \frac{\Gamma(1+\alpha/2)}{\Gamma(1-\alpha/2)} \right) \quad l \in \mathbb{N},$$

and the eigenvectors are the spherical harmonics

$$Y_{l,m}(\hat{k}) = Y_{l,m}(\theta, \varphi) := \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos(\theta)) e^{im\varphi}, \quad (l, m) \in \mathbb{N} \times \{-l, \dots, l\},$$

where the P_l^m are the associated Legendre polynomials. In order to derive a semi-analytical solution, we Fourier transform (25) w.r.t. x , and introduce

$$\hat{u}(t, q, k) = \int_{\mathbb{R}^3} u(t, x, k) e^{-iq \cdot x} dx.$$

Above, $q = \tilde{q} := (0, 0, \xi)$ so that $\tilde{u}(t, \xi, \hat{k}) = \hat{u}(t, \tilde{q}, \hat{k})$ solves

$$\partial_t \tilde{u} + i \hat{k} \cdot \tilde{q} \tilde{u} = \mathcal{Q} \tilde{u}. \quad (26)$$

Writing \hat{k} in spherical coordinates with $(0, 0, 1)$ as north-pole, this latter equation reads,

$$\partial_t \tilde{u}(t, \xi, \theta, \varphi) = (\mathcal{Q} - i\xi \cos(\theta)) \tilde{u}(t, \xi, \theta, \varphi), \quad (t, \xi, \theta, \varphi) \in (0, \infty) \times \mathbb{R} \times (0, \pi) \times (0, 2\pi).$$

We now decompose \tilde{u} on the basis of spherical harmonics

$$\tilde{u}(t, \xi, \theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \hat{u}_{l,m}(t, \xi) Y_{l,m}(\theta, \varphi),$$

resulting in

$$\frac{d}{dt} \hat{u}_{l,m} = \lambda_l \hat{u}_{l,m} - i\xi (d_{l,m}^+ \hat{u}_{l+1,m} + d_{l,m}^- \hat{u}_{l-1,m}) \quad \text{for } l \geq 1, \quad \frac{d}{dt} \hat{u}_{0,0} = -i\xi d_{0,0}^+ \hat{u}_{1,0} \quad \text{for } l = 0. \quad (27)$$

Above, we have used the fact that

$$\langle Y_{m',l'}, \cos(\theta) Y_{m,l} \rangle_{L^2(\mathbb{S}^2)} = \begin{cases} d_{l,m}^+ := \sqrt{\frac{(l+m+1)(l-m+1)}{(2l+1)(2l+3)}} & \text{if } m = m' \text{ and } l' - l = 1, \\ d_{l,m}^- := \sqrt{\frac{(l+m)(l-m)}{(2l-1)(2l+1)}} & \text{if } m = m' \text{ and } l' - l = -1, \\ 0 & \text{otherwise.} \end{cases}$$

For computational purposes, we introduce a cutoff in the variable l ($l \in \{0, \dots, L\}$), and consider a truncated version of (27) as the vector differential equation

$$\frac{d}{dt} \hat{u}^L(t, \xi) = (D^L - i\xi A^L) \hat{u}^L(t, \xi), \quad \hat{u}^L(t, \xi) = (\hat{u}_{l^2+j}^L(t, \xi))_{l \in \{0, \dots, L\}, j \in \{0, \dots, 2l\}} \in \mathbb{C}^{(L+1)^2}, \quad (28)$$

where D^L and A^L are two $(L+1)^2 \times (L+1)^2$ matrices defined by

$$\begin{cases} D_{l^2+j+1, l^2+j+1}^L := \lambda_l & \text{for } l \in \{0, \dots, L\}, j \in \{0, \dots, 2l\}, \\ A_{l^2+j+1, (l+1)^2+j+2}^L := d_{l,j-l}^+ & \text{for } l \in \{0, \dots, L-1\}, j \in \{0, \dots, 2l\} \\ A_{l^2+j+2, (l-1)^2+j+1}^L := d_{l,j-l+1}^- & \text{for } l \in \{1, \dots, L\}, j \in \{0, \dots, 2(l-1)\}. \end{cases}$$

All other coefficients in both D^L and A^L are set to 0. Note that the indexing of the matrices starts at 0 for simplicity. The solution to (28) reads $\hat{u}^L(t, \xi) = e^{(D^L - i\xi A^L)t} \hat{u}^L(0, \xi)$, where the matrix exponential is computed numerically. For our test case, we consider the following initial condition

$$u(t=0, x, \hat{k}) = \frac{1}{\sqrt{2\pi}} e^{-|x|^2/2} \cdot 2 \cos^2(\theta/2) = \frac{1}{\sqrt{2\pi}} e^{-|x|^2/2} (2\sqrt{\pi} Y_{0,0}(\theta, \varphi) + 2\sqrt{\pi/3} Y_{0,1}(\theta, \varphi)),$$

so that

$$\hat{u}_{l^2+j}^L(t=0, \xi) = \begin{cases} 2\sqrt{\pi} e^{-\xi^2/2} & \text{for } l = j = 0, \\ 2\sqrt{\pi/3} e^{-\xi^2/2} & \text{for } l = j = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, an approximation of \tilde{u} , solution to (26), is given by

$$\tilde{u}^L(t=0, \xi, \theta, \varphi) = \sum_{l=0}^L \sum_{j=0}^{2l} \left[e^{(D^L - i\xi A^L)t} \hat{u}^L(0, \xi) \right]_{l^2+j} Y_{l,j-l}(\theta, \varphi).$$

For numerical comparisons with our MC method, we introduce a discretization of the unit sphere \mathbb{S}^2 via the polar and azimuthal angles $(\theta_m)_m$ and $(\varphi_m)_m$, with respective stepsize $\Delta\theta$ and $\Delta\varphi$. We then compare

$$\tilde{u}^L(t, \xi, \theta_m, \varphi'_m) \simeq \frac{1}{\Delta\theta \Delta\varphi} \int_{\theta_m}^{\theta_{m+1}} \int_{\varphi'_m}^{\varphi'_{m+1}} \tilde{u}^L(t, \xi, \theta, \varphi) \sin(\theta) d\theta d\varphi$$

with its MC approximation

$$\tilde{u}_N^L(t, \xi, m, m') = \frac{1}{\Delta\theta\Delta\varphi N} \sum_{n=1}^N e^{-i\xi X_{3,h,\varepsilon}^n(t)} \mathbf{1}_{(\theta_{h,\varepsilon}^n(t) \in (\theta_m, \theta_{m+1}), \varphi_{h,\varepsilon}^n(t) \in (\varphi_{m'}, \varphi_{m'+1}))}$$

where $\theta_{h,\varepsilon}^n$ and $\varphi_{h,\varepsilon}^n$ are respectively the polar and azimuthal angles for $\hat{K}_{h,\varepsilon}^n$, and where $(D_{h,\varepsilon}^n)_n = (X_{h,\varepsilon}^n, \hat{K}_{h,\varepsilon}^n)_n$ is a sample of $D_{h,\varepsilon}$ introduced in Section 3.

In the following numerical illustrations we consider $a = 0.002$ in the RTE, and set $\Delta\theta = \Delta\varphi = 0.05$, $\varepsilon = 0.1$ and $h = 0.5/\bar{\Lambda}_\varepsilon \simeq 12.6$ for the approximation parameters. Note that these choices for ε and h are providing us with a good accuracy at a very low computational cost as we will see. Such values may have to be decreased in other setups and when considering different observables. For instance, in Section 5.2 where α is varying, smaller values of ε and h are needed to capture correctly the solution.

Also, in the context of singular scattering kernels, the classical notion of scattering mean free time is not informative since it is equal to 0 (see (3)). Instead, we define a characteristic time using the inverse of the second eigenvalue of \mathcal{Q} , i.e. the first non zero eigenvalue, and set $t_c = -\frac{1}{\lambda_1}$. We refer to Figure 2 for the evolution of t_c w.r.t. α .

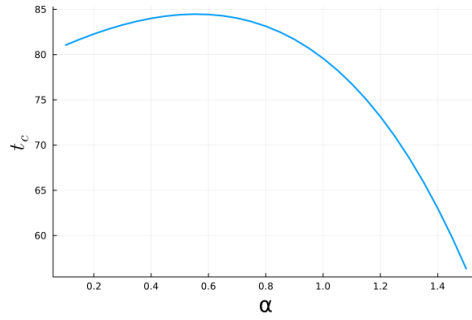


Figure 2: Illustration of the evolution of t_c w.r.t. α .

In our setting, $t_c \simeq 79.6$ (for $\alpha = 1$), which is about six times the stepsize h needed to capture the diffusive correction. Also, since ε is not too small, this correction plays a significant role in obtaining the correct dynamics.

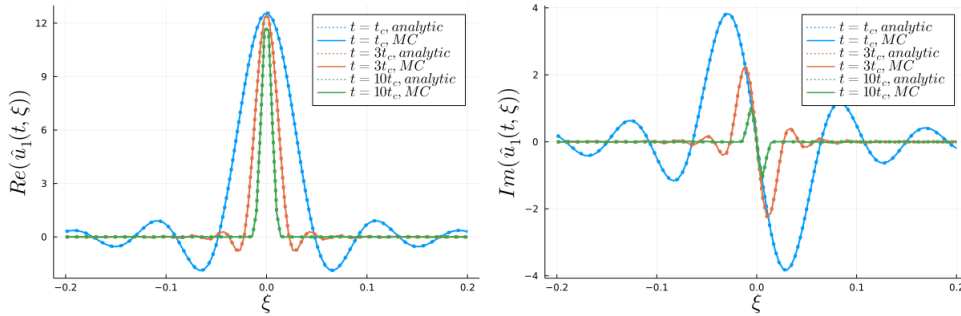


Figure 3: Comparisons of the real and imaginary parts of $\hat{u}_1(t, \xi)$ for three observation times and for $\alpha = 1$. The grid in ξ range from -0.2 to 0.2 with 100 discretization points and we run $N = 2.4 \times 10^6$ particles.

In Figure 3, we compare, for $\alpha = 1$, the real and imaginary parts of the observable

$$\hat{u}_1(t, \xi) := \int_0^\pi \int_0^{2\pi} \tilde{u}^L(t, \xi, \theta, \varphi) \sin(\theta) d\theta d\varphi \quad \text{with} \quad 4\pi \Delta\theta\Delta\varphi \sum_{m,m'} \tilde{u}_N^L(t, \xi, m, m'),$$

for three values of t . In Figure 4, we compare the real and imaginary parts of

$$\hat{u}_2(t, \xi, \theta) := \int_0^{2\pi} \tilde{u}^L(t, \xi, \theta, \varphi) d\varphi \sin(\theta) \quad \text{with} \quad 4\pi \Delta\varphi \sum_{m'} \tilde{u}_N^L(t, \xi, m, m'),$$

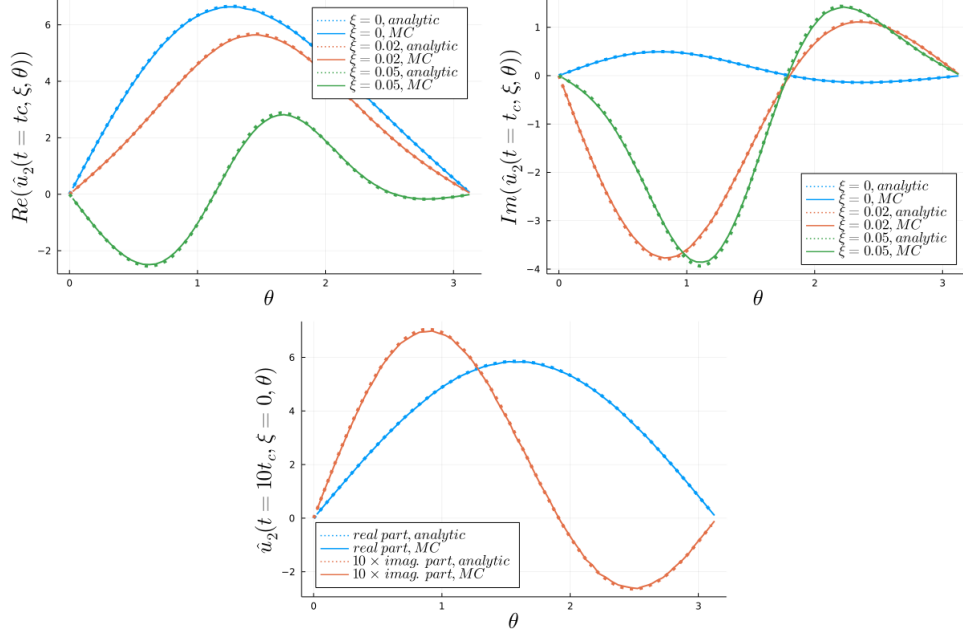


Figure 4: Comparisons of the real and imaginary parts of $\hat{u}_2(t, \xi, \theta)$ for $\alpha = 1$, for three values of ξ if $t = t_c$, and for $\xi = 0$ if $t = 10t_c$. We run $N = 2.4 \times 10^6$ particles for the top two pictures and $N = 24 \times 10^6$ for the third one.

for three values of ξ .

In Figure 5, we compare

$$\hat{u}_3(t = 2t_c, \xi = 0.02, \theta) := \int_0^{2\pi} \tilde{u}^L(t = 2t_c, \xi = 0.02, \theta, \varphi) d\varphi \sin(\theta) \quad \text{with} \quad 4\pi \Delta\varphi \sum_{m'} \tilde{u}_N^L(t = 2t_c, \xi = 0.02, m'),$$

for three values of α .

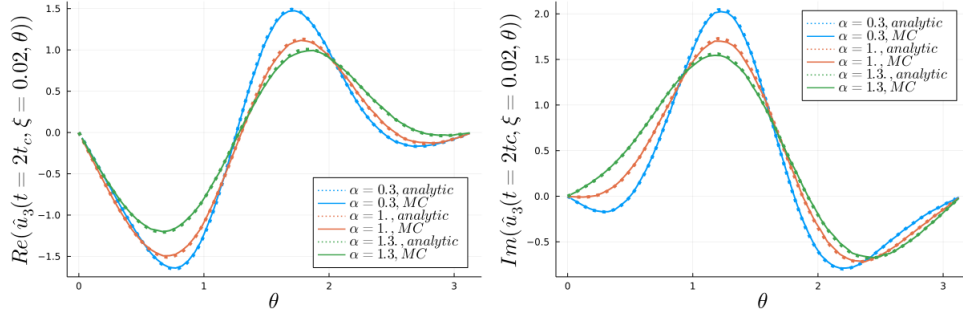


Figure 5: Comparisons of the real and imaginary parts of the observable $\hat{u}_3(t = 2t_c, \xi = 0.02, \theta)$ for three values of α . We run $N = 24 \times 10^6$ particles.

