A numerical approach for a system of transport equations in the field of radiotherapy

Teddy Pichard [1,*], Stéphane Brull[2], Bruno Dubroca[3]

[1] *Sorbonne Université, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, 4 place Jussieu, F-75005, Paris, France.*
[2] *INP Bordeaux, UMR 5251, Institut de Mathématiques de Bordeaux, 351 cours de la libération, F-33400, Talence, France.*
[3] *Université de Bordeaux, UMR 5801, Laboratoire des Composites Thermostructuraux, 3 allée de la Boétie, F-33600, Pessac, France.*

Abstract. Numerical schemes for systems of transport equations are commonly constrained by a stability condition of Courant-Friedrichs-Lewy (CFL) type. We consider a system modeling the steady transport of photons and electrons in the field of radiotherapy. Naive discretizations of such a system are commonly constrained by a very restrictive CFL condition. This issue is circumvented by constructing an implicit scheme based on a relaxation approach.

We use an entropy-based moment model, namely the M_1 model. Such a system of equations possesses the non-linear flux terms of a hyperbolic system but no time derivative. The flux terms are well-defined only under a condition on the unknowns, called realizability, which corresponds to the positivity of an underlying kinetic distribution function.

The present numerical approach is applicable to non-linear systems which possess no hyperbolic operator, and it preserves the realizability property. However, the discrete equations are non-linear, and we propose a numerical method to solve such non-linear systems.

Our approach is tested on academic and practical cases in 1D, 2D, and 3D and it is shown to require significantly less computational power than reference methods.

AMS subject classifications: 35A35, 65M22, 35L65, 82C40

Key words: Implicit scheme, Relaxation scheme, M_1 model, Radiotherapy dose computation

1 Introduction

The present work aims to construct a numerical solver for systems of steady transport equations emerging in the field of radiotherapy. It is a follow-up to [5, 23, 54] and it analyses the numerical methods used [6, 14, 50–53, 55].

[*] Corresponding author. *Email addresses:* pichard@ljll.math.upmc.fr (T. Pichard)

The motion of energetic particles in radiotherapy can be modeled by a system of coupled linear kinetic equations over the fluences of the particles, *i.e.* over the densities, or distribution functions, of the particles in a phase space composed of position $x \in \mathbb{R}^3$, energy $\epsilon \in \mathbb{R}^+$, and direction of flight $\Omega \in \mathcal{S}^2$ on the unit sphere. Due to the high dimensionality of the phase space, solving directly such systems of equations, through either Monte Carlo methods ([15, 34, 37]) or discrete ordinates methods ([43]; see also [44] and references therein for a review on numerical approaches for dose computation) commonly requires much higher numerical powers than the standard available in medical centers. Recent technological advances lead to the development of industrial codes based on those methods which require considerably lower computational power, *i.e.* the so-called fast Monte Carlo methods (see *e.g.* [65]) and Acuros$^{®}$ code ([26,48,63]).

As an alternative, in this paper, we use an angular moment extraction technique. The resulting system is under-determined, and we use an entropy minimization procedure, leading to the so-called M_1 model. We chose such a closure because it is known to preserve the main features of the underlying kinetic model (especially positivity, hyperbolicity, and entropy dissipation), and it models accurately beams of particles. This method is widely used for diverse applications in physics and biology *e.g.* in astrophysics ([16,17,30]), radiative transfer ([21,56]), in fluid dynamics ([29,42,45]), for semiconductors ([31,57]) or chemotaxis ([7]) modeling, and showed a considerable reduction of the numerical costs.

Numerical approaches for solving moment equations are typically constrained by a stability condition. Such a condition becomes very restrictive when considering low density media. Typically, the step size (see [5, 54] or Section 3 below) needs to be taken proportional to the minimum density in the medium and therefore many steps are required. This problem was first studied for application in radiotherapy in [5] and it was circumvented by the use of a clever change of variables. The previous work [54] showed another approach based on a relaxation method (based on [1,11,47], see also recent work [18]) and on the method of characteristics. However, both those approaches are inappropriate to model the motion of photons. Indeed, those numerical schemes are applicable only to hyperbolic systems, but the transport of photons is ill-modeled by such equations ([49]).

In the present paper, we present an implicit scheme based on a relaxation method, preserving the realizability property and efficient with large steps. However, the discretized equations are non-linear, and we construct an iterative solver to solve such equations.

The paper is constructed as follows. In the next section, models of transport of photons and electrons are presented, first a kinetic model, then the angular moment extraction is described. A first numerical scheme is described for 1D problems in Section 3, an iterative algorithm adapted to this scheme is constructed and tested on an academic test case. This numerical scheme is completed and adapted to multi-D problems in Section 4 and tested on academic test cases in 2D and 3D. Section 5 is devoted to conclusion.

2 Models of transport of photons and electrons

Photons and electrons are characterized their position $x \in \mathbb{R}^3$, their energy $\epsilon \in \mathbb{R}^+$ and their direction of flight $\Omega \in \mathcal{S}^2$ on the unit sphere. We assume that the transported particles interact only with the atoms of the background medium, and we neglect the influence of such interactions on the medium. In particular, the density of atoms in the medium is given constant data.

2.1 A kinetic model

The motion of transported photons γ and electrons e is modeled by their fluence ψ_γ and ψ_e, *i.e.* their density in the (x,ϵ,Ω) space. The fluences satisfy the following steady kinetic equations (see *e.g.* [25,34])

$$\begin{aligned} \Omega.\nabla_x \psi_\gamma(x,\epsilon,\Omega) &= \rho(x)\left[Q_{\gamma\to\gamma}(\psi_\gamma)(x,\epsilon,\Omega) + Q_{e\to\gamma}(\psi_e)(x,\epsilon,\Omega)\right], & (2.1a)\\ \Omega.\nabla_x \psi_e(x,\epsilon,\Omega) &= \rho(x)\left[Q_{e\to e}(\psi_e)(x,\epsilon,\Omega) + Q_{\gamma\to e}(\psi_\gamma)(x,\epsilon,\Omega)\right], & (2.1b)\end{aligned}$$

composed of time-independent free transport terms on the left-hand side and collisions operators on the right-hand side. The collision operator $Q_{\alpha\to\beta}$ models the variations of the fluence ψ_β due to the collisions involving incident particles α. As a first approximation, we neglect the influence of the composition of the medium on the collisions. The collision operator is only proportional to the relative density $\rho > 0$ compared to the density of water.

We consider collision operators of the form

$$\begin{aligned} Q_{\gamma\to\gamma}(\psi_\gamma) &= [G_{\gamma\to\gamma} - P_\gamma](\psi_\gamma), & (2.2a)\\ Q_{\gamma\to e}(\psi_\gamma) &= G_{\gamma\to e}(\psi_\gamma), & (2.2b)\\ Q_{e\to\gamma}(\psi_e) &= 0, & (2.2c)\\ Q_{e\to e}(\psi_e) &= \partial_\epsilon(S\psi_e) + [G_{e\to e} - P_e](\psi_e), & (2.2d)\end{aligned}$$

where the terms $G_{\alpha\to\beta}$ and P_β are linear Boltzmann gain and loss terms and are given by

$$\begin{aligned} G_{\alpha\to\beta}(\psi_\alpha)(x,\epsilon,\Omega) &= \int_\epsilon^{\epsilon_{\max}} \int_{\mathcal{S}^2} \sigma_{\alpha\to\beta}(\epsilon',\epsilon,\Omega'.\Omega)\psi_\alpha(x,\epsilon',\Omega')d\epsilon' d\Omega', & (2.2e)\\ P_\beta(\psi_\beta)(x,\epsilon,\Omega) &= \sigma_{T,\beta}(\epsilon)\psi_\beta(x,\epsilon,\Omega). & (2.2f)\end{aligned}$$

The stopping power $S > 0$, the differential cross sections $\sigma_{\alpha\to\beta} \geq 0$ and the total cross sections $\sigma_{T,\beta} > 0$ are *a priori* given functions characterizing the collisions in a medium. As we assume that the composition of the medium do not influence the collision behaviors, the coefficients S, $\sigma_{T,\beta}$ and $\sigma_{\alpha\to\beta}$ do not depend on position x. Only the density ρ in front of the collision operators in (2.1) depends on x. We chose those collision operators (2.2) because they accurately model Compton, Møller, and elastic nuclear scattering (see *e.g.* [25,34]) which are the predominant effects at the considered ranges of energy.

Remark 2.1. The computation of the integrals in the collision operators (2.2e) at energy ϵ requires the value of the fluences ψ_γ and ψ_e at all energy $\epsilon' \in [\epsilon,\epsilon_{max}]$. In practice, we solve the equations from a maximum energy ϵ_{max} to a minimum energy ϵ_{min}.

This is relevant for well-posedness consideration. Indeed, under additional conditions on the physical parameters S, $\sigma_{\alpha\to\beta}$ and $\sigma_{T,\beta}$, the system (2.1) with an initial condition at $\epsilon_{\max}$ and appropriate boundary conditions (incoming ones) is well-posed (see *e.g.* [20,60]).

This is also relevant to the considered physics, the particles are injected with a certain energy (given data) which is deposited in the medium. Thus, their energy only decreases.

In medical physics, the function of interest is the energy deposited by the particles per mass unit, so-called dose. At the kinetic level, this dose D is simply given by

$$D(x)=\int_{\epsilon_{\min}}^{\epsilon_{\max}}\int_{S^2}(-\epsilon)\sum_{\alpha,\beta=\gamma,e} Q_{\alpha\to\beta}(\psi_\alpha)(x,\epsilon,\Omega)d\Omega d\epsilon. \tag{2.3}$$

2.2 A moment model

Solving numerically kinetic equations of the form (2.1-2.2) requires high computational power. Instead, we use the method of moments as it requires lower computational power (see *e.g.* comparisons in the previous work [51,53,54]).

In the following, the construction of the M_1 model associated to the kinetic model (2.1-2.2) is recalled.

The moments ψ^0, ψ^1 and ψ^2 of a fluence ψ are

$$\begin{gathered}\psi^0(x,\epsilon)=\int_{S^2}\psi(x,\epsilon,\Omega)d\Omega, \qquad \psi^1(x,\epsilon)=\int_{S^2}\Omega\psi(x,\epsilon,\Omega)d\Omega, \\ \psi^2(x,\epsilon)=\int_{S^2}\Omega\otimes\Omega\psi(x,\epsilon,\Omega)d\Omega.\end{gathered} \tag{2.4}$$

Extracting the moments of (2.1) up to order 1 yields, for $i=0,1$,

$$\nabla_x.\psi_\gamma^{i+1}(x,\epsilon) = \rho(x)\left[Q^i_{\gamma\to\gamma}(\psi^i_\gamma)+Q^i_{e\to\gamma}(\psi^i_e)\right](x,\epsilon), \tag{2.5a}$$

$$\nabla_x.\psi_e^{i+1}(x,\epsilon) = \rho(x)\left[Q^i_{e\to e}(\psi^i_e)+Q^i_{\gamma\to e}(\psi^i_\gamma)\right](x,\epsilon), \tag{2.5b}$$

and the moments of the collision operators of order i are

$$Q^i_{\gamma\to\gamma}(\psi_\gamma) = \left[G^i_{\gamma\to\gamma}-P^i_\gamma\right](\psi^i_\gamma), \tag{2.5c}$$

$$Q^i_{\gamma\to e}(\psi_\gamma) = G^i_{\gamma\to e}(\psi^i_\gamma), \tag{2.5d}$$

$$Q^i_{e\to\gamma}(\psi_e) = 0, \tag{2.5e}$$

$$Q^i_{e\to e}(\psi_e) = \partial_\epsilon(S\psi^i_e)+\left[G^i_{e\to e}-P^i_e\right](\psi^i_e), \tag{2.5f}$$

where the terms $G^i_{\alpha\to\beta}$ and P^i_β read

$$\begin{aligned} G^i_{\alpha\to\beta}(\psi^i_\alpha)(x,\epsilon) &= \int_\epsilon^{\epsilon_{\max}} \sigma^i_{\alpha\to\beta}(\epsilon',\epsilon)\psi^i_\alpha(x,\epsilon')d\epsilon', && (2.5g)\\ P^i_\beta(\psi^i_\beta)(x,\epsilon) &= \sigma_{T,\beta}(\epsilon)\psi^i_\beta(x,\epsilon), && (2.5h)\\ \sigma^i_{\alpha\to\beta}(\epsilon',\epsilon) &= 2\pi\int_{-1}^{+1}\mu^i\sigma_{\alpha\to\beta}(\epsilon',\epsilon,\mu)d\mu. && (2.5i)\end{aligned}$$

The system (2.5) is under-determined, and we add one more relation, so-called M_1 closure ([42,46]), to close the system. This relation expresses ψ^2_α as a function of ψ^0_α and ψ^1_α for $\alpha=\gamma,e$. The M_1 closure consists in reconstructing the unique function ([8–10,32, 33,42,59]) of the form

$$\psi_{M_1}(\Omega)=\exp(\boldsymbol{\lambda}.\mathbf{m}(\Omega)), \qquad \mathbf{m}(\Omega)=(1,\,\Omega_1,\,\Omega_2,\,\Omega_3), \qquad \boldsymbol{\lambda}=(\lambda_0,\,\lambda_1,\,\lambda_2,\,\lambda_3), \quad (2.6)$$

i.e. by computing the unique vector $\boldsymbol{\lambda}\in\mathbb{R}^4$ such that

$$\int_{\mathcal{S}^2}\psi_{M_1}(\Omega)d\Omega=\psi^0, \qquad \int_{\mathcal{S}^2}\Omega\psi_{M_1}(\Omega)d\Omega=\psi^1. \quad (2.7)$$

Then, the last moment ψ^2 is expressed as

$$\psi^2=\int_{\mathcal{S}^2}\Omega\otimes\Omega\psi_{M_1}(\Omega)d\Omega. \quad (2.8)$$

Among the positive functions satisfying (2.7), the reconstruction (2.6) is the one minimizing Boltzmann entropy

$$\mathcal{H}(f)=\int_{\mathcal{S}^2}(f\log f-f)(\Omega)d\Omega.$$

This choice of closure is often used because it provides desirable properties in mathematics and in physics. Indeed, with such a closure, the flux of system (2.5) is the one of a symmetric hyperbolic system ([27]), this system has a positive kinetic reconstruction and it dissipates an entropy ([42]).

