Sparse Adaptive Finite Elements
for
Radiative Transfer[1]

G. Widmer, R. Hiptmair and C. Schwab

Sparse Adaptive Finite Elements for Radiative Transfer[1]

G. Widmer, R. Hiptmair and C. Schwab

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Abstract

The linear radiative transfer equation, a partial differential equation for the radiation intensity $u(\mathbf{x}, \mathbf{s})$, with independent variables $\mathbf{x} \in D \subset \mathbb{R}^n$ in the physical domain D of dimension $n = 2, 3$, and angular variable $\mathbf{s} \in S^2 := \{\mathbf{y} \in \mathbb{R}^3 : |\mathbf{y}| = 1\}$, is solved in the $n+2$-dimensional computational domain $D \times S^2$.

We propose an adaptive multilevel Galerkin FEM for its numerical solution. Our approach is based on a) a stabilized variational formulation of the transport operator and b) on so-called sparse tensor products of two hierarchic families of Finite Element spaces in $H^1(D)$ and in $L^2(S^2)$, respectively.

An a-priori error analysis shows, under strong regularity assumptions on the solution, that the method converges with essentially optimal asymptotic rates while its complexity grows essentially only as that for a linear transport problem in $\mathbb{R}^n$. Numerical experiments for $n = 2$ on a set of example problems agree with the convergence and complexity analysis of the method and show its performance in terms of accuracy vs. number of degrees of freedom to be superior to the discrete ordinates method.

---

1 Introduction

We are concerned with the numerical solution of the monochromatic radiative transfer equation [1] on a bounded Lipschitz domain $D \subset \mathbb{R}^n$, where $n = 2, 3$, without scattering.

We identify a direction $\mathbf{s}$ with a point on the sphere S^2 and are looking for the intensity $u(\mathbf{x}, \mathbf{s})$, satisfying

$$\mathbf{s} \cdot \boldsymbol{\nabla}_x u(\mathbf{x}, \mathbf{s}) + \kappa(\mathbf{x}) u(\mathbf{x}, \mathbf{s}) = \kappa(\mathbf{x}) f(\mathbf{x}), \quad (\mathbf{x}, \mathbf{s}) \in D \times S^2 \tag{1}$$

$$u(\mathbf{x}, \mathbf{s}) = g(\mathbf{x}, \mathbf{s}), \qquad \mathbf{x} \in \partial D, \ \mathbf{s} \cdot \mathbf{n}(\mathbf{x}) < 0 \ . \tag{2}$$

$\mathbf{n}(\mathbf{x})$ is the outer unit normal on the boundary, $\kappa \geq 0$ the absorption coefficient, $f \geq 0$ the blackbody intensity and $g \geq 0$ the wall emission. More general models involve scattering terms [1], which we ignore in this paper, because we focus on the discretization of the transport equation. We will assume "cold walls" leading to $g = 0$.

Regarding the direction $\mathbf{s}$ as a mere parameter, the equation can be solved by line integration for any given position $(\mathbf{x}, \mathbf{s})$. However, as the equation is stated in five (respectively four for $n = 2$) dimensions, this strategy is too expensive in order to compute the intensity field $u(\mathbf{x}, \mathbf{s})$ with a fine resolution.

Popular methods to solve the radiative transfer problem are, apart from Monte Carlo schemes, the method of spherical harmonics (in particular the P_1 approximation) or the discrete ordinates method. Overviews of numerical methods for radiative transfer can eg. be found in [1] and some recent developments in [2].

The method of spherical harmonics is based on a semi-discretization in the solid angle by expanding the intensity into a truncated series of spherical harmonics, which leads to a coupled system of equations in space only. For the P_1 approximation, the equations boil down to a diffusion equation. The P_N approximation is only suitable in diffuse regimes, where the intensity function is smooth with respect to the solid angle, as the approximation rate for non-smooth functions with respect to the number of spherical harmonics is very poor.

In the discrete ordinates method (often referred to as S_N), the equation is solved for N fixed directions. The method is very popular due to its simplicity, but suffers from so-called *ray effects* which require a fine angular resolution if localized emissive areas are present.

In most applications, the system of equations arising from a P_N- or an S_N approximation are then solved with finite difference or finite element schemes. In [3], for example, a least squares formulation is discretized with spherical harmonics in the solid angle and finite elements in space. Kanschat [4] uses the discrete ordinates method with FE discretization and streamline diffusion stabilization in the physical domain D.

A severe problem when solving the radiative transfer equation is the scaling of the accuracy of the solution with respect to the number of degrees of freedom in the discretization due to the high dimensionality of the problem. We will present a method to overcome this "curse of dimension" already observed in [5] for radiative transfer problems with sufficiently smooth absorption coefficients $\kappa(\mathbf{x})$ and blackbody intensity $f(\mathbf{x})$. Unlike some other methods for radiative transfer, our method does not require κ to be large.

The paper is structured as follows:

In section 2, we describe the problem setting considered in this paper. Sections 3 and 4 contain the stabilized variational formulation and the Galerkin discretization of the radiative transfer equation. For stabilization, we use a scaled least squares variational formulation. Such formulations have been used e.g. in [6] and in particular for the linear Boltzmann equation arising from the neutron transport problem in [7], [3] or [8]. While in [7] and [3] the absorption coefficient is assumed to be bounded away from zero and the authors in [3] optimize the scaling parameter to balance the absorption and scattering effects for constant coefficients, we tailor the scaling parameter to provide coercivity and continuity estimates for more general absorption coefficient functions. Our space for the ansatz- and test functions is a tensor product space of functions in space and functions in the solid angle. With piecewise constant functions in the solid angle, this formulation is equivalent to the discrete ordinates method.

The core development of this paper is contained in sections 4.1 to 4.3. There, we describe the sparse tensor product method, applied to the radiative transfer equation, including the wavelet bases we use in the underlying spaces. Our approach is based on the idea of *sparse grids*, introduced by Zenger in [9]. This method has been used for solving a wide range of high-dimensional problems, such as e.g. numerical integration [10], the Schrödinger-equation in quantum chemistry [11], elliptic [12], [13] and parabolic [14] partial differential equations, high-order FE methods [15] or integral equations [16] among others.

Our sparse tensor product method reduces the total number of degrees of freedom to the number of degrees of freedom in physical space only (up to logarithmic terms). For solutions of sufficient smoothness, we also give an error bound for the solution in the $H^{1,0}(D \times S^2)$-norm.

For problems like light beams, the sparse tensor product method is not optimal in the sense of a *best-N-term approximation* [17]. In order to also cover this class of problems, we describe a particular adaptive sparse tensor product method in section 5. Adaptive sparse grid methods have for example been applied to the Helmholtz equation [18], elliptic PDE's [12] or a singular perturbation model [19]. More general adaptive wavelet methods can e.g. be found in [20], [21], [22] and [23] and the references therein.

In contrast to other adaptive approaches in the field of radiative transfer like eg. [4], where only the mesh in the physical domain D is refined according to an a-posteriori error estimator, our method is adaptive with respect to space as well as solid angle with no a-priori knowledge of the solution.

Section 6 contains numerical results. We compare the discrete ordinates method, the sparse tensor product method and the adaptive sparse tensor product method for some model problems.

2 Radiative Transfer Equation

The general monochromatic radiative transfer equation on a bounded Lipschitz polyhedron $D \subset \mathbb{R}^3$ reads :

$$\left(\frac{1}{c} \frac{\partial}{\partial t} + \mathbf{s} \cdot \boldsymbol{\nabla}_x + \beta(\mathbf{x}, t) \right) u(\mathbf{x}, \mathbf{s}, t) = \mathcal{K}u(\mathbf{x}, \mathbf{s}, t) + q(\mathbf{x}, \mathbf{s}, t) \quad \text{in} \quad D \times S^2 \times (0, T). \qquad (3)$$

Here $u(\mathbf{x}, \mathbf{s}, t)$ is the radiation intensity at position $\mathbf{x} \in D$ into direction $\mathbf{s} \in S^2$ at time $t \in (0, T)$. If we assume that $\frac{\partial u(\mathbf{x})}{\partial z} = 0$, where $\mathbf{x} = (x, y, z)' \in \tilde{D} \times \mathbb{R}^2, \tilde{D} \subset \mathbb{R}$, the equation reduces to a 2D-problem for $u(\mathbf{x}, \mathbf{s}, t)) = u(\tilde{\mathbf{x}}, \mathbf{s}, t), \tilde{\mathbf{x}} = (x, y, 0)'$, in physical space. In that case, we ignore the third component of the inner product $\mathbf{s} \cdot \boldsymbol{\nabla}_x u$. As the numerical experiments are carried out for $n = 2$, we include that case explicitly in our analysis. For the sake of simplicity, we will denote $\tilde{D}$ by D. In (3), c denotes the signal propagation speed (typically, the speed of light) and $\beta \geq 0$ the total extinction coefficient. The quantity

$$\mathcal{K}u(\mathbf{x}, \mathbf{s}, t) := \frac{\sigma(\mathbf{x}, t)}{4\pi} \int_{S^2} P(\mathbf{x}, \mathbf{s}, \mathbf{s}') u(\mathbf{x}, \mathbf{s}', t) d\mathbf{s}'$$

where $d\mathbf{s}'$ denotes the surface measure on S^2 is the emission due to scattering with $\sigma \geq 0$ and a scattering kernel $P(\mathbf{x}, \mathbf{s}, \mathbf{s}') \geq 0$ that satisfies

$$\frac{1}{4\pi} \int_{S^2} P(\mathbf{x}, \mathbf{s}, \mathbf{s}') d\mathbf{s}' = 1 \qquad \forall \, \mathbf{s} \in S^2, \, \mathbf{x} \in D.$$

In (3), $q(\mathbf{x}, \mathbf{s}, t)$ is a known radiating source. The radiative transfer equation is complemented by initial conditions

$$u(\mathbf{x}, \mathbf{s}, 0) = u_0(\mathbf{x}, \mathbf{s}) \qquad (4)$$

and by boundary conditions. Since the differential operator on the left hand side of (3) is a linear transport operator in the direction $\mathbf{s} \in S^2$, for well-posedness boundary conditions must be

prescribed on the "inflow" part $\partial_-(D \times S^2)$ of $\partial(D \times S^2) = \partial D \times S^2$, defined by

$$\partial_-(D \times S^2) := \{(\mathbf{x}, \mathbf{s}) \in \partial D \times S^2 : \mathbf{x} \in \Gamma_-(\mathbf{s})\}, \tag{5}$$

where

$$\Gamma_-(\mathbf{s}) := \{\mathbf{x} \in \partial D : \mathbf{s} \cdot \mathbf{n}(\mathbf{x}) < 0\} \subset \partial D, \quad \mathbf{s} \in S^2, \tag{6}$$

$\mathbf{n}(\mathbf{x})$ being the outer unit normal to D at the point $\mathbf{x} \in \partial D$.

The "inflow" boundary condition on $\partial_-(D \times S^2)$ reads then

$$u(\mathbf{x}, \mathbf{s}, t) = g(\mathbf{x}, \mathbf{s}, t), \text{ for } \mathbf{x} \in \Gamma_-(\mathbf{s}), \quad \mathbf{s} \in S^2, \quad t \in (0, T). \tag{7}$$

In most applications P does not depend on the position $\mathbf{x}$ and the external source q is a product of the blackbody intensity $f(\mathbf{x}, t)$ and the absorption coefficient $\kappa := \beta - \sigma$. This leads to the simplified problem

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \mathbf{s} \cdot \boldsymbol{\nabla}_x + \beta(\mathbf{x}, t)\right) u(\mathbf{x}, \mathbf{s}, t) = \frac{\sigma(\mathbf{x})}{4\pi} \int_{S^2} P(\mathbf{s}, \mathbf{s}') u(\mathbf{x}, \mathbf{s}', t) d\mathbf{s}' + \kappa(\mathbf{x}, t) f(\mathbf{x}) \tag{8}$$

$$u(\mathbf{x}, \mathbf{s}, 0) = u_0(\mathbf{x}, \mathbf{s}) \tag{9}$$

$$u(\mathbf{x}, \mathbf{s}, t) = g(\mathbf{x}, \mathbf{s}, t), \; \mathbf{x} \in \Gamma_-(\mathbf{s}), \; \mathbf{s} \in S^2, \; t \in (0, T). \tag{10}$$

If we semidiscretize this problem in time by the implicit Euler scheme with time step Δt, we obtain the stationary radiative transfer problem

$$\left(\mathbf{s} \cdot \boldsymbol{\nabla}_x + \kappa^{k+1} + \sigma^{k+1} + \frac{1}{c\Delta t}\right) u^{k+1} = \mathcal{K}^{k+1} u^{k+1} + \kappa^{k+1} f^{k+1} + \frac{u^k}{c\Delta t}, \tag{11}$$

in each time step k, where u^k is the solution from the previous time step.

Equation (11) is a mere perturbation of the steady state solution of (8). When c is much larger than the characteristic velocity of a process, the changes subject to radiative effects can be considered instantaneous. In this case, the perturbation can be neglected.

For the sake of simplicity, we also neglect scattering. This assumption is valid for many applications. However, we are convinced that the method will also work for scattering media - with minor changes to the stabilized variational formulation - as long as the scattering kernel has a smoothing effect on the solution. For the remainder of the paper, we consider the *simplified, stationary radiation transfer problem*:

Find the intensity $u(\mathbf{x}, \mathbf{s}) : D \times S^2 \to \mathbb{R}$ such that

$$(\mathbf{s} \cdot \boldsymbol{\nabla}_x + \kappa(\mathbf{x})) u(\mathbf{x}, \mathbf{s}) = \kappa(\mathbf{x}) f(\mathbf{x}) \tag{12}$$

$$u(\mathbf{x}, \mathbf{s}) = g(\mathbf{x}, \mathbf{s}), \quad \mathbf{x} \in \Gamma_-(\mathbf{s}), \; \mathbf{s} \in S^2. \tag{13}$$

3 Stabilized Variational Formulation

When regarding $\mathbf{s}$ as a mere parameter, the radiative transfer equation (12) reduces to a linear convection equation for the directed intensity $u(\cdot, \mathbf{s})$. It is well known that its standard, continuous Galerkin discretization is unstable (e.g., [24]). We use the stabilized variational formulation of (12) proposed in [3].