In all these illustrations, and despite somewhat fairly large values for ε and h , we observe a very good agreement between the Monte Carlo results and the semi-analytic calculations.

4.2 Role of the correction

In this section, we highlight the role of the correction provided by the diffusion over the unit sphere w.r.t. the \hat{k} -variable. To this end, we compare the following observables obtained from the semi-analytic solution

$$u_4(x_3) = \int_0^{3t_c} \tilde{u}^L(t, x_\perp, x_3, \theta, \varphi) \sin(\theta) dt dx_\perp d\theta d\varphi, \quad x := (x_\perp, x_3) \in \mathbb{R}^2 \times \mathbb{R},$$

with the ones obtained with our MC method, with and without this diffusive correction, and for various values of α , ε and h . The grid in z range from -300 to 300 with size 2^8 and we run $N = 300 \times 10^6$ particles. In Figure

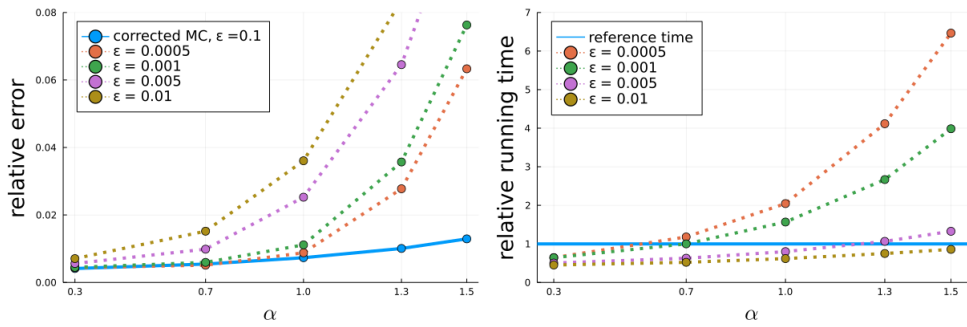


Figure 6: Illustration of the relative error $Err_{\varepsilon,\alpha}$ and running time of the MC method with and without a diffusive correction. The reference time in the right picture is the one of corrected method with $\varepsilon = 0.1$.

6, we represent the relative error

$$Err_{\varepsilon,\alpha} := \max_z \frac{|u_4(z) - u_{4,MC}(z)|}{\max_z u_4(z)},$$

for various sizes of the cutoff ε , and where $u_{4,MC}(z)$ is the MC approximation to $u_4(z)$. The left picture illustrates the evolution of the relative error for various ε . The blue curve corresponds to the corrected MC with $\varepsilon = 0.1$ (with still a fairly large stepsize $h = 0.5/\bar{\Lambda}_\varepsilon$) providing at most a relative error slightly larger than 1%. The other curves correspond to the noncorrected MC method for several values of ε . The corrected MC consistently yields a better accuracy than the noncorrected version, and even in weakly singular cases where α is less than one, a very small value of ε (red and green curves) is necessary to match the accuracy of the corrected method. The right picture illustrates the evolution of the relative running time of the noncorrected method w.r.t. the corrected one. For values of α less than 0.7 (weakly singular kernels), corrected and noncorrected methods have similar computational times for comparable accuracy, while in the case of singular kernels with $\alpha \geq 1$, the noncorrected methods yield a much larger cost and a much lower accuracy.

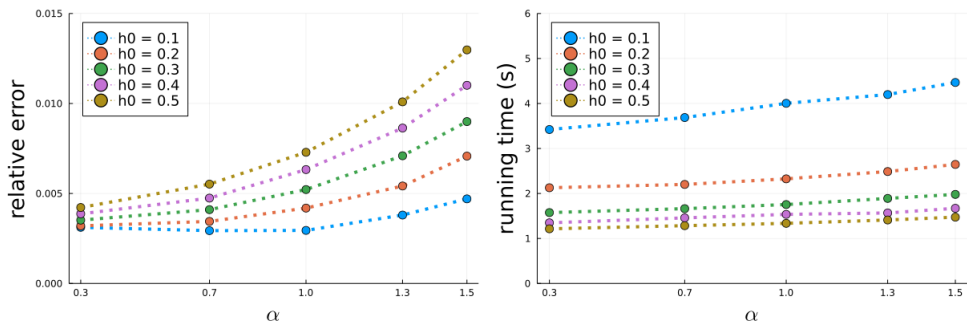


Figure 7: Illustration of the relative error $Err_{\varepsilon,\alpha}$ and running time of the (corrected) Monte Carlo method.

In Figure 7, we illustrate the precision and running time sensitivity of the (corrected) MC method w.r.t. the stepsize $h = h_0/\bar{\Lambda}_\varepsilon$. As expected, we obtain a better precision for smaller stepsizes but at the price of a longer running time. These effects are amplified as α increases due to the increasing strength of the diffusion correction. In what follows, we select $h_0 = 0.3$ since this yields a relative error less than 1% for a wide range of α 's while not changing significantly the running time.

In Figure 8, we depict the precision and running time sensitivity w.r.t. the cutoff parameter ε , and observe the same phenomena as in the case of the stepsize h . The parameter ε defines not only the accuracy of the diffusion correction, but also the average number of jumps, and as a consequence the running time increases as ε decreases as in the case of the noncorrected Monte Carlo method.

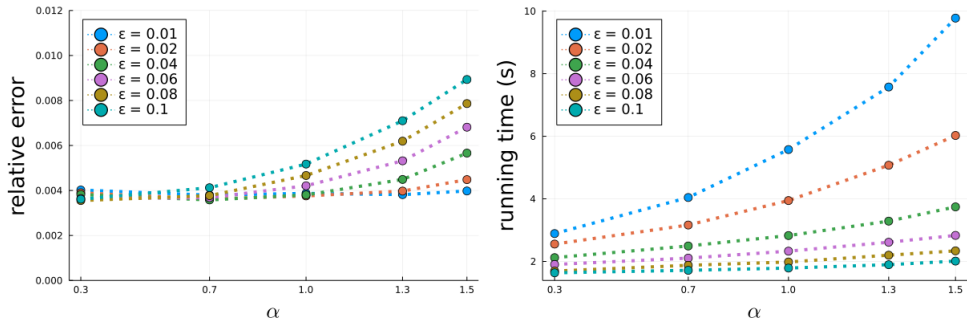


Figure 8: Illustration of the relative error $Err_{\varepsilon, \alpha}$ and running time of the (corrected) MC method.

5 Numerical illustrations

5.1 The role of α

In this section, we highlight the effects of the kernel singularity on the energy density. We consider a constant α , with $a = 0.002$ in this section. Our setting is depicted in Figure 9. The spatial variable x is decomposed into a main propagation axis x_3 and a transverse plane x_\perp , i.e. $x = (x_\perp, x_3) \in \mathbb{R}^2 \times \mathbb{R}$. The same notation holds for the direction variable $\hat{k} = (\hat{k}_\perp, \hat{k}_3) \in \mathbb{S}^2$. We choose an initial condition for (1) of the form

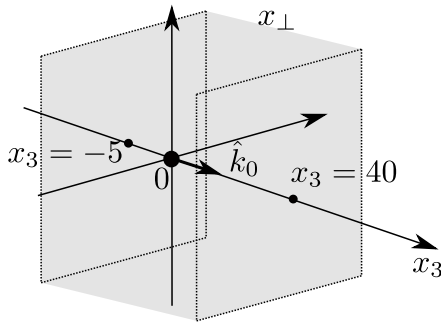


Figure 9: Illustration of the numerical setting.

$$u_0(x, \hat{k}) = \delta(x) \delta(\hat{k} - \hat{k}_0), \quad \hat{k}_0 = (0, 0, 1),$$

modeling a source located at $x = 0$ and embedded in the random medium, and emitting in the forward x_3 -direction. We set a function λ of the form $\lambda(x) = \mathbf{1}_{(-5, 40)}(x_3)$, that defines a scattering layer between $x_3 = -5$ and $x_3 = 40$. In such a configuration, both transmitted and reflected quantities at $x_3 = 40$ and $x_3 = -5$ are of interest. With our particular choice for \hat{k}_0 , what is obtained at $x_3 = -5$ is purely due to backscattering.

In the following two subsections, the MC estimations are obtained using $N = 1 \times 10^9$ particles and a diffusion stepsize $h = 0.3/\Lambda_\varepsilon$. We set $\varepsilon = 0.01$ for the calculation of transmitted quantities, and $\varepsilon = 0.1$ for the reflected ones. For any value of α , the observation time we consider is $T = 4t_c$, for t_c the critical time computed for $\alpha = 1$.

In the transmission case and when ε is too large, the mean free time is large as well and it is possible that particles escape the slab without undergoing any jumps, leading to inaccurate results. Hence the choice $\varepsilon = 0.01$. A larger value of ε is acceptable in the calculation of the reflected quantities since the particles exiting early would not have traveled to the plane located at $x_3 = -5$, and the error is reduced compared to the transmission case.

The running times for the time-integrated transmitted ($\varepsilon = 0.01$) and reflected ($\varepsilon = 0.1$) signals for different values of α are the following:

running time (s)	$\alpha = 0.3$	$\alpha = 0.7$	$\alpha = 1.0$	$\alpha = 1.3$	$\alpha = 1.5$
$\varepsilon = 0.01$	9.38	15.21	24.54	42.74	65.39
$\varepsilon = 0.1$	3.61	3.72	4.12	4.92	6.01

All these running time measurements account also for the transfer of the resulting arrays from the device to the host. We clearly observe a significantly larger running time for smaller values of ε and large values of α . This is due to the increase in scattering events as the mean free time decreases.

5.1.1 Energy at the boundaries of the transverse plane

In what follows, the (time-integrated) transverse reflected and transmitted energy are defined by

$$F_{tr}^T(x_\perp) := \int_0^T dt \int_{\mathbb{S}^2} \sigma(d\hat{k}) u(t, x_\perp, x_3 = 40, \hat{k}) \quad \text{and} \quad F_{ref}^T(x_\perp) := \int_0^T dt \int_{\mathbb{S}^2} \sigma(d\hat{k}) u(t, x_\perp, x_3 = -5, \hat{k}).$$

The MC estimators for these quantities are given respectively by

$$\hat{F}_{tr}^T(m, n) := \frac{1}{\Delta x_\perp N} \sum_{j=1}^N \mathbf{1}_{(X_{h,\varepsilon}^{j,\perp}(\tau^j) \in \square_{mn}, X_{3,h,\varepsilon}^j(\tau^j) > 40)}, \quad \hat{F}_{ref}^T(m, n) := \frac{1}{\Delta x_\perp N} \sum_{j=1}^N \mathbf{1}_{(X_{h,\varepsilon}^{j,\perp}(\tau^j) \in \square_{mn}, X_{3,h,\varepsilon}^j(\tau^j) < -5)}$$

where

$$\tau^j := \inf(t \in [0, T] : X_{3,h,\varepsilon}^j(t) > 40 \quad \text{or} \quad X_{3,h,\varepsilon}^j(t) < -5),$$

is the first time the j -th particle exits the slab. Note that once a particle escapes, it cannot reenter it since it propagates freely. Above, $(\square_{mn})_{m,n}$ is a uniform square grid of the transverse plane to the x_3 -axis. All squares in the grid have area Δx_\perp . Note that the grid can be different for the transmitted and reflected signals. We have considered for the transverse variable of the transmitted energy a uniform grid over a detector of size $[-10, 10] \times [-10, 10]$ centered around the x_3 -axis, and over a detector of size $[-50, 50] \times [-50, 50]$ for the reflected energy. For both cases, we chose 128×128 grid points. The principle of these estimators is simply to count the number of particles that exit the slab before time T and to record their position in the transverse plane.

In Figure 10, we illustrate the transmitted and reflected energy flux, for several values of α . We represent the variations w.r.t. the first coordinate of $x_\perp = (x_1, x_2)$, and for two values of x_2 .

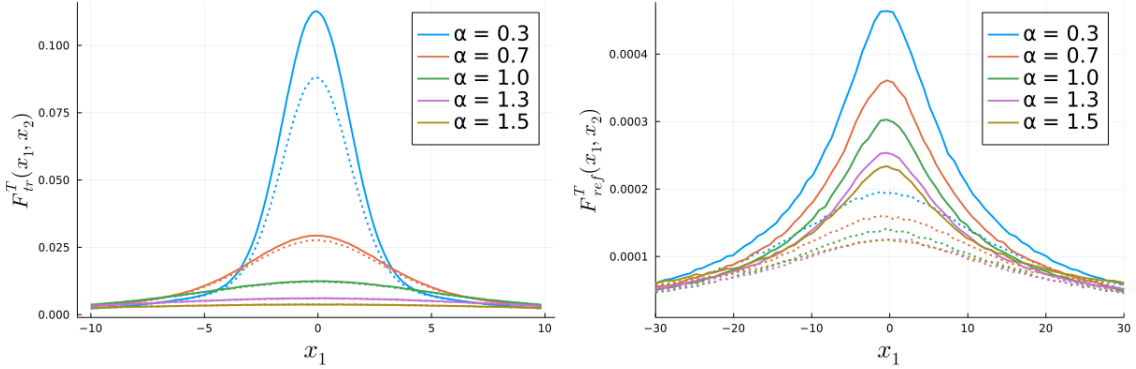


Figure 10: Illustration of the energy at the boundaries $F_{tr}^T(x_\perp)$ and $F_{ref}^T(x_\perp)$ w.r.t. x_1 with $x_2 = 0$ (solid line) and $x_2 = 1$ (dotted line) for the left picture and $x_2 = 10$ (dotted line) for the right one.

One can observe that at fixed times, the larger the α , the more diffuse are the signals. Indeed, as α increases, the jump intensity Λ_ε (in other words the number of scattering events) increases as well as the strength of the diffusive correction σ_ε (see Figure 11).