Using (2.6), the closure (2.8) can be rewritten (see *e.g.* [41,53])

$$\begin{aligned}\psi^2 &= \psi^0\left(\frac{1-\chi}{2}Id+\frac{3\chi-1}{2}\frac{\psi^1\otimes\psi^1}{|\psi^1|^2}\right), && (2.9)\\ \chi &= 1+\frac{2}{|\alpha|}(1+\coth(|\alpha|)), \qquad \alpha=\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2},\end{aligned}$$

where χ is a scalar called Eddington factor depending on one unique scalar α. Using (2.6) and (2.7), one computes

$$\frac{|\psi^1|}{\psi^0}=\frac{|\alpha|\coth(|\alpha|)-1}{|\alpha|}.$$

which can be inverted to compute α and therefore to obtain the closure (2.9).

Definition 2.1. The M_1 closure is defined only if there exists a function of the form (2.6) satisfying (2.7). Let us define the following set $\mathcal{R}$ called the realizability domain ([36])

$$(\psi^0,\psi^1)\in\mathcal{R} \quad = \quad \Big\{(f^0,f^1)\in\mathbb{R}^4, \quad \text{s.t.} \quad |f^1|\leq f^0\Big\}\cup(0,0_{\mathbb{R}^3}). \tag{2.10}$$

The realizability domain $\mathcal{R}$ is the closure set in $\mathbb{R}^4$ of the moments of the functions of the form (2.6). The moments of all non-negative integrable functions ([8–10]) and measures ([19,36]) belong to the set $\mathcal{R}$.

The requirement (2.10) needs to be kept in mind when constructing numerical schemes because the flux in (2.5) becomes ill-defined if this property is violated. In particular, in the next section, we use the following remark to prove that a numerical scheme preserves the realizability property.

Remark 2.2. The realizability domain is a convex cone. Therefore, any scheme constructed based on a positive combination of realizable vector preserves the realizability property.

Since the differential cross sections $\sigma_{\alpha\to\beta}$ are also positive scalars, their moments need to satisfy the following realizability condition ([36])

$$|\sigma^1_{\alpha\to\beta}|\leq\sigma^0_{\alpha\to\beta}. \tag{2.11}$$

3 A discretization for 1D problems

We aim to construct a solver for the moment system (2.5).

For the sake of simplicity, we describe the numerical approach for problems in one spatial dimension. The results are generalized to multidimensional problems in the next section.

In the first subsection, we present the main problem with the discretization of (2.5) and we describe and test a first numerical method in the remaining subsections.

3.1 The problem with fast characteristics in 1D

First, we rewrite the problem (2.5), then we present the main difficulty with the discretization of the rewritten equations.

3.1.1 Problem settings

In 1D, the system (2.5) can be rewritten under the vectorial form

$$\partial_x\mathbf{F}(\boldsymbol{\psi})(x,\epsilon) \quad = \quad \rho(x)\mathbf{Q}(\boldsymbol{\psi})(x,\epsilon), \tag{3.1a}$$

where the unknown is $\boldsymbol{\psi} = (\boldsymbol{\psi}_\gamma, \boldsymbol{\psi}_e) \in \mathcal{R}^2$, and $\boldsymbol{\psi}_\alpha = (\psi_\alpha^0, \psi_\alpha^1) \in \mathcal{R}$ are the moments of the fluence of particles $\alpha = \gamma, e$. The fluxes $\mathbf{F}(\boldsymbol{\psi})$ and the collision operator $\mathbf{Q}(\boldsymbol{\psi})$ are defined over $\mathcal{R}^2$ by

$$\mathbf{F}(\boldsymbol{\psi}) = \Big(\mathbf{F}_{M_1}(\boldsymbol{\psi}_\gamma), \quad \mathbf{F}_{M_1}(\boldsymbol{\psi}_e)\Big), \tag{3.1b}$$

$$\mathbf{Q}(\boldsymbol{\psi}) = \Big(\mathbf{Q}_{\gamma\to\gamma}(\boldsymbol{\psi}_\gamma) + \mathbf{Q}_{e\to\gamma}(\boldsymbol{\psi}_e), \quad \mathbf{Q}_{e\to e}(\boldsymbol{\psi}_e) + \mathbf{Q}_{\gamma\to e}(\boldsymbol{\psi}_\gamma)\Big), \tag{3.1c}$$

and are composed of the moments of the kinetic flux and collision operator

$$\mathbf{F}_{M_1}(\boldsymbol{\psi}_\alpha) = (\psi_\alpha^1, \quad \psi_\alpha^2), \tag{3.1d}$$

$$\mathbf{Q}_{\alpha\to\beta}(\boldsymbol{\psi}_\alpha) = \Big(Q_{\alpha\to\beta}^0(\psi_\alpha^0), \quad Q_{\alpha\to\beta}^1(\psi_\alpha^1)\Big), \tag{3.1e}$$

where ψ_α^2 is given by the closure relation (2.8). Before describing the numerical approaches, we introduce the following notations.

Notation 1. The superscript n refers to the discretization in energy ϵ and the subscript l to the discretization in the x variable. In the next section, the subscript m will refer to the discretization in the second space variable y.

According to Remark 2.1, the energy grid is such that $\epsilon^n > \epsilon^{n+1}$.

3.1.2 Position of the problem

Standard methods to solve (3.1a) create stiff terms at the discrete level, and are therefore very time-consuming for practical applications in medical physics. Such stiffness arises in weakly collisional media, *e.g.* when the background medium has a low density ρ. This problem was illustrated in [5,54] through a 1D electron transport equation of the form

$$\partial_x \mathbf{F}_{M_1}(\boldsymbol{\psi}_e) = \rho[\partial_\epsilon(S\boldsymbol{\psi}_e) + A\boldsymbol{\psi}_e], \qquad A = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}, \tag{3.2}$$

with $T \in \mathbb{R}^+$ by using the scheme

$$\frac{\mathbf{F}_{e,l+\frac{1}{2}}^n - \mathbf{F}_{e,l-\frac{1}{2}}^n}{\Delta x} - \rho_l \frac{S^n \boldsymbol{\psi}_{e,l}^n - S^{n+1}\boldsymbol{\psi}_{e,l}^{n+1}}{\Delta\epsilon^n} = \rho_l A^n \boldsymbol{\psi}_{e,l}^n, \tag{3.3a}$$

with numerical fluxes of Lax-Friedrichs type ([40,61])

$$\mathbf{F}_{e,l+\frac{1}{2}}^n = \frac{1}{2}\Big[\mathbf{F}_{M_1}(\boldsymbol{\psi}_{e,l+1}^n) + \mathbf{F}_{M_1}(\boldsymbol{\psi}_{e,l}^n) + (\boldsymbol{\psi}_{e,l+1}^n - \boldsymbol{\psi}_{e,l}^n)\Big]. \tag{3.3b}$$

Such a scheme is consistent with (3.2), however it is only stable under the following Courant-Friedrichs-Lewy (CFL) condition ([40,54,61])

$$\Delta\epsilon^n \leq S^n \Delta x \min_l(\rho_l). \tag{3.4}$$

In the domain of approximate Riemann solvers, (3.4) corresponds to imposing that the waves emerging from two different interfaces do not cross each other. The solutions of the Riemann problems at two interfaces do not ovelap (see *e.g.* computations in [54]). In the domain of relaxation schemes, it is a consequence of the subcharacteristic condition and of the standard CFL on the relaxed equations.

The condition (3.4) turns very restrictive when considering low collisional media, here when ρ is small. In such a case, one requires a very large number of energy steps and therefore considerably long computational times are necessary.

A first solution to this problem was proposed in [5] by the use of a change of variable. We proposed an alternative in [54] through a method of characteristic applied on a relaxed system for (3.2). However, both of these methods can be used only on hyperbolic systems, and they are therefore not applicable to the coupled photons-electrons transport equations (2.5).

We present our discretization of the system (3.1) in three parts:

- **Step 1** (Subsection 3.2): the discretization of the advection operator. This corresponds to the discretization over the position variable x.
- **Step 2** (Subsection 3.3): the discretization of the collision operator. This corresponds to the discretization over the energy variable ϵ.
- **Step 3** (Subsection 3.4): both discretizations are gathered to construct a numerical scheme, and we construct an iterative algorithm to solve the resulting discrete equations.

3.2 Discretization of the advection operator

In the spirit of [54], we construct a numerical scheme for (2.5) based on the relaxation method developed in [1,11,12,47]. We first recall the principle of the relaxation method.

The relaxation method is presented to justify the construction of the implicit scheme (3.11) below. The aim is to isolate the non-linearity in (3.1a) into a relaxation term and to solve more easily the remaining linear part by the use of an implicit scheme.

We perform a vectorial BGK approximation of the system (2.5). In practice, this consists in replacing the non-linear flux term by a linear advection term, and a relaxation term is added the right-hand side. Let us introduce the following system of two relaxed equations

$$
\begin{aligned}
c^{-}\partial_x \mathbf{f}_\tau^{-} - \rho \mathbf{Q}(\mathbf{f}_\tau^{-}) &= \frac{\mathbf{M}^{-} - \mathbf{f}_\tau^{-}}{\tau}, && (3.5a)\\
c^{+}\partial_x \mathbf{f}_\tau^{+} - \rho \mathbf{Q}(\mathbf{f}_\tau^{+}) &= \frac{\mathbf{M}^{+} - \mathbf{f}_\tau^{+}}{\tau}, && (3.5b)
\end{aligned}
$$

where $\mathbf{f}_\tau^{\pm}$ are the unknowns relaxing toward the equilibrium represented by the Maxwellians $\mathbf{M}^{\pm}(\boldsymbol{\psi}) \in \mathcal{R}^2$, and τ is a relaxation parameter. The unknowns $\mathbf{f}_\tau^{\pm}$ are only related to the original $\boldsymbol{\psi}$ in the limit $\tau \to 0$.

We chose the Maxwellians $\mathbf{M}^{\pm}(\boldsymbol{\psi})\in\mathcal{R}^2$ such that they relate to the original system through the following consistency formulae

$$\mathbf{M}^{+}+\mathbf{M}^{-}=\boldsymbol{\psi}\in\mathcal{R}^2, \qquad c^{+}\mathbf{M}^{+}+c^{-}\mathbf{M}^{-}=\mathbf{F}(\boldsymbol{\psi}), \tag{3.6a}$$

and the relaxation velocities $c^{\pm}(\boldsymbol{\psi})\in\mathbb{R}$ such that they bound the physical velocities. This leads to the following stability requirement ([1,11,12])

$$Sp\big(\mathbf{F}'(\boldsymbol{\psi})\big) \quad \subset \quad [c^{-},c^{+}]. \tag{3.7}$$

Formally, at the limit $\tau\to 0$ in (3.5), one obtains $\mathbf{f}_0^{\pm}=\mathbf{M}^{\pm}$. Then, replacing $\mathbf{f}^{\pm}$ by $\mathbf{M}^{\pm}$ and summing the two equations (3.5) yields (3.1). Therefore, one recovers the solution of the original equation (3.1) in the limit case $\tau\to 0$ as

$$\boldsymbol{\psi}=\lim_{\tau\to 0}\big(\mathbf{f}_\tau^{+}+\mathbf{f}_\tau^{-}\big). \tag{3.8}$$

We refer to [1,2,11,12,47] for a proper analysis of this asymptotic limit.

For the sake of simplicity, in the 1D case, we use the following classical result (see *e.g.* [4,13,54]), to choose the relaxation parameters.