We seek $u : D \times S^2 \mapsto \mathbb{R}$ as the minimizer of the quadratic least squares functional

$$J(u) := \left(\epsilon(\mathbf{s} \cdot \boldsymbol{\nabla}_x u + \kappa u - \kappa f), \mathbf{s} \cdot \boldsymbol{\nabla}_x u + \kappa u - \kappa f\right)_{L^2}, \tag{14}$$

where $\epsilon \in L^\infty(D)$ is a strictly positive scaling function which will be specified later on. In (14), we adopted the notation

$$(u, v)_{L^2} := (u, v)_{L^2(D \times S^2)} = \int_D \int_{S^2} u\, v\, d\mathbf{s}\, d\mathbf{x}$$

and the associated L^2-norm will be denoted by $||\cdot||$.

For the proper setting of this minimization problem as well as of the FEM below, we define the Hilbert spaces

$$\mathcal{V} := \{u \in L^2(D \times S^2) : \mathbf{s} \cdot \boldsymbol{\nabla}_x u \in L^2(D \times S^2)\} \tag{15}$$

and

$$\mathcal{V}_0 := \left\{u \in L^2(D \times S^2) : \mathbf{s} \cdot \boldsymbol{\nabla}_x u \in L^2(D \times S^2); \; u = 0 \text{ on } \Gamma_-(\mathbf{s}), \; \mathbf{s} \in S^2\right\}. \tag{16}$$

For $n = 2$ we further require that there is a constant $C < \infty$ such that

$$||u||^2 \leq C|| \sin \theta u||^2 \tag{17}$$

with $\mathbf{s} = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)'$, $0 \leq \phi \leq 2\pi$, $0 \leq \theta \leq \pi$.

We equip $\mathcal{V}$ in (15) with the norm $||.||_S$, defined by

$$||u||_S^2 := ||\mathbf{s} \cdot \boldsymbol{\nabla}_x u||^2 + ||u||^2. \tag{18}$$

Functions $v \in \mathcal{V}$ admit traces on certain parts of $\partial(D \times S^2)$. Following [3], we define for $(\mathbf{x}, \mathbf{s}) \in \partial_-(D \times S^2)$

$$l(\mathbf{x}, \mathbf{s}) := \left\{ \begin{array}{ll} \{\sup_t : (\mathbf{x}, \mathbf{x} + t \, (\cos(\phi), \sin(\phi))') \subset D\}, & n = 2, \\ \{\sup_t : (\mathbf{x}, \mathbf{x} + t\mathbf{s} \subset D\}, & n = 3, \end{array} \right.$$

as the length of the line segment in D that starts at $\mathbf{x}$ on the inflow boundary and proceeds in the direction $\mathbf{s}$ (or its projection onto D for $n = 2$). We furthermore define the weighted norm $||.||_{\partial_-}$ on the inflow boundary:

$$||v||_{\partial_-}^2 := \int_{\mathbf{s} \in S^2} \int_{\mathbf{x} \in \Gamma_-(\mathbf{s})} |v(\mathbf{x}, \mathbf{s})|^2 \, l(\mathbf{x}, \mathbf{s}) |\, \mathbf{n}(\mathbf{x}) \cdot \mathbf{s}| \, d\mathbf{x} \, d\mathbf{s}. \tag{19}$$

Proposition 3.1 *For $v \in \mathcal{V}$, the trace $v|_{\partial_-(D \times S^2)}$ exists in $L^2(\partial_-(D \times S^2))$ and for any bounded domain D ($\mathrm{diam}(D) < d < \infty$) it holds that*

$$||v||_{\partial_-}^2 \leq (d+1) \, ||v||_S^2.$$

Proof:

a) For $n = 3$, we follow the proof of [3] theorem 2.2 and obtain

$$\begin{aligned} ||v||_{\partial_-}^2 &\leq 2d \int_{S^2} \int_D |(\mathbf{s} \cdot \boldsymbol{\nabla}_x v(\mathbf{x}, \mathbf{s})) \, v(\mathbf{x}, \mathbf{s})| \, d\mathbf{x} \, d\mathbf{s} + ||v||^2 \\ &\leq 2d \left(||\mathbf{s} \cdot \boldsymbol{\nabla}_x v(\mathbf{x}, \mathbf{s})|| \, ||v||\right) + ||v||^2 \leq (d+1)||v||_S^2. \end{aligned}$$

b) By definition of the radiative transfer equation in two dimensions $\frac{\partial u(\mathbf{x}, \mathbf{s})}{\partial z} = 0$ with $\mathbf{x} = (x, y, z)'$. We define $\hat{\mathbf{x}} := (x, y, 0)'$ and introduce spherical coordinates $\mathbf{s} = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)'$, $0 \leq \phi \leq 2\pi$, $0 \leq \theta \leq \pi$ and $\hat{\mathbf{s}} := (\cos \phi, \sin \phi, 0)'$.

We adapt the proof in a) to the two-dimensional case and first state that for a given direction $\mathbf{s}$ and integrable function $f(\hat{\mathbf{x}})$ the following holds:

$$\int_D f(\hat{\mathbf{x}}) \, d\hat{\mathbf{x}} = \int_{\hat{\mathbf{x}} \in \Gamma_-(\mathbf{s})} \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} f(\hat{\mathbf{x}} + t\hat{\mathbf{s}}) |\mathbf{n}(\hat{\mathbf{x}}) \cdot \hat{\mathbf{s}}| \, dt \, d\hat{\mathbf{x}}.$$

We define

$$\tilde{t} := \min\{\arg\min_t\{|v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s})|; \hat{\mathbf{x}} + t \sin \theta \, \hat{\mathbf{s}} \in \bar{D}\}\}$$

and

$$\mathbf{x}_{min} := \hat{\mathbf{x}} + \tilde{t}\mathbf{s}$$

4

as the point where $v(\hat{\mathbf{x}} + t\mathbf{s})$ takes on its minimum the first time.

Then

$$
\begin{aligned}
v(\hat{\mathbf{x}}, \mathbf{s})^2 &= v(\mathbf{x}_{min}, \mathbf{s})^2 - \int_0^{\tilde{t}} \frac{\partial}{\partial t} v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s})^2 \, dt \\
&= v(\mathbf{x}_{min}, \mathbf{s})^2 - 2 \int_0^{\tilde{t}} \left(\frac{\partial}{\partial t} v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s}) \right) v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s}) \, dt \\
&= v(\mathbf{x}_{min}, \mathbf{s})^2 - 2 \int_0^{\tilde{t}} (\mathbf{s} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s})) \, v(\hat{\mathbf{x}} + t\mathbf{s}, \mathbf{s}) \, dt.
\end{aligned}
$$

With the substitution $\hat{t} = t \sin\theta$, we obtain

$$
\begin{aligned}
v(\hat{\mathbf{x}}, \mathbf{s})^2 &\leq v(\mathbf{x}_{min}, \mathbf{s})^2 + 2 \int_0^{\tilde{t}\sin\theta} |\hat{\mathbf{s}} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s}) \, d\hat{t} \\
&\leq v(\mathbf{x}_{min}, \mathbf{s})^2 + 2 \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} |\hat{\mathbf{s}} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})||v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, d\hat{t}
\end{aligned}
$$

and

$$
v(\mathbf{x}_{min}, \mathbf{s})^2 \, l(\hat{\mathbf{x}}, \mathbf{s}) \leq \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})^2 \, d\hat{t}.
$$

Therefore

$$
\begin{aligned}
&\int_{S^2} \int_{\Gamma_-(\mathbf{s})} |v(\hat{\mathbf{x}}, \mathbf{s})|^2 l(\hat{\mathbf{x}}, \mathbf{s}) |\hat{\mathbf{n}}(\hat{\mathbf{x}}) \cdot \mathbf{s}| \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
\leq \quad &\int_{S^2} \int_{\Gamma_-(\mathbf{s})} \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})^2 |\hat{\mathbf{n}}(\hat{\mathbf{x}}) \cdot \hat{\mathbf{s}}| \sin\theta \, d\hat{t} \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
+ \quad &2d \int_{S^2} \int_{\Gamma_-(\mathbf{s})} \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} |\hat{\mathbf{s}} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, |v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, |\hat{\mathbf{n}}(\hat{\mathbf{x}}) \cdot \mathbf{s}| \, d\hat{t} \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
\leq \quad &\int_{S^2} \int_D v(\hat{\mathbf{x}}, \mathbf{s})^2 \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
+ \quad &2d \int_{S^2} \int_{\Gamma_-(\mathbf{s})} \int_0^{l(\hat{\mathbf{x}}, \mathbf{s})} |\mathbf{s} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, |v(\hat{\mathbf{x}} + \hat{t}\hat{\mathbf{s}}, \mathbf{s})| \, |\hat{\mathbf{n}}(\hat{\mathbf{x}}) \cdot \hat{\mathbf{s}}| \, d\hat{t} \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
= \quad &||v||^2 + 2d \int_{\mathbf{s} \in S^2} \int_D |\mathbf{s} \cdot \boldsymbol{\nabla}_x v(\hat{\mathbf{x}}, \mathbf{s})| \, |v(\hat{\mathbf{x}}, \mathbf{s})| \, d\hat{\mathbf{x}} \, d\mathbf{s} \\
\leq \quad &||v||^2 + 2d||v|| ||\mathbf{s} \cdot \boldsymbol{\nabla}_x v|| \leq (d+1)||v||_S^2
\end{aligned}
$$

$\blacksquare$

Accordingly, the subset $\mathcal{V}_0$ of functions which vanish on $\Gamma_-(\mathbf{s})$ for almost every $\mathbf{s} \in S^2$ is a linear subspace of $\mathcal{V}$ which is closed in the norm $||.||_S$. We minimize J over the affine space $\widetilde{g} + \mathcal{V}_0$, where $\widetilde{g} \in \mathcal{V}$ is an extension (we assume to exist) of the nonzero boundary data from (2) into D. The Finite Element discretization will be based on the first order stationarity conditions of $J(u)$ satisfied by the minimizer.

We first state the Poincaré-Friedrichs inequalities for $n = 2, 3$ that we will need later on.

Lemma 3.2 *(Poincaré-Friedrichs inequality for $D \subset \mathbb{R}^n$, $n = 2, 3$)*
Let $D \subset \mathbb{R}^n$, $n = 2, 3$, be bounded $(diam(D) < d < \infty)$. Then for all $u \in \mathcal{V}_0$ the following holds:

$$
||u||^2 \leq c \, ||\mathbf{s} \cdot \boldsymbol{\nabla}_x u||^2,
$$

where

$$c = \begin{cases} Cd^2, & n = 2, \\ d^2, & n = 3. \end{cases}$$

The constant C is given in (17) below.

Proof:

a) $n = 3$:

The proof is analogous to that of Lemma 3.1 in ([3]), taking into account zero inflow boundary conditions.

b) $n = 2$:

As we assume that there exists a constant $C > 0$ such that $||u||^2 \leq C|| \sin\theta u||^2$ for all $u \in \mathcal{V}$, it is sufficient to prove that

$$|| \sin\theta\, u||^2 \leq d^2 ||\mathbf{s} \cdot \boldsymbol{\nabla}_x u||^2.$$

We use the notation introduced in the proof of proposition (3.1) and adapt the proof in a) to the two-dimensional case. For fixed $(\hat{\mathbf{x}}, \mathbf{s}) \in D \times S^2$ let $\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}) \in D \times \mathbb{R}$ denote the position, where the ray along $\mathbf{s}$ through position $\hat{\mathbf{x}}$ enters the domain $D \times \mathbb{R}$ and $\mathbf{x}_o(\hat{\mathbf{x}}, \mathbf{s}) \in D \times \mathbb{R}$ the position, where the ray along $\mathbf{s}$ through position $\hat{\mathbf{x}}$ leaves the domain $D \times \mathbb{R}$:

$$\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}) := \{\hat{\mathbf{x}} + t_i \mathbf{s}; \, t_i = \min_t \{\hat{\mathbf{x}} + t\mathbf{s} \in D \times \mathbb{R}\}\}$$

$$\mathbf{x}_o(\hat{\mathbf{x}}, \mathbf{s}) := \{\hat{\mathbf{x}} + t_o \mathbf{s}; \, t_o = \max_t \{\hat{\mathbf{x}} + t\mathbf{s} \in D \times \mathbb{R}\}\}.$$

For $u \in \mathcal{V}_0$

$$
\begin{aligned}
u(\hat{\mathbf{x}}, \mathbf{s}) &= u(\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}), \mathbf{s}) + \int_0^{-\tilde{t}} \frac{\partial}{\partial t} u(\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}) + t\mathbf{s}, \mathbf{s})\, dt \\
&= \int_0^{-\tilde{t}} \frac{\partial}{\partial t} u(\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}) + t\mathbf{s}, \mathbf{s})\, dt
\end{aligned}
$$

due to the boundary condition.