5.1.2 Time evolution of the exiting energy

Here, we are interested of the time evolution of the energy exiting the slab, and we define the (integrated) reflected and transmitted energy by

$$\mathcal{F}_{tr}(t) := \int_{\mathbb{R}^2} dx_\perp \int_{\mathbb{S}^2} \sigma(d\hat{k}) u(t, x_\perp, x_3 = 40, \hat{k}) \quad \text{and} \quad \mathcal{F}_{ref}(t) := \int_{\mathbb{R}^2} dx_\perp \int_{\mathbb{S}^2} \sigma(d\hat{k}) u(t, x_\perp, x_3 = -5, \hat{k}).$$

The MC estimators for these two quantities are given by

$$\hat{\mathcal{F}}_{tr}(n) := \frac{1}{dt N} \sum_{j=1}^N \mathbf{1}_{(\tau^j \in (t_n, t_{n+1}], X_{3,h,\varepsilon}^j(\tau^j) > 40)} \quad \text{and} \quad \hat{\mathcal{F}}_{ref}(n) := \frac{1}{dt N} \sum_{j=1}^N \mathbf{1}_{(\tau^j \in (t_n, t_{n+1}], X_{3,h,\varepsilon}^j(\tau^j) < -5)}.$$

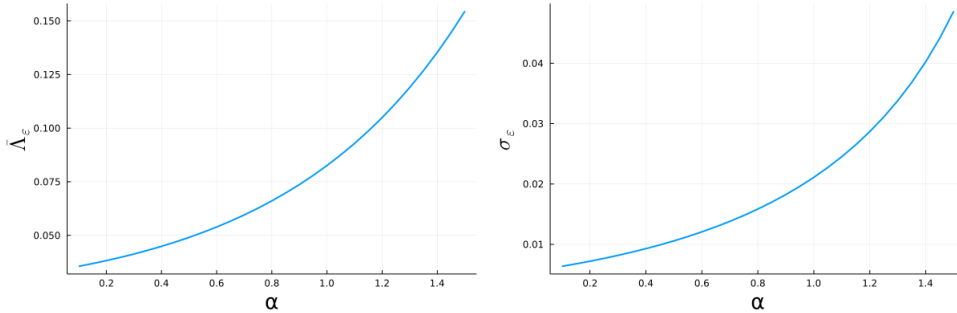


Figure 11: Illustrations of the evolution w.r.t. α of Λ_ε defined by (14) and σ_ε defined by (10). Here, $a = 0.002$ and $\varepsilon = 0.01$.

Here, $(t_n)_n$ is a uniform grid of the time interval with stepsize dt . For the transmitted signal, we have considered the time interval $[40, 45]$ with a stepsize $dt = 0.02$, and have set $[0, 4t_c]$ with a stepsize $dt = 0.4$ for the backscattered signal. Note that the time interval starts at 40 for the transmitted energy, which is the travel time of the wave (traveling at speed $c_0 = 1$) from the source to the plane $x_3 = 40$. These estimators count the number of particles that exit the slab in the time interval $(t_n, t_{n+1}]$ at each side of the slab.

In Figure 12, we illustrate the evolution of the transmitted and reflected energy, for several values of α .

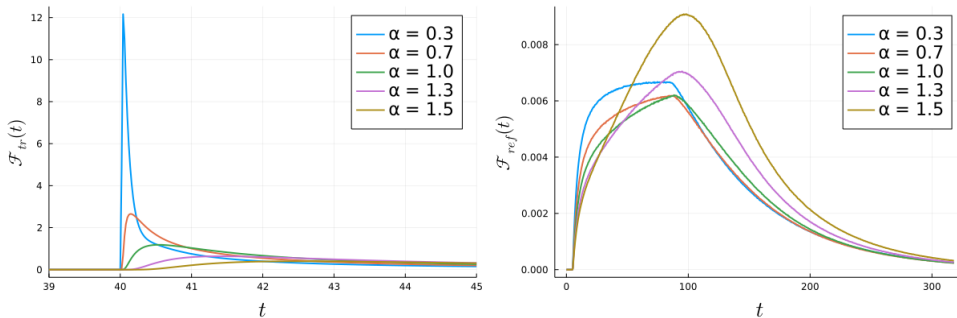


Figure 12: Time evolution of the energy at the boundaries $\mathcal{F}_{tr}(t)$ and $\mathcal{F}_{ref}(t)$.

In the case of the transmitted signal (left), and for small values of α , we see the arrival of the coherent wave at the proper travel time followed by the coda. When α increases, one notices the stronger impact of scattering and of the diffusive correction that smoothes the signal out and damps its amplitude. For the largest α , we only observe a coda. Regarding the reflected signal (right), there is only a coda for all α due our choice of \hat{k}_0 , and one can observe two stages in the dynamics: backscattering increases up to a time of order t_c , about which exponentially decay due to the operator \mathcal{Q} takes over.

5.1.3 Comparison with the Henyey-Greenstein scattering kernel

In this section, we compare the solutions to the RTE with Henyey-Greenstein scattering kernel (6) for an anisotropy factor g close to one with the solutions to (9) with singular kernel derived from (7), that is by setting $\underline{a} \equiv (1 - g)/(2\pi)$ and $\alpha = 1$ in (4). Note that the value of the constant a changes with g , and as a consequence Λ_ε , h , and σ_ε vary accordingly. To illustrate this approximation, we still consider the setting depicted in Figure 9 and the various observables introduced in the previous sections, but now at a time $T = 300$.

We observe in Figure 13 the very good agreement between the two solutions. The reflected signal is well captured by our method despite fairly large values of ε and h . Also, let us mention that the computational cost is decreasing as the anisotropic parameter g is getting close to 1, as the overall jump intensity decreases in this case in the highly peaked regime $g \rightarrow 1$. Regarding the transmitted signal, ε (and then h) needs to be lowered for an accurate approximation, as explained at the beginning of Section 5.1.

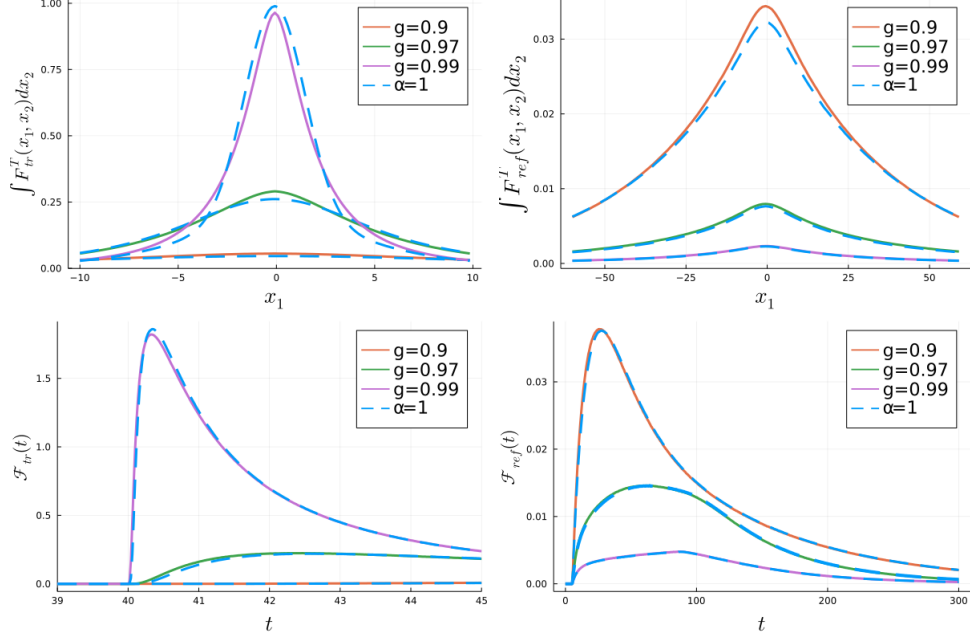


Figure 13: Comparison of the observables obtained using the Henyey-Greenstein scattering kernel and our singular kernel with $\alpha = 1$, $T = 300$, $\varepsilon = 0.01$ for the transmitted observables (left panels), and $\varepsilon = 0.1$ for the reflected ones (right panels).

5.2 Varying α function

In this section, we investigate the influence of a varying α function that characterizes the strength of the singularity. We consider two situations, one inspired from optical tomography, and the second one from wave propagation through atmospheric turbulence.

5.2.1 A two-stage model with a sphere

We keep the setting introduced in Section 5.1, and add a defect with a different value of α to the setting. This defect is modeled by ball of radius 3 centered at the origin and where α is equal to α_1 . We set $\alpha \equiv 1$ in the exterior of the ball, corresponding to the peak forward regime of the Henyey-Greenstein scattering kernel. See Figure 14. This situation models a biological tissue in which statistical properties are changing and define a region of interest for imaging.

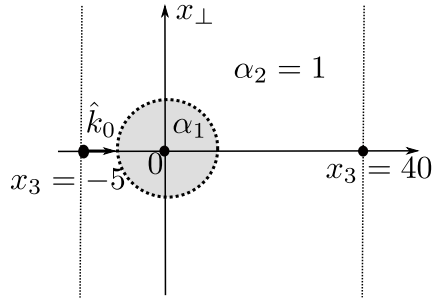


Figure 14: Illustration of the setting with $\lambda = \mathbf{1}_{\{x_3 \in (-5, 40)\}}$ and $\alpha(x) = \alpha_1 \mathbf{1}_{x \in B} + 1 \cdot \mathbf{1}_{x \notin B}$ where B is a ball centered at 0 with radius 3.

We illustrate in Figure 15 the impact of the introduction of the defect on the observables introduced in Section 5.1. The impact is stronger on transmitted observables and quite significant, giving then the possibility to identify the defect with $\alpha = \alpha_1$ inside the scattering medium. Reflected quantities tend to be less sensitive to the presence of the defect since a fraction of the signal is backscattered before reaching it.

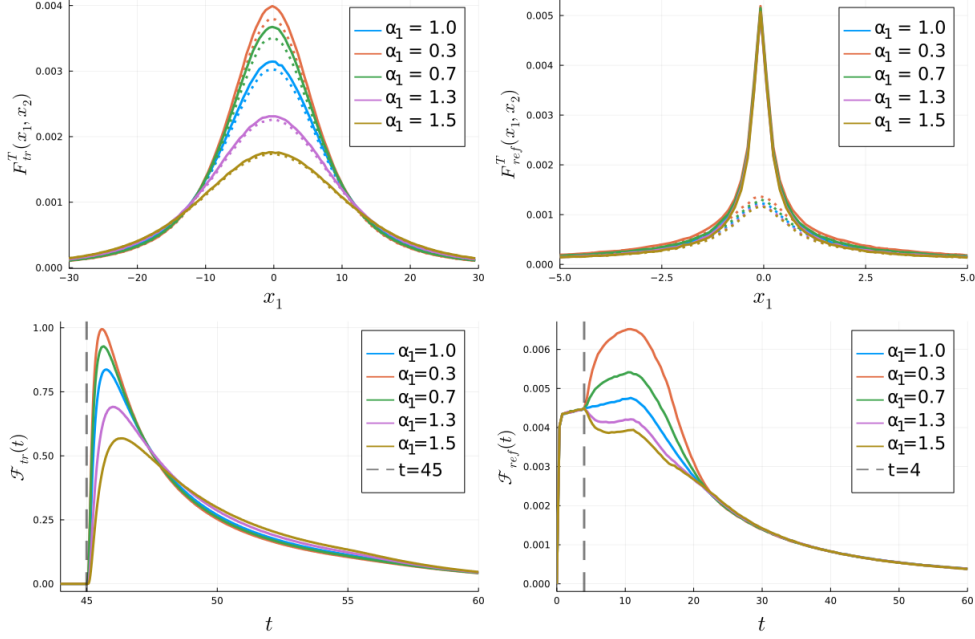


Figure 15: Illustrations the transmitted (left-hand-side) and reflected (right-hand-side) observables with $T = 300$, and $\varepsilon = 0.01$. For the top two pictures we set $x_2 = 0$ (solid lines) and $x_2 = 1.5$ for the top-left and $x_2 = 0.5$ for top-right picture (dotted lines).

5.2.2 Non-Kolmogorov turbulences

In this section, we keep once more the setting introduced in Section 5.1, with the difference that α takes three different large values depending on the altitude parametrized by x_3 , see Figure 16:

$$\alpha(x_3) = 5/3 \cdot \mathbf{1}_{\{x_3 \leq 2\}} + 4/3 \cdot \mathbf{1}_{\{2 < x_3 \leq 8\}} + 1.9 \cdot \mathbf{1}_{\{8 < x_3\}}.$$

The value $5/3$ corresponds to standard Kolmogorov turbulences, while other values are associated with non-Kolmogorov turbulence models [3, 41, 44]. In these models, it is considered that for altitudes higher than 8km, the atmospheric turbulence yields larger statistical patterns (which tend to be created by singular kernels) than around the ground (0-2km). Hence, we set $\alpha = 1.9$ for altitudes greater than 8km. The function a is no longer constant in these models, and for our illustrations we chose

$$a(r) = 0.002 \cdot \exp(-r^2/(2 \times 0.8^2)).$$

In Figure 17, one can notice that non-Kolmogorov turbulence yields quite different signals compared to Kolmogorov turbulence, in particular for reflected quantities. As we saw in Section 5.1, the higher the α , the more diffuse is the signal which then enhances reflected signals. This explains the increased reflections in the non-Kolmogorov case.

6 Proofs

6.1 Proof of Proposition 2.1

Let $v_\varepsilon := u_\varepsilon - u$, so that $v_\varepsilon(t=0) = 0$. We have

$$\begin{aligned} \frac{d}{dt} \|v_\varepsilon(t)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)}^2 &= 2 \langle \partial_t v_\varepsilon(t), v_\varepsilon(t) \rangle_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \\ &= 2 \langle (\sigma_\varepsilon^2 \Delta_{\mathbb{S}^2} + \mathcal{L}_>^\varepsilon) v_\varepsilon(t), v_\varepsilon(t) \rangle_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} + 2 \langle (\sigma_\varepsilon^2 \Delta_{\mathbb{S}^2} - \mathcal{L}_<^\varepsilon) u, v_\varepsilon(t) \rangle_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)}. \end{aligned}$$

Since $\Delta_{\mathbb{S}^2}$ is a nonpositive operator, we have

$$\begin{aligned} \langle (\sigma_\varepsilon^2 \Delta_{\mathbb{S}^2} + \mathcal{L}_>^\varepsilon) v_\varepsilon(t), v_\varepsilon(t) \rangle_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} &\leq -\frac{1}{2} \iiint_{\mathbb{R}^3 \times \mathbb{S}^2 \times \mathbb{S}^2} dx \sigma(d\hat{p}) \sigma(d\hat{k}) \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{S_\varepsilon^\xi} \rho(x, \hat{p} \cdot \hat{k}) (f(x, \hat{p}) - f(x, \hat{k}))^2 \\ &\leq 0. \end{aligned}$$

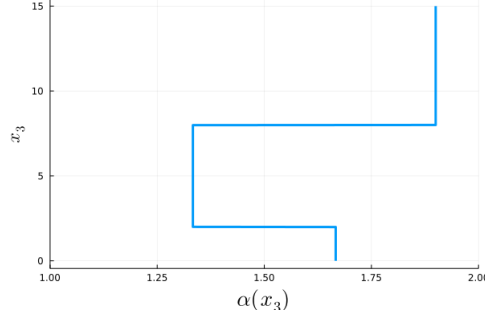


Figure 16: Illustration of a three stages α -profile for a non-Kolmogorov phase function.