Lemma 3.1. *The eigenvalues of the Jacobian* $\mathbf{F}'_{M_1}(\boldsymbol{\psi}_\alpha)$ *of the* M_1 *fluxes are bounded by* 1 *for all* $\boldsymbol{\psi}_\alpha\in\mathcal{R}$*, that is*

$$\forall\boldsymbol{\psi}_\alpha\in\mathcal{R}, \quad Sp(\mathbf{F}'_{M_1}(\boldsymbol{\psi}_\alpha))\subset]-1,1[.$$

Furthermore, for all realizable moments $\boldsymbol{\psi}_\alpha\in\mathcal{R}$*, one has*

$$\boldsymbol{\psi}_\alpha\pm\mathbf{F}_{M_1}(\boldsymbol{\psi}_\alpha)\in\mathcal{R}.$$

Thus, in 1D, we use the following parameters

$$c^{+}=1=-c^{-}, \qquad \mathbf{M}^{\pm}=\frac{\boldsymbol{\psi}\pm\mathbf{F}(\boldsymbol{\psi})}{2}\in\mathcal{R}^2, \tag{3.9a}$$

$$\pm\partial_x\mathbf{f}_\tau^{\pm}-\rho\mathbf{Q}(\mathbf{f}_\tau^{\pm})=\frac{\mathbf{M}^{\pm}-\mathbf{f}_\tau^{\pm}}{\tau}, \tag{3.9b}$$

which satisfy the requirements (3.6).

Then, we use upwind fluxes on (3.9b) leading to the scheme

$$\frac{\mathbf{f}_{\tau,l}^{\pm,n}-\mathbf{f}_{\tau,l\mp 1}^{\pm,n}}{\Delta x}-\rho\mathbf{Q}(\mathbf{f}_{\tau,l}^{\pm,n})=\frac{\mathbf{M}^{\pm}-\mathbf{f}_{\tau,l}^{\pm,n}}{\tau}. \tag{3.10}$$

Then, summing these equations over ± and having $\tau\to 0$ leads to define the following scheme over $\boldsymbol{\psi}$

$$\frac{\mathbf{F}^n_{l+\frac{1}{2}}-\mathbf{F}^n_{l-\frac{1}{2}}}{\Delta x}-[\rho\mathbf{Q}(\boldsymbol{\psi})]_l^n=0, \tag{3.11a}$$

$$\mathbf{F}^n_{l+\frac{1}{2}}=\frac{1}{2}\left[\mathbf{F}(\boldsymbol{\psi}^n_{l+1})+\mathbf{F}(\boldsymbol{\psi}^n_{l})-(\boldsymbol{\psi}^n_{l+1}-\boldsymbol{\psi}^n_{l})\right]. \tag{3.11b}$$

The term $[\rho\mathbf{Q}(\boldsymbol{\psi})]_l^n$ will be defined in the next subsection.

Remark 3.1. Since the collision operator $\mathbf{Q}$ is linear, one should recover

$$\mathbf{Q}(\boldsymbol{\psi})^n_l = \mathbf{Q}(\mathbf{M}^- + \mathbf{M}^+)^n_l = \mathbf{Q}(\mathbf{M}^-)^n_l + \mathbf{Q}(\mathbf{M}^+)^n_l.$$

3.3 Discretization of the collision operator

We simply discretize the collision terms with a quadrature rule for the integrals in ϵ and an implicit Euler discretization for the term $\partial_\epsilon(S\boldsymbol{\psi}_e)$. This reads

$$\begin{aligned}
[\rho\mathbf{Q}(\boldsymbol{\psi})]^n_l &= \rho_l \mathbf{Q}(\boldsymbol{\psi})^n_l, && \text{(3.12a)}\\
\mathbf{Q}(\boldsymbol{\psi})^n_l &= \Big(\mathbf{Q}_{\gamma\to\gamma}(\boldsymbol{\psi}_\gamma)^n_l + \mathbf{Q}_{e\to\gamma}(\boldsymbol{\psi}_e)^n_l, \quad \mathbf{Q}_{e\to e}(\boldsymbol{\psi}_e)^n_l + \mathbf{Q}_{\gamma\to\gamma}(\boldsymbol{\psi}_\gamma)^n_l\Big), && \text{(3.12b)}\\
\mathbf{Q}_{\alpha\to\beta}(\boldsymbol{\psi}_\alpha)^n_l &= \Big(Q^0_{\alpha\to\beta}(\psi^0_\alpha)^n_l, \quad Q^1_{\alpha\to\beta}(\psi^1_\alpha)^n_l\Big), && \text{(3.12c)}
\end{aligned}$$

where each discrete collision operators is, for $i=0,1$

$$\begin{aligned}
Q^i_{\gamma\to\gamma}(\psi^i_\gamma)^n_l &= \sum_{n'=1}^{n} \sigma^{i,n',n}_{\gamma\to\gamma} \psi^{i,n'}_{\gamma,l} \Delta\epsilon^{n'} - \sigma^n_{T,\gamma} \psi^{i,n}_{\gamma,l}, && \text{(3.12d)}\\
Q^i_{\gamma\to e}(\psi^i_\gamma)^n_l &= \sum_{n'=1}^{n} \sigma^{i,n',n}_{\gamma\to e} \psi^{i,n'}_{\gamma,l} \Delta\epsilon^{n'}, && \text{(3.12e)}\\
Q^i_{e\to\gamma}(\psi^i_e)^n_l &= 0, && \text{(3.12f)}\\
Q^i_{e\to e}(\psi^i_e)^n_l &= \frac{S^{n-1}\psi^{i,n-1}_{e,l} - S^n \psi^{i,n}_{e,l}}{\Delta\epsilon^n} + \sum_{n'=1}^{n} \sigma^{i,n',n}_{e\to e} \psi^{i,n'}_{e,l} \Delta\epsilon^{n'} - \sigma^n_{T,e} \psi^{i,n}_{e,l}. && \text{(3.12g)}
\end{aligned}$$

Remark 3.2.
- We choose this particular discretization because it leads to an implicit scheme of the form (3.11-3.12), in the sense that the fluxes $\mathbf{F}^n_{l+\frac{1}{2}}$ are evaluated at the latest energy step ϵ^n. In practice, the obtained scheme is efficient even without imposing a restriction on the step size $\Delta\epsilon^n$, which circumvents the problem presented in Subsection 3.1.

- By construction, we expect the discretization to be of order one in Δx and in $\Delta\epsilon^n$ when the solution is smooth. Thus, the scheme (3.11-3.12) is consistent with the continuous equation (3.1a).

- In order to use the present scheme, one needs to compute $\boldsymbol{\psi}^n_l$ for all l based on $\boldsymbol{\psi}^{n'}_l$ for $n' < n$. Here, this implies solving the non-linear equation (3.11-3.12) over the vector $(\boldsymbol{\psi}^n_l)_{l=1,\ldots,l_{\max}} \in (\mathcal{R}^2)^{l_{\max}}$.

3.4 An iterative solver for the 1D scheme

Writing together the discretization of the 1D advection and the collision term with the relaxation parameters (3.9) yields the following numerical scheme

$$-\mathbf{L}(\boldsymbol{\psi}_{l-1}^n)+\mathbf{D}(\boldsymbol{\psi}_l^n)-\mathbf{U}(\boldsymbol{\psi}_{l+1}^n)=\rho_l\mathbf{R}_l^n, \tag{3.13a}$$

where the operators **L** and **U** are non-linear, and **D** is linear and invertible. Those operators are

$$\mathbf{L}(\boldsymbol{\psi}_{l-1}^n)=\frac{\boldsymbol{\psi}_{l-1}^n+\mathbf{F}(\boldsymbol{\psi}_{l-1}^n)}{2\Delta x}, \qquad \mathbf{U}(\boldsymbol{\psi}_{l+1}^n)=\frac{\boldsymbol{\psi}_{l+1}^n-\mathbf{F}(\boldsymbol{\psi}_{l+1}^n)}{2\Delta x}, \tag{3.13b}$$

$$\mathbf{D}(\boldsymbol{\psi}_l^n)=\left(\frac{Id}{\Delta x}+\rho_l A^n\right)\boldsymbol{\psi}_l^n, \tag{3.13c}$$

and the matrix A^n and the source $\mathbf{R}_l^n$ are

$$\begin{aligned} A^n\boldsymbol{\psi}_l^n &= (B_0^n\boldsymbol{\psi}_l^n,\quad B_1^n\boldsymbol{\psi}_l^n,\quad D_0^n\boldsymbol{\psi}_l^n,\quad D_1^n\boldsymbol{\psi}_l^n), & (3.13\text{d})\\ \mathbf{R}_l^n &= (C_0^n,\quad C_1^n,\quad E_0^n,\quad E_1^n)+BC_l^n, & (3.13\text{e})\end{aligned}$$

with, for $i=0,1$,

$$\begin{aligned} B_i^n\boldsymbol{\psi}_l^n &= (\sigma_{T,\gamma}^n-\sigma_{\gamma\to\gamma}^{i,n,n}\Delta\epsilon^n)\psi_{\gamma,l}^{i,n}, \qquad C_i^n=\sum_{n'=1}^{n-1}\sigma_{\gamma\to\gamma}^{i,n',n}\psi_{\gamma,l}^{i,n'}\Delta\epsilon^{n'}, & (3.13\text{f})\\ D_i^n\boldsymbol{\psi}_l^n &= \left(\frac{S^n}{\Delta\epsilon^n}+\sigma_{T,e}^n-\sigma_{e\to e}^{i,n,n}\Delta\epsilon^n\right)\psi_{e,l}^{i,n}-\sigma_{\gamma\to e}^{i,n,n}\psi_{\gamma,l}^{i,n}\Delta\epsilon^n, & (3.13\text{g})\\ E_i^n &= \frac{S^{n-1}}{\Delta\epsilon^n}\psi_{e,l}^{i,n-1}+\sum_{n'=1}^{n-1}\left(\sigma_{\gamma\to\gamma}^{i,n',n}\psi_{\gamma,l}^{i,n'}+\sigma_{e\to e}^{i,n',n}\psi_{e,l}^{i,n'}\right)\Delta\epsilon^{n'}. & (3.13\text{h})\end{aligned}$$

Defining properly boundary conditions for moment models based on the underlying kinetic ones remains an open problem (see *e.g.* [28,38,58] for linear moment equations). For the sake of simplicity, we use here discrete boundary conditions defined as a source term in (3.13e) with

$$BC_l^n=\boldsymbol{\psi}_0^n\delta_{1,l}+\boldsymbol{\psi}_{l_{\max}+1}^n\delta_{l_{\max},l},$$

with given $\boldsymbol{\psi}_0^n\in\mathcal{R}^2$ and $\boldsymbol{\psi}_{l_{\max}+1}^n\in\mathcal{R}^2$.

In order to use this scheme, one needs to solve (3.13) at each energy step n, which is a non-linear equation on the vector $(\boldsymbol{\psi}^n)_{l=1,\ldots,l_{\max}}$. For this purpose, we propose an iterative solver inspired of [22], which is tested in [6,50,52,53,55].

Algorithm 1. *Initialization:* At energy step n, set $\boldsymbol{\psi}_l^{n,(0)}=\boldsymbol{\psi}_l^{n-1}$ for all l.
Iteration: Compute iteratively

$$\boldsymbol{\psi}_l^{n,(k+1)}=\mathbf{D}^{-1}\left(\mathbf{L}(\boldsymbol{\psi}_{l-1}^{n,(k)})+\mathbf{U}(\boldsymbol{\psi}_{l+1}^{n,(k)})+\rho_l\mathbf{R}_l^n\right), \tag{3.14}$$

until convergence.

Proposition 3.1. Suppose that $\mathbf{R}_l^n \in \mathcal{R}^2$ is realizable for all l, and that

$$\min_l \rho_l \Delta x m_A^n > CFL, \qquad CFL := \sup_{\boldsymbol{\psi}\in\mathcal{R}^2} \left(\frac{(M_F - m_F)(\boldsymbol{\psi})}{2} \right), \tag{3.15}$$

where

$$m_A^n = \min\, Sp(A^n), \quad m_F(\boldsymbol{\psi}) = \min\, Sp(\mathbf{F}'(\boldsymbol{\psi})), \quad M_F(\boldsymbol{\psi}) = \max\, Sp(\mathbf{F}'(\boldsymbol{\psi}))$$

are respectively the minimum eigenvalue of the matrix A^n, of the Jacobian $\mathbf{F}'(\boldsymbol{\psi})$ and the maximum eigenvalue of $\mathbf{F}'(\boldsymbol{\psi})$.
Then, there exists a unique solution $(\boldsymbol{\psi}^n)_{l=1,\dots,l_{\max}} \in (\mathcal{R}^2)^{l_{\max}}$ satisfying (3.13) for all l. Moreover, Algorithm 1 converges to this solution.

Remark 3.3.

- In practice, the physical parameters S, $\sigma_{T,\beta}$ and $\sigma_{\alpha\to\beta}$ satisfy additional conditions that come either from the physics or from the study of the well-posedness of (2.1). These conditions lead to imposing that $m_A(\Delta\epsilon^n)$ is a strictly positive strictly decreasing function of $\Delta\epsilon^n$. We assume this holds in the rest of the paper, and we refer to [20,51,60] for more details on those conditions on the physical parameters.
Under such requirements, the condition (3.15) corresponds to a CFL-like condition, which restricts the size of $\Delta\epsilon^n$ based on Δx.
- Furthermore, we add the condition (3.15) here in order to prove that Algorithm 1 converges. However, the bound (3.18) used in the proof below is not optimal. We have not yet found any theoretical nor experimental test case violating (3.15) that lead to a non-converging sequence $(\boldsymbol{\psi}^{n,(k)})_{k=1,\dots,\infty}$, even with very low collisional media (for small ρ_l). In the test cases below, Δx and $\Delta\epsilon^n$ do not necessarily respect the condition (3.15) and we always verify that Algorithm 1 has converged at every step n in the experiments below. We refer *e.g.* to [3,24,35,64] and references therein for more complete study on convergence for such algorithms.