In the spherical coordinates introduced earlier on,

$$u(\mathbf{x}_i(\hat{\mathbf{x}}, \mathbf{s}) + t\mathbf{s}, \mathbf{s}) = u(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \hat{\mathbf{s}}) + t\, sin\theta\, \hat{\mathbf{s}}, \mathbf{s}),$$

where we used that $\frac{\partial u}{\partial z} = 0$. With the substitution $\hat{t} = t \sin\theta$

$$|u(\hat{\mathbf{x}}, \mathbf{s})| \leq \int_0^{-\tilde{t}\sin\theta} |\cos\phi\, u_x(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \mathbf{s}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s}) + \sin\phi\, u_y(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \mathbf{s}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s})|\, d\hat{t} \qquad (20)$$

$$\leq \int_0^{(t_o - t_i)\sin\theta} |\cos\phi\, u_x(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \mathbf{s}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s}) + \sin\phi\, u_y(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \mathbf{s}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s})|\, d\hat{t}. \qquad (21)$$

Applying Hölder's inequality leads to

$$|u(\hat{\mathbf{x}}, \mathbf{s})|^2 \leq d \int_0^{(t_o - t_i)\sin\theta} |\cos\phi\, u_x(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \hat{\mathbf{s}}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s}) + \sin\phi\, u_y(\hat{\mathbf{x}}_i(\hat{\mathbf{x}}, \hat{\mathbf{s}}) + \hat{t}\hat{\mathbf{s}}, \mathbf{s})|^2\, d\hat{t} \quad (22)$$

Therefore

$$| \sin\theta\, u(\hat{\mathbf{x}}, \mathbf{s})|^2 \leq d \int_{\hat{\mathbf{x}}_i}^{\hat{\mathbf{x}}_o} |\sin\theta \cos\phi\, u_x(\hat{\mathbf{x}}) + \sin\theta \sin\phi\, u_y(\hat{\mathbf{x}})|^2\, d\hat{\mathbf{x}}. \qquad (23)$$

After integrating over $D \times S^2$ and using Fubini's theorem, we obtain

$$\int_{D \times S^2} | \sin\theta\, u(\hat{\mathbf{x}}, \mathbf{s})|^2 d\mathbf{s}d\hat{\mathbf{x}} \leq d^2 ||\mathbf{s} \cdot \boldsymbol{\nabla}_x u||^2. \qquad (24)$$

6

■

We now introduce the bilinear form

$$a(u,v) := (\epsilon \mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v)_{L^2} + (\epsilon \mathbf{s} \cdot \boldsymbol{\nabla}_x u, \kappa v)_{L^2} + (\epsilon \kappa u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v)_{L^2} + (\epsilon \kappa u, \kappa v)_{L^2}. \qquad (25)$$

We adapt the ideas of [7] and [3] for the more general case where $\kappa(\mathbf{x})$ is not constant and set

$$\epsilon(\mathbf{x}) = \begin{cases} 1, & \kappa(\mathbf{x}) < \kappa_0, \\ \frac{1}{\kappa(\mathbf{x})}, & \kappa(\mathbf{x}) \geq \kappa_0, \end{cases} \qquad (26)$$

where $0 < \kappa_0 < 1$ is given below.

Proposition 3.3 *Assume that*

$$\kappa \in L^\infty(D), \ \kappa \geq 0.$$

Then the bilinear form $a(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ defined in (25) is continuous and coercive in the norm $\|.\|_S$, i.e. there exist constants $c_1, c_2 > 0$ that depend only on $\|\kappa\|_{L^\infty(D)}$ and $d := diam(D)$ such that for $u, v \in \mathcal{V}_0$ holds that

$$\begin{aligned} |a(u,v)| &\leq c_1 \|u\|_S \|v\|_S \\ |a(u,u)| &\geq c_2 \|u\|_S^2 \end{aligned}$$

with $c_1 = (\max\{\frac{1}{\kappa_0}, d\|\kappa\|_\infty\} + 1)$ and $c_2 = \min\{\frac{1}{d\|\kappa\|_\infty}, \kappa_0\}$ for $n = 3$,
$c_1 = (\max\{\frac{1}{\kappa_0}, \sqrt{C}d\|\kappa\|_\infty\} + 1)$ and $c_2 = \min\{\frac{1}{\sqrt{C}d\|\kappa\|_\infty}, \kappa_0\}$ for $n = 2$.

Proof: Without loss of generality we assume $d = 1$ for $n = 3$ (respectively $Cd^2 = 1$ for $n = 2$). (If $d \neq 1$, we can choose a coordinate system where the origin is centered in the domain and scale the equation by replacing $\mathbf{x} \to \frac{\mathbf{x}}{d}$, $\kappa \to d\kappa$, $f \to f$ for $n = 3$ and $\mathbf{x} \to \frac{\mathbf{x}}{d\sqrt{C}}$, $\kappa \to d\sqrt{C}\kappa$, $f \to f$ for $n = 2$.)

We state the proof for the regimes $(\kappa < \kappa_0 < 1)$ and $(\kappa \geq \kappa_0 < 1)$, where κ_0 still has to be determined.

- *Continuity for $\kappa < \kappa_0 < 1$:*

$$\begin{aligned} a(u,v) &= (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (\kappa u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \kappa v) + (\kappa u, \kappa v) \\ &\leq (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, v) + (u, v) \\ &\leq 2\|u\|_S \|v\|_S, \end{aligned}$$

  where we used Cauchy-Schwarz inequality and the fact that $|\kappa| < 1$.

- *Continuity for $\kappa \geq \kappa_0 > 0$:*

$$\begin{aligned} a(u,v) &= (\frac{1}{\kappa} \mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, v) + (u, \kappa v) \\ &\leq \frac{1}{\kappa_0}(\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, v) + \|\kappa\|_\infty(u, v) \\ &\leq (\max\{\frac{1}{\kappa_0}, \|\kappa\|_\infty\} + 1)\|u\|_S \|v\|_S \end{aligned}$$

- *Coercivity for $\kappa < \kappa_0 < 1$:*

$$\begin{aligned} a(u,u) &= (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + (\kappa u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \kappa u) + (\kappa u, \kappa u) \\ &\geq \|\mathbf{s} \cdot \boldsymbol{\nabla}_x u\|^2 - 2|(\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \kappa u)| \\ &\geq (1-\beta)\|\mathbf{s} \cdot \boldsymbol{\nabla}_x u\|^2 + \beta\|\mathbf{s} \cdot \boldsymbol{\nabla}_x u\|^2 - \alpha\|\mathbf{s} \cdot \boldsymbol{\nabla}_x u\|^2 - \frac{1}{\alpha}\kappa_0\|u\|^2, \end{aligned}$$

7

for all $\alpha > 0$, where we applied the inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$. Using the Poincaré-Friedrichs inequality from Lemma 3.2 we obtain

$$a(u,u) \geq (1 - \alpha - \beta)||\mathbf{s} \cdot \boldsymbol{\nabla}_x u||^2 + (\beta - \frac{1}{\alpha}\kappa_0)||u||^2,$$

- *Coercivity for $\kappa \geq \kappa_0 > 0$:*

$$
\begin{aligned}
a(u,u) &= (\frac{1}{\kappa}\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x v) + (u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, v) + (u, \kappa v) \\
&\geq \frac{1}{||\kappa||_\infty}(\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + (u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + (\mathbf{s} \cdot \boldsymbol{\nabla}_x u, u) + \kappa_0(u, u) \\
&\geq \frac{1}{||\kappa||_\infty}(\mathbf{s} \cdot \boldsymbol{\nabla}_x u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) + 2(\mathbf{s} \cdot \boldsymbol{\nabla}_x u, u) + \kappa_0(u, u) \\
&\geq \min\{\frac{1}{||\kappa||_\infty}, \kappa_0\}||u||_S^2,
\end{aligned}
$$

taking into account that $2(u, \mathbf{s} \cdot \boldsymbol{\nabla}_x u) = \int_{\partial D \times S^2} \mathbf{n} \cdot \mathbf{s}\, u^2\, d\mathbf{x}\, d\mathbf{s} = \int_{S^2} \int_{\Gamma_+(\mathbf{s})} \mathbf{n} \cdot \mathbf{s}\, u^2\, d\mathbf{x}\, d\mathbf{s} \geq 0$, where

$$\Gamma_+(\mathbf{s}) := \{\mathbf{x} \in \partial D : \mathbf{s} \cdot \mathbf{n}(\mathbf{x}) > 0\} \subset \partial D, \quad \mathbf{s} \in S^2. \tag{27}$$

We now choose $\alpha > 0$, $\beta > 0$ and $\kappa_0 > 0$ as the solution of the optimization problem

$$\min_{\alpha, \beta, \kappa_0} \{1 - \alpha - \beta, \beta - \frac{1}{\alpha}\kappa_0, \kappa_0\} \to \max.$$

leading to $\alpha \approx 0.366$, $\beta = 0.5$ and $\kappa_0 \approx 0.134$ as optimal value.

The bilinear form therefore satisfies

$$
\begin{aligned}
a(u,v) &\leq (\max\{\frac{1}{\kappa_0}, ||\kappa||_\infty\} + 1)||u||_S||v||_S \\
a(u,u) &\geq \min\{\frac{1}{||\kappa||_\infty}, \kappa_0\}||u||_S^2.
\end{aligned}
$$

with $\kappa_0 \approx 0.134$.

For $d \neq 1$ we replace $||\kappa||_{L^\infty}$ by $d||\kappa||_{L^\infty}$ for $n = 3$ and by $\sqrt{C}d||\kappa||_{L^\infty}$ for $n = 2$. ∎

For the variational formulation, we assume that the nonzero Dirichlet data g on $\partial_-(D \times S^2)$ admit an extension $\widetilde{g} \in \mathcal{V}$ and define the "load" functional

$$l(v) := (\epsilon \kappa^2 f, v)_{L^2} + (\epsilon \kappa f, \mathbf{s} \cdot \boldsymbol{\nabla}_x v)_{L^2}. \tag{28}$$

Proposition 3.4 *Assume that*
$$\kappa \in L^\infty(D), \ \kappa \geq 0,$$
$f \in L^2(D \times S^2)$ and that ϵ is defined as in (26). Then the load functional $l(v)$ defined in (28) is continuous in with respect to the $||.||_S$-norm, i.e. there exists a constant c such that

$$|l(v)| \leq c\,||v||_S$$

with $c = \max\{||\sqrt{\kappa}||_\infty, \kappa_0\}\sqrt{c_1}$, where c_1 and κ_0 are given in (3.3).

Proof:

$$
\begin{aligned}
|l(v)| &= (\sqrt{\epsilon}\kappa f, \sqrt{\epsilon}\kappa v)_{L^2} + (\sqrt{\epsilon}\kappa f, \sqrt{\epsilon}\mathbf{s} \cdot \boldsymbol{\nabla}_x v)_{L^2} \leq ||\sqrt{\epsilon}\kappa||_\infty ||f||\,||\sqrt{\epsilon}\,(\kappa v + \mathbf{s} \cdot \boldsymbol{\nabla}_x v)|| \\
&\leq \max\{||\sqrt{\kappa}||_\infty, \kappa_0\}\sqrt{c_1}||v||_S.
\end{aligned}
$$

8

Then the resulting linear variational problem reads: seek $\widetilde{u} \in \mathcal{V}_0$ such that

$$a(\widetilde{u}, v) = l(v) - a(\widetilde{g}, v) \quad \forall v \in \mathcal{V}_0 . \tag{29}$$

Here, the right hand side is well defined and the well-posedness of the variational problem is a consequence of the Lax-Milgram Lemma, due to Propositions 3.3 and 3.4

Theorem 3.5 *Under the assumptions on $\kappa(x)$ in Propositions 3.3, 3.4, for every $f \in L^2(D \times S^2)$ and for every g defined on $\partial_-(D \times S^2)$ admitting an extension $\tilde{g}$ to the domain $D \times S^2$ which belongs to the space $\mathcal{V}$ defined in (15), equipped with the norm $\| \circ \|_S$ in (18), there exists a unique weak solution $u = \tilde{g} + \tilde{u} \in \mathcal{V}$ of the stabilized variational form (29) of the radiative transfer problem (12), (13).*

As the bilinear form $a(.,.)$ is symmetric and positive definite on $\mathcal{V}_0$, we can define the energy norm

$$\|u\|_A := \sqrt{a(u,u)} . \tag{30}$$

Remark 1. An alternative to obtain a stable formulation would be to relax the continuity of the Galerkin discretization. Passing to nonconforming, discontinuous Galerkin FEM has become very popular in recent years; here, stability is achieved by incorporating into the interelement multipliers Finite Volume type fluxes. As we will derive the sparse Galerkin FEM for (12) by projection on a family of so-called *sparse tensor product* hierarchical FE spaces in $D \times S^2$ which are subspaces of $\mathcal{V}_0$ in (16), we need a variational formulation which allows to use continuous FE spaces in the physical domain D and which does not require the specification of fluxes at interelement boundaries.

4 Galerkin Discretization

From now on, the variational problem (29) will be considered on the space

$$V_0 := H^{1,0}(D \times S^2) \cap \mathcal{V}_0, \quad H^{1,0}(D \times S^2) = H^1(D) \otimes L^2(S^2).$$

Since V_0 is a proper, closed subspace of $\mathcal{V}_0$, the variational problem (29), *restricted to V_0, admits a unique weak solution $\bar{u} \in V_0$. In what follows, we shall assume that the two weak solutions, $\tilde{u} \in \mathcal{V}_0$ and $\bar{u} \in V_0$, of (29) with homogeneous Dirichlet data $g = 0$, coincide and denote this solution by u.* This is a regularity assumption stating that the weak solution $u \in \mathcal{V}$ of (12), (13) with $g = 0$ belongs, in fact, to $H^1(D) \otimes L^2(S^2)$. Note that this assumption precludes line discontinuities of u in D which may arise due to transport along rays of discontinuous boundary data. The Galerkin discretization of (29) with homogeneous Dirichlet data, i.e. with

$$g = 0 \quad \text{on} \quad \partial_-(D \times S^2)$$

is obtained, as usual, by restricting $u = \tilde{u}$ and v in the weak formulation (29) to a one-parameter family of finite dimensional subspaces $\{V_0^L\}_L$ of $\mathcal{V}_0$, where the superscript L will denote "level" of mesh refinement below. This yields

$$u_L \in V_0^L : \quad a(u_L, v) = l(v) \quad \forall v \in V_0^L. \tag{31}$$

Due to the coercivity and continuity of $a(\cdot, \cdot)$ on $\mathcal{V}_0 \times \mathcal{V}_0$, Proposition 3.3, (31) admits a unique solution which satisfies the Galerkin orthogonality

$$\forall v \in V_0^L : \qquad a(u - u_L, v) = 0. \tag{32}$$

Therefore, the error $e_L = u - u_L$ is quasioptimal in the $\|.\|_S$-norm, i.e. for every subspace V_0^L in the sequence we obtain

$$\|u - u_L\|_S \le C(\kappa, D) \inf_{v_L \in V_0^L} \|u - v_L\|_S. \tag{33}$$

9

It remains to specify subspace sequences V_0^L. Since the computational domain $D \times S^2$ is a product domain, we shall build V_0^L out of tensor products of "component" finite element spaces in D and in S^2, respectively. Note, however, that due to the $\mathbf{s}$ dependence of the Dirichlet boundary $\Gamma_-(\mathbf{s}) \subset \partial D$, the subspaces V_0^L will generally *not* be of tensor product type, once the boundary condition (13) is imposed.