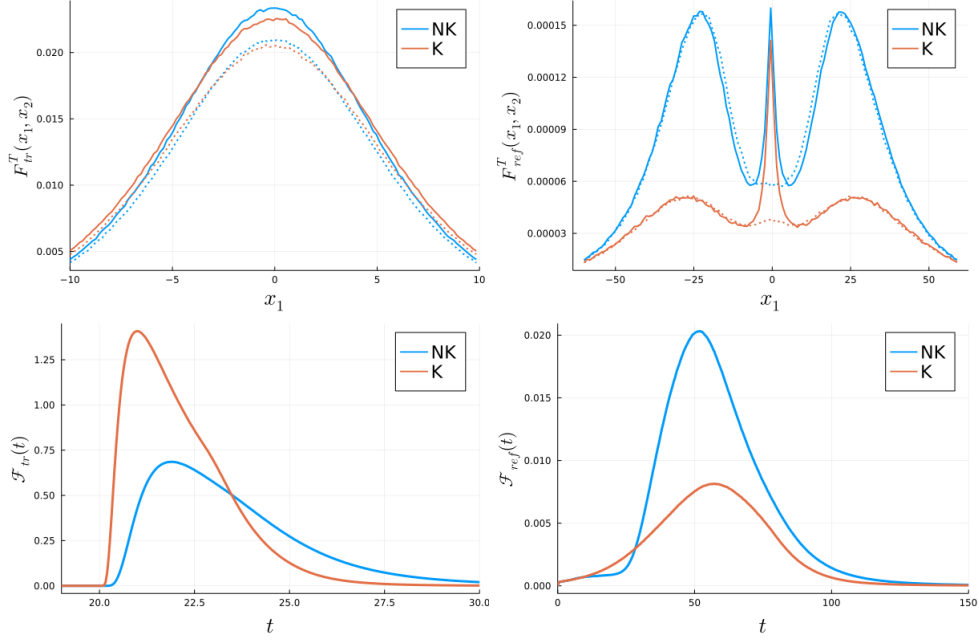


Figure 17: Illustrations the transmitted (left-hand-side) and reflected (right-hand-side) observables with $T = 300$, and $\varepsilon = 0.01$. For the top two pictures we illustrate $x_2 = 0$ (solid lines) and $x_2 = 2$ for the top-left and $x_2 = 5$ for top-right picture (dot line).

We then obtain

$$\sup_{t \in [0, T]} \|v_\varepsilon(t)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)}^2 \leq 2 \int_0^T \|(\sigma_\varepsilon^2 \Delta_{\mathbb{S}^2} - \mathcal{L}_<^\varepsilon)u(t)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)}^2 dt,$$

which concludes the proof using the following lemma.

Lemma 6.1 *Let $0 < \varepsilon < \varepsilon_0 < 1$. Then, for any $f \in L_x^2(\mathbb{R}^3, \mathcal{C}_k^4(\mathbb{S}^2))$, we have*

$$\|(\mathcal{L}_<^\varepsilon - \sigma_\varepsilon^2 \Delta_{\mathbb{S}^2})f\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \leq \varepsilon^{2-\alpha_M/2} E(f)$$

where, with $\check{f}(x, v) := f(x, v/|v|)$ for $v \in \mathbb{R}^3$,

$$\begin{aligned} E(f) := & \frac{\pi}{3(1-\varepsilon_0)^6} \sup \lambda \sup a \sup_{|h| \leq r_{\varepsilon_0}} \|D_k^4 \check{f}(\cdot, \cdot + h)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \\ & + \left(\frac{6\pi}{(1-\varepsilon_0)^3} \sup a + 2^4 \pi \sup_{v \in [0, 2\sqrt{2\varepsilon_0}]} |a''(v)| \right) \sup \lambda \|\Delta_{\mathbb{S}^2} f\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)}. \end{aligned} \quad (29)$$

Proof Before starting the proof, we introduce the *retraction* $R_{\hat{k}}$ at \hat{k} onto the sphere $R_{\hat{k}}(h) := \frac{\hat{k}+h}{|\hat{k}+h|}$, and

$$B_{\varepsilon, \hat{k}} := R_{\hat{k}}^{-1}(S_{<}^\varepsilon) = \{h = \beta_1 \hat{k}_1^\perp + \beta_2 \hat{k}_2^\perp : \beta = (\beta_1, \beta_2) \in \mathbb{R}^2 \text{ with } |\beta| < r_\varepsilon\},$$

where $(\hat{k}_1^\perp, \hat{k}_2^\perp)$ stands for an orthonormal basis of \hat{k}^\perp . We also recall that $r_\varepsilon = \sqrt{1 - (1 - \varepsilon)^2}/(1 - \varepsilon)$, coming from the relation $\tan(\arccos(s)) = \sqrt{1 - s^2}/s$ and (8). In different terms, $B_{\varepsilon, \hat{k}}$ is a ball centered at 0 with radius r_ε on the tangent plane to the unit sphere at \hat{k} , and the retraction $R_{\hat{k}}$ holds from $B_{\varepsilon, \hat{k}}$ onto $S_\varepsilon^\varepsilon$.

To prove the lemma, we start with the following change of variables $\hat{p} = R_{\hat{k}}(h)$ in $\mathcal{L}_<^\varepsilon$, so that

$$\mathcal{L}_<^\varepsilon f(x, \hat{k}) = \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{B_{\varepsilon, \hat{k}}} \rho(x, R_{\hat{k}}(h) \cdot \hat{k}) (f(x, R_{\hat{k}}(h)) - f(x, R_{\hat{k}}(0))) |\det \text{Jac} R_{\hat{k}}(h)| dh.$$

Using that $\check{f}(x, \hat{k} + h) = f(x, R_{\hat{k}}(h))$ and $\check{f}(x, \hat{k}) = f(x, \hat{k})$, one can decompose $\mathcal{L}_<^\varepsilon f$ as

$$\mathcal{L}_<^\varepsilon f(x, \hat{k}) = D_1 + D_2 + D_3 + D_4,$$

where the terms D_j follow with obvious notations from the Taylor expansion

$$\begin{aligned} \check{f}(x, \hat{k} + h) - \check{f}(x, \hat{k}) &= D_k \check{f}(x, \hat{k})(h) + \frac{1}{2!} D_k^2 \check{f}(x, \hat{k})(h, h) + \frac{1}{3!} D_k^3 \check{f}(x, \hat{k})(h, h, h) \\ &+ \frac{1}{3!} \int_0^1 (1-s)^3 D_k^4 \check{f}(x, \hat{k} + sh)(h, h, h, h) ds. \end{aligned}$$

The terms D_1 and D_3 . Using that the ball $B_{\varepsilon, \hat{k}}$ in the tangent plane is symmetric with respect to 0, we just make the change of variables $h \rightarrow -h$, so that $D_1 = -D_1$ and $D_3 = -D_3$ leading to $D_1 = D_3 = 0$.

The term D_4 . We have

$$|D_4| \leq \frac{\lambda(x)}{3! 2^{1+\alpha(x)/2}} \int_0^1 ds (1-s)^3 \int_{B_{\varepsilon, \hat{k}}} dh \rho(x, R_{\hat{k}}(h) \cdot \hat{k}) \|D_k^4 \check{f}(x, \hat{k} + sh)\| |h|^4 |\det \text{Jac} R_{\hat{k}}(h)| dh.$$

Since $R_{\hat{k}}(h) = \frac{\hat{k} + h}{\sqrt{1 + |h|^2}}$, we have

$$R_{\hat{k}}(h) \cdot \hat{k} = \frac{1}{\sqrt{1 + |h|^2}} \quad \text{and} \quad |\det \text{Jac} R_{\hat{k}}(h)| = \frac{1}{\sqrt{1 + |h|^2}}.$$

As a consequence, we find

$$\begin{aligned} \|D_4\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} &\leq \frac{1}{4!} \sup_v a(v) \sup_{|h| \leq r_\varepsilon} \|D^4 \check{f}(\cdot, \cdot + h)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \\ &\times \sup_x \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{\{|h| \leq r_\varepsilon\}} \frac{|h|^4}{(1 - 1/\sqrt{1 + |h|^2})^{1+\alpha(x)/2}} dh. \end{aligned}$$

Changing to polar coordinates in the last integral gives

$$\begin{aligned} \int_{\{|h| \leq r_\varepsilon\}} \frac{|h|^4}{(1 - 1/\sqrt{1 + |h|^2})^{1+\alpha(x)/2}} dh &= 2\pi \int_0^{r_\varepsilon} \frac{r^5}{(1 - 1/\sqrt{1 + r^2})^{1+\alpha(x)/2}} dr \\ &= 2\pi \int_0^{\varepsilon/(1-\varepsilon)} \frac{v^{1-\alpha(x)/2} (2+v)^2 (v+1)^{2+\alpha(x)/2}}{v} dv \end{aligned}$$

where we used the change of variables $v = \sqrt{1 + r^2} - 1$ and that $\sqrt{1 + r_\varepsilon^2} - 1 = \varepsilon/(1 - \varepsilon)$. This gives finally

$$\|D_4\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \leq \frac{\pi}{3} \sup_v a(v) \sup_{h \in B_\varepsilon} \|D^4 \hat{f}(\cdot, \cdot + h)\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} \sup_x \lambda(x) \frac{\varepsilon^{2-\alpha_M/2}}{(1 - \varepsilon)^6}.$$

The term D_2 . For this last term, we have

$$D_2 = \frac{\lambda(x)}{2^{2+\alpha(x)/2}} \int_{B_{\varepsilon, \hat{k}}} \rho(x, R_{\hat{k}}(u) \cdot \hat{k}) D^2 \check{f}(x, \hat{k})(h, h) |\det \text{Jac} R_{\hat{k}}(h)| dh,$$

with

$$D^2 \check{f}(x, \hat{k})(h, h) = h_1^2 \partial_{k_1 k_1}^2 \check{f}(x, \hat{k}) + h_2^2 \partial_{k_2 k_2}^2 \check{f}(x, \hat{k}) + 2h_1 h_2 \partial_{k_1 k_2}^2 \check{f}(x, \hat{k}),$$

and, accordingly, the following decomposition $D_2 = D_{21} + D_{22} + 2D_{23}$. Applying the change of variables $h = (h_1, h_2) \rightarrow (-h_1, h_2)$ leads to $D_{23} = 0$. Setting $h = (h_1, h_2) \rightarrow (h_2, h_1)$ leads to

$$D_2 = \frac{\lambda(x)}{2^{3+\alpha(x)/2}} \int_{B_{\varepsilon, \hat{k}}} \rho(x, R_{\hat{k}}(h) \cdot \hat{k}) |h|^2 |\det \text{Jac} R_{\hat{k}}(h)| dh \text{Trace}(\text{Hess} \check{f}(x, \hat{k})),$$

where

$$\text{Trace}(\text{Hess} \check{f}(x, \hat{k})) = \Delta_p f \left(x, \frac{p}{|p|} \right) \Big|_{p=\hat{k}} = \Delta_{\mathbb{S}^2} f(x, \hat{k}).$$

Furthermore, with the change of variables $\hat{p} = R_{\hat{k}}(h)$, we find

$$\int_{B_{\varepsilon, \hat{k}}} \rho(x, R_{\hat{k}}(h) \cdot \hat{k}) |h|^2 |\det \text{Jac} R_{\hat{k}}(h)| dh = \int_{S_{\varepsilon}^2} \rho(x, \hat{p} \cdot \hat{k}) |R_{\hat{k}}^{-1}(\hat{p})|^2 \sigma(d\hat{p}),$$

and note that for $\hat{p} \cdot \hat{k} = \cos(\theta)$, we have $|R_{\hat{k}}^{-1}(\hat{p})|^2 = \tan^2(\theta)$. As a result, moving to spherical coordinates, and performing the change of variables $v = \tan(\theta/2)$ together with the relation

$$\arccos(s) = 2 \arctan \left(\frac{\sqrt{1-s^2}}{1+s} \right) \quad \text{for } s \in (-1, 1],$$

we find, with $r'_\varepsilon = \sqrt{1 - (1-\varepsilon)^2}/(2-\varepsilon)$,

$$\begin{aligned} \int_{S_{\varepsilon}^2} \rho(x, \hat{p} \cdot \hat{k}) |R_{\hat{k}}^{-1}(\hat{p})|^2 \sigma(d\hat{p}) &= 2\pi \int_0^{\arccos(1-\varepsilon)} \rho(x, \cos(\theta)) \tan^2(\theta) \sin(\theta) d\theta \\ &= 2^{3-\alpha(x)/2} \pi \int_0^{r'_\varepsilon} \frac{a(2v/\sqrt{1+v^2})(1+v^2)^{\alpha(x)/2}}{(1-v^2)^2 v^{\alpha(x)-1}} dv, \end{aligned}$$

leading to

$$D_2 = \tilde{D}_2 \Delta_{\mathbb{S}^2} f(\hat{k}) := 2^{1-\alpha(x)} \pi \lambda(x) \int_0^{r'_\varepsilon} \frac{a(2v/\sqrt{1+v^2})(1+v^2)^{\alpha(x)/2}}{(1-v^2)^2 v^{\alpha(x)-1}} dv \Delta_{\mathbb{S}^2} f(x, \hat{k}).$$

Now, let us introduce

$$\tilde{\sigma}_\varepsilon^2(x) := 2^{1-\alpha(x)} \pi \lambda(x) \int_0^{r'_\varepsilon} a(2v/\sqrt{1+v^2}) \frac{dv}{v^{\alpha(x)-1}},$$

and remark that

$$\begin{aligned} |\tilde{D}_2 - \tilde{\sigma}_\varepsilon^2(x)| &\leq 3 \cdot 2^{3-\alpha(x)} \pi \lambda(x) \sup_v a(v) \frac{1}{(1-r'_\varepsilon)^3} \int_0^{r'_\varepsilon} v^{3-\alpha(x)} dv \leq 3 \cdot 2^{2-\alpha(x)} \pi \lambda(x) \sup_v a(v) \frac{r'_\varepsilon{}^2}{(1-r'_\varepsilon)^3} \\ &\leq 6\pi \sup_x \lambda(x) \sup_v a(v) \frac{\varepsilon^{2-\alpha_M/2}}{(1-\varepsilon)^3}, \end{aligned}$$

since $r'_\varepsilon \leq \sqrt{2\varepsilon}$, $0 \leq \alpha_m \leq \alpha(x) \leq \alpha_M < 2$ and $1 - r'_\varepsilon{}^2 = \frac{2(1-\varepsilon)}{2-\varepsilon} > 2(1-\varepsilon)$. With the definition of σ_ε given in (10), we obtain using $a'(0) = 0$,

$$\begin{aligned} |\tilde{\sigma}_\varepsilon^2(x) - \sigma_\varepsilon^2(x)| &\leq 2^5 \pi \lambda(x) \sup_{\tilde{v} \in [0, 2\sqrt{2\varepsilon}]} |a''(\tilde{v})| \int_0^{r'_\varepsilon} v^{3-\alpha(x)} dv, \\ &\leq 2^4 \pi \sup_x \lambda(x) \sup_{v \in [0, 2\sqrt{2\varepsilon}]} |a''(v)| \varepsilon^{2-\alpha_M/2}. \end{aligned}$$

Collecting the various estimates on the D_j and using that $\varepsilon \leq \varepsilon_0 < 1$ concludes the proof of Lemma 6.1 and therefore of Proposition 2.1. \square

6.2 Proof of Proposition 2.2

We first show that the infinitesimal generator of the Markov process D_ε is \mathcal{A}_ε .