Proof. Define the operator J over $\boldsymbol{\psi} \in (\mathcal{R}^2)^{l_{\max}}$ by

$$\boldsymbol{\psi}^{n,(k+1)} \;=\; J(\boldsymbol{\psi}^{n,(k)}),$$

where the l-th component $J(\boldsymbol{\psi}^{n,(k+1)})_l$ is given by (3.14), *i.e.*

$$J(\boldsymbol{\psi}^{n,(k+1)})_l = \left(\frac{Id}{\Delta x} + \rho_l A^n \right)^{-1} \left[\rho_l \mathbf{R}_l^n + \frac{\boldsymbol{\psi}_{l+1}^{n,(k)} - \mathbf{F}(\boldsymbol{\psi}_{l+1}^{n,(k)})}{2\Delta x} + \frac{\boldsymbol{\psi}_{l-1}^{n,(k)} + \mathbf{F}(\boldsymbol{\psi}_{l-1}^{n,(k)})}{2\Delta x} \right]. \tag{3.16}$$

First, we verify that J preserves the realizability from one step to another.

Let us suppose $\boldsymbol{\psi}^{n,(k)} \in (\mathcal{R}^2)^{l_{max}}$. Then, $\boldsymbol{\psi}_{l+1}^{n,(k)} - \mathbf{F}(\boldsymbol{\psi}_{l+1}^{n,(k)}) \in \mathcal{R}^2$ and $\boldsymbol{\psi}_{l-1}^{n,(k)} + \mathbf{F}(\boldsymbol{\psi}_{l-1}^{n,(k)}) \in \mathcal{R}^2$ are realizable according to the second part of Lemma 3.1. Thus, the term between square brackets in (3.16) is realizable according to Remark 2.2.

Now, we need to prove that the operator $(\frac{Id}{\Delta x}+\rho_l A^n)^{-1}$ preserves the realizability property. Using its definition (3.13), the matrix A^n can be rewritten

$$A^n = \begin{pmatrix} a_0 & 0 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ b_0 & 0 & c_0 & 0 \\ 0 & b_1 & 0 & c_1 \end{pmatrix}, \qquad \left(\frac{Id}{\Delta x}+\rho_l A^n\right)^{-1} = \begin{pmatrix} \alpha_0 & 0 & 0 & 0 \\ 0 & \alpha_1 & 0 & 0 \\ \beta_0 & 0 & \gamma_0 & 0 \\ 0 & \beta_1 & 0 & \gamma_1 \end{pmatrix}$$

where

$$\begin{aligned} a_i &= \sigma^n_{T,\gamma} - \sigma^{i,n,n}_{\gamma\to\gamma}\Delta\epsilon^n, & \alpha_i &= \frac{1}{\frac{1}{\Delta x}+\rho_l a_i}, \\ b_i &= -\sigma^{i,n,n}_{\gamma\to e}\Delta\epsilon^n, & \beta_i &= -\frac{b_i}{(\frac{1}{\Delta x}+\rho_l a_i)(\frac{1}{\Delta x}+\rho_l c_i)}, \\ c_i &= \frac{S^n}{\Delta\epsilon^n} + \sigma^n_{T,e} - \sigma^{i,n,n}_{e\to e}\Delta\epsilon^n, & \gamma_i &= \frac{1}{\frac{1}{\Delta x}+\rho_l c_i}. \end{aligned}$$

Using (2.11) leads to

$$0 \le a_0 \le a_1, \qquad 0 \le c_0 \le c_1 \quad \text{and} \quad -b_0 \ge -b_1 \ge 0$$

and so

$$\alpha_0 \ge \alpha_1 > 0, \qquad \beta_0 \ge \beta_1 > 0 \qquad \gamma_0 \ge \gamma_1 > 0.$$

Then, one verifies that if $\boldsymbol{\psi}\in\mathcal{R}^2$, then $\left(\frac{Id}{\Delta x}+\rho_l A^n\right)^{-1}\boldsymbol{\psi}$ is a positive combination of vectors satisfying the criteria (2.10). Thus, the operator $\left(\frac{Id}{\Delta x}+\rho_l A^n\right)^{-1}$ preserves the realizability, and J is an operator from $(\mathcal{R}^2)^{l_{\max}}$ into itself.

Now, in order to prove that Algorithm 1 converges, we prove that J is a contraction. Differentiating $J(\boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$ reads

$$d_{\boldsymbol{\psi}} J(\boldsymbol{\psi}) = \begin{pmatrix} 0 & J^n_{1,2} & 0 & \cdots & 0 \\ J^n_{2,1} & 0 & J^n_{2,3} & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & J_{l_{\max}-1,l_{\max}} \\ 0 & \cdots & 0 & J_{l_{\max},l_{\max}-1} & 0 \end{pmatrix}, \tag{3.17a}$$

$$J^n_{l,l-1} = \left(\frac{Id}{\Delta x}+\rho_l A^n\right)^{-1}\left(\frac{Id+\mathbf{F}'(\boldsymbol{\psi}_{l-1})}{2\Delta x}\right), \tag{3.17b}$$

$$J^n_{l,l+1} = \left(\frac{Id}{\Delta x}+\rho_l A^n\right)^{-1}\left(\frac{Id-\mathbf{F}'(\boldsymbol{\psi}_{l+1})}{2\Delta x}\right), \tag{3.17c}$$

so $d_{\boldsymbol{\psi}} J(\boldsymbol{\psi})$ is a block matrix with non-zero blocks on the super- and sub-diagonal.

Using a Gershgorin theorem for block matrices (see *e.g.* [62]) provides

$$Sp\left(d_{\boldsymbol{\psi}}J(\boldsymbol{\psi})\right)\subset[-r,r],$$

with a spectral radius satisfying

$$r\leq\max_{l}\frac{|||Id-\mathbf{F}'(\boldsymbol{\psi}_l)|||+|||Id+\mathbf{F}'(\boldsymbol{\psi}_l)|||}{2\Delta x\min Sp\left(\frac{Id}{\Delta x}+\rho_l A^n\right)}.$$

Using Lemma 3.1 and the fact that A^n is positive definite, we obtain upper and lower bounds on the eigenvalues of $Id\pm\mathbf{F}'(\boldsymbol{\psi})$ and of $Id+\rho_l A^n\Delta x$ that leads to

$$r\ \leq\ \max_{l}\frac{1+\frac{(M_F-m_f)(\boldsymbol{\psi}_l)}{2}}{1+\rho_l m_A^n\Delta x}. \tag{3.18}$$

Thus, $r<1$ under condition (3.15), J is a contraction and Algorithm 1 converges to the unique fixed point of J. □

Remark 3.4.

- The iterative method proposed in Algorithm 1 can be interpreted as a Jacobi method with non-linear extra-diagonal term. Indeed, the desired solution $\boldsymbol{\psi}^n$ solves (3.13) which can be rewritten

$$\begin{pmatrix} \mathbf{D} & -\mathbf{U} & 0 & \dots & \dots & \dots & 0\\ -\mathbf{L} & \mathbf{D} & -\mathbf{U} & 0 & \dots & \dots & 0\\ 0 & -\mathbf{L} & \mathbf{D} & -\mathbf{U} & 0 & \dots & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \dots & \dots & 0 & -\mathbf{L} & \mathbf{D} & -\mathbf{U}\\ 0 & \dots & \dots & \dots & 0 & -\mathbf{L} & \mathbf{D} \end{pmatrix}\boldsymbol{\psi}^n=\begin{pmatrix} \rho_1\mathbf{R}_1^n\\ \rho_2\mathbf{R}_2^n\\ \vdots\\ \vdots\\ \vdots\\ \rho_{l_{\max}-1}\mathbf{R}_{l_{\max}-1}^n\\ \rho_{l_{\max}}\mathbf{R}_{l_{\max}}^n \end{pmatrix}, \tag{3.19}$$

where $\mathbf{L}$ and $\mathbf{U}$ are non-linear and $\mathbf{D}$ is linear, and they are defined in (3.13).

Similarly, we implemented and tested Gauss-Seidel and successive over-relaxation (SOR) methods for the non-linear problem (3.19). For example, the non-linear Gauss-Seidel type method consists in solving alternatively

$$\boldsymbol{\psi}_l^{n,(k+1)}=\mathbf{D}^{-1}\left(\mathbf{L}(\boldsymbol{\psi}_{l-1}^{n,(k+1)})+\mathbf{U}(\boldsymbol{\psi}_{l+1}^{n,(k)})+\rho_l\mathbf{R}_l^n\right), \tag{3.20a}$$

$$\boldsymbol{\psi}_l^{n,(k+1)}=\mathbf{D}^{-1}\left(\mathbf{L}(\boldsymbol{\psi}_{l-1}^{n,(k)})+\mathbf{U}(\boldsymbol{\psi}_{l+1}^{n,(k+1)})+\rho_l\mathbf{R}_l^n\right), \tag{3.20b}$$

in Algorithm 1 instead of (3.14). One can easily adapt the proof of Proposition 3.1 for such algorithms.

- The convergence rate of Algorithm 1 depends on the eigenvalues of $d_{\boldsymbol{\psi}} J(\boldsymbol{\psi})$. In the computations (3.18), the worst possible convergence rate corresponds to the case where $(M_F - m_F)(\boldsymbol{\psi}_l)$ has the highest value. Such highest value is obtained in the limit case of a purely anisotropic distribution (see *e.g.* computations in [4]) modeled by a fluence
$$\psi(\Omega) = K\delta(\Omega_1 - 1).$$
Thus, Algorithm 1 is slower if the expected solution of (2.1) possesses purely anisotropic regions.

- Algorithm 1 is based on a fixed point theorem. One may think of using Newton's method to accelerate the convergence. We did not develop such a method here since we did not compute numerically the Jacobian $\mathbf{F}'(\boldsymbol{\psi})$. We only used upper bounds of its eigenvalues in the proof of Proposition 3.1. The construction of such Newton-like method would require an approximation of $\mathbf{F}'$ adapted to the approximation of $\mathbf{F}$ that we used in the numerical test cases. We leave this development as a perspective for improving the convergence speed of Algorithm 1.

3.5 Numerical experiments

We study experimentally the convergence of the present method. Especially, we consider two convergence rates:

- The convergence with respect to the number $k_{\max}$ of iterations in Algorithm 1.
- The convergence with respect to the $\Delta\epsilon^n$ and Δx of the numerical scheme (3.13).

For this purpose, we consider the following test case.

We consider a 1D domain $Z = [0\text{ cm}, 8\text{ cm}]$ uniformly composed of water (*i.e.* $\rho = 1$) and impose a source of electrons of $\epsilon_0 = 5$ MeV modeled by the initial condition

$$\psi_e(x, 5\,\text{MeV}, \Omega) \;=\; K\mathbf{1}_{[3\text{ cm},5\text{ cm}]}(x), \qquad \psi_\gamma(x, 5\,\text{MeV}, \Omega) = 0, \tag{3.21a}$$

and we use a zero condition flux on the boundary which corresponds to extracting the moments of distributions of the form

$$\text{for}\quad \Omega_1 > 0, \qquad \psi_e(0\text{ cm}, \epsilon, \Omega) = 0, \qquad \psi_\gamma(0\text{ cm}, \epsilon, \Omega) = 0, \tag{3.22a}$$
$$\text{for}\quad \Omega_1 < 0, \qquad \psi_e(8\text{ cm}, \epsilon, \Omega) = 0, \qquad \psi_\gamma(8\text{ cm}, \epsilon, \Omega) = 0. \tag{3.22b}$$

Remark that the equations of system (2.1) are decoupled. As we impose no source of photons, and due to the considered physics, no photons are created in the system. Thus, the solution of (2.1a) is simply $\psi_\gamma = 0$, and the discretization (3.13) can be simplified when considering beams of electrons only. This is no longer true when considering more complex physics, *e.g.* when taking into account Bremsstrahlung effect [39,44].

The mesh is composed of 800 cells in x uniformly distributed. The step size $\Delta\epsilon^n$ and the grid in ϵ are such that

$$\Delta\epsilon^n = 5S^n\Delta x. \tag{3.23}$$

This corresponds approximately to fixing

$$m_A^n \Delta x = 5.$$

With such a grid size, we aim to avoid having numerical diffusion effects depending on $\Delta\epsilon^n$.

This test case is academic and aims only to study experimentally the numerical convergence rates. We provide the converged dose normalized by its maximum value computed with Algorithm 1 on Fig. 1 as an indication. For more examples with different applications in medical physics, we refer *e.g.* to [6,50,53,55].

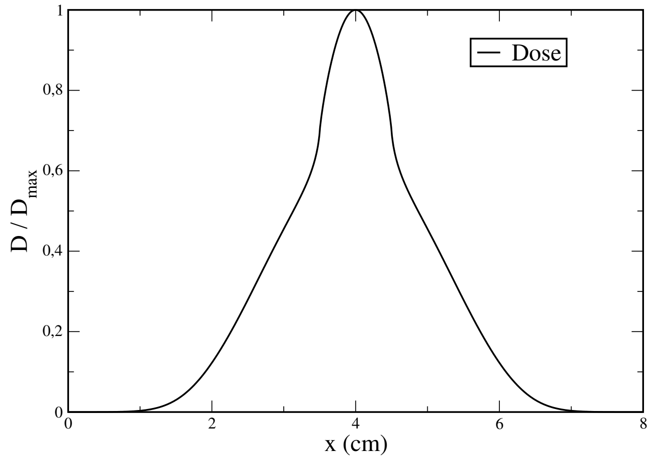

Figure 1: Normalized dose obtained with Algorithm 1 for the beam of electron (3.22).