Let us start by giving the construction of the component spaces without boundary conditions. To this end, we equip both the domain D as well as the sphere S^2 each with a coarse mesh denoted by $\mathcal{T}_D^0$ and by $\mathcal{T}_{S^2}^0$, respectively. Both meshes are assumed to be quasi-uniform meshes of shape-regular simplices. The hierarchic mesh sequences $\mathcal{T}_D^l$, $\mathcal{T}_{S^2}^l$, $l = 1, ..., L$, are then obtained by dyadic refinement of the coarse meshes.

On the hierarchic mesh sequences, we specify Finite Element spaces. In physical space D, the finite element space $V_D^L := S^{p,1}(D, \mathcal{T}_D^L) \subset H^1(D)$ consists of piecewise polynomial functions of degree $p \geq 1$ on the finest triangulation $\mathcal{T}_D^L$ which are continuous in the physical domain D. In the solid angle $\mathbf{s} \in S^2$, we use $V_{S^2}^L := S^{q,0}(S^2, \mathcal{T}_{S^2}^L) \subset L^2(S^2)$ of discontinuous, piecewise polynomials of degree $q \geq 0$ on the spherical triangles of $\mathcal{T}_{S^2}^L$. In the implementation ahead, we realized the simplest case $p = 1$ and $q = 0$.

Based on the FE spaces V_D^L and $V_{S^2}^L$ in the 'component domains' D and S^2, we define the tensor product Finite Element space $V_0^L \subset V_0$ on the cartesian product mesh $\mathcal{T}_D^L \times \mathcal{T}_{S^2}^L$ at refinement level L by

$$V_0^L := (V_D^L \otimes V_{S^2}^L) \cap \mathcal{V}_0 = (S^{p,1}(D, \mathcal{T}_D^L) \otimes S^{q,0}(S^2, \mathcal{T}_{S^2}^L)) \cap \mathcal{V}_0. \tag{34}$$

The Galerkin discretized problem then reads: find $u_L(\mathbf{x}, \mathbf{s}) \in V_0^L$ such that

$$a(u_L, v_L) = l(v_L) \qquad \forall v_L \in V_0^L. \tag{35}$$

Let $\alpha_i(\mathbf{x})$, $i = 1, ..., M_L = \dim V_D^L$ be a basis of V_D^L and $\beta_j(\mathbf{s})$, $j = 1, ..., N_L = \dim V_{S^2}^L$ a basis of $V_{S^2}^L$. Then the approximate intensity $u_L \in V^L$ can be expressed in the tensor product form

$$u_L(\mathbf{x}, \mathbf{s}) = \sum_{i=1}^{M_L} \sum_{j=1}^{N_L} u_{ij} \alpha_i(\mathbf{x}) \beta_j(\mathbf{s}) \in V^L. \tag{36}$$

Then (35) leads to a linear system of equations for the $M_L \cdot N_L$ unknowns u_{ij}.

A natural choice of bases in the simplest case $p = 1$ (continuous, piecewise linear elements in D) and $q = 0$ (discontinuous, piecewise constant elements in S^2) is

- locally supported piecewise linear "hat functions" for V_D^L that is $\alpha_i(x_j) = \delta_{ij}$, where $\{x_1, ..., x_M\}$ is the set of vertices of $\mathcal{T}_D^L$.

- β_j are chosen as the characteristic functions of the triangles of $\mathcal{T}_{S^2}^L$.

In order to impose the boundary conditions in our implementation based on $p = 1$ in D and on $q = 0$ on S^2, we take the nodal basis in physical space and the characteristic functions of triangles $T \in \mathcal{T}_{S^2}^L$ as a basis in the solid angle S^2. We then set all degrees of freedom u_{ij} to zero, if i corresponds to a 'inflow' boundary node $x_i \in \Gamma_-(\mathbf{s})$ for $\mathbf{s} \in \text{supp}(\beta_j) \subset S^2$.

The Galerkin tensor product discretization with $q = 0$ is equivalent to the discrete ordinates method [1], if the centers of gravity of the spherical triangles are chosen as discrete ordinates and the resulting M_L transport equations in space are discretized as described in sections 3 and 4. This discrete ordinates method is usually referred to as S_N, where N is the number of discrete ordinates [1, Sect. 16].

As the number of degrees of freedom is $M \cdot N$, where M and N both are the number of basis functions arising from a two- or three- dimensional discretization, this method turns out to be very expensive.

Thus, one should try to keep both N and M small by picking *adapted bases* that offer a good representation of the solution with only a few degrees of freedom. The construction of such basis functions can be pursued via local adaptive mesh refinement starting from $\mathcal{T}_D^0$ and $\mathcal{T}_{S^2}^0$. Then, the

ultimate trial and test functions are built according to (36). The limitations of this approach are evident: since the space $V_{S^2}^L$ is expected to provide good resolution of the radiation everywhere, $\mathcal{T}_{S^2}^L$ will usually have to be a fairly uniform mesh.

Yet, using all the basis functions $\alpha_i(\mathbf{x})\beta_j(\mathbf{s})$, $i = 1, \ldots, M_L$, $j = 1, \ldots, N_L$, as in (36) may not be necessary at all, because only a few of them may really make a significant contribution to representing the final solution. Hence, a promising approach to obtaining efficient trial spaces is to select a few important *combined basis functions* of the form $\alpha_i(\mathbf{x})\beta_j(\mathbf{s})$ and let them span V^L. The *component basis functions* α_i and β_j can be chosen from large sets, which will not translate into prohibitively large discrete problems. This idea underlies the present approach to the Galerkin discretization of the radiative transfer problem which is based on *sparse tensor products* of the component finite element spaces V_D^L and $V_{S^2}^L$. In the following, we shall discuss two choices of the sparse tensor product space – the *a-priori* selection of combinations of basis functions, and the *a-posteriori*, adaptive selection of such combinations.

4.1   Sparse Tensor Product Space

The selection of important combined basis functions needs to follow strict rules in order to enable efficient computational procedures. Such a set of rules for selecting basis functions *a-priori* is offered by the framework of sparse grids [9], [25]. In the following, we adapt these techniques to sparse tensor products of FE spaces in $H^{1,0}(D \times S^2) \simeq H^1(D) \otimes L^2(S^2)$.

Based on the nested triangulations $\mathcal{T}_D^l$ and $\mathcal{T}_{S^2}^l$, $l = 0, 1, 2, \ldots$, we recall the corresponding nested sequences of Finite Element spaces

$$V_D^l := S^{p,1}(D, \mathcal{T}_D^l) \subset H^1(D), \qquad V_{S^2}^l := S^{q,0}(S^2, \mathcal{T}_{S^2}^l) \subset L^2(S^2).$$

Here, $S^{p,1}(D, \mathcal{T}_D^l)$ denotes the continuous, piecewise polynomial functions of degree $p \geq 1$ on $\mathcal{T}_D^l$ and $S^{q,0}(S^2, \mathcal{T}_{S^2}^l)$ denotes the space of possibly discontinuous, piecewise polynomial functions of degree $q \geq 0$.

As the triangulations, these sequences of spaces are in turn nested, and there are so-called "detail spaces" W_D^l, $W_{S^2}^l$ such that

$$V_D^l = W_D^l \oplus V_D^{l-1}, \qquad V_{S^2}^l = W_{S^2}^l \oplus V_{S^2}^{l-1}, \tag{37}$$

where $\oplus$ is the orthogonal direct sum with respect to the corresponding L^2 inner products.

Iterating (37), we see that for $l \geq 1$ the spaces V_D^l and $V_{S^2}^l$ possess an L^2- orthogonal decomposition into the detail subspaces W_D^l and $W_{S^2}^l$, respectively:

$$V_D^l = \bigoplus_{i=0}^{l} W_D^i, \qquad V_{S^2}^l = \bigoplus_{i=0}^{l} W_{S^2}^i \tag{38}$$

where we defined $W_D^0 := V_D^0$ and $W_{S^2}^0 := V_{S^2}^0$, respectively. The continuous multiscale basis functions used in [25] do not exhibit L^2-orthogonality. Although the following approximation properties hold also (almost) for those basis functions, we prefer L^2-orthogonal constructions for stability reasons, in particular with respect to mesh adaptation (section 5).

With these definitions, the full tensor product space $V^L \subset H^1(D) \otimes L^2(S^2)$ at mesh refinement level L is easily seen to coincide with

$$V^L = V_D^L \otimes V_{S^2}^L = \bigoplus_{0 \leq l_1, l_2 \leq L} W_D^{l_1} \otimes W_{S^2}^{l_2}. \tag{39}$$

We shall consider the sparse tensor product space $\hat{V}^L \subset V^L$ defined by

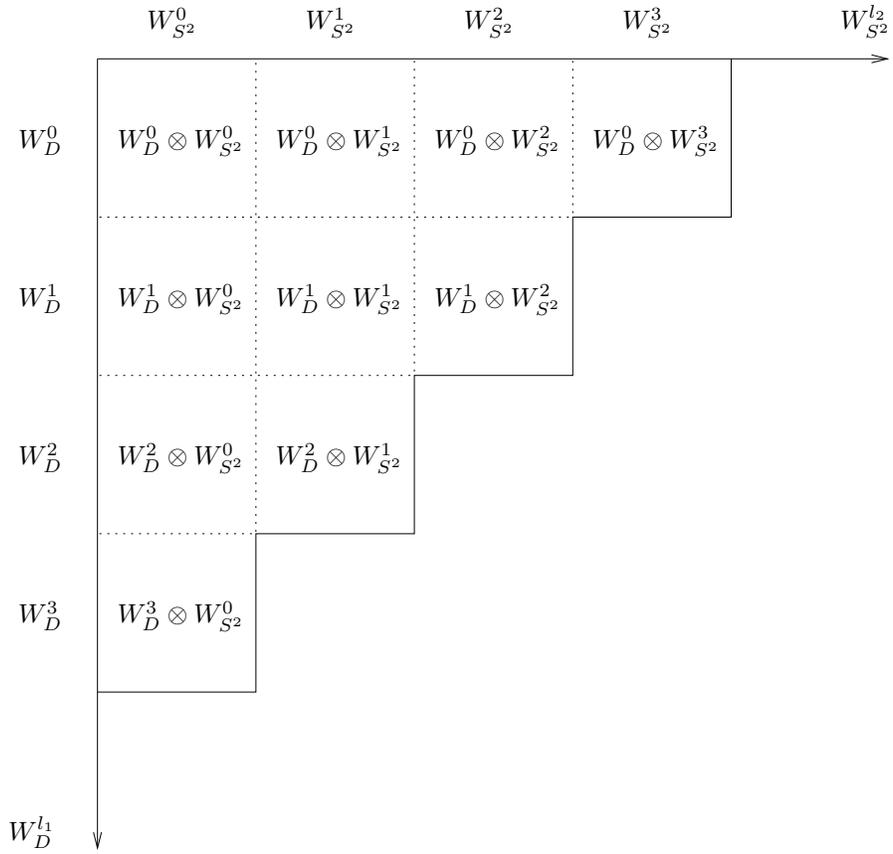$$\widehat{V}^L := \bigoplus_{0 \leq l_1 + l_2 \leq L} W_D^{l_1} \otimes W_{S^2}^{l_2} \tag{40}$$

$$W_{S^2}^0 \qquad W_{S^2}^1 \qquad W_{S^2}^2 \qquad W_{S^2}^3 \qquad\qquad W_{S^2}^{l_2}$$

PSfrag replacements

W_D^0 : $W_D^0 \otimes W_{S^2}^0$, $W_D^0 \otimes W_{S^2}^1$, $W_D^0 \otimes W_{S^2}^2$, $W_D^0 \otimes W_{S^2}^3$

W_D^1 : $W_D^1 \otimes W_{S^2}^0$, $W_D^1 \otimes W_{S^2}^1$, $W_D^1 \otimes W_{S^2}^2$

W_D^2 : $W_D^2 \otimes W_{S^2}^0$, $W_D^2 \otimes W_{S^2}^1$

W_D^3 : $W_D^3 \otimes W_{S^2}^0$

$W_D^{l_1}$

Figure 1: Component spaces of the sparse tensor product space $\widehat{V}_L$ (for $L = 3$).

which is illustrated schematically in Fig. 1 and set

$$V_0^L := V^L \cap \mathcal{V}_0, \qquad \hat{V}_0^L = V^L \cap \mathcal{V}_0.$$

An easy counting argument based on Fig. 1 also shows that

$$\dim V_0^L = O(N_L M_L), \qquad \hat{\mathcal{N}}_L := \dim \hat{V}_0^L = O(N_L \log M_L + M_L \log N_L) \qquad \text{as} \quad L \to \infty.$$

This means that basis functions of level l in subspaces in D are tensorized only with basis functions in S^2 up to level $L - l$ and vice versa (see also Fig. 2). As explained above, the full tensor product Galerkin discretization with piecewise constants on S^2 combined with one-point quadrature in S^2 is equivalent to the discrete ordinates method. The approximation in the corresponding sparse tensor product space $\hat{V}^L$ can be viewed as discrete ordinates method where the number of ordinates used for each physical degree of freedom varies according to the mesh level in the physical domain.

4.2   Approximation Properties

Since the sparse tensor product space $\hat{V}^L$ is substantially smaller than the full tensor product space V^L, accuracy may be lost. We shall now estimate the rate of convergence of the sparse Galerkin approximations to u which show that, at least for smooth solutions, both spaces achieve the same asymptotic convergence rate.

To this end, we define the L^2-projection operators $P_D^l : L^2(D) \longrightarrow V_D^l$ and $P_{S^2}^l : L^2(S^2) \longrightarrow V_{S^2}^l$ with the convention that $P_D^{-1} = P_{S^2}^{-1} = 0$. Then the projector $\hat{P}_L$ onto the sparse tensor product space of level L reads

$$\hat{u}_L(\mathbf{x}, \mathbf{s}) := \hat{P}_L u(\mathbf{x}, \mathbf{s}) := \sum_{0 \le l_1 + l_2 \le L} \left(P_D^{l_1} - P_D^{l_1 - 1} \right) \otimes \left(P_{S^2}^{l_2} - P_{S^2}^{l_2 - 1} \right) u(\mathbf{x}, \mathbf{s}) \tag{41}$$

and the projection $P_L = P_D^L \otimes P_{S^2}^L$ onto V_0^L can be represented analogously, if the summation is over $0 \le l_1, l_2 \le L$ instead.