Infinitesimal generator for D_ε . Let f be a smooth bounded function on $\mathbb{R}^3 \times \mathbb{S}^2$. The goal of this section is to prove that

$$\lim_{h \rightarrow 0^+} \frac{1}{h} (\mathbb{E}_z[f(D_\varepsilon(h))] - f(z)) = \mathcal{A}_\varepsilon f(z). \quad (30)$$

To this end, we introduce the first jump time T_1 to obtain

$$\mathbb{E}_z[f(D_\varepsilon(t))] = \mathbb{E}_z[f(D_\varepsilon(t))\mathbf{1}_{(T_1 > t)}] + \mathbb{E}_z[f(D_\varepsilon(t))\mathbf{1}_{(T_1 \leq t)}]. \quad (31)$$

Using conditional expectations, we find for the first term

$$\begin{aligned} \mathbb{E}_z[f(D_\varepsilon(t))\mathbf{1}_{(T_1 > t)}] &= \mathbb{E}_z[f(\psi_0^z(t))\mathbf{1}_{(T_1 > t)}] = \mathbb{E}_z[\mathbb{E}_z[f(\psi_0^z(t))\mathbf{1}_{(T_1 > t)} | \psi_0^z(s), s \in [0, t]]] \\ &= \mathbb{E}_z\left[f(\psi_0^z(t)) \mathbb{P}_z(T_1 > t | \psi_0^z(s), s \in [0, t])\right] = \mathbb{E}_z\left[f(\psi_0^z(t)) e^{-\int_0^t \Lambda_\varepsilon(\psi_0^z(s)) ds}\right]. \end{aligned}$$

With the following notations for the flow ψ_n^z ,

$$\psi_n^z = (X_n^x, K_n^{\hat{k}}) = ((X_{j,n}^x)_{j=1,2,3}, (K_{j,n}^{\hat{k}})_{j=1,2,3}) \in \mathbb{R}^3 \times \mathbb{S}^2,$$

together with $\check{f}(x, k) = f(x, k/|k|)$ for $(x, k) \in \mathbb{R}^3 \times \mathbb{R}^3$, the Itô formula yields

$$\begin{aligned} df(\psi_n^z(t)) &= d\check{f}(\psi_n^z(t)) = \nabla_x \check{f}(\psi_n^z(t)) \cdot dX_n^x(t) + \nabla_k \check{f}(\psi_n^z(t)) \cdot dK_n^{\hat{k}}(t) \\ &\quad + \frac{1}{2} \sum_{j,l=1,2,3} \partial_{k_j k_l}^2 \check{f}(\psi_n^z(t)) d\langle K_{j,n}^{\hat{k}}, K_{l,n}^{\hat{k}} \rangle \\ &= K_n^{\hat{k}}(t) \cdot \nabla_x \check{f}(\psi_n^z(t)) dt + \sigma_\varepsilon^2(X_n^x(t)) \Delta_{\mathbb{S}^2} f(\psi_n^z(t)) dt \\ &\quad + \sqrt{2} \sigma_\varepsilon(X_n^x(t)) \nabla_k \check{f}(\psi_n^z(t)) \cdot (K_n^{\hat{k}}(t) \times dW_n(t)). \end{aligned}$$

Above, we have used the fact that

$$\Delta_{\mathbb{S}^2} f(x, \hat{k}) = \Delta_k \check{f}(x, \hat{k}) - \sum_{j,l=1,2,3} \hat{k}_j \hat{k}_l \partial_{k_j k_l}^2 \check{f}(x, \hat{k}) - 2 \sum_{j=1,2,3} \hat{k}_j \partial_{k_j} \check{f}(x, \hat{k}). \quad (32)$$

Therefore, we have for $n = 0$,

$$d\left((f(\psi_0^z(t)) - f(z)) e^{-\int_0^t \Lambda_\varepsilon(\psi_0^z(s)) ds}\right) = \left(df(\psi_0^z(t)) - \Lambda_\varepsilon(\psi_0^z(t))(f(\psi_0^z(t)) - f(z))\right) e^{-\int_0^t \Lambda_\varepsilon(\psi_0^z(s)) ds}$$

so that

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \mathbb{E}_z[(f(D_\varepsilon(h)) - f(z))\mathbf{1}_{(T_1 > h)}] = \hat{k} \cdot \nabla_x f(z) + \sigma_\varepsilon^2(x) \Delta_{\mathbb{S}^2} f(z), \quad \forall z = (x, \hat{k}) \in \mathbb{R}^3 \times \mathbb{S}^2.$$

Regarding the second term in (31), we find, using the Markov property in the third line,

$$\begin{aligned} \mathbb{E}_z[f(D_\varepsilon(h))\mathbf{1}_{(T_1 \leq h)}] &= \mathbb{E}_z[\mathbb{E}_z[f(D_\varepsilon(h)) | T_1] \mathbf{1}_{(T_1 \leq h)}] \\ &= \mathbb{E}_z\left[\int_0^h \mathbb{E}_z[f(D_\varepsilon(h)) | T_1 = v] \Lambda_\varepsilon(\psi_0^z(v)) e^{-\int_0^v \Lambda_\varepsilon(\psi_0^z(s)) ds} dv\right] \\ &= \mathbb{E}_z\left[\int_0^h \int_{\mathbb{R}^3 \times \mathbb{S}^2} \mathbb{E}_{z'}[f(D_\varepsilon(h-v))] \Pi_\varepsilon(\psi_0^z(v), dz') \Lambda_\varepsilon(\psi_0^z(v)) e^{-\int_0^v \Lambda_\varepsilon(\psi_0^z(s)) ds} dv\right] \\ &= \mathbb{E}_z\left[\int_0^h \int_{\mathbb{S}^2} \mathbb{E}_{(X_0^x(h-w), \hat{p})}[f(D_\varepsilon(w))] \pi_\varepsilon(\psi_0^z(h-w), \hat{p}) \sigma(d\hat{p}) \Lambda_\varepsilon(\psi_0^z(h-w)) e^{-\int_0^{h-w} \Lambda_\varepsilon(\psi_0^z(s)) ds} dw\right], \end{aligned}$$

where the probability Π_ε and the density π_ε are defined respectively in (16) and (17). As a consequence,

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \mathbb{E}_z[f(D_\varepsilon(h))\mathbf{1}_{(T_1 \leq h)}] = \Lambda_\varepsilon(z) \int_{\mathbb{S}^2} f(x, \hat{p}) \pi_\varepsilon(z, \hat{p}) \sigma(d\hat{p}).$$

Moreover, we have

$$\mathbb{P}_z(T_1 \leq h) = \mathbb{E}_z[\mathbb{P}_z(T_1 \leq h | \psi_0^z(s), s \in [0, h])] = \mathbb{E}_z\left[1 - e^{-\int_0^h \Lambda_\varepsilon(\psi_0^z(s)) ds}\right],$$

and it is then direct to see that

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}_z(T_1 \leq h) = \Lambda_\varepsilon(z).$$

This finally yields

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \mathbb{E}_z[(f(D_\varepsilon(h)) - f(z))\mathbf{1}_{(T_1 \leq h)}] = \Lambda_\varepsilon(z) \int_{\mathbb{S}^2} (f(x, \hat{p}) - f(x, \hat{k})) \pi_\varepsilon(z, \hat{p}) \sigma(d\hat{p}),$$

which gives (30) collecting all results.

Proof of (18). Since D_ε is a solution to the martingale problem associated to \mathcal{A}_ε (see [12, Proposition 1.7 pp. 162]), we have, for any smooth bounded function f on $\mathbb{R}^3 \times \mathbb{S}^2$,

$$\mu_t(f) = \mu_0(f) + \int_0^t \mu_s(\mathcal{A}_\varepsilon f) ds, \quad \text{where} \quad \mu_t(f) := \mathbb{E}_{\mu_0}[f(D_\varepsilon(t))].$$

Let

$$\nu_t(f) := \frac{1}{\bar{u}_0} \int_{\mathbb{R}^3 \times \mathbb{S}^2} u_\varepsilon(t, x, \hat{k}) f(x, \hat{k}) dx \sigma(d\hat{k}).$$

Since u_ε solves (9), we have

$$u_\varepsilon(t) = u_0 + \int_0^t \mathcal{A}_\varepsilon^* u_\varepsilon(s) ds,$$

where $\mathcal{A}_\varepsilon^*$ stands for the adjoint operator of \mathcal{A}_ε in $L^2(\mathbb{R}^3 \times \mathbb{S}^2)$. Then,

$$\begin{aligned} \nu_t(f) &= \mu_0(f) + \frac{1}{\bar{u}_0} \int_{\mathbb{R}^3 \times \mathbb{S}^2} \int_0^t \mathcal{A}_\varepsilon^* u_\varepsilon(s, x, \hat{k}) ds f(x, \hat{k}) dx \sigma(d\hat{k}) \\ &= \mu_0(f) + \int_0^t \frac{1}{\bar{u}_0} \int_{\mathbb{R}^3 \times \mathbb{S}^2} u_\varepsilon(s, x, \hat{k}) \mathcal{A}_\varepsilon f(x, \hat{k}) dx \sigma(d\hat{k}) ds = \mu_0(f) + \int_0^t \nu_s(\mathcal{A}_\varepsilon f) ds. \end{aligned}$$

Therefore, according to [12, Proposition 9.18 pp. 251], we have $\mu_t = \nu_t$ for any $t \geq 0$, which concludes the proof.

6.3 Proof of Theorem 3.1

6.3.1 Step 1

We first introduce an equivalent formulation (in the statistical sense) for the process D_ε in terms of a stochastic differential equation (SDE) with jumps. This representation is useful when comparing with the discrete scheme. Let $\tilde{D}_\varepsilon = (\tilde{X}_\varepsilon, \tilde{K}_\varepsilon)$ be the solution to the following SDE with jumps:

$$\begin{aligned} d\tilde{X}_\varepsilon(t) &= \tilde{K}_\varepsilon(t) dt \\ d\tilde{K}_\varepsilon(t) &= \sqrt{2} \sigma_\varepsilon(\tilde{X}_\varepsilon(t^-)) \tilde{K}_\varepsilon(t^-) \times dW - 2 \sigma_\varepsilon^2(\tilde{X}_\varepsilon(t^-)) \tilde{K}_\varepsilon(t^-) dt \\ &\quad + \int_{(0, \pi) \times (0, 2\pi) \times (0, 1)} \tilde{R}(\theta, \varphi, \tilde{D}_\varepsilon(t^-), v) P(dt, d\theta, d\varphi, dv) \end{aligned} \tag{33}$$

where the function \tilde{R} is defined by, for $z = (x, \hat{k})$,

$$\tilde{R}(\theta, \varphi, z, v) = \begin{cases} R(\theta, \varphi, \hat{k}) - \hat{k} & \text{if } v \leq \Lambda_\varepsilon(z) \pi_\varepsilon(z, R(\theta, \varphi, \hat{k})) / \bar{\Lambda}_\varepsilon \\ 0 & \text{otherwise.} \end{cases}$$

Above, R is defined in (23), Λ_ε in (14), $\bar{\Lambda}_\varepsilon$ in (20), π_ε in (17), and P is a random Poisson measure with intensity measure

$$\mu(dt, d\theta, d\varphi, dv) = \bar{\Lambda}_\varepsilon \mathbf{1}_{(0, \infty) \times (0, \pi) \times (0, 2\pi) \times (0, 1)}(t, \theta, \varphi, v) \sin(\theta) dt d\varphi d\theta / (4\pi). \tag{34}$$

See e.g. [1, Chapter 2] for more details on Poisson random measures. The notation t^- is standard and refers to the left limit when approaching t before a jump. With this construction, the infinitesimal generator for the Markov process \tilde{D}_ε is

$$\tilde{\mathcal{A}}_\varepsilon f(z) = \hat{k} \cdot \nabla_x f(z) + \sigma_\varepsilon^2(x) \Delta_{\mathbb{S}^2} f(z) + \bar{\Lambda}_\varepsilon \int_{(0, \pi) \times (0, 2\pi) \times (0, 1)} (f(x, \hat{k} + \tilde{R}(\theta, \varphi, z, v)) - f(z)) \sigma(d\hat{p}) dv,$$

and we have the following result.