3.5.1 Convergence results of the iterative algorithm

The iterative method of Algorithm 1 requires a criterion to stop.

A first naive criterion consists in fixing the number of iterations k_{max}. This is not optimal, neither in terms of precision nor in terms of computational costs. At each energy step, the desired solution follows (3.13) for all l. Then, one better stopping criterion consists in defining the residual

$$r^{n,(k)} = \max_l \left\| -\mathbf{L}(\boldsymbol{\psi}_{l-1}^{n,(k)}) + \mathbf{D}(\boldsymbol{\psi}_l^{n,(k)}) - \mathbf{U}(\boldsymbol{\psi}_{l+1}^{n,(k)}) - \rho_l \mathbf{R}_l^n \right\|_\infty, \tag{3.24}$$

and choose to stop Algorithm 1 as soon as k satisfies

$$r^{n,(k)} \leq r_{\max}. \tag{3.25}$$

Fig. 2 depicts the minimum number of iterations k required in Algorithm 1 such that the residual $r^{n,(k)}$ satisfies the criterion (3.25) with $r_{\max}$ fixed at 1, 10^{-1} and 10^{-2} as a function of the energy step n.

Similarly, Fig. 3 depicts the final residual $r^{n,(k_{\max})}$ obtained by fixing $k_{\max}$ to 30, 50 or 70 as a function of the energy step n.

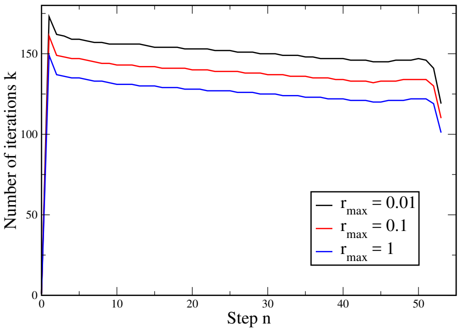

Figure 2: Number of iterations k as a function of the energy step n for a given maximum residual $r_{\max}$.

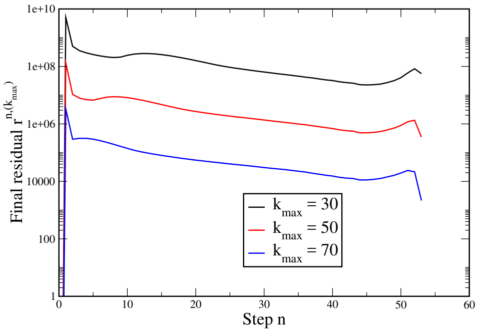

Figure 3: Final residual $r^{n,(k_{\max})}$ as a function of the energy step n for a given number of iteration $k_{\max}$.

In the first steps where the values of ψ_e^b, Algorithm 1 requires numerous iterations to converge. In this range of energy, $\partial_\epsilon \boldsymbol{\psi} \equiv (\boldsymbol{\psi}^{n-1} - \boldsymbol{\psi}^n)/\Delta\epsilon^n$ is large. Thus, the initialization $\boldsymbol{\psi}^{n,(0)} = \boldsymbol{\psi}^{n-1}$ of Algorithm 1 is far from the desired solution. Therefore, the present algorithm requires more iterations to converge.

The convergence rate progressively raises, *i.e.* the final residual $r^{n,(k_{\max})}$ and the number of iterations k reduce.

The drop of $r^{n,(k_{\max})}$ and of iterations k near the end of the simulation is due to the physical parameters used. The stopping power S skyrockets near the threshold $\epsilon = \epsilon_{\min}$. According to the definition of $\mathbf{D}$ in (3.13), this implies that the eigenvalues of $\mathbf{D}$ also raise at the lowest energy. Thus, the eigenvalues of J drop to zero, and so the convergence rate of Algorithm 1 skyrockets. This explains the shape of the curves k and $r^{n,(k_{\max})}$ in the last steps n.

In all the remaining test cases, we fix the parameters $r_{\max}$ and $k_{\max}$ sufficiently high such that the residual $r_{\max}$ is always reached during the computations, and we present only converged results.

3.5.2 Convergence results of the numerical scheme

In this subsection, we verify experimentally that the numerical scheme is converging. The maximum residual $r_{\max}$ is fixed at 5.10^{-3}.

Since there are no analytic solution, we use the solution obtained with the largest number of cells as reference solution. The spatial domain $Z=[0,8\text{ cm}]$ is uniformly meshed. We choose the number of spatial cells $l_{\max}$ at 100, 200, 400, 800 and 1600 cells and 3200 cells for the reference solution. We represent the convergence rate in Δx through the discrete L^1, L^2 and L^∞ errors between the reference solution, *i.e.* the most refined one, and the less refined ones

$$\text{Error}_{L^1}(\Delta x)=\sum_{l=1}^{l_{\max}}\sum_{n=1}^{n_{\max}}|\boldsymbol{\psi}_l^n-\tilde{\boldsymbol{\psi}}_l^n|\tilde{\Delta}x\Delta\tilde{\epsilon}^n, \qquad \text{Error}_{L^2}(\Delta x)=\sqrt{\sum_{l=1}^{l_{\max}}\sum_{n=1}^{n_{\max}}(\boldsymbol{\psi}_l^n-\tilde{\boldsymbol{\psi}}_l^n)^2\tilde{\Delta}x\Delta\tilde{\epsilon}^n},$$
$$\text{Error}_{L^\infty}(\Delta x)=\max_{l=1,\dots,l_{\max}}\max_{n=1,\dots,n_{\max}}|\boldsymbol{\psi}_l^n-\tilde{\boldsymbol{\psi}}_l^n|,$$

where $\boldsymbol{\psi}_l^n$ is the solution with mesh cells of size Δx and an energy step size $\Delta\epsilon^n$ given by (3.23) approximated by piecewise linear polynomials at the points (x_l,ϵ^n) on the finest mesh (of grid size Δx) and $\tilde{\boldsymbol{\psi}}_l^n$ is the most refined solution at the same points (x_l,ϵ^n). Those errors are plotted on Fig. 4 as a function of Δx. As expected, we observe a convergence rate of order 1, the mean slope of the curves obtained on Fig. 4 being of 1.230901 for the L^1 error, 1.213013 for the L^2 error and 0.974655 for the L^∞ error.

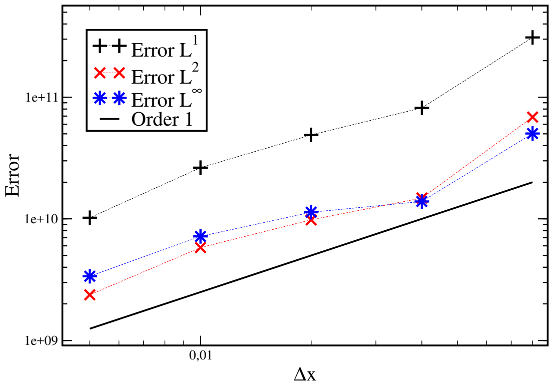

Figure 4: Discrete L^1, L^2 and L^∞ errors compared to the most refined solution as a function of Δx with $\Delta\epsilon^n$ given by (3.23).

4 A numerical approach for multi-D problems

We extend here the previous approach for coupled electrons and photons transport in multi-D media. The next three subsections describe the difficulties and the method in 2D through the model

$$\partial_x\mathbf{F}_1(\boldsymbol{\psi})+\partial_y\mathbf{F}_2(\boldsymbol{\psi})=\rho\mathbf{Q}(\boldsymbol{\psi}). \tag{4.1}$$

However, the method is also valid in 3D and a 3D test case is provided in Subsection 4.4.

4.1 A correction of the numerical transverse diffusion

As described in [4,55] and through the experimental results below, using the relaxation parameters (3.9) when considering 2D photon beams leads to a numerical overestimation of the diffusion effects in the direction orthogonal to the beams.

In multi-D, the velocities $c^{\pm}$ are vectors instead of scalars. We consider two relaxation velocities $c_i^{\pm}=\pm|c_i^{\pm}|e_i$ per Cartesian direction e_i. The relaxation parameters need to satisfy ([1,11,47])

$$\forall d\in\mathcal{S}^2, \qquad Sp\left(\mathbf{F}_d'(\boldsymbol{\psi})\right)\subset\left[\min_{i,\pm}(c_i^{\pm}.d),\, \max_{i,\pm}(c_i^{\pm}.d)\right], \tag{4.2a}$$

$$\sum_{i,\pm}\mathbf{M}_i^{\pm}=\boldsymbol{\psi}, \qquad \sum_{i,\pm}(c_i^{\pm}.d)\mathbf{M}_i^{\pm}=\mathbf{F}_d(\boldsymbol{\psi}), \tag{4.2b}$$

where

$$\mathbf{F}_d(\boldsymbol{\psi})=d_1\mathbf{F}_1(\boldsymbol{\psi})+d_2\mathbf{F}_2(\boldsymbol{\psi})=(\psi_\gamma^1.d,\, \psi_\gamma^2.d,\, \psi_e^1.d,\, \psi_e^2.d)$$

is the flux in the direction d. In practice, we use the relaxation velocities $c_i^{\pm}$ defined in [55] that approximate numerically the maximum physical velocities in each Cartesian directions

$$|c_i^{\pm}|\approx\max\left(\delta,\, \max\left[Sp\left(\pm\mathbf{F}_i'(\boldsymbol{\psi})\right)\right]\right).$$

Here, $\delta=10^{-8}$ is a constant chosen arbitrarily small to avoid numerical divisions by zero. We use Maxwellians of the form

$$\mathbf{M}_i^{\pm}=\mu_i^{\pm}\boldsymbol{\psi}+\lambda_i^{\pm}\mathbf{F}_i(\boldsymbol{\psi}).$$

In order to satisfy (4.2b), one finds that $\mu_i^{\pm}$ and $\lambda_i^{\pm}$ need to satisfy

$$\sum_{i,\pm}\mu_i^{\pm}=1, \tag{4.3a}$$

$$\lambda_i^-+\lambda_i^+=0, \qquad -|c_i^-|\mu_i^-+|c_i^+|\mu_i^+=0, \qquad -|c_i^-|\lambda_i^-+|c_i^+|\lambda_i^+=1. \tag{4.3b}$$

The last three equations (4.3b) can be rewritten

$$\mu_i^-=\frac{|c_i^+|}{|c_i^-|}\mu_i^+, \qquad \lambda_i^{\pm}=\pm\frac{1}{|c_i^+|+|c_i^-|}.$$

In practice, in 2D, we choose to fix the last degrees of freedom by

$$\mu_i^{\pm}=\frac{|c_i^{\mp}|}{2(|c_i^+|+|c_i^-|)},$$

which satisfies (4.3a). This leads to write the Maxwellians

$$\mathbf{M}_i^{\pm}=\frac{|\tilde{c}_i^{\mp}|\boldsymbol{\psi}\pm 2\mathbf{F}_i(\boldsymbol{\psi})}{2(|\tilde{c}_i^{+}|+|\tilde{c}_i^{-}|)}\in\mathcal{R}^2, \tag{4.4}$$

where $|\tilde{c}_i^{\pm}|$ is either $|c_i^{\pm}|$ or the minimum scalar such that (4.4) is realizable.