In order to describe the approximation properties of the sparse tensor product space $\hat{V}_L$, we follow [16] and [14] and introduce anisotropic Sobolev spaces with fractional derivatives. We start by defining, for $m, n \in \mathbb{N}_0$, the anisotropic Sobolev spaces

$$H^{s,t}(D \times S^2) := H^s(D) \otimes H^t(S^2) \tag{42}$$

which are, for integer values of s and of t, given by

$$\left\{ u \in L^2(D \times S^2) | D_{\mathbf{x}}^\alpha D_{\mathbf{s}}^\beta u \in L^2(D \times S^2), 0 \le |\alpha| \le s, \ 0 \le |\beta| \le t \right\},$$

where for $\alpha \in \mathbb{N}_0^n$, $D_{\mathbf{x}}^\alpha$ denotes the α-th weak derivative with respect to $\mathbf{x} \in D$; we denote its order by $|\alpha| = \alpha_1 + ... + \alpha_n$. Analogously, for $\beta \in \mathbb{N}_0^2$, $D_{\mathbf{s}}^\beta$ denotes the weak derivative with respect to $\mathbf{s} \in S^2$ and we denote its order $|\beta| = \beta_1 + \beta_2$.

We equip the anisotropic space with the norm

$$||u||_{H^{s,t}}^2 := \sum_{\substack{0 \le |\alpha| \le s \\ 0 \le |\beta| \le t}} ||D_{\mathbf{x}}^\alpha D_{\mathbf{s}}^\beta u||_{L^2(D \times S^2)}^2. \tag{43}$$

For arbitrary $s, t \ge 0$, we define $H^{s,t}(D \times S^2)$ by interpolation and tensorization.

For functions $v(\mathbf{x}) \in H^1(D)$ and $w(\mathbf{s}) \in L^2(S^2)$ that are sufficiently smooth, the following approximation properties hold for $l \in \mathbb{N}_0$ (see eg. [26]):

$$||v - P_D^l v||_{H^1(D)} \quad \lesssim \quad 2^{-ls} ||v||_{H^{s+1}(D)}, \ s \in [0, \ p], \tag{44}$$

$$||w - P_{S^2}^l w||_{L^2(S^2)} \quad \lesssim \quad 2^{-lt} ||w||_{H^t(S^2)}, \ t \in [0, \ q + 1]. \tag{45}$$

(a) physical space level 0   (b) physical space level 1   (c) physical space level 2   (d) physical space level 3

(e) solid angle level 0   (f) solid angle level 1   (g) solid angle level 2   (h) solid angle level 3
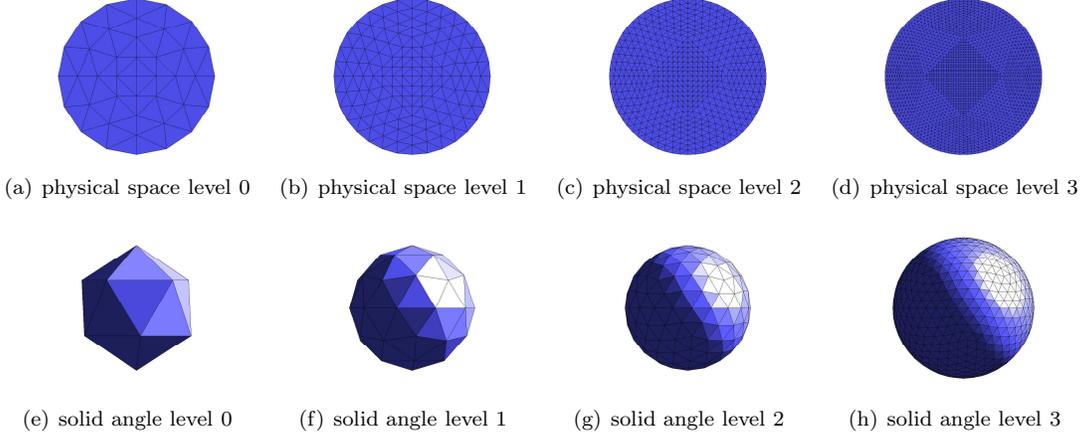
Figure 2: Mesh hierarchies used in the sparse tensor product space.

Here and in what follows, we use the notations $a \lesssim b$ ($a \simeq b$) if there exists a constant $C < \infty$ with $a \leq Cb$ ($a \leq Cb$ and $a \geq C^{-1}b$). The constants in these estimates may only depend on the angles in the meshes $\mathcal{T}_D^0$ and $\mathcal{T}_{S^2}^0$.

The asymptotic density of the discrete subspace sequences in $H^1(D) \otimes L^2(S^2)$ permits us to write any function $v \in H^1(D) \otimes L^2(S^2)$ uniquely as

$$u(\mathbf{x}, \mathbf{s}) = \sum_{l_1, l_2 = 0}^{\infty} u_{l_1, l_2}(\mathbf{x}, \mathbf{s}), \quad u_{l_1, l_2} \in W_D^{l_1} \otimes W_{S^2}^{l_2}. \tag{46}$$

Then for the best approximation $\widehat{v}_L$ of u in the sparse tensor product space, the following holds:

$$
\min_{\widehat{v}_L \in \widehat{V}_L} ||u - \widehat{v}_L||_{H^{1,0}(D \times S^2)} \leq ||u - \widehat{P}_L u||_{H^{1,0}(D \times S^2)} = ||\sum_{l_1=0}^{\infty} \sum_{l_2 = max\{0, L - l_1 + 1\}}^{\infty} u_{l_1, l_2}||_{H^{1,0}(D \times S^2)}
$$

$$
\leq ||\underbrace{\sum_{l_1=0}^{L} \sum_{l_2 = L - l_1 + 1}^{\infty} u_{l_1, l_2}||_{H^{1,0}(D \times S^2)}}_{:=I} + ||\underbrace{\sum_{l_1=L+1}^{\infty} \sum_{l_2=0}^{\infty} u_{l_1, l_2}||_{H^{1,0}(D \times S^2)}}_{:=II}
$$

For the first part, we have the following estimate:

$$
\begin{aligned}
I &:= ||\sum_{l_1=0}^{L} \sum_{l_2 = L - l_1 + 1}^{\infty} u_{l_1, l_2}||_{H^{1,0}(D \times S^2)} \\
&= ||\sum_{l_1=0}^{L} \left(P_D^{l_1} - Id + Id - P_\Omega^{l_1 - 1}\right) \otimes \left(Id - P_{S^2}^{L - l_1 + 1}\right) u||_{H^{1,0}(D \times S^2)} \\
&\leq \sum_{l_1=0}^{L} \left(||\left(Id - P_D^{l_1}\right) \otimes \left(Id - P_{S^2}^{L - l_1 + 1}\right) u||_{H^{1,0}(D \times S^2)} \right. \\
&+ \left. ||\left(Id - P_D^{l_1 - 1}\right) \otimes \left(Id - P_{S^2}^{L - l_1 + 1}\right) u||_{H^{1,0}(D \times S^2)}\right).
\end{aligned}
$$

Assuming sufficient smoothness of the solution in space and solid angle and using (44) and (45), we find for $0 \leq s \leq p$ and for $0 \leq t \leq q + 1$

$$
\begin{aligned}
||(Id - P_D^{l_1 - 1}) \otimes (Id - P_{S^2}^{L - l_1 + 1})u|| &\lesssim 2^{-L \min(s,t)} ||u||_{H^{1+s,t}(D \times S^2)}, \\
||(Id - P_D^{l_1}) \otimes (Id - P_{S^2}^{L - l_1 + 1})u|| &\lesssim 2^{-L \min(s,t)} ||u||_{H^{1+s,t}(D \times S^2)}.
\end{aligned}
$$

14

Under the same assumptions, the following estimate holds for the second summand

$$II := \left\| \sum_{l_1=L+1}^{\infty} \sum_{l_2=0}^{\infty} u_{l_1,l_2} \right\|_{H^{1,0}} = \|(Id - P_D^L)u\|_{H^{1,0}} \lesssim 2^{-sL} \|u\|_{H^{1+s,0}}. \tag{47}$$

The best approximation converges then as

$$\min_{\widehat{v}_L \in \widehat{V}_L} \|u - \widehat{v}_L\|_{H^{1,0}(D\times S^2)} \lesssim (L+2)2^{-L\min(s,t)} \|u\|_{H^{1+s,t}(D\times S^2)}. \tag{48}$$

As $h_L \simeq 2^{-L}$ implies that $L \simeq |\log_2(h_L)|$, we get for $0 \le s \le p$ and $0 \le t \le q$ the error bounds

$$\min_{\hat{v}_L \in \widehat{V}_0^L} \|u - \widehat{v}_L\|_{H^{1,0}(D\times S^2)} \le \|u - \widehat{P}_L u\|_{H^{1,0}(D\times S^2)} \lesssim |\log_2 h_L| h_L^{\min(s,t)} \|u\|_{H^{1+s,t}(D\times S^2)}. \tag{49}$$

In the same way, one obtains for $0 \le s \le p$ and $0 \le t \le q$ and the full tensor product approximation the error estimates

$$\min_{v_L \in V_0^L} \|u - v_L\|_{H^{1,0}(D\times S^2)} \le \|u - P_L u\|_{H^{1,0}(D\times S^2)} \lesssim h^{\min(s,t)} \|u\|_{(H^{1+s,0}\cap H^{1,t})(D\times S^2)}. \tag{50}$$

We collect the above observations in

Theorem 4.1 *Under the assumption that $p = q + 1 \ge 1$ and that $u \in H^{p+1,p}(D \times S^2)$, the Galerkin FEM approximations u_L in (31) based on the full tensor product space V_0^L satisfy the asymptotic error estimate*

$$\|u - u_L\|_A \lesssim \|u - P_L u\|_S \lesssim \|u - P_L u\|_{H^{1,0}(D\times S^2)} \lesssim h_L^p \|u\|_{(H^{p+1,0}\cap H^{1,q+1})(D\times S^2)}. \tag{51}$$

and the Galerkin FEM approximations $\hat{u}^L$ obtained from (31) based on the sparse tensor product space $\hat{V}_0^L$ satisfy the error estimate

$$\|u - \hat{u}_L\|_A \lesssim \|u - \widehat{P}_L u\|_S \lesssim \|u - \widehat{P}_L u\|_{H^{1,0}(D\times S^2)} \lesssim h_L^p |\log h_L| \|u\|_{H^{p+1,p}(D\times S^2)}. \tag{52}$$

The number of degrees of freedom required for these solutions behaves, as $L \to \infty$, as

$$\dim(V_0^L) \sim N_L M_L \quad and \quad \dim(\hat{V}_0^L) \sim N_L \log M_L + M_L \log N_L,$$

respectively.

We see that, up to a logarithmic factor, the convergence rates attainable with the full and the sparse tensor product discretizations are identical, while the number of degrees of freedom necessary in the sparse tensor product space is, again up to logarithmic factors, the same as that of the component spaces. In effect, the use of the sparse tensor product space $\hat{V}_0^L$ in (31) reduces the FEM complexity from $3 + 2$ dimensional domain to essentially that of a FE computation in a 3 dimensional domain while retaining (up to logarithmic terms) the asymptotic rate of convergence *provided that $u \in \mathcal{V}_0$ is sufficiently smooth.* This cannot be guaranteed for general absorption and emission data (see also section 5).

4.3 Wavelet Finite Element Bases

Whereas the definitions of the increment spaces do not require bases for these increments, any implementation of the sparse tensor product method does require bases of multilevel subspaces W_D^l, $W_{S^2}^l$. Best suited for our purposes are either hierarchical bases or wavelet Finite Element bases. The latter offer the following important advantages over the former:

a) there hold multilevel norm equivalences in a scale of Sobolev spaces which allows diagonal preconditioning of the discrete problem, and

b) wavelets exhibit vanishing moments which allows to characterize the regularity of the intensity u in terms of the coefficient decay in its wavelet expansion which is the basis for adaptive refinements, both in D and in S^2,

c) (not exploited or addressed yet in the present paper) for nonlocal operators $\mathcal{K}$ in (3), wavelet type basis functions allow for compression of the corresponding stiffness matrices without sacrificing convergence orders which means that even in this case the computational costs are of order $\mathcal{O}(N \log M + M \log N)$, where M is the number of degrees of freedom in physical space and N the number of degrees of freedom in solid angle.

To date, several constructions of wavelet FEM in general domains D are available. In particular, we use in D the isotropic, piecewise linear and continuous finite element wavelet basis described in [26]. They are constructed level-wise. On level 0, the basis functions $\psi_0^m(\mathbf{x})$, $m \in \mathcal{I}(\mathcal{T}_D^0)$, are the standard hat functions on the coarsest mesh. With $\mathcal{I}(\mathcal{T}_D^l)$ we denote the index set of vertices of the mesh $\mathcal{T}_D^l$ and with $\widehat{\mathcal{I}}(\mathcal{T}_D^l)$ the index set of vertices of $\mathcal{T}_D^l$ that do not belong to $\mathcal{T}_D^{l-1}$ (see Figs. 3 and 4).

PSfrag replacements
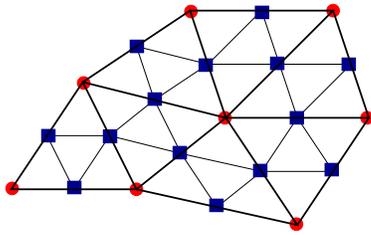
Figure 3: Index sets $\mathcal{I}(\mathcal{T}_D^0)$ and $\widehat{\mathcal{I}}(\mathcal{T}_D^1)$.
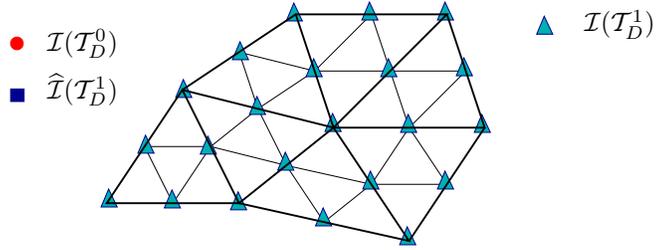
Figure 4: Index set $\mathcal{I}(\mathcal{T}_D^1)$.