Lemma 6.2 *The processes \tilde{D}_ε and D_ε have the same distribution.*

Proof \tilde{D}_ε and D_ε are both Markov processes and are therefore characterized by their generators. We just have then to prove that $\mathcal{A}_\varepsilon g = \tilde{\mathcal{A}}_\varepsilon g$ for any bounded smooth function g . This is a direct consequence of the definition to \tilde{R} . Indeed, denoting by $\tilde{\mathcal{I}}_\varepsilon$ and \mathcal{I}_ε the integral operators in respectively $\tilde{\mathcal{A}}_\varepsilon$ and \mathcal{A}_ε , we have with $z = (x, \hat{k})$,

$$\begin{aligned}\tilde{\mathcal{I}}_\varepsilon g(z) &= \bar{\Lambda}_\varepsilon \int_{(0,\pi) \times (0,2\pi) \times (0,1)} (g(x, \hat{k} + \tilde{R}(\theta, \varphi, z, v)) - g(z)) \sin(\theta) d\varphi d\theta dv / (4\pi) \\ &= \bar{\Lambda}_\varepsilon \int_{(0,\pi) \times (0,2\pi)} \int_0^{\Lambda_\varepsilon(z) \pi_\varepsilon(z, R(\theta, \varphi, \hat{k})) / \bar{\Lambda}_\varepsilon} dv (g(x, R(\theta, \varphi, \hat{k})) - g(z)) \sin(\theta) d\varphi d\theta / (4\pi) \\ &= \bar{\Lambda}_\varepsilon \int_{\mathbb{S}^2} \frac{\Lambda_\varepsilon(z) \pi_\varepsilon(z, \hat{p})}{\bar{\Lambda}_\varepsilon} (g(x, \hat{p}) - g(z)) \sigma(d\hat{p}) = \mathcal{I}_\varepsilon g(z),\end{aligned}$$

which concludes the proof. \square

6.3.2 Step 2

The goal is now to prove that the discretized process $D_{h,\varepsilon}$ approximates \tilde{D}_ε in a statistical sense. We use for this the notations of Section 3.2 for $X_{n,m}$, $K_{n,m}$ and $\hat{K}_{n,m}$. For simplicity, we suppose that the Gaussian vectors $(W_{n,m})$ in (21) are obtained from a single 3D standard Brownian motion W as follows: for $n \geq 0$ and $m = 0, \dots, m_n$, we set $W_{n,m} = (W(T_{n,m} + h_{n,m}) - W(T_{n,m})) / \sqrt{h_{n,m}}$. In the sequel, we will use the following process, defined by, for $t \in [T_{n,m}, T_{n,m} + h_{n,m}]$, $m = 0, \dots, m_n$,

$$\begin{cases} \mathcal{X}_{n,m}(t) = X_{n,m} + \int_{T_{n,m}}^t \hat{K}_{n,m} ds \\ \mathcal{K}_{n,m}(t) = \hat{K}_{n,m} - 2 \int_{T_{n,m}}^t \sigma_\varepsilon(X_{n,m}) \hat{K}_{n,m} ds + \sqrt{2} \int_{T_{n,m}}^t \sigma_\varepsilon(X_{n,m}) \hat{K}_{n,m} \times dW(s). \end{cases} \quad (35)$$

For $t \geq 0$, we then combine the $(\mathcal{X}_{n,m}, \mathcal{K}_{n,m})$ into

$$\mathfrak{D}_{h,\varepsilon}(t) = (\mathfrak{X}_{h,\varepsilon}(t), \mathfrak{K}_{h,\varepsilon}(t)) = \sum_{n=0}^{\infty} \sum_{m=0}^{m_n} \mathbf{1}_{[T_{n,m}, T_{n,m} + h_{n,m})}(t) \Psi_{n,m}(t),$$

where $\Psi_{n,m}(t) = (\mathcal{X}_{n,m}(t), \mathcal{K}_{n,m}(t))$ is the solution to (35) with initial condition $\Psi_{n,m}(T_{n,m}) = (X_{n,m}, \hat{K}_{n,m})$. Note that $\mathfrak{D}_{h,\varepsilon}$ is simply an interpolation of $D_{h,\varepsilon}$ in the intervals $[T_{n,m}, T_{n,m} + h_{n,m}]$ that will allow us to use the Itô formula.

For any smooth function f , we now introduce b the (smooth) solution to the following backward RTE,

$$\partial_t b + \hat{k} \cdot \nabla_x b + \frac{\lambda(x)}{2^{1+\alpha(x)/2}} \int_{\mathbb{S}^2} \rho(x, \hat{k} \cdot \hat{p}) (b(\hat{p}) - b(\hat{k})) \sigma(d\hat{p}) = 0, \quad (36)$$

with terminal condition $b(T, x, \hat{k}) = f(x, \hat{k})$ and use the notation $\check{b}(t, x, k) = b(t, x, k/|k|)$, $(t, x, k) \in [0, T] \times \mathbb{R}^3 \times \mathbb{R}^3$. We have the following result.

Proposition 6.1 *For any $T > 0$, any smooth bounded function f on $\mathbb{R}^3 \times \mathbb{S}^2$, and any $(x, \hat{k}) \in \mathbb{R}^3 \times \mathbb{S}^2$, we have*

$$|\mathbb{E}_{(x, \hat{k})}[f(D_{h,\varepsilon}(T))] - \mathbb{E}_{(x, \hat{k})}[f(\tilde{D}_\varepsilon(T))]| \leq \varepsilon^{2-\alpha_M/2} 2T E_\infty(b) + hF_1(b),$$

where $E_\infty(b)$ is defined as in (29) with L^2 norms replaced by L^∞ norms in all variables, and where $F_1(b)$ is an explicit function independent of ε and h that depends on derivatives of b up to order 4.

The notation $\mathbb{E}_{(x, \hat{k})}$ above indicates that the process under the expectation starts at the point (x, \hat{k}) .

Proof The proof consists in analyzing the discretization error of the diffusion process in a weak sense following the ideas of [42].

We first write

$$E_{h,\varepsilon} := \mathbb{E}_z[f(D_{h,\varepsilon}(T))] - \mathbb{E}_z[f(\tilde{D}_\varepsilon(T))] = \mathbb{E}_z[b(T, D_{h,\varepsilon}(T))] - \mathbb{E}_z[b(T, \tilde{D}_\varepsilon(T))],$$

with $z = (x, \hat{k})$. Using the Itô formula, see e.g. [1, Th 4.4.7 p. 226], together with the expression of $\Delta_{\mathbb{S}^2}$ given in (32), we have

$$\begin{aligned} \mathbb{E}_z[b(T, \tilde{D}_\varepsilon(T))] - b(0, z) &= \mathbb{E}_z \left[\int_0^T \partial_t b(t, \tilde{D}_\varepsilon(t)) + \tilde{K}_\varepsilon(t) \cdot \nabla_x b(t, \tilde{D}_\varepsilon(t)) \right. \\ &\quad \left. + \sigma_\varepsilon^2(\tilde{X}_\varepsilon(t)) \Delta_{\mathbb{S}^2} b(t, \tilde{D}_\varepsilon(t)) + \mathcal{I}_>^\varepsilon(t) dt \right], \end{aligned}$$

where

$$\mathcal{I}_>^\varepsilon(t) = \frac{\lambda(\tilde{X}_\varepsilon(t))}{2^{1+\alpha(\tilde{X}_\varepsilon(t))/2}} \int_{S_\varepsilon^\xi} \rho(\tilde{X}_\varepsilon(t), \tilde{K}_\varepsilon(t) \cdot \hat{p})(b(t, \tilde{X}_\varepsilon(t), \hat{p}) - b(t, \tilde{X}_\varepsilon(t), \tilde{K}_\varepsilon(t))) \sigma(d\hat{p}).$$

Using the fact that b satisfies (36), we find

$$\mathbb{E}_z[b(T, \tilde{D}_\varepsilon(T))] = b(0, z) + \mathbb{E}_z \left[\int_0^T \sigma_\varepsilon^2(\tilde{X}_\varepsilon(t)) \Delta_{\mathbb{S}^2} b(t, \tilde{D}_\varepsilon(t)) - \mathcal{I}_<^\varepsilon(t) dt \right],$$

where $\mathcal{I}_<^\varepsilon$ is as $\mathcal{I}_>^\varepsilon$ with S_ε^ξ replaced by S_ε^ξ . Following the lines of the proof of Lemma 6.1, we obtain

$$|\mathbb{E}[b(T, \tilde{D}_\varepsilon(T))] - b(0, z)| \leq \varepsilon^{2-\alpha_M/2} T E_\infty(b),$$

where $E_\infty(b)$ is defined as in (29) with L^2 norms replaced by L^∞ norms. We move now to the term $\mathbb{E}_z[b(T, D_{h,\varepsilon}(T))]$ which requires more work. Decomposing the interval $[0, T]$ according to the grid $(T_{n,m})$, we have

$$\begin{aligned} b(T, D_{h,\varepsilon}(T)) - b(0, z) &= \check{b}(T, D_{h,\varepsilon}(T)) - \check{b}(0, z) \\ &= \sum_{n \geq 0} \sum_{m \geq 0} \check{b}(T_{n,m+1} \wedge T, D_{h,\varepsilon}(T_{n,m+1} \wedge T)) - \check{b}(T_{n,m} \wedge T, D_{h,\varepsilon}(T_{n,m} \wedge T)) \\ &\quad + \check{b}(T_{n,m} \wedge T, D_{h,\varepsilon}(T_{n,m} \wedge T)) - \check{b}(T_{n,m} \wedge T, D_{h,\varepsilon}(T_{n,m} \wedge T^-)) \\ &=: B_1 + B_2, \end{aligned}$$

with obvious notations and where B_1 is meant to capture the dynamic between jumps while B_2 captures that at the jumps. The double sum and the $T_{n,m} \wedge T$ are only here to simplify the proof. Note that in order to define the sum for all $m \geq 0$, we set $T_{n,m} = T$ for $m > m_n$, and note also that there is only a finite number of terms in the sums. We are then led to estimate the differences in B_1 and B_2 for which we will use the process $\mathfrak{D}_{h,\varepsilon}$. We introduced \check{b} since $\hat{\mathfrak{K}}_{h,\varepsilon}$ is not necessarily on the sphere between the grid points. Since T is on the grid, we have by definition $D_{h,\varepsilon}(T_{n,m} \wedge T) = \mathfrak{D}_{h,\varepsilon}(T_{n,m} \wedge T)$. Consider now the notation

$$\hat{\mathfrak{D}}_{\varepsilon,h} := (\mathfrak{X}_{\varepsilon,h}, \hat{\mathfrak{K}}_{\varepsilon,h}) \quad \text{with} \quad \hat{\mathfrak{K}}_{\varepsilon,h} := \frac{\hat{\mathfrak{K}}_{\varepsilon,h}}{|\hat{\mathfrak{K}}_{\varepsilon,h}|}.$$

By construction, the process $\hat{\mathfrak{D}}_{h,\varepsilon}$ is continuous at the times $T_{n,m}$ that do not correspond to jump times, so that

$$\check{b}(T_{n,m}, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) = b(T_{n,m}, \hat{\mathfrak{D}}_{h,\varepsilon}(T_{n,m})) = b(T_{n,m}, \hat{\mathfrak{D}}_{h,\varepsilon}(T_{n,m}^-)) = \check{b}(T_{n,m}, \mathfrak{D}_{h,\varepsilon}(T_{n,m}^-))$$

for those $T_{n,m}$. As a consequence, B_2 indeed only accounts for jumps. We will then estimate B_1 using the Itô formula for $\mathfrak{D}_{h,\varepsilon}$ between $T_{n,m}$ and $T_{n,m+1}$, and the properties of Poisson random measures for B_2 . For the latter, we notice that we have by construction $D_{h,\varepsilon}(T_n^-) = \hat{\mathfrak{D}}_{h,\varepsilon}(T_n^-)$. Using then the random Poisson measure P with intensity measure μ introduced in (33) and (34), we can write

$$B_2 = \int_0^T ((\check{b}(t, \mathfrak{D}_{h,\varepsilon}(t^-)) + \bar{R}(\theta, \varphi, \hat{\mathfrak{D}}_{h,\varepsilon}(t^-), v)) - \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t^-))) P(dt, d\theta, d\varphi, dv),$$

so that, together with the fact that $P - \mu$ is a measure-valued martingale, see e.g. [1, Chapter 2],

$$\mathbb{E}_z[B_2] = \sum_n \sum_m \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \tilde{\mathcal{I}}_>^\varepsilon(t) dt \right]$$

with

$$\tilde{\mathcal{I}}_>^\varepsilon(t) = \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_{S_\varepsilon^\xi} \rho(\mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t) \cdot \hat{p})(\check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{p}) - \check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t))) \sigma(d\hat{p}).$$

For B_1 , we have from the Itô formula

$$\mathbb{E}_z[B_1] = \sum_n \sum_m \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) + \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_x \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) + \sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) \mathcal{B}_{n,m}^\varepsilon(t) dt \right]$$

where

$$\mathcal{B}_{n,m}^\varepsilon(t) = \Delta_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \mathfrak{K}_{h,\varepsilon}(T_{n,m})^T D_k^2 \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) \mathfrak{K}_{h,\varepsilon}(T_{n,m}) - 2\mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)).$$

As a result,

$$\begin{aligned} \mathbb{E}_z[B_1 + B_2] &= \sum_n \sum_m \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) + \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_x \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) \right. \\ &\quad \left. + \sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) \mathcal{B}_{n,m}^\varepsilon(t) + \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) dt \right]. \end{aligned}$$

Using again the fact that b satisfies (36), we have

$$\partial_t b(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) + \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_x b(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) + \mathcal{Q}b(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) = 0,$$

which also holds true for \check{b} since the variable $\mathfrak{K}_{h,\varepsilon}(T_{n,m})$ has norm 1 at the grid points $T_{n,m}$. As a consequence,

$$\mathbb{E}_z[b(T, D_{h,\varepsilon}(T))] - b(0, z) = \sum_n \sum_m \sum_{j=1}^7 E_{n,m}^{(j)}$$

where

$$\begin{aligned} E_{n,m}^{(1)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) dt \right] \\ E_{n,m}^{(2)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot (\nabla_x \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \nabla_x \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m}))) dt \right] \\ E_{n,m}^{(3)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) (\Delta_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \Delta_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m}))) dt \right] \\ E_{n,m}^{(4)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} -\sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) \mathfrak{K}_{h,\varepsilon}(T_{n,m})^T (D_k^2 \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - D_k^2 \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m}))) \mathfrak{K}_{h,\varepsilon}(T_{n,m}) dt \right] \\ E_{n,m}^{(5)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} -2\sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot (\nabla_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \nabla_k \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m}))) dt \right] \\ E_{n,m}^{(6)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \sigma_\varepsilon^2(\mathfrak{X}_\varepsilon(T_{n,m})) \Delta_{\mathbb{S}^2} b(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) - \tilde{\mathcal{I}}_{<}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(T_{n,m})) dt \right] \\ E_{n,m}^{(7)} &:= \mathbb{E}_z \left[\int_{T_{n,m} \wedge T}^{T_{n,m+1} \wedge T} \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) - \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(T_{n,m})) dt \right] \end{aligned}$$

with

$$\tilde{\mathcal{I}}_{<}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) = \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_{S_\varepsilon^\alpha} \rho(\mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t) \cdot \hat{p}) (\check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{p}) - \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) \sigma(d\hat{p})). \quad (37)$$

In the estimates below involving \check{b} , derivatives involving $k \mapsto \hat{k} = k/|k|$ produce terms of the form $1/|k|^p$ for some $p > 0$. These terms are due to the fact that $\mathfrak{K}_{n,m}$ does not stay on the sphere at all times. However, these terms can be bounded uniformly thanks to the following lemma.