Applying the method described in Subsection 3.2 to the 2D equation (4.1) leads to write the scheme

$$\frac{\mathbf{F}^n_{l+\frac{1}{2},m}-\mathbf{F}^n_{l-\frac{1}{2},m}}{\Delta x}+\frac{\mathbf{F}^n_{l,m+\frac{1}{2}}-\mathbf{F}^n_{l,m-\frac{1}{2}}}{\Delta y}-[\rho\mathbf{Q}(\boldsymbol{\psi})]^n_{l,m}=0, \tag{4.5a}$$

$$\begin{aligned}\mathbf{F}^n_{l+\frac{1}{2},m}=&c^{-,n}_{l+\frac{1}{2},m}\lambda^{-,n}_{l+\frac{1}{2},m}\mathbf{F}_1(\boldsymbol{\psi}^n_{l+1,m})+c^{+,n}_{l+\frac{1}{2},m}\lambda^{+,n}_{l+\frac{1}{2},m}\mathbf{F}_1(\boldsymbol{\psi}^n_{l,m})\\&+\left(c^{-,n}_{l+\frac{1}{2},m}\mu^{-,n}_{l+\frac{1}{2},m}\boldsymbol{\psi}^n_{l+1,m}+c^{+,n}_{l+\frac{1}{2},m}\mu^{+,n}_{l+\frac{1}{2},m}\boldsymbol{\psi}^n_{l,m}\right),\end{aligned} \tag{4.5b}$$

$$\begin{aligned}\mathbf{F}^n_{l,m+\frac{1}{2}}=&c^{-,n}_{l,m+\frac{1}{2}}\lambda^{-,n}_{l,m+\frac{1}{2}}\mathbf{F}_2(\boldsymbol{\psi}^n_{l,m+1})+c^{+,n}_{l,m+\frac{1}{2}}\lambda^{+,n}_{l,m+\frac{1}{2}}\mathbf{F}_2(\boldsymbol{\psi}^n_{l,m})\\&+\left(c^{-,n}_{l,m+\frac{1}{2}}\mu^{-,n}_{l,m+\frac{1}{2}}\boldsymbol{\psi}^n_{l,m+1}+c^{+,n}_{l,m+\frac{1}{2}}\mu^{+,n}_{l,m+\frac{1}{2}}\boldsymbol{\psi}^n_{l,m}\right),\end{aligned} \tag{4.5c}$$

where the coefficients in the discrete fluxes are

$$\begin{aligned}&\lambda^{\pm,n}_{j,k}=\pm\frac{1}{|c^{+,n}_{j,k}|+|c^{-,n}_{j,k}|},\qquad \mu^{\pm,n}_{j,k}=\frac{|c^{\mp,n}_{j,k}|}{2(|c^{+,n}_{j,k}|+|c^{-,n}_{j,k}|)},\\&c^{\pm,n}_{l+\frac{1}{2},m}=\pm\max\left(|\tilde{c}_1^{\pm}(\boldsymbol{\psi}^n_{l+1,m})|,|\tilde{c}_1^{\pm}(\boldsymbol{\psi}^n_{l,m})|\right),\\&c^{\pm,n}_{l,m+\frac{1}{2}}=\pm\max\left(|\tilde{c}_2^{\pm}(\boldsymbol{\psi}^n_{l,m+1})|,|\tilde{c}_2^{\pm}(\boldsymbol{\psi}^n_{l,m})|\right),\end{aligned}$$

for $(j,k)=(l+\frac{1}{2},m)$ or $(j,k)=(l,m+\frac{1}{2})$.

4.2 An iterative solver for the multi-D scheme with the transverse diffusion correction

We propose to adapt Algorithm 1 when the coefficients $c_i^{\pm}$ are not constants. In this case, the scheme (4.5) can be rewritten under the form

$$\begin{aligned}\rho_l\mathbf{R}^n_l\ =\ &-\mathbf{L}_1(\boldsymbol{\psi}^n_{l-1,m})-\mathbf{L}_2(\boldsymbol{\psi}^n_{l,m-1})+\mathbf{D}(\boldsymbol{\psi}^n_{l,m})\\&-\mathbf{U}_1(\boldsymbol{\psi}^n_{l+1,m})-\mathbf{U}_2(\boldsymbol{\psi}^n_{l,m+1}),\end{aligned} \tag{4.6a}$$

where the operators $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{D}$ yield

$$\begin{aligned}
\mathbf{L}_1(\boldsymbol{\psi}^n_{l-1,m}) &= \frac{c^{+,n}_{l-\frac{1}{2},m}}{\Delta x}\left[\mu^{+,n}_{l-\frac{1}{2},m}\boldsymbol{\psi}^n_{l-1,m}+\lambda^{+,n}_{l-\frac{1}{2},m}\mathbf{F}_1(\boldsymbol{\psi}^n_{l-1,m})\right], && (4.6b)\\
\mathbf{L}_2(\boldsymbol{\psi}^n_{l,m-1}) &= \frac{c^{+,n}_{l,m-\frac{1}{2}}}{\Delta y}\left[\mu^{+,n}_{l,m-\frac{1}{2}}\boldsymbol{\psi}^n_{l,m-1}+\lambda^{+,n}_{l,m-\frac{1}{2}}\mathbf{F}_2(\boldsymbol{\psi}^n_{l,m-1})\right], && (4.6c)\\
\mathbf{U}_1(\boldsymbol{\psi}^n_{l+1,m}) &= \frac{-c^{-,n}_{l+\frac{1}{2},m}}{\Delta x}\left[\mu^{-,n}_{l+\frac{1}{2},m}\boldsymbol{\psi}^n_{l+1,m}+\lambda^{-,n}_{l+\frac{1}{2},m}\mathbf{F}_1(\boldsymbol{\psi}^n_{l+1,m})\right], && (4.6d)\\
\mathbf{U}_2(\boldsymbol{\psi}^n_{l,m+1}) &= \frac{-c^{-,n}_{l,m+\frac{1}{2}}}{\Delta y}\left[\mu^{-,n}_{l,m+\frac{1}{2}}\boldsymbol{\psi}^n_{l,m+1}+\lambda^{-,n}_{l,m+\frac{1}{2}}\mathbf{F}_2(\boldsymbol{\psi}^n_{l,m+1})\right], && (4.6e)\\
\mathbf{D}(\boldsymbol{\psi}^n_{l,m}) &= \left(\rho_l A^n+\beta^n_{l,m}Id\right)\boldsymbol{\psi}^n_{l,m}+\gamma^n_{1,l,m}\mathbf{F}_1(\boldsymbol{\psi}^n_{l,m})+\gamma^n_{2,l,m}\mathbf{F}_2(\boldsymbol{\psi}^n_{l,m}), && (4.6f)
\end{aligned}$$

where the coefficients $\beta^n_{l,m}$, $\gamma^n_{1,l,m}$ and $\gamma^n_{2,l,m}$ read

$$\begin{aligned}
\beta^n_{l,m} &= \frac{c^{+,n}_{l+\frac{1}{2},m}\mu^{+,n}_{l+\frac{1}{2},m}-c^{-,n}_{l-\frac{1}{2},m}\mu^{-,n}_{l-\frac{1}{2},m}}{\Delta x}+\frac{c^{+,n}_{l,m+\frac{1}{2}}\mu^{+,n}_{l,m+\frac{1}{2}}-c^{-,n}_{l,m-\frac{1}{2}}\mu^{-,n}_{l,m-\frac{1}{2}}}{\Delta y},\\
\gamma^n_{1,l,m} &= \frac{c^{+,n}_{l+\frac{1}{2},m}\lambda^{+,n}_{l+\frac{1}{2},m}-c^{-,n}_{l-\frac{1}{2},m}\lambda^{-,n}_{l-\frac{1}{2},m}}{\Delta x},\\
\gamma^n_{2,l,m} &= \frac{c^{+,n}_{l,m+\frac{1}{2}}\lambda^{+,n}_{l,m+\frac{1}{2}}-c^{-,n}_{l,m-\frac{1}{2}}\lambda^{-,n}_{l,m-\frac{1}{2}}}{\Delta y}.
\end{aligned}$$

The difficulty here emerges from the non-linearity of the operator $\mathbf{D}$ to invert and from the realizability requirements (4.2).

Let us decompose the operator $\mathbf{D}$ into

$$\begin{aligned}
\mathbf{D}(\boldsymbol{\psi}) &= \mathbf{D}_{imp}(\boldsymbol{\psi})-\mathbf{D}_{exp}(\boldsymbol{\psi}), && (4.7)\\
\mathbf{D}_{imp}(\boldsymbol{\psi}) &= \left[\rho_l A^n+(\alpha^n_{l,m}+\beta^n_{l,m})Id\right]\boldsymbol{\psi},\\
\mathbf{D}_{exp}(\boldsymbol{\psi}) &= \alpha^n_{l,m}\boldsymbol{\psi}+\gamma^n_{1,l,m}\mathbf{F}_1(\boldsymbol{\psi})+\gamma^n_{2,l,m}\mathbf{F}_2(\boldsymbol{\psi}),
\end{aligned}$$

such that $\mathbf{D}_{imp}$ is linear and invertible. Here, we choose the coefficient $\alpha^n_{l,m}$ non-negative such that the operator $\mathbf{D}_{exp}$ preserves the realizability. In practice, we choose

$$\alpha^n_{l,m}=|\gamma^n_{1,l,m}|+|\gamma^n_{2,l,m}|.$$

Finally, Algorithm 1 is rewritten by modifying (3.14). This leads to the following algorithm.

Algorithm 2. *Initialization:* Set $\boldsymbol{\psi}^{n,(0)}_{l,m}=\boldsymbol{\psi}^{n-1}_{l,m}$ for all l,m.
Iteration: Compute iteratively

$$\begin{aligned}
\boldsymbol{\psi}^{n+1,(k+1)}_{l,m} &= \mathbf{D}^{-1}_{imp}\Big(\mathbf{R}^n_{l,m}+\mathbf{L}_1(\boldsymbol{\psi}^{n+1,(k)}_{l-1,m})+\mathbf{L}_2(\boldsymbol{\psi}^{n+1,(k)}_{l,m-1}) && (4.8)\\
&\qquad +\mathbf{D}_{exp}(\boldsymbol{\psi}^{n+1,(k)}_{l,m})+\mathbf{R}_1(\boldsymbol{\psi}^{n+1,(k)}_{l-1,m})+\mathbf{R}_2(\boldsymbol{\psi}^{n+1,(k)}_{l,m-1})\Big),
\end{aligned}$$

until convergence.

Remark 4.1.
- We artificially add the parameter $\alpha^n Id$ on both sides of (4.7) when splitting the operator $\mathbf{D}$ in two parts. This enforces the preservation of the realizability property and makes the algorithm more stable. However, this reduces the convergence rate of Algorithm 2.
- One may reproduce the computations from the proof of Proposition 3.1 with (4.8) to show that the realizability property is preserved from one iteration to another in Algorithm 2 and to show it is convergent.

4.3 Numerical experiment in 2D: a photon beam in water

In this test case, photons are injected in a 2D homogeneous domain composed of water. The size of the medium is 2 cm × 10 cm. We inject a 0.5 cm large beam of 500 keV photons on the left boundary modeled by the following incoming boundary condition

$$\begin{aligned}
\text{for}\quad (X,\Omega)\in\Gamma^- &= \left\{(X,\Omega)\in\partial Z\times\mathcal{S}^2 \quad \text{s.t.}\quad n(X).\Omega<0\right\},\\
\psi_\gamma(X,\epsilon,\Omega) &= 10^{10}\exp\left(-\alpha_\epsilon(\epsilon-\epsilon_0)^2\right)\exp\left(-\alpha_\mu(\Omega_1-1)^2\right)\mathbf{1}_B(X)+\delta\mathbf{1}_{\partial Z\backslash B}(X),\\
\psi_e(X,\epsilon,\Omega) &= \delta,\\
B &= \left\{(x,y),\quad x=0,\quad y\in[0.75\text{ cm},1.25\text{ cm}]\right\},
\end{aligned}$$

where $n(X)$ is the outgoing normal, $\epsilon_0=500$ keV, $\alpha_\epsilon=20000$, $\alpha_\mu=500000$ and $\delta=10^{-15}$. And we used the moments of those distributions as boundary conditions for the moment equations.

Through this test case, we aim to highlight the influence of the choice of the parameters $c^\pm$ on the convergence rate of Algorithm 1. The influence of $c^\pm$ on the dose results were studied in [55] and are only recalled here for completeness. We test the present method using two sets of parameters $c_i^\pm$:

- First, we fix $|c_i^\pm|=2$ at a sufficiently large value (afterward called large c) such that the conditions (4.2) are satisfied. Those parameters were shown to provide an overestimated numerical angular diffusion in [55] (see also Fig. 5 below).
- Second, we fix $c_i^\pm$ at a value closer to the value actual eigenvalues of the Jacobian of the flux as proposed in [55] (afterward called small c). Those parameters were shown to reduce the diffusion effects of the numerical method, especially in the direction orthogonal to the beam.

We compare Algorithm 2 to a reference Monte Carlo solver ([25]). The dose results obtained with those methods are gathered on Fig. 5 with the computational times in Table 1. A cut of the doses along the axis of the beam $y=1$ cm and in the transverse direction at depth $x=2$ cm and $x=8$ cm are shown on Fig. 6.

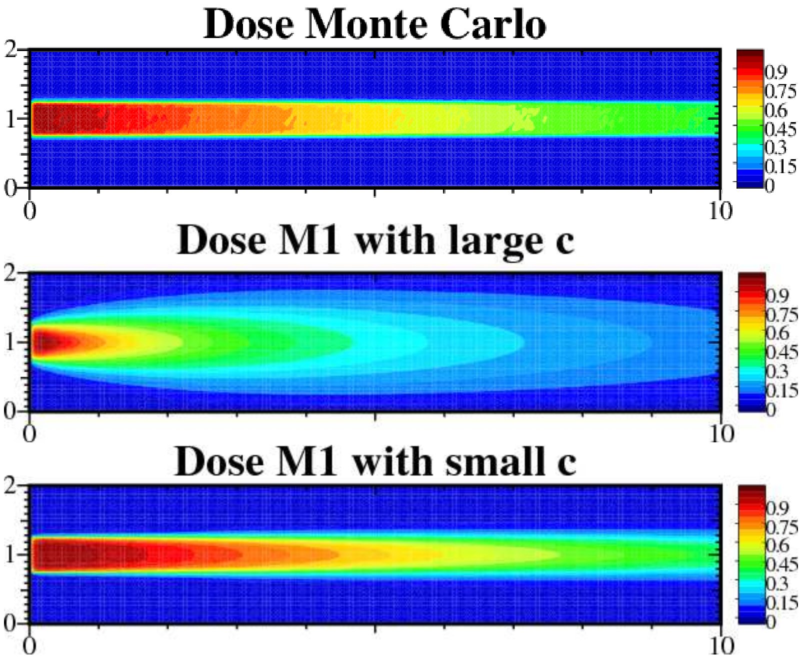

Figure 5: Doses obtained with the Monte Carlo solver (top) and the M_1 solver with large c (middle) and small c (below) relaxation parameters, normalized by their maximum value.