- $\bullet$ $\mathcal{I}(\mathcal{T}_D^0)$
- $\blacksquare$ $\widehat{\mathcal{I}}(\mathcal{T}_D^1)$
- $\blacktriangle$ $\mathcal{I}(\mathcal{T}_D^1)$

On a higher level $l > 0$, the construction of the wavelet functions $\psi_l^m(\mathbf{x})$, $m \in \widehat{\mathcal{I}}(\mathcal{T}_D^l)$, is based on the meshes $\mathcal{T}_D^l$ and $\mathcal{T}_\Omega^{l-1}$. For $j \in \widehat{\mathcal{I}}(\mathcal{T}_D^l)$, we define $\phi_l^j(\mathbf{x})$ to be the hat function of vertex j on mesh $\mathcal{T}_D^l$.

We now construct a family of functions $\theta_l^i(\mathbf{x}) \in S^{1,1}(\mathcal{T}_D^l)$, $i \in \mathcal{I}(\mathcal{T}_D^{l-1})$, that satisfy $(\theta_l^i, \phi_{l-1}^k)_{L^2(D)} \simeq \delta_{ik}$.

As it can easily be verified, the piecewise linear functions $\theta_l^i(\mathbf{x})$ with

$$\theta_l^i(\mathbf{v}) = \begin{cases} 14, & \mathbf{v} \text{ is vertex } \mathbf{v}_i \text{ of mesh } \mathcal{T}_D^{l-1} \\ -1, & \mathbf{v} \text{ is neighbouring vertex of } \mathbf{v}_i \text{ on mesh } \mathcal{T}_D^l \\ 0, & \mathbf{v} \text{ is any other vertex of mesh } \mathcal{T}_D^l \end{cases}$$

fulfill this condition. Examples of the functions $\phi_{l-1}^k(\mathbf{x})$, $\phi_l^j(\mathbf{x})$ and $\theta_l^k(\mathbf{x})$ are displayed in Figs. 5, 6 and 7.

The wavelets on level l are then obtained by

$$\psi_l^m(\mathbf{x}) = \phi_l^m(\mathbf{x}) - \sum_{k \in \mathcal{I}(\mathcal{T}_D^{l-1})} \frac{(\phi_l^m, \phi_{l-1}^k)_{L^2(\Omega)}}{(\theta_l^k, \phi_{l-1}^k)_{L^2(D)}} \theta_l^k(\mathbf{x}), \ m \in \widehat{\mathcal{I}}(\mathcal{T}_D^l). \tag{53}$$

(An example is shown is Fig. 8.)

The functions $\frac{\psi_l^m(\mathbf{x})}{\|\psi_l^m\|}$, $m \in \widehat{\mathcal{I}}(\mathcal{T}_D^l)$, form a uniform L^2 Riesz basis for $V_D^l \cap S^{1,1}(\mathcal{T}_\Omega^{j-1})^{\perp_{L^2}}$.

The scaled functions $\tilde{\psi}_l^m(\mathbf{x}) = 2^{-l} \frac{\psi_l^m(\mathbf{x})}{\|\psi_l^m\|}$, $l \in \mathbb{N}$, $m \in \widehat{\mathcal{I}}(\mathcal{T}_D^l)$, then form a Riesz basis for $H^1(D)$, i.e. in addition to the unique the decomposition of every $u \in H^1(D)$,

$$u = \sum_{l,m} u_{lm} \tilde{\psi}_m^l,$$
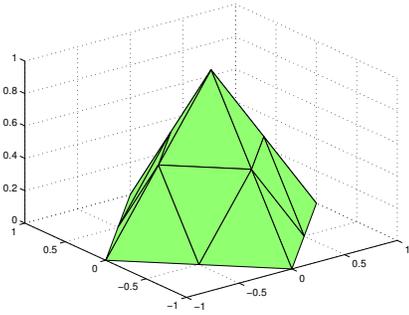
Figure 5: $\phi_{l-1}^k(\mathbf{x})$
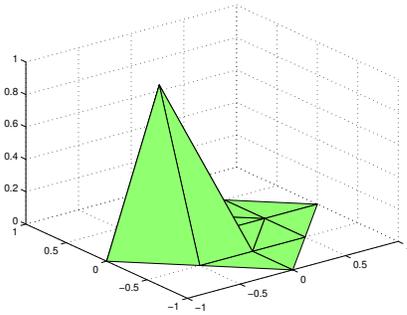
Figure 6: $\phi_l^j(\mathbf{x})$

Figure 7: $\theta_l^k(\mathbf{x})$

17
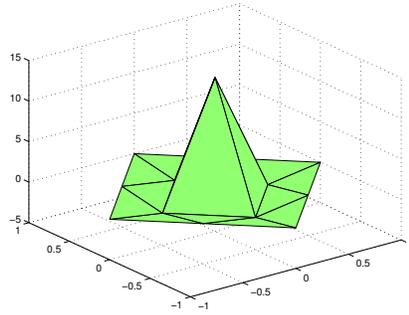
the basis functions are stable in the sense that there exist so-called Riesz constants c, C such that

$$c||c_{ij}||_{l^2}^2 \leq ||\sum_{l,m} c_{lm}\tilde{\psi}_l^m||_{H^1(D)}^2 \leq C||c_{ij}||_{l^2}^2, \ \forall \mathbf{c} \in l^2.$$

(see eg. [23] and references therein). With $W_D^j = \text{span}\{\psi_j^m \ , \ m \in \hat{\mathcal{I}}(\mathcal{T}_D^j)\}$, $V_D^j = \text{span}\{\psi_i^m \ , \ 0 \leq i \leq j, \ m \in \hat{\mathcal{I}}(\mathcal{T}_D^i)\}$, they fit into the framework of section 4.1.
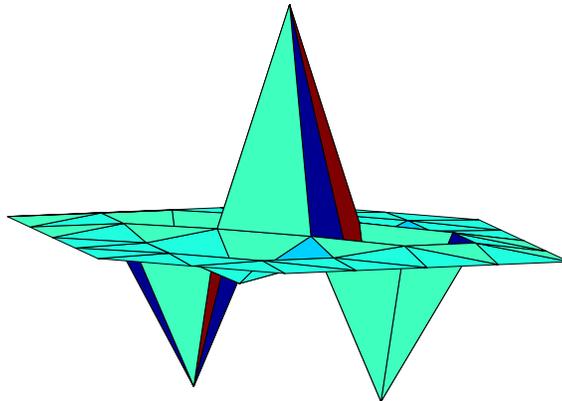
Figure 8: Piecewise linear FE wavelet.

On the sphere, we use agglomerated Haar wavelets (see e.g. [27]) that are slightly adapted for the sphere. As we use piecewise constant functions, each degree of freedom corresponds to a spherical triangle.

On level 0, χ_0^n, are the characteristic functions on the triangles T_n of the coarsest triangulation. On higher levels $l > 0$, the basis functions χ_l^n are based on the meshes $\mathcal{T}_{S^2}^l$ and $\mathcal{T}_{S^2}^{l-1}$, where $n \in$ index set on level l. The support of χ_l^n is a triangle on $\mathcal{T}_{S^2}^{l-1}$. On each sub-triangle T_i, $i = 1, ..., 4$,

$$\chi_l^n = \pm \frac{\frac{1}{|T_i|}}{\sqrt{\sum_{k=1}^4 \frac{1}{|T_k|}}},$$

where $|T_i|$ denotes the area of the spherical triangle T_i. (See Fig. 9.) This ensures L^2-orthogonality between the different levels.

PSfrag replacements
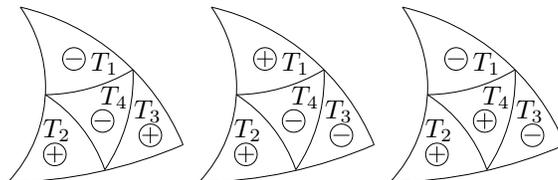
Figure 9: Haar wavelet basis functions on refined spherical triangle.

18

5 Adaptivity

5.1 Theory

The sparse tensor method described in the previous section is a powerful tool to reduce the number of degrees of freedom in the discretization in the case of an intensity function that is smooth with respect to physical space as well as solid angle.

However, for a large number of problems, this a-priori choice of degrees of freedom is not asymptotically optimal in the sense of a *best N-term approximation*: assuming that we have at hand a wavelet basis $\{\psi_\lambda(\mathbf{x}, \mathbf{s})\}_{\lambda \in \nabla}$ indexed by a multiindex λ in the index set ∇ which spans the space $\mathcal{V}_0$, such as the product wavelet basis constructed in the previous section, the best N-term approximation u_Λ^* of the intensity $u(\mathbf{x}, \mathbf{s})$ is an approximation of u from the space $V_0^\Lambda \subset \mathcal{V}_0$ given by

$$V_0^\Lambda := \operatorname{span}\{\psi_\lambda : \lambda \in \Lambda\}$$

for some index set $\Lambda \subset \nabla$ with $\hat{N}_L = \#\Lambda = \dim V_0^\Lambda$ many indices chosen such that

$$\|u - u_\Lambda^*\|_A \leq \inf_{\substack{\Lambda \subset \nabla \\ \#\Lambda = N}} \min_{v_\Lambda \in V_0^\Lambda} \|u - v_\Lambda\|_A.$$

In other words, the best N-term approximation u_Λ^* is the best approximation to u if we are willing to spend N degrees of freedom in $D \times S^2$. It is obvious that the best N-term approximation u_Λ^* of u converges at least as fast as the sparse tensor product approximation $\hat{u}^L$ with $N = \dim \widehat{V}_0^L$ degrees freedom. Therefore, by Theorem 4.1, also the best N-term approximation u_Λ^* of the intensity will be able to break the curse of dimension (see [28] for theoretical background on this). What is more, however, is that the set of solutions u for which u_Λ^* attains the convergence rate of the sparse tensor product approximation $\hat{u}_L$ with $N = \dim \widehat{V}_0^L$ degrees of freedom is much larger than $H^{p+1,p}(D \times S^2)$ – it is typically some Besov space (see [29] and the references there for details).

Consider a light beam, for example. On the one hand a very fine resolution in the direction of the ray as well as a fine spatial resolution, where the ray crosses the domain, is required. On the other hand, most degrees of freedom in physical space as well as solid angle can be neglected as the intensity there is zero. In some real-world applications, the intensity will neither be smooth everywhere in the domain nor only consist of a light beam in vacuum. We therefore opt for an adaptive algorithm that selects the degrees of freedom that are relevant for the given problem.

Since the supports of the tensorized basis functions in $D \times S^2$ are highly anisotropic subsets in $\mathbb{R}^{n+2}$, $n = 2, 3$, the usual residual based error indicators and estimators are not well suited to steer adaptive mesh-refinements in both coordinates, $\mathbf{x}$ and $\mathbf{s}$.

Therefore, in the ensuing implementation, we exploit another important feature of best N-term approximation, namely that a *near best N-term approximation* can be computed by *wavelet thresholding*, i.e. by simply keeping the N largest contributions to the solution, measured in the energy norm $\|\circ\|_A$ (cf. (30)) of our problem, of the wavelet expansion of u (e.g. [30], Thm. 4.3.1).

Adaptive sparse tensor product methods based on this idea have been applied to various problems, see eg. [20], [21], [22], [19]. A survey of adaptive wavelet techniques can be found, for example, in [23] and the references there.

Note carefully that while the norm equivalence of the wavelet coefficients to the norm $\|\circ\|_{H^{1,0}(D \times S^2)}$ is rather straightforward to establish, the norm equivalence of the wavelet coefficients to the energy norm $\|\circ\|_A$ in (30) is to date open. Nevertheless, we propose to use thresholding of wavelet coefficients also in the present context and describe next the implementation of an adaptive solver based on the above ideas.

5.2 Implementation

In order to describe the algorithm, we introduce a partial order ("parent-child relationship") of the basis functions as follows:

$$\psi(\mathbf{x}) \quad <_{\Omega} \quad \psi'(\mathbf{x}) \text{ if } \psi(\mathbf{x}) \text{ is a child of } \psi'(\mathbf{x}), \tag{54}$$

$$\chi(\mathbf{s}) \quad <_{S^2} \quad \chi'(\mathbf{s}) \text{ if } \chi(\mathbf{s}) \text{ is a child of } \chi'(\mathbf{s}), \tag{55}$$

$\psi(\mathbf{x})$, $\psi(\mathbf{x})'$ being wavelets in physical space and $\chi(\mathbf{s})$, $\chi(\mathbf{s})'$ being wavelets on the sphere.

We restrict ourselves to the two-dimensional case, as the numerical experiments are carried out for $n = 2$. Generalizations of all concepts below to $n = 3$ are straightforward.

In physical space, the children of a basis function corresponding to a given vertex correspond to the edge midpoints of the neighboring triangles of the vertex (see Fig. 10).

■ degree of freedom to be refined on level 3

· refined degrees of freedom on level 4 (children)
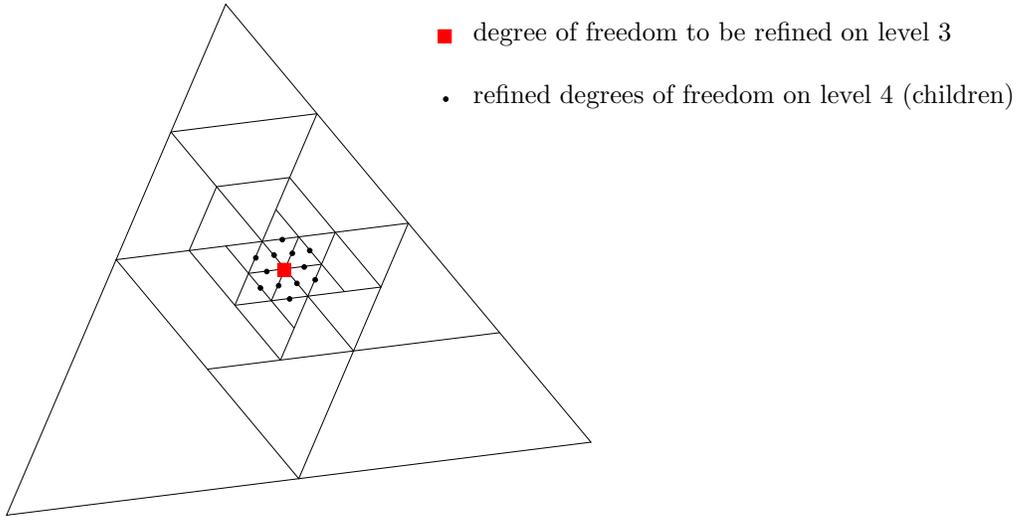
Figure 10: Refinement of a degree of freedom in 2D space.