Lemma 6.3 *We have for any $n \geq 0$ and $m \leq m_n$,*

$$\inf_{s \in [T_{n,m}, T_{n,m} + h_{n,m})} |\mathcal{K}_{n,m}(s)| \geq (1 - 2h\sigma_{\varepsilon,\infty}^2) \geq \frac{1}{2},$$

for h and ε small enough, and

$$\sup_{s \in [T_{n,m}, T_{n,m} + h_{n,m})} \mathbb{E} [|\mathcal{K}_{n,m}(s)|^2 | T_{n,m}, T_{n,m+1}] \leq 1 + 4h\sigma_{\varepsilon,\infty}^2 \leq 2.$$

Proof For $s \in [T_{n,m}, T_{n,m} + h_{n,m}]$, $\mathcal{K}_{n,m}$ can be rewritten as the sum of two orthogonal components

$$\mathcal{K}_{n,m}(s) = (1 - 2(t - T_{n,m})\sigma_\varepsilon^2(X_{n,m}))\hat{K}_{n,m} + \sqrt{2}\sigma_\varepsilon(X_{n,m})\hat{K}_{n,m} \times (W(s) - W(T_{n,m}))$$

so that

$$|\mathcal{K}_{n,m}(s)|^2 \geq (1 - 2h\sigma_{\infty,\varepsilon}^2)^2.$$

Now for the upper bound, using that W_n and P are independent, we have

$$\begin{aligned} \mathbb{E}[|\mathcal{K}_{n,m}(s)|^2 | T_{n,m}, T_{n,m+1}] &\leq (1 + 2h\sigma_{\varepsilon,\infty}^2) + 2\sigma_{\varepsilon,\infty}^2 \mathbb{E}[|W(s) - W(T_{n,m})|^2 | T_{n,m}, T_{n,m+1}] \\ &\leq 1 + 4h\sigma_{\varepsilon,\infty}^2 \\ &\leq 2, \end{aligned}$$

where we used that $4h\sigma_{\varepsilon,\infty}^2 \leq 1$. This concludes the proof. \square

For $E_{n,m}^{(1)}$. Using the Itô formula between $T_{n,m}$ and $T_{n,m+1}$, we find

$$\begin{aligned} \mathbb{E}\left[\partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(t)) - \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(T_{n,m})) \mid T_{n,m}, T_{n,m+1}\right] \\ = \int_{T_{n,m}}^t \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_x \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(s)) \\ + \sigma_\varepsilon^2(\mathfrak{X}_{h,\varepsilon}(T_{n,m})) \left(\Delta_k \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(s)) - \mathfrak{K}_{h,\varepsilon}(T_{n,m})^T D_k^2 \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(s)) \mathfrak{K}_{h,\varepsilon}(T_{n,m}) \right. \\ \left. - 2\mathfrak{K}_{h,\varepsilon}(T_{n,m}) \cdot \nabla_k \partial_t \check{b}(t, \mathfrak{D}_{h,\varepsilon}(s)) \right) ds, \end{aligned}$$

so that

$$|E_{n,m}^{(1)}| \leq h \left(\|D_{t,x}^2 \check{b}\|_\infty + 2\sigma_{\varepsilon,\infty}^2 (\|D_{t,k,k}^3 \check{b}\|_\infty + \|D_{t,k}^2 \check{b}\|_\infty) \right),$$

where $\sigma_{\infty,\varepsilon} = \sup_x \sigma_\varepsilon(x)$. When $\varepsilon \leq \varepsilon_0$, we have $\sigma_{\infty,\varepsilon} \leq \sigma_{\infty,\varepsilon_0}$ since r'_ε is an increasing function of ε . As a result, we obtain

$$\sum_n \sum_m |E_{n,m}^{(1)}| \leq hT \left(\|D_{t,x}^2 \check{b}\|_\infty + 2\sigma_{\infty,\varepsilon_0}^2 (\|D_{t,k,k}^3 \check{b}\|_\infty + \|D_{t,k}^2 \check{b}\|_\infty) \right) =: hF_2(b).$$

For $E_{n,m}^{(j)}$ with $j = 2, \dots, 5$. Following the same lines as above, we have

$$\begin{aligned} \sum_n \sum_m \sum_{j=2}^5 |E_{n,m}^{(j)}| &\leq 2hT \left(\|D_{x,x}^2 \check{b}\|_\infty + 2\sigma_{\infty,\varepsilon_0}^2 (\|D_{x,k,k}^3 \check{b}\|_\infty + \|D_{x,k}^2 \check{b}\|_\infty) \right) \\ &\quad + 4hT \sigma_{\infty,\varepsilon_0}^2 \left(\|D_{x,k}^2 \check{b}\|_\infty + \|D_{x,k,k}^3 \check{b}\|_\infty + 2\sigma_{\infty,\varepsilon_0}^2 (2\|D_{k,k,k}^3 \check{b}\|_\infty + \|D_{k,k,k,k}^4 \check{b}\|_\infty + \|D_{k,k}^2 \check{b}\|_\infty) \right) \\ &=: hF_3(b). \end{aligned}$$

For $E_{n,m}^{(6)}$. For this term, we follow the proof of Lemma 6.1 and find

$$\sum_n \sum_m |E_{n,m}^{(6)}| \leq T \varepsilon^{2-\alpha_M/2} E_\infty(b),$$

where $E_\infty(b)$ is defined as before.

For $E_{n,m}^{(7)}$. Starting from (37), we have

$$\begin{aligned} \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) &= \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_{S_{>}^\varepsilon} \rho(\mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t) \cdot \hat{p}) (\check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{p}) - \check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t))) \sigma(d\hat{p}) \\ &= \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_0^{2\pi} d\varphi \int_{-1}^{1-\varepsilon} ds \rho(\mathfrak{X}_{h,\varepsilon}(t), s) \\ &\quad \times \left(\check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), s\hat{\mathfrak{K}}_{h,\varepsilon}(t) + \sqrt{1-s^2}\mathcal{G}(\varphi, \hat{\mathfrak{K}}_{h,\varepsilon}(t))) - \check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), \hat{\mathfrak{K}}_{h,\varepsilon}(t)) \right) \end{aligned}$$

where the last line is obtained by changing to spherical coordinates with $s = \cos(\theta)$, and

$$\mathcal{G}(\varphi, k) := \cos(\varphi)\hat{k}_1^\perp + \sin(\varphi)\hat{k}_2^\perp.$$

Above, $(\hat{k}_1^\perp, \hat{k}_2^\perp)$ forms an orthonormal basis of the plane \hat{k}^\perp . Note that the choice of $(\hat{k}_1^\perp, \hat{k}_2^\perp)$ does not play any role since the variable φ is integrated. Now, writing

$$\begin{aligned} \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) &= \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_0^{2\pi} d\varphi \int_{-1}^{1-\varepsilon} ds \int_0^1 dv \rho(\mathfrak{X}_{h,\varepsilon}(t), s) \\ &\quad \times ((s-1)\hat{\mathfrak{R}}_{h,\varepsilon}(t) + \sqrt{1-s^2}\mathcal{G}(\varphi, \hat{\mathfrak{R}}_{h,\varepsilon}(t))) \\ &\quad \cdot \nabla_k \check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), (1+v(s-1))\hat{\mathfrak{R}}_{h,\varepsilon}(t) + v\sqrt{1-s^2}\mathcal{G}(\varphi, \hat{\mathfrak{R}}_{h,\varepsilon}(t))), \end{aligned}$$

and using that $\mathcal{G}(\varphi + \pi, k) = -\mathcal{G}(\varphi, k)$, we just have to focus on

$$\begin{aligned} \tilde{\mathcal{I}}_{>}^\varepsilon(\mathfrak{D}_{h,\varepsilon}(t)) &= \frac{\lambda(\mathfrak{X}_{h,\varepsilon}(t))}{2^{1+\alpha(\mathfrak{X}_{h,\varepsilon}(t))/2}} \int_0^{2\pi} d\varphi \int_{-1}^{1-\varepsilon} ds \int_0^1 dv \rho(\mathfrak{X}_{h,\varepsilon}(t), s)(s-1)\hat{\mathfrak{R}}_{h,\varepsilon}(t) \\ &\quad \cdot \nabla_k \check{b}(t, \mathfrak{X}_{h,\varepsilon}(t), (1+v(s-1))\hat{\mathfrak{R}}_{h,\varepsilon}(t) + v\sqrt{1-s^2}\mathcal{G}(\varphi, \hat{\mathfrak{R}}_{h,\varepsilon}(t))). \end{aligned}$$

Before applying the Itô formula to this term, we rewrite \mathcal{G} as

$$\mathcal{G}(\varphi, \hat{k}) = \left(I_3 + \sin(\varphi)Q(\hat{k}) + (1 - \cos(\varphi))Q^2(\hat{k}) \right) \mathcal{H}_1(\hat{k})$$

where I_3 is the 3×3 identity matrix, Q is defined by (24), and where

$$\mathcal{H}_1(\hat{k}) := \frac{1}{\sqrt{\hat{k}_1^2 + \hat{k}_2^2}} \begin{pmatrix} \hat{k}_2 \\ -\hat{k}_1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{k_1^2 + k_2^2}} \begin{pmatrix} k_2 \\ -k_1 \\ 0 \end{pmatrix},$$

which is orthogonal to $\hat{k} = k/|k|$. In fact, $\mathcal{G}(\varphi, \hat{k})$ corresponds to the rotation of $\mathcal{H}_1(\hat{k}) \in \hat{k}^\perp$ with angle φ and axis \hat{k} . This choice simplifies calculations. Now, note that

$$\int_{-1}^{1-\varepsilon} \rho(x, s)(s-1)ds \leq \int_{-1}^1 \rho(x, s)(s-1) \leq \|a\|_\infty \frac{2^{1-\eta_M/2}}{1-\eta_M/2},$$

so that $E_{n,m}^{(7)}$ does not depend on ε . Applying the Itô formula, we obtain

$$\begin{aligned} \sum_n \sum_m |E_{n,m}^{(7)}| &\leq hT(1 + (1 + 2^7 \cdot 3^3)\sigma_{\infty, \varepsilon_0}^2) \|a\|_\infty \frac{2^{1-\eta_M/2}}{1-\eta_M/2} \left(1 + \frac{1}{1-\eta_M/2} \right) \\ &\quad \times (\|\nabla_x \lambda\|_\infty + \|\lambda\|_\infty + \|\nabla_x \alpha\|_\infty) (\|D_k^1 \check{b}\|_\infty + \|D_{k,k}^2 \check{b}\|_\infty) =: hF_4(b) \end{aligned}$$

Setting finally $F_1 := F_2 + F_3 + F_4$ and gathering all previous results concludes the proof of Proposition 6.1. \blacksquare

6.3.3 Step 3 and conclusion

We remark first that the error bound in Proposition 6.1 does not depend on the starting point (x, \hat{k}) . Then, from this *pointwise* result, we find

$$|\mathbb{E}_{\mu_0}[f(D_{h,\varepsilon}(T))] - \mathbb{E}_{\mu_0}[f(\tilde{D}_\varepsilon(T))]| \leq \int_{\mathbb{R}^3 \times \mathbb{S}^2} |\mathbb{E}_{x,\hat{k}}[f(D_{h,\varepsilon}(T))] - \mathbb{E}_{x,\hat{k}}[f(\tilde{D}_\varepsilon(T))]| \mu_0(dx, d\hat{k}),$$

where μ_0 is the probability measure given by (19). Let now

$$\mu(t, f) = \int_{\mathbb{R}^3 \times \mathbb{S}^2} u(t, x, \hat{k}) f(x, \hat{k}) dx \sigma(d\hat{k}) \quad \text{and} \quad \mu_{h,\varepsilon}(t, f) = \mathbb{E}_{\mu_0}[f(D_{h,\varepsilon}(t))].$$

Using Propositions 2.1 and 6.1, and that $\int_{\mathbb{R}^3 \times \mathbb{S}^2} u_\varepsilon(T, x, \hat{k}) f(x, \hat{k}) dx \sigma(d\hat{k}) = \mathbb{E}_{\mu_0}[f(D_\varepsilon(T))] = \mathbb{E}_{\mu_0}[f(\tilde{D}_\varepsilon(T))]$ according to Lemma 6.2, we have

$$|\mu_{h,\varepsilon}(T, f) - \mu(T, f)| \leq \varepsilon^{2-\alpha_M/2} \sqrt{2TE}(u) \|f\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} + \varepsilon^{2-\alpha_M/2} 2TE_\infty(b) + hF_1(b).$$

In order to end the proof of Theorem 3.1, it suffices to remark now that

$$|\mu_{N,h,\varepsilon}(T, f) - \mu(T, f)| \leq |\mu_{N,h,\varepsilon}(T, f) - \mu_{h,\varepsilon}(T, f)| + |\mu_{h,\varepsilon}(T, f) - \mu(T, f)|,$$

so that

$$\begin{aligned} & \mathbb{P}\left(|\mu_{N,h,\varepsilon}(T, f) - \mu(T, f)| > \frac{\eta \Sigma_{h,\varepsilon}}{\sqrt{N}} + \varepsilon^{2-\alpha_M/2} F_0(u, b, f) + \varepsilon^{2-\alpha_M/2} 2TE_\infty(b) + h F_1(b)\right) \\ & \leq \mathbb{P}\left(|\mu_{N,h,\varepsilon}(T, f) - \mu_{h,\varepsilon}(T, f)| > \frac{\eta \Sigma_{h,\varepsilon}}{\sqrt{N}} + \underbrace{\varepsilon^{2-\alpha_M/2} F_0(u, b, f) + h F_1(b)}_{\geq 0} - |\mu_{h,\varepsilon}(T, f) - \mu(T, f)|\right) \\ & \leq \mathbb{P}\left(|\mu_{N,h,\varepsilon}(T, f) - \mu_{h,\varepsilon}(T, f)| > \frac{\eta \Sigma_{h,\varepsilon}}{\sqrt{N}}\right) \end{aligned}$$

where

$$F_0(u, b, f) := \sqrt{2TE}(u) \|f\|_{L^2(\mathbb{R}^3 \times \mathbb{S}^2)} + 2TE_\infty(u).$$

We conclude by applying the central limit theorem [13] together with the Portmanteau theorem [5, Theorem 2.1 pp.16]. ■

7 Conclusion

We have derived an efficient MC method for the resolution of the RTE with non-integrable scattering kernels. It is based on a small jumps/large jumps decomposition that allows us to simulate the small jumps part at a low cost by solving a standard SDE. The large jumps are obtained by using the stochastic collocation technique with a candidate distribution function that captures the singular behavior of the kernel. We have moreover demonstrated the necessity to include the small jumps component in order to obtain a good accuracy at a manageable computational cost, and investigated practical situations in optical tomography and atmospheric turbulence where the singular RTE is of interest. We in particular highlighted the role of the singularity strength α on the qualitative behavior of the solution.