Solver	Monte Carlo	M_1 with large c	M_1 with small c
Computation times	14 hours	49.78699 sec	204.1239 sec

Table 1: Computational times with the Monte Carlo solver and the implicit solver with the different c.

The dose results with the modified relaxation parameters are much closer to the reference Monte Carlo results. As expected, the dose is less diffused with the small c.

Due to the noise in the Monte Carlo and the normalization by $\max D$, the M_1 dose curves with the modified relaxation parameters are slightly above the Monte Carlo reference on Fig. 6.

Following Remark 4.1, we observe that the computational time is higher with the small c than with the large ones. Those times remain much lower than the one with the Monte Carlo reference.

4.4 Numerical experiment in 3D: a photon beam in a chest

This test case aims to exhibit the efficiency of our method when considering more complex density maps. For this purpose, we use a density map obtained from a computed tomography (CT) scan of a chest. This map is depicted on Fig. 7. This domain is a 29.5

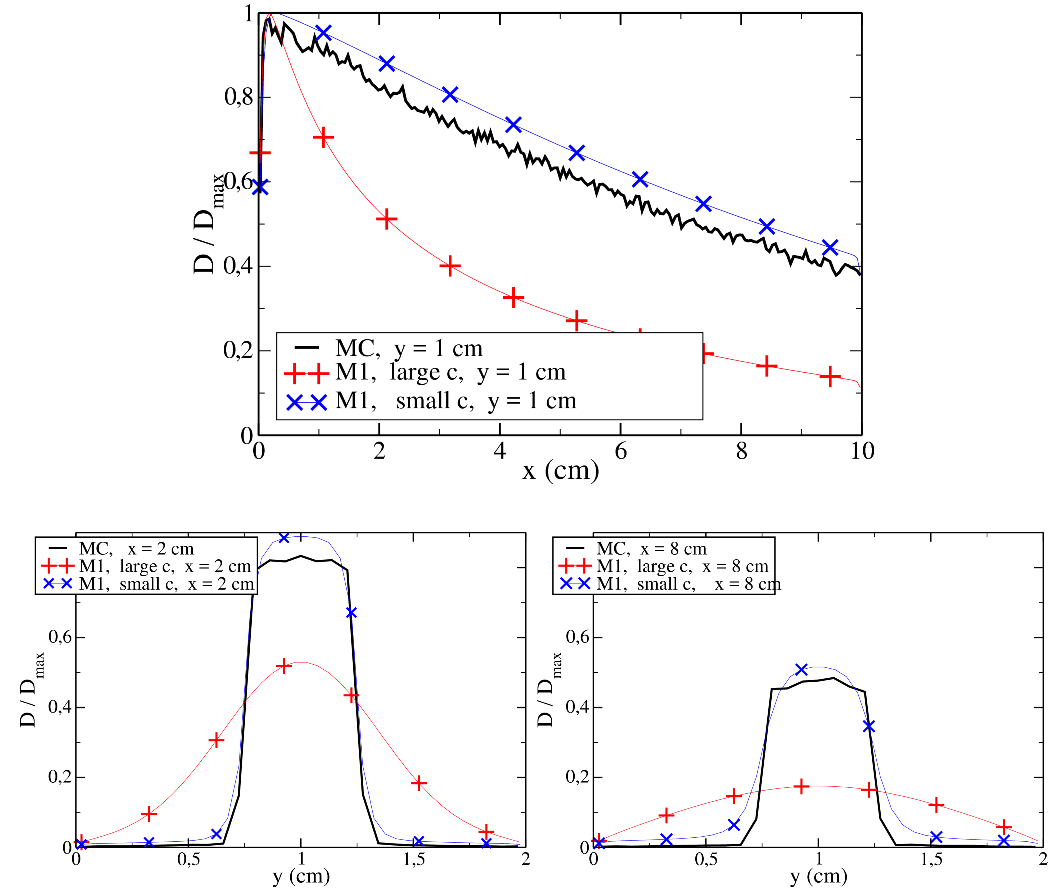

Figure 6: Doses obtained with the Monte Carlo solver and the M_1 solver with large and small c along the axis of the beam (top) and the axis transverse to the beam at 2 cm depth (below left) and 8 cm depth (below right), normalized by their maximum value.

cm deep cube.

We impose a beam of photons on the boundary of the medium to pass through the ribs. In order to reduce the computational time, we perform the computations on a smaller domain of size 14 cm × 25 cm × 11.35 cm in which the density of photons is non-negligible. This domain is meshed with $140 \times 220 \times 50$ cells.

The beam is modeled by the following condition over Γ^-

$$\begin{aligned}\psi_\gamma(X,\epsilon,\Omega) &= 10^{10}\exp\left(-\alpha_\epsilon(\epsilon-\epsilon_0)^2\right)\exp\left(-\alpha_\mu(\Omega_2-1)^2\right)\mathbf{1}_B(X)+\delta\mathbf{1}_{\partial Z\setminus B}(X),\\ \psi_e(X,\epsilon,\Omega) &= \delta,\end{aligned}$$

$$B=\left\{(x,y,z),\quad x\in[6\text{ cm},8\text{ cm}],\ y=0\text{ cm},\ z\in[4\text{ cm},6\text{ cm}]\right\},$$

where $n(x)$ is the outgoing normal, $\epsilon_0 = 1$ MeV, $\alpha_\epsilon = 20000$, $\alpha_\mu = 3000$ and $\delta = 10^{-15}$.

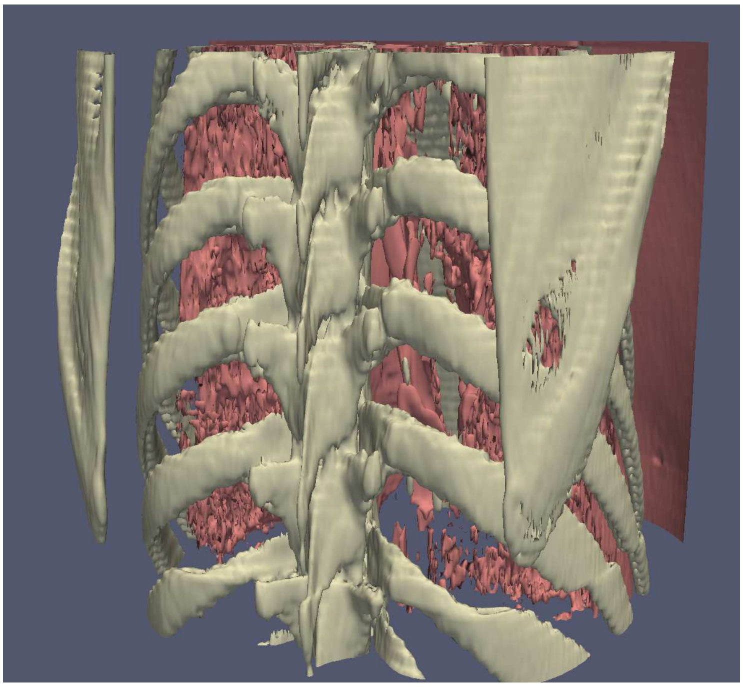

Figure 7: Density map represented by isosurfaces of density 1.8 (ivory; equivals to bone density) and 0.3 (flesh color; equivals to lung density).

And we use the moments of those distributions as boundary conditions for the moment equations.

The maximum residual is fixed at $r_{\max} = 10^{-1}$.

We perform the computations with the small parameters c. Those dose results are depicted on Fig. 8 as isodose surfaces cut along the surfaces parameterized by $\{x = 5$ cm$\}$, $\{y = 12.5$ cm$\}$ and $\{z = 5.675$ cm$\}$, that are along the axis of the beam or at half depth in the domain.

This test case is a rather complex problem with a large 3D mesh, a complete physics (photons and electrons together) is considered and we used the large c. This corresponds to the most complex settings with the present approach. We perform the computations in parallel on four cores and the computations required 2h and 53 min. Such a computational time remains too long for practical applications in medical physics. Several features in our approach can be improved to reduce drastically the computational times. We present in the next paragraph one common idea to improve it.

4.5 Mesh adaptation

In practice, the moments $\boldsymbol{\psi}$ are computed in the whole space-energy domain, even in cells where they are negligible. This results in solving (4.8) over a very large vector $\boldsymbol{\psi}^n$.

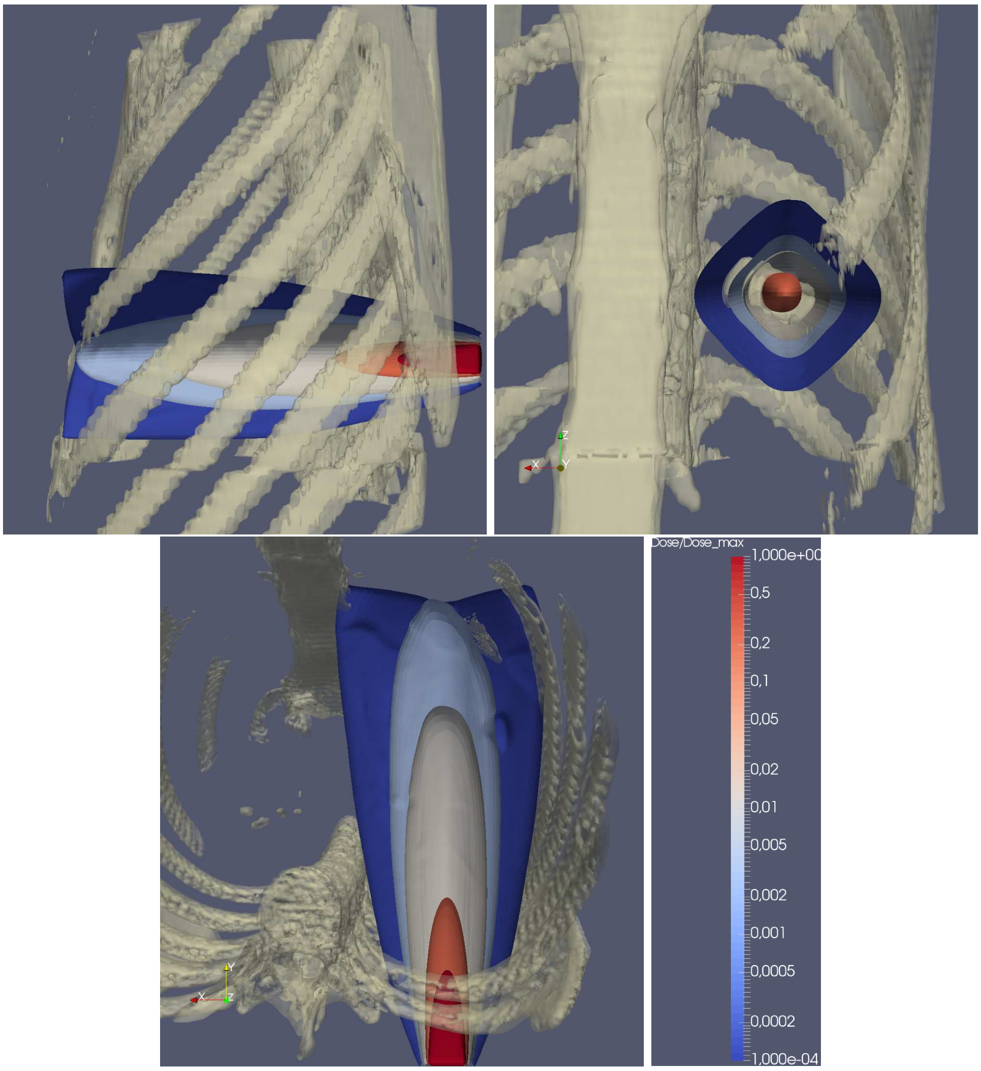

Figure 8: Isosurfaces of dose at 60%, 30%, 6%, 3% and 0.6% of the maximum dose in a chest.

Those non-necessary computations raise the size of the system to solve and reduce the convergence rate of our algorithm (see Remark 3.4).

The computational times can be considerably reduced using common code optimization techniques. The most common idea for such a problem is the mesh adaptation, here over the space and the energy. Instead of computing $\boldsymbol{\psi}$ in the whole domain, one may solve (4.8) on a smaller domain where $\boldsymbol{\psi}$ is expected not to be negligible. This would lead to a considerable reduction of both the computational times and the memory requirements.

As a comparison, this adaptation of the mesh size is used in the industrial code

Acuros® [26, 48, 63]. This code is based on a S_N discretization of (2.1), *i.e.* a direct discretization with respect to the Ω variable. Such a discretization is expected to require more computational power than the present moment method. This code commonly requires only a few minutes to compute the dose deposited by a unique beam in 3D. We expect better time comparisons after performing similar code optimization. The comparison of our code with Acuros® and the adaptation of the mesh size are parts of our perspectives.

5 Conclusion and perspectives

The present approach aims to circumvent restrictive stability conditions of numerical schemes for transport equations in the field of radiotherapy dose computations. This restriction arises when considering low collisional media.