In the solid angle, the children are the twelve basis functions on the next level that overlap with the basis function to be refined. Fig. 11 shows three parent wavelet functions with their twelve children. Each parent has twelve children and each child three parents (except on the coarsest level).

The adaptive algorithm proceeds levelwise (see also Fig 12).

1. We start with the computation of the intensity on the coarsest tensor product space and

2. select the degrees of freedom where the wavelet coefficients are above a given threshold.

3. We impose the constraint that complete trees have to be maintained. We therefore ensure that all ancestors of the selected degrees of freedom are also included in the set, i.e. if the tensor product of basis functions $\psi(\mathbf{x})\chi(\mathbf{s})$ is in the active set, we add recursively all degrees of freedom $\psi'(\mathbf{x})\chi'(\mathbf{s})$ with $(\psi = \psi', \chi <_{S^2} \chi')$ or $(\psi <_{\Omega} \psi', \chi = \chi')$.

4. We then recompute the solution with the active set of degrees of freedom before

5. we threshold the coefficients again.

6. We now add the degrees of freedom that correspond to refinements in physical space or solid angle of the active degrees of freedom, i.e. in order to refine a tensor product degree of freedom $\psi(\mathbf{x})\chi(\mathbf{s})$, we add all degrees of freedom $\psi'(\mathbf{x})\chi'(\mathbf{s})$ with either $(\psi' = \psi, \chi' <_{S^2} \chi)$ or $(\psi' <_{\Omega} \psi, \chi' = \chi)$.

20
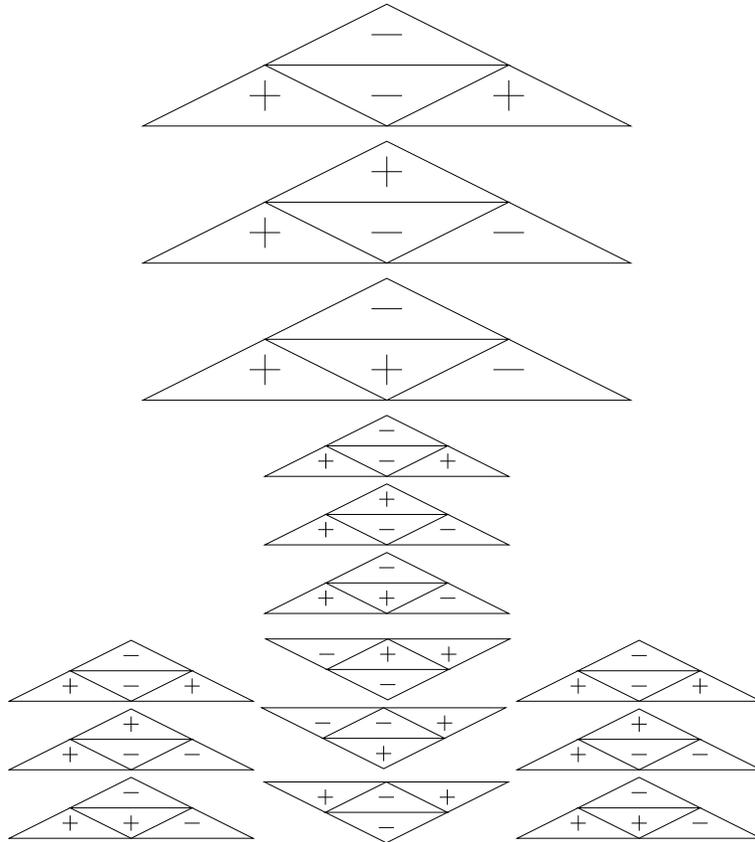
Figure 11: Parents-children relationship of Haar wavelet functions on nested triangulations on the sphere.

7. Again, we add all coarser degrees of freedom.

8. After recomputing the solution, we repeat this procedure on higher levels until the finest level L is reached.

Although the solution has to be computed twice in each iteration step, the method is not too expensive, as a good initial guess is available from the previous step.

We select the active degrees of freedom by simply thresholding the wavelet coefficients. A residual based strategy would be an alternative, see eg. [22].

PSfrag replacements
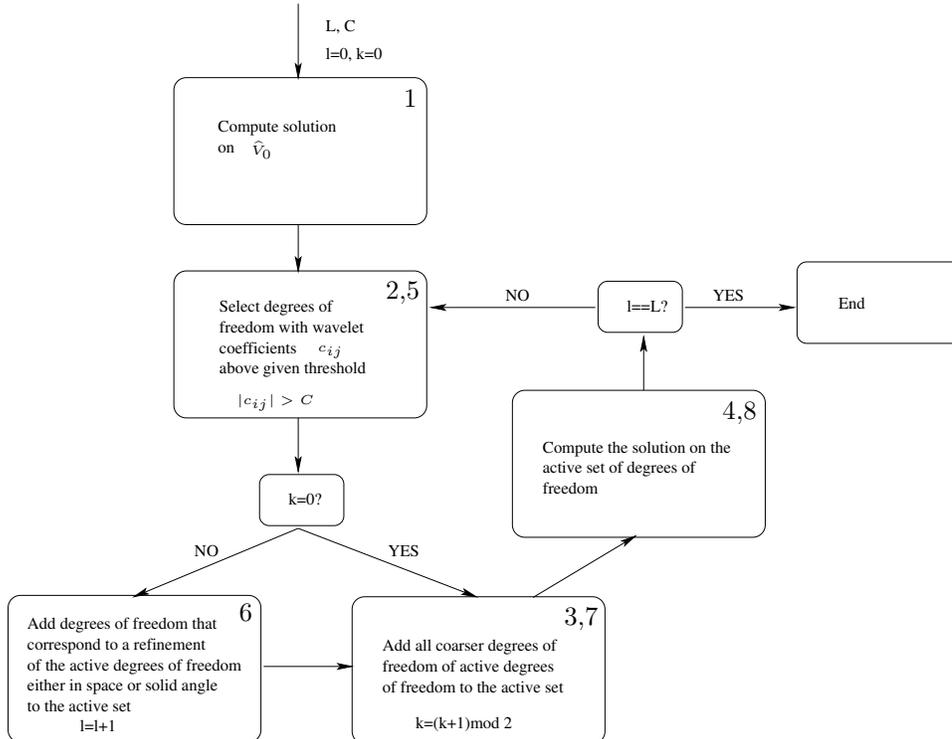
Figure 12: Flowchart of the algorithm.

In Fig. 13, one can see which coarser degrees of freedom have to be active, if the set contains a certain vertex on the finest level.

A consistent refinement of a degree of freedom is shown in Fig. 14.

As the set of active degrees of freedom is a subset of the degrees of freedom of a sparse tensor product space, the adaptive algorithm can be seen as an additional sparsification. However, in extreme examples as in the one of a narrow light beam, this algorithm allows to go up to much higher levels and the selected degrees of freedom will have little in common with the original sparse tensor product space.

6   Numerical Experiments

We test our method on the model circular 2D spatial domain

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^2, \ |\mathbf{x}| \le 1 \right\}, \tag{56}$$

with different sets of emission/ absorption data. For all examples, we use non-emitting walls, i.e. $g = 0$.
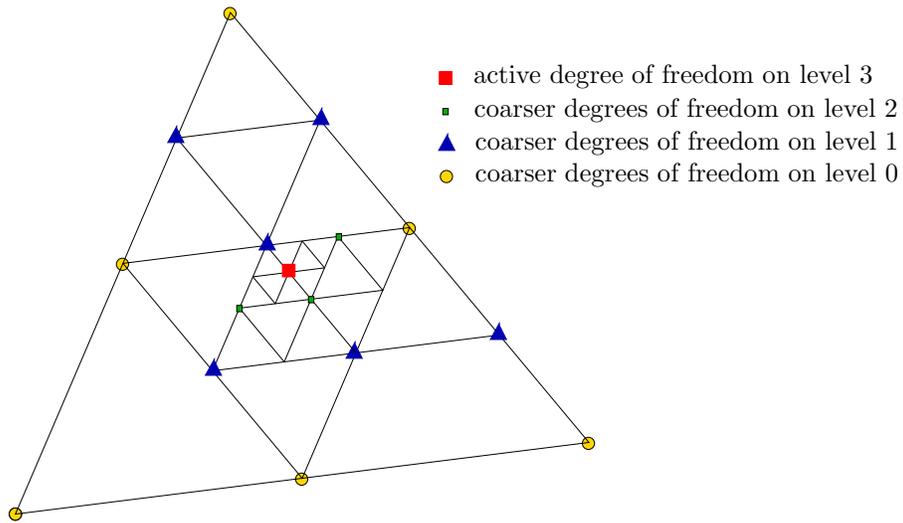
Figure 13: Coarser degrees of freedom in 2D space that have to be contained in the mesh when a degree of freedom on the finest level is active.
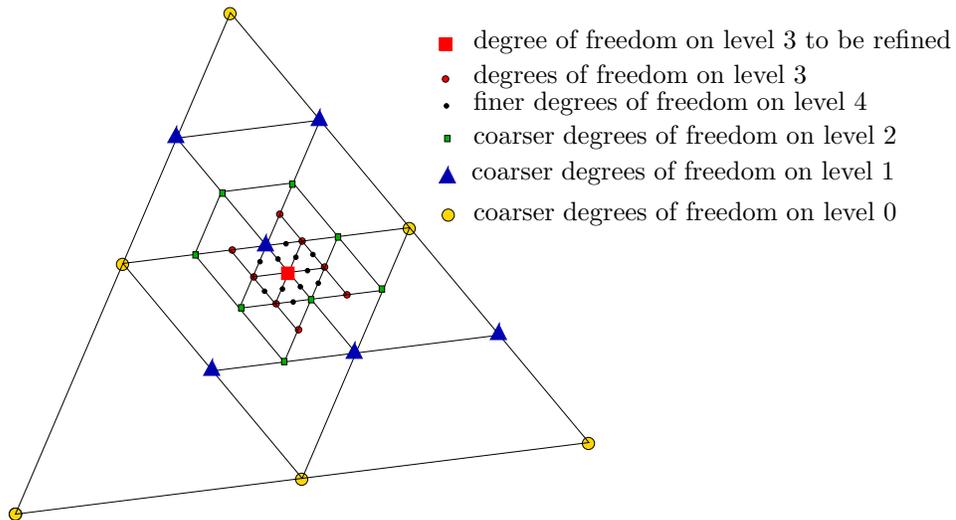
Figure 14: Active degrees of freedom in 2D space after refining a degree of freedom on level 3.

The nested meshes we use for the definition of the multilevel hierarchies are shown in Fig. 2. In physical space, they consist of 41, 145, 545 and 2113 degrees of freedom, in solid angle of 20, 80, 320 and 1280 degrees of freedom. In order to compute the solution for a given set of degrees of freedom, we compute the stiffness matrices with respect to physical space and solid angle and use the conjugate gradient method with diagonal preconditioning to solve the linear system. We stop the iteration when the relative error in the energy norm is smaller than 10^{-6}. We impose the boundary conditions by projecting the solution onto the subspace described in section 4 in each iteration step.

Incident radiation

$$G(\mathbf{x}) = \int_{S^2} u(\mathbf{x}, \mathbf{s}) \, d\mathbf{s}$$

and net emission

$$E(\mathbf{x}) = \kappa(\mathbf{x}) \left(4\pi f(\mathbf{x}) - G(\mathbf{x})\right).$$

are important quantities in radiative transfer simulations. We plot profiles along the (positive) x_1-axis of those quantities for the different methods and levels.

In order to obtain error estimates, we compute reference solutions by line integration of the transport-reaction equation on level 3 in physical space and solid angle. We measure the error of the radiation intensity, the incident radiation and the heat flux in suitable norms.

We therefore define the following errors:

$$I_A := \frac{||u - I_{ref}||_{A(D \times S^2)}}{||u||_{A(D \times S^2)}},$$

$$I_{H^{1,0}} := \frac{||u - I_{ref}||_{H^{1,0}(D \times S^2)}}{||u||_{H^{1,0}(D \times S^2)}},$$

$$G_{L^2} := \frac{||G - G_{ref}||_{L^2(D)}}{||G||_{L^2(D)}},$$

and

$$q_{L^2} := \frac{||\mathbf{q} - \mathbf{q}_{ref}||_{L^2(D)}}{||\mathbf{q}||_{L^2(D)}},$$

where

$$\mathbf{q}(\mathbf{x}) := \int_{S^2} u(\mathbf{x}, \mathbf{s}) \, \mathbf{s} \, d\mathbf{s}$$

is the heat flux and I_{ref}, G_{ref} and $\mathbf{q}_{ref}$ are the reference solutions of the intensity, incident radiation and heat flux.

We visualize the efficiency of the (adaptive) sparse tensor product approximation by plotting the error in the intensity (resp. in the incident radiation or heat flux) versus the number of active degrees of freedom on levels 0 to 3.