Future investigations include the estimation of the scattering kernel, with an emphasis on the parameter α , from either simulated or experimental data obtained e.g. from light propagation in biological tissues. This problem is of practical interest in biomedical applications and will require the development of appropriate inverse techniques.

A Stochastic collocation

In this section, we describe the stochastic collocation method, see e.g. [20], and consider the situation of Section 5.2.2 as an illustration. The goal is to simulate a real-valued random variable W (for which direct simulation is not possible or too costly) from an auxiliary variable V that can be generated efficiently. In our context, we want to simulate W with probability density function (PDF)

$$f_W(w) := \frac{a(\sqrt{2w})}{C_W w^{1+\alpha/2}} \mathbf{1}_{(\varepsilon, 2)}(w),$$

where C_W is a normalization constant. As already noticed in Section 3.4, a direct method is available when $a \equiv 1$. Therefore, we take V with PDF

$$f_V(v) := \frac{1}{C_V v^{1+\alpha/2}} \mathbf{1}_{(\varepsilon, 2)}(v),$$

that can be simulated with

$$V = F_V^{-1}(U) = \varepsilon(1 - (1 - (\varepsilon/2)^{\alpha/2})U)^{-2/\alpha},$$

where $U \sim \mathcal{U}(0, 1)$ and where F_V is the cumulative distribution function (CDF) of V . The stochastic collocation method is based on the following three observations. First, we have $F_V(V) \sim \mathcal{U}(0, 1)$. Second, denoting by F_W the CDF of W , we note that W can be (theoretically) simulated with

$$F_W^{-1}(U) = F_W^{-1}(F_V(V)) =: G(V),$$

with $G = F_W^{-1} \circ F_V$ and $U = F_V(V)$. Last, we only need to approximate G and not F_W^{-1} , and with a good candidate V , G behaves better than F_W^{-1} . In order to approximate G , we use Gauss polynomial interpolation and only need to invert F_W at a small number of points.

In our example, V captures the "singular" behavior of W , and is as a consequence a good candidate. The function G is then direct to approximate with just a few quadrature points for a reduced computational cost. Because G needs only to be approximated over $(\varepsilon, 2)$, with known values at the extremes, we rather use a Gauss-Lobatto-Jacobi quadrature rule. In Figure 18, we illustrate the polynomial approximation of G , with 5 and 10 interpolation points for $\alpha = 5/3$, $\varepsilon = 0.01$, and $a(r) = \exp(-r^2/(2 \times 0.8^2))$. Because of our choice for V , one can observe that the overall behavior of the PDF f_W is well captured with just 5 quadrature points, even for strongly singular kernels with $\alpha = 5/3$. However, the fast decay of the function a , which is the main source of error between F_W and F_V , requires more quadrature points for an accurate approximation and 10 points seem sufficient.

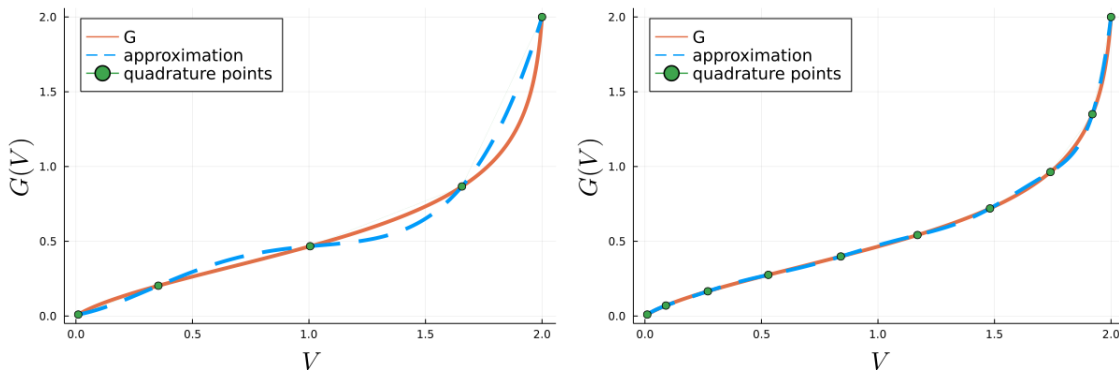


Figure 18: Illustration of the polynomial approximation to G with 5 (left picture) and 10 (right picture) interpolations points. We use the library Jacobi.jl to compute these quadrature points.

References

- [1] D. APPLEBAUM, *Lévy processes and stochastic calculus*, Cambridge University Press, New York, 2004
- [2] S. ASMUSSEN AND J. ROSIŃSKI, *Approximations of small jumps of lévy processes with a view towards simulation*, J. Appl. Probab., 38 (2001), pp. 482–493.
- [3] M.S. BELEN’KII, S.J. KARIS, C.L. OSMON, J.M. BROWN II, AND R. Q. FUGATE, *Experimental evidence of the effects of non-Kolmogorov turbulence and anisotropy of turbulence*, Proc. SPIE 3749, 18th Congress of the International Commission for Optics (1999).
- [4] M. VAN DEN BERG AND J.T. LEWIS, *Brownian motion on hypersurface*, Bulletin of the London Mathematical Society, 17 (1985), pp. 144–150.
- [5] P. BILLINGSLEY, *Convergence of probability measures*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, 1999.
- [6] X.BLANC, C.BORDIN, G.KLUTH AND G.SAMBA, *Variance reduction method for particle transport equation in spherical geometry*, Journal of Computational Physics, 364 (2018), pp. 274–297.
- [7] T. BESARD, C. FOKET, AND B. DE SUTTER, *Effective Extensible Programming: Unleashing Julia on GPUs*, IEEE Transactions on Parallel and Distributed Systems, 30 (2019), pp. 827–841.
- [8] T. BESARD, V. CHURAVY, A. EDELMAN, AND B. DE SUTTER, *Rapid software prototyping for heterogeneous and distributed platforms*, Advances in Engineering Software, 132 (2019), pp. 29–46.
- [9] S. CHANDRASEKHAR, *Radiative transfer*, Dover Publications, New York, 1960.
- [10] S. COHEN AND J. ROSIŃSKI, *Gaussian approximation of multivariate lévy processes with applications to simulation of tempered stable processes*, Bernoulli, 13 (2007), pp. 195–210.
- [11] M. DEHAES, L. GAGNON, A. VIGNAUD, R. VALABR, R. GREBE, F. WALLOIS, AND H. BENALI, *Quantitative investigation of the effect of the extra-cerebral vasculature in diffuse optical imaging : A simulation study*, Biomed. Opt. Express (2011), pp. 680–695.
- [12] S.N. ETHIER AND T.G KURTZ, *Markov processes characterization and convergence*, Wiley series in probability and mathematical statistics, 1986.

- [13] W. FELLER, *An Introduction to Probability Theory and Its Applications, Volume II*, 2nd Edition, WILEY, 1991.
- [14] H. FUJII, S. OKAWA, Y. YAMADA, Y. HOSHI AND M. WATANADE, *Renormalization of the highly forward-peaked phase function using the double exponential formula for radiative transfer*, J. Math. Chem., 54 (2016), pp. 2048–2061.
- [15] H. GAO AND H. ZHAO, *A fast-forward solver of radiative transfer equation*, Transport Theory and Statistical Physics, 38(3) (2009), pp. 149–192.
- [16] C.T. GERMER, A. ROGGAN, J.P. RITZ, C. ISBERT, D. ALBRECHT, G. MÜLLER, AND J.H. BUHR, *Optical properties of native and coagulated human liver tissue and liver metastases in the near infrared range*, Lasers Surg. Med.(1998), pp. 194–203.
- [17] C. GOMEZ, O. PINAUD, AND L. RYZHIK, *Hypoelliptic estimates in radiative transfer*, Commun. Part. Diff. Eq, 41 (2015), pp. 150–184.
- [18] C. GOMEZ, O. PINAUD, AND L. RYZHIK, *Radiative transfer with long-range interactions regularity and asymptotics*, SIAM Multiscale Model. Simul., 15 (2017), pp. 1048–1072.
- [19] C. GOMEZ AND O. PINAUD, *Monte Carlo methods for radiative transfer with singular kernels*, SIAM J. Sci. Comput., 40 (2018), pp.A1714–A1741.
- [20] L. A. GRZELAK, J. WITTEVEEN, M. SUAREZ-TABOADA, AND C. W. OOSTERLEE, *The stochastic collocation monte carlo sampler: Highly efficient sampling from "expensive" distributions*, Quantitative Finance, 2018, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=2529691> or <http://dx.doi.org/10.2139/ssrn.2529691>
- [21] L.G. HENYEY AND J. GREENSTEIN, *Diffuse radiation in the galaxy*, Astrophys. J. (1941), pp. 70–83.
- [22] S.L. JACQUES, *Properties of Biological Tissues: A Review*, Phys. Med. Biol. 58 (2013), R37.
- [23] J.H JEANS, *Stars, Gaseous, radiative transfer of energy*, Monthly Notices of the Royal Astronomical Society, 78 (1917), pp. 28–36
- [24] A.D. KIM, *Transport theory for light propagation in biological tissue*, J. Opt. Soc. Am. 21 (2004), pp. 820–827.
- [25] A.D. KIM AND M. MOSCOSO, *Beam propagation in sharply peaked forward scattering media*, J. Opt. Soc. Am. 21 (2004), pp. 797–803.
- [26] O. KOROTKOVA AND I. TOSELLI, *Non-Classic Atmospheric Optical Turbulence: Review*, Appl. Sci., 11 (2021), 8487.
- [27] B. LAPEYRE, E. PARDOUX AND R. SENTIS, *Introduction to Monte-Carlo Methods for Transport and Diffusion Equations*, Oxford University Press (2003).
- [28] C.L. LEAKEAS AND E.W. LARSEN, *Generalized Fokker-Planck approximations of particle transport with highly forward-peak scattering*, Nucl. Sci. Eng., 137 (2001), pp. 236–250.
- [29] H. LOUVIN, E. DUMONTEIL, T. LELIÈVRE, M. ROUSSET, AND C.M. DIOP, *Adaptive multilevel splitting for Monte Carlo particle transport*, EPJ Nuclear Sci. Technol., 3 (2017).
- [30] I. LUX AND LÁSZLÓ KOBLINGER, *Monte Carlo particle transport methods : neutron and photon calculations*, CRC Press, 1991.
- [31] L. MARGERIN, M. CAMPILLO, AND B.A. VAN TIGGELEN, *Monte Carlo simulation of multiple scattering of elastic waves*, J. Geophys. Res., 105 (2000), pp. 7873–7892.
- [32] L. MARGERIN AND G. NOLET, *Multiple scattering of high-frequency seismic waves in the deep Earth: modeling and numerical examples*, J. Geophys. Res., 108 (2003), pp. 2234–2251.
- [33] L. MARGERIN, A. BAJARAS, AND M. CAMPILLO, *A scalar radiative transfer model including the coupling between surface and body waves*, Geophys. J. Int., 219 (2019), pp. 1092–1108.
- [34] U.M. NOEBAUER AND S.A. SIM, *Monte Carlo radiative transfer*, Living Reviews in Computational Astrophysics, Springer, 2019.
- [35] S. POWELL, B.T. COX, AND S.R. ARRIDGE, *A pseudospectral method for solution of the radiative transport equation*, Journal of Computational Physics, 384 (2019), pp. 376–382.
- [36] J. PRZYBILLA AND M. KORN, *Monte Carlo simulation of radiative energy transfer in continuous elastic random media - three-component envelopes and numerical validation*, Geophys. J. Int. (2008), pp. 566–576.
- [37] L. REYNOLDS AND N. MCCORMICK, *Approximate two-parameter phase function for light scattering*, J. Opt. Soc. Am. 70 (1980), pp. 1206–1212
- [38] L. RYZHIK, G. PAPANICOLAOU, AND J. B. KELLER, *Transport equations for elastic and other waves in random media*, Wave Motion (1996), pp. 327–370.
- [39] S. SAMKO, *On inversion of fractional spherical potentials by spherical hypersingular operators*, in Singular integral operators, factorization and applications, vol. 142 of Oper. Theory Adv. Appl., Birkhäuser, Basel, 2003, pp. 357–368.
- [40] J. SPANIER AND E.M. GELBARD, *Monte Carlo principles and neutron transport problems*, Dover Publications, New York, 2008.

- [41] B.E. STRIBLING, B.M. WELSH, AND M.C. ROGGEMANN, *Optical propagation in non-Kolmogorov atmospheric turbulence*, Proc. SPIE 2471, Atmospheric Propagation and Remote Sensing IV (1995).
- [42] D. TALAY AND L. TUBARO, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Analysis and Applications, 8 (1991), pp. 94–120.
- [43] R. VISKANTA AND M. MENGÜÇ, *Radiation heat transfer in combustion systems*, Prog. Energy Combust. Sci. 13 (1987), pp. 97–160.
- [44] A. ZILBERMAN, E. GOLBRAIKH, N.S. KOPEIKA, A. VIRTSER, I. KUPERSHMIDT, Y. SHTEMLER, *Lidar study of aerosol turbulence characteristics in the troposphere: Kolmogorov and non-Kolmogorov turbulence*, Atmospheric Research, 88 (2008), pp. 66-77.