First, we performed an angular moment extraction leading to a M_1 system of equations. Such models require lower computational cost to solve, but are valid only under a realizability condition.

Then, we proposed a numerical scheme for M_1 systems by the use of a relaxation method. Using an implicit scheme on the relaxed equations leads to a numerical method that does not present such a stiffness. However, the resulting discrete equations are nonlinear on the unknowns. We constructed an iterative method to compute this solution. During the construction of the scheme and of the iterative method, we specially focused on the preservation of the realizability property.

Numerical experiments show that our method behaves appropriately in academic and on practical cases in 1D, 2D and 3D. The convergence of the method is tested in 1D. Despite removing the restrictive stability constraints, the present approach was compared a reference Monte Carlo method and it requires much lower computational costs.

These computational costs remain too high for practical application in medical centers. Several features in our approach may improved in order to accelerate the computations. First, the iterative method used to solve non-linear discrete equations can be improved by initializing it with more clever values, and by using higher order methods (here only a first order method was tested). Second, the size of the computational grid being rather large for 3D problems, such a computational domain may be automatically adapted to the domain of interest where the solution is non-negligible. Together with comparisons to industrial codes used in medical centers, these are the main objectives of the project at that stage.

Acknowledgement

The authors are grateful to Gabriele Birindelli (Université de Bordeaux) and Kerstin Küpper (RWTH Aachen) for performing the PENELOPE simulations, and to Martin Frank

(RWTH Aachen) for fruitful discussions. This work is partially funded by Aquitaine Region and FEDER fund through IOPRA interface. T. Pichards PhD and postdoctoral research are funded respectively by IdEx Bordeaux and Aquitaine Region; and by CEA and LJLL through LRC Manon.

References

[1] D. Aregba-Driollet and R. Natalini. Discrete kinetic schemes for multidimensional systems of conservation laws. *SIAM J. Numer. Anal.*, 6:1973–2004, 2000.

[2] D. Aregba-Driollet, R. Natalini, and S. Tang. Explicit diffusive kinetic schemes for nonlinear degenerate parabolic systems. *Math. Comp.*, 73:63–94, 2004.

[3] O. Axelsson. *Iterative solution methods*. Cambridge University Press, New York, NY, USA, 1994.

[4] C. Berthon, P. Charrier, and B. Dubroca. An HLLC scheme to solve the M_1 model of radiative transfer in two space dimensions. *J. Sci. Comput.*, 31(3):347–389, 2007.

[5] C. Berthon, M. Frank, C. Sarazin, and R. Turpault. Numerical methods for balance laws with space dependent flux: Application to radiotherapy dose calculation. *Commun. Comput. Phys.*, 10(5), 2011.

[6] G. Birindelli, J.-L. Feugeas, J. Caron, B. Dubroca, G. Kantor, J. Page, T. Pichard, V. T. Tikhonchuk, and Ph. Nicolaï. High performance modelling of the transport of energetic particles for photon radiotherapy. *Phys. Med., spec. issue SFPM 2016 conference*, to appear.

[7] R. Borsche, J. Kall, A. Klar, and T.N.H. Pham. Kinetic and related macroscopic models for chemotaxis on networks. *Math. Mod. Meth. Appl. S.*, 26(6):1219–1242, 2016.

[8] J. Borwein and A. Lewis. Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.*, 29(2):325–338, 1991.

[9] J. Borwein and A. Lewis. Partially finite convex programming, part I: Quasi relative interiors and duality theory. *Math. Program.*, 57:15–48, 1992.

[10] J. Borwein and A. Lewis. Partially finite convex programming: Part II. *Math. Program.*, 57:49–83, 1992.

[11] F. Bouchut. Construction of BGK models with a family of kinetic entropies for a given system of conservation laws. *J. Stat. Phys.*, 95(1):113–170, 1998.

[12] F. Bouchut, F. R. Guarguaglini, and R. Natalini. Diffusive BGK approximations for nonlinear multidimensional parabolic equations. *Indiana Univ. Math. J.*, 49:723–749, 2000.

[13] T. A. Brunner and J. P. Holloway. One-dimensional riemann solvers and the maximum entropy closure. *J. Quant. Spectros. Radiat. Transfer*, 69(5):543 – 566, 2001.

[14] J. Caron, J.-L. Feugeas, B. Dubroca, G. Kantor, C. Dejean, G. Birindelli, T. Pichard, Ph. Nicolaï, E. d'Humières, M. Frank, and V. Tikhonchuk. Deterministic model for the transport of energetic particles: Application in the electron radiotherapy. *Phys. Medica*, 31(8):912–921, 2015.

[15] CERN. *Geant4 user's guide for application developers*, 2015.

[16] S. Chandrasekhar. On the radiative equilibrium of a stellar atmosphere. *Astrophys. J.*, 99:180 – 190, 1943.

[17] S. Chandrasekhar. On the radiative equilibrium of a stellar atmosphere X. *Astrophys. J.*, 103:351 – 370, 1946.

[18] D. Coulette, E. Franck, P. Helluy, M. Mehrenberger, and L. Navoret. Palindromic discontinuous Galerkin method for kinetic equations with stiff relaxation. *Hal archives ouvertes*, 2016.

[19] R. Curto and L. A. Fialkow. Recusiveness, positivity, and truncated moment problems. *Houston J. Math.*, 17(4):603–634, 1991.

[20] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology: Volume 6, Evolution problems II.* Springer, 2000.

[21] B. Dubroca and J.-L. Feugeas. Hiérarchie des modèles aux moments pour le transfert radiatif. *C. R. Acad. Sci. Paris*, 329:915–920, 1999.

[22] B. Dubroca and M. Frank. An iterative method for transport equations in radiotherapy. *Progress in Industrial Mathematics at ECMI 2008*, pages 407–412, 2010.

[23] R. Duclous, B. Dubroca, and M. Frank. A deterministic partial differential equation model for dose calculation in electron radiotherapy. *Phys. Med. Biol.*, 55:3843–3857, 2010.

[24] L. Elsner and V. Mehrmann. Convergence of block iterative methods for linear systems arising in the numerical solution of euler equations. *Numer. Math.*, 59(1):541–559, 1991.

[25] J. Sempau F. Salvat, J. M. Fernández-Varea. *PENELOPE-2011: A code system for Monte Carlo simulation of electron and photon transport*, 2011.

[26] G.A. Failla, T.A. Wareing, Y. Archambault, and S. Thompson. Acuros XB ® advanced dose calculation for the Eclipse ^{TM} treatement planning system. *Clinical perspectives*, 2010.

[27] K. O. Friedrichs and P. D. Lax. Systems of conservation equations with a convex extension. *Proc. Nat. Acad. Sci.*, 68(8):1686–1688, 1971.

[28] B. D. Ganapol, C. T. Kelley, and G. C. Pomraning. Asymptotically exact boundary conditions for the P_N equations. *Nucl. Sci. Eng.*, 114, 1993.

[29] C. Groth and J. McDonald. Towards physically realizable and hyperbolic moment closures for kinetic theory. *Cont. Mech. Therm.*, 21(6):467–493, 2009.

[30] T. Hanawa and E. Audit. Reformulation of the M_1 model of radiative transfer. *J. Quant. Spectros. Radiat. Transfer*, 145:9 – 16, 2014.

[31] C. Hauck. *Entropy-based moment closures in semiconductor models*. PhD thesis, University of Maryland, 2006.

[32] C. D. Hauck, C. D. Levermore, and A. L. Tits. Convex duality and entropy-based moment closures: Characterizing degenerate densities. *SIAM J. Control Optim.*, 2007.

[33] M. Junk. Maximum entropy for reduced moment problems. *Math. Mod. Meth. Appl. S.*, 10(1001–1028):2000, 1998.

[34] I. Kawrakow and D. W. Rogers. *The EGSnrc code system*, 2013.

[35] C. Kelley. *Iterative methods for linear and nonlinear equations.* SIAM, 1995.

[36] D. Kershaw. Flux limiting nature's own way. Technical report, Lawrence Livermore Laboratory, 1976.

[37] Los Alamos National Laboratory. *MCNP - A general Monte Carlo N-particle transport code, Version 5*, 2003.

[38] E. W. Larsen and G. C. Pomraning. The P_N theory as an asymptotic limit of transport theory in planar geometry I: Analysis. *Nucl. Sci. Eng.*, 109(49), 1991.

[39] T. Leroy, R. Duclous, B. Dubroca, V.T. Tikhonchuk, S. Brull, A. Decoster, M. Lobet, and L. Gremillet. Deterministic and stochastic kinetic descriptions of electron-ion Bremsstrahlung: From thermal to non-thermal regimes, 2014. Project at CEMRACS.

[40] R. J. LeVeque. *Finite Volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[41] C. D. Levermore. Relating eddington factors to flux limiters. *J. Quant. Spectrosc. Radiat. Transfer*, 31(2):149–160, 1984.

[42] C. D. Levermore. Moment closure hierarchies for kinetic theories. *J. Stat. Phys.*, 83(5–6):1021–1065, 1996.

[43] E. E. Lewis and W. F. Miller. *Computational methods of neutron transport*. American nuclear society, 1993.

[44] P. Mayles, A. Nahum, and J.C. Rosenwald, editors. *Handbook of radiotherapy physics: Theory and practice*. Taylor & Francis, 2007.

[45] J. Mcdonald and M. Torrilhon. Affordable robust moment closures for CFD based on the maximum-entropy hierarchy. *J. Comput. Phys.*, 251:500–523, 2013.

[46] G. N. Minerbo. Maximum entropy Eddington factors. *J. Quant. Spectros. Radiat. Transfer*, 20:541–545, 1978.

[47] R. Natalini. A discrete kinetic approximation of entropy solutions to multidimensional scalar conservation laws. *J. Differ. Equations*, 148(2):292 – 317, 1998.

[48] J. Ojala. The accuracy of acuros xb algorithm in external beam radiotherapy - a comprehensive review. Technical report, Varian Medical System, 2014.

[49] E. Olbrant and M. Frank. Generalized Fokker-Planck theory for electron and photon transport in biological tissues: Application to radiotherapy. *Comput. Math. Methods Med.*, 11(4):313–339, 2010.

[50] J. Page, J.-L. Feugeas, G. Birindelli, J. Caron, B. Dubroca, G. Kantor, T. Pichard, V.T. Tikhonchuk, and Ph. Nicolaï. New fast entropic algorithm for MRI-guided radiotherapy. *Phys. Med., spec. issue SFPM 2016 conference*, to appear.

[51] T. Pichard. *Mathematical modelling for dose deposition in photontherapy*. PhD thesis, Université de Bordeaux and RWTH Aachen University, 2016.

[52] T. Pichard, G.W. Alldredge, S. Brull, B. Dubroca, and M. Frank. The M_2 model for dose simulation in radiation therapy. *J. Comput. Theor. Transport, spec. issue ICTT 2015 conference*, 2016.

[53] T. Pichard, G.W. Alldredge, S. Brull, B. Dubroca, and M. Frank. An approximation of the M_2 closure: Application to radiotherapy dose simulation. *J. Sci. Comput*, 71(1):71–108, 2017.

[54] T. Pichard, D. Aregba-Driollet, S. Brull, B. Dubroca, and M. Frank. Relaxation schemes for the M_1 model with space-dependent flux: Application to radiotherapy dose calculation. *Commun. Comput. Phys.*, 19:168–191, 2016.

[55] T. Pichard, S. Brull, B. Dubroca, and M. Frank. On the transverse diffusion of beams of photons in radiation therapy. *Proc. HYP2016 conference*, to appear.

[56] G. C. Pomraning. *The equations of radiation hydrodynamics*. Pergamon Press, 1973.

[57] S. La Rosa, G. Mascali, and V. Romano. Exact maximum entropy closure of the hydrodynamical model for Si semiconductors: The 8-moment case. *SIAM J. Appl. Math.*, 70(3):710–734, 2009.

[58] R. P. Rulko, E. W. Larsen, and G. C. Pomraning. The P_N theory as an asymptotic limit of transport theory in planar geometry II: Numerical results. *Nucl. Sci. Eng.*, 109(76), 1991.

[59] J. Schneider. Entropic approximation in kinetic theory. *ESAIM-Math. Model. Num.*, 38(3):541–561, 2004.

[60] J. Tervo, P. Kokkonen, M. Frank, and M. Herty. On existence of L^2-solutions of coupled Boltzmann continuous slowing down transport equation system. *arxive*, 2016.

[61] E. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer, 2009.

[62] A. van der Sluis. Gershgorin domains for partitioned matrices. *Linear Algebra Appl.*, 26:265 – 280, 1979.

[63] O. N. Vassiliev, T. A. Wareing, J. McGhee, G. Failla, M. R. Salehpour, and F. Mourtada. Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. *Phys. Med. Biol.*, 55:581–598, 2010.

[64] M.N. Vrahatis, G.D. Magoulas, and V.P. Plagianakos. From linear to nonlinear iterative

methods. *Appl. Numer. Math.*, 45(1):59 – 77, 2003.

[65] C. Zankowski, M. Laitinen, and H. Neuenschwander. Fast electron Monte Carlo for Eclipse$^{\text{TM}}$. Technical report, Varian Medical System.