Example (1) illustrates the performance of the sparse tensor product method, when the intensity is a smooth function with respect to physical space as well as solid angle due to a large absorption coefficient. The blackbody intensity is given in Fig. 15, while the absorption coefficient is 10 everywhere in the domain. In Fig. 16 the absolute values of the coefficients with respect to the tensor product wavelet basis of the full tensor product solution are displayed, while Fig. 17 shows the $\hat{\mathcal{N}}_L$ largest wavelet coefficients. Here, $\hat{\mathcal{N}}_L$ is the number of degrees of freedom in the sparse tensor product space. In both figures, the sparse tensor product space corresponds to the area to the left and above the blue line. As expected, most of the $\hat{\mathcal{N}}_L$ largest coefficients are contained in the sparse tensor product space.
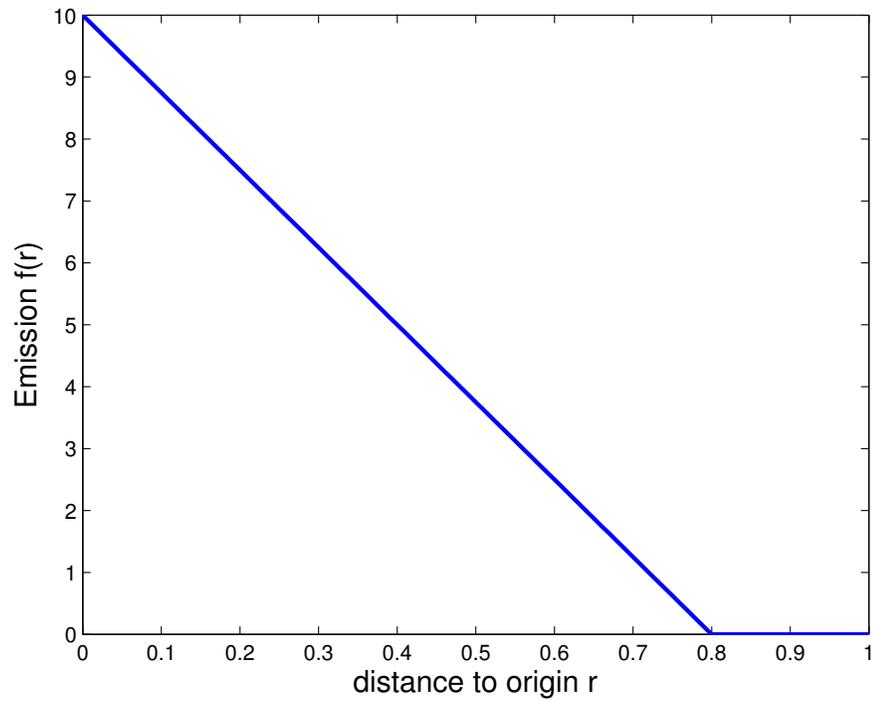
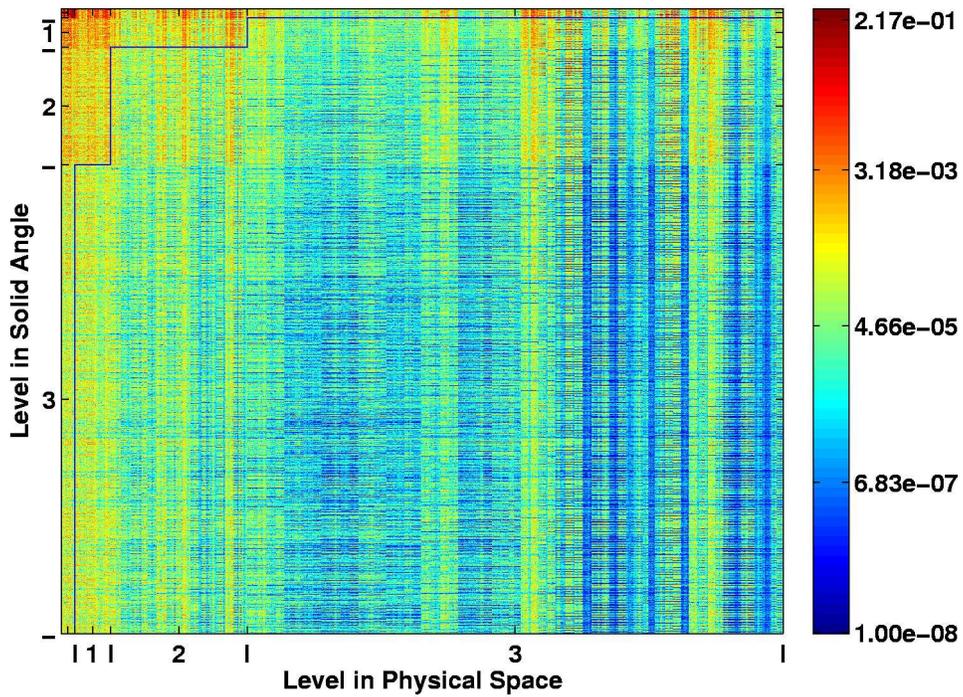Figure 15: Blackbody intensity for example 1.
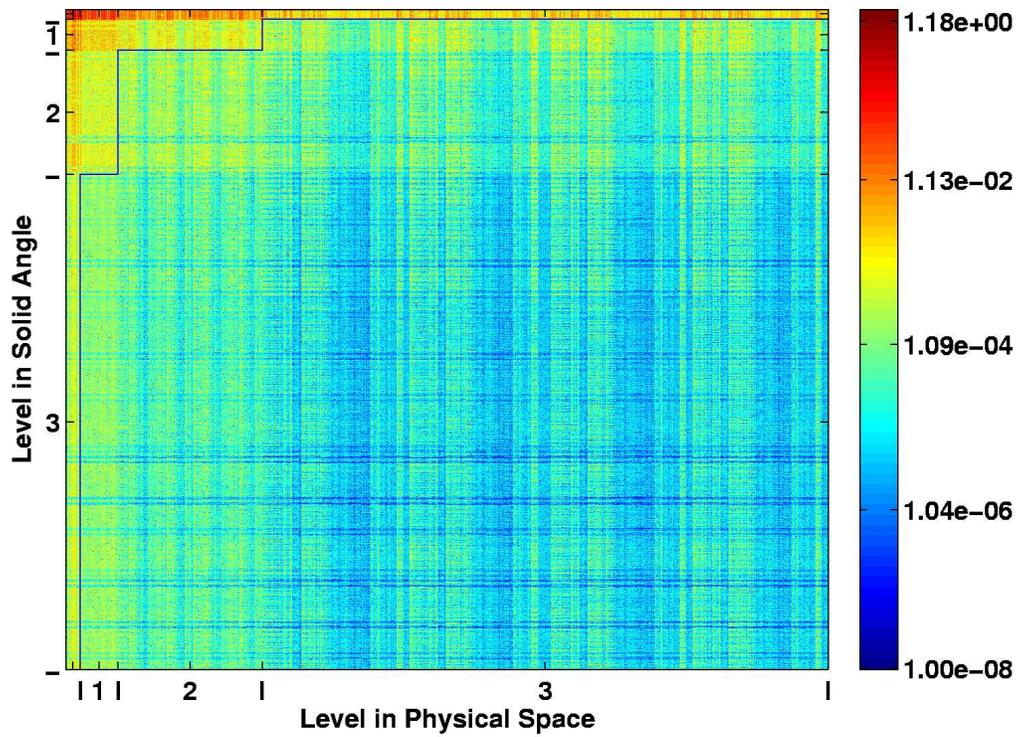
Figure 16: Size of wavelet coefficients of the full tensor product solution for example 1.
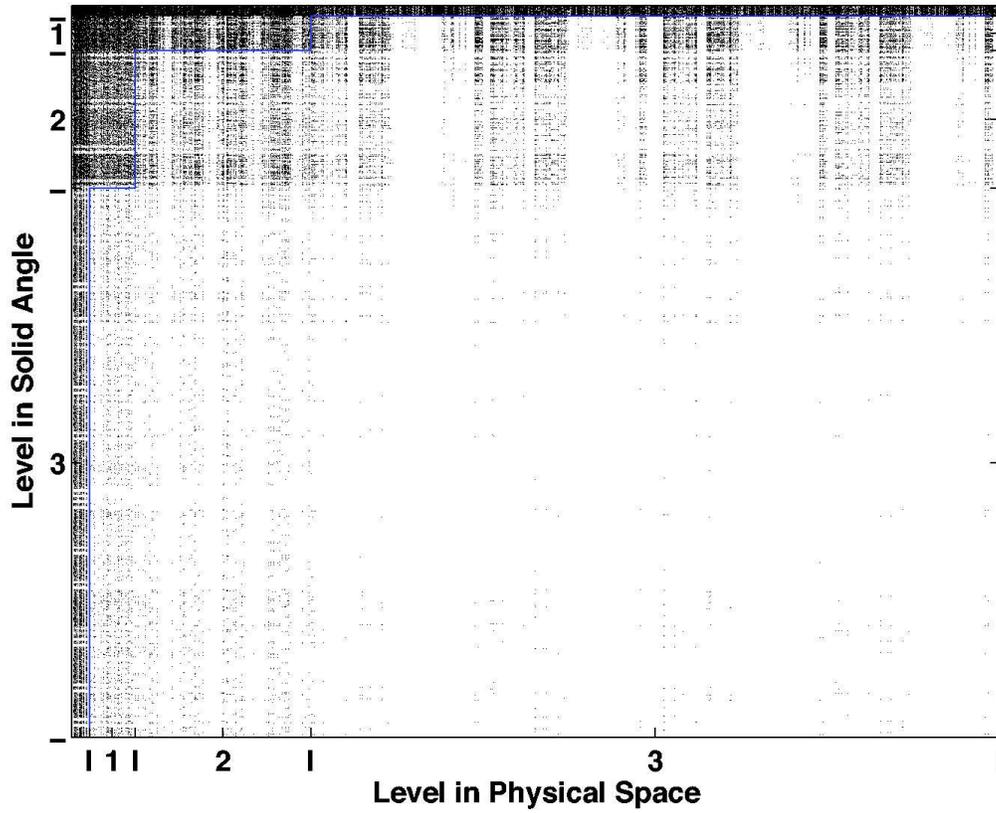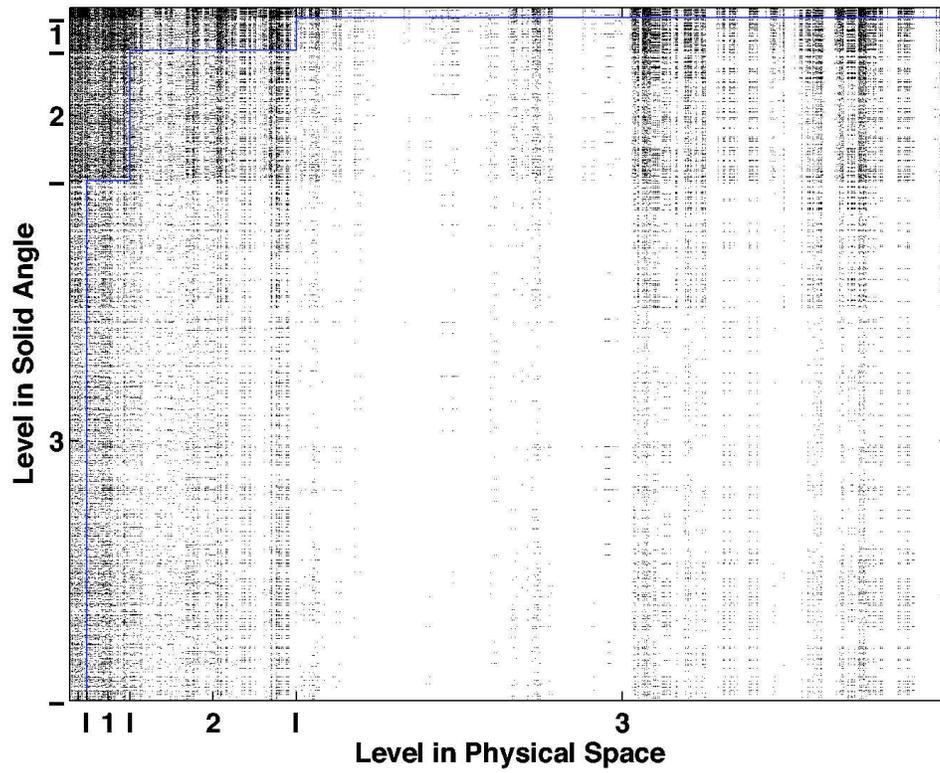
Figure 17: $\hat{\mathcal{N}}_L = 149120$ largest wavelet coefficients of the full tensor product solution for example 1 at level $L = 3$.

The excellent approximation properties of the sparse tensor product space are also confirmed by the profiles of the incident radiation and net emission that are shown in Figs. 18 and 19
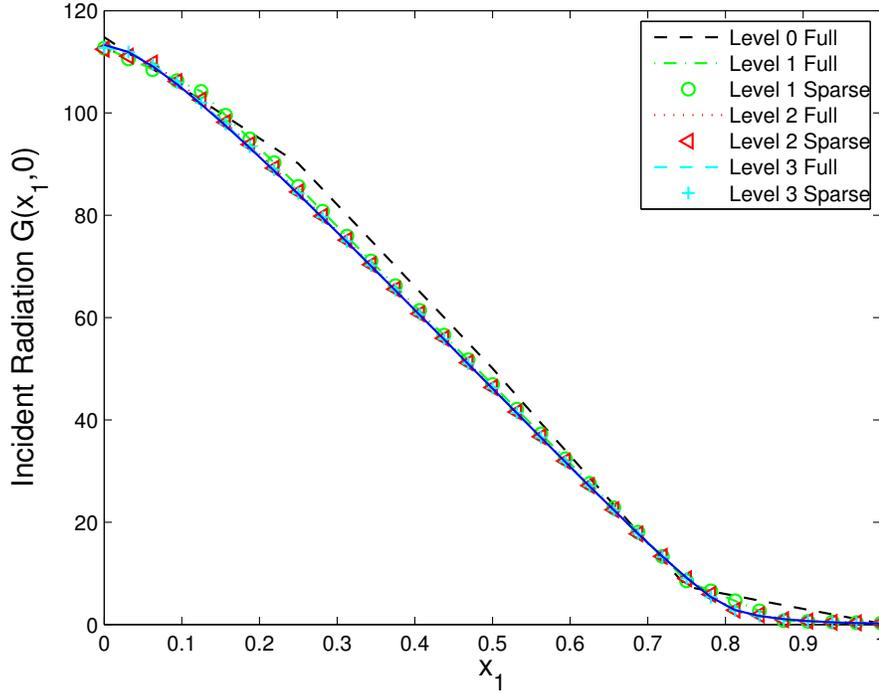
Figure 18: Incident radiation for example 1.

and the convergence results in Figs. (20) to (23): The sparse tensor product approximation is (almost) as accurate as the full tensor product approximation, while the number of degrees of freedom is reduced from $N \cdot M$ to $\mathcal{O}(N \log M + M \log N)$, where M is the number of degrees of freedom in physical space D and N the number of degrees of freedom in solid angle S^2.

In **example (2)**, the domain is divided into an absorbing and non-absorbing area with absorption coefficient

$$\kappa(\mathbf{x}) = \left\{ \begin{array}{ll} 10, & x_1 < -\frac{1}{15}, \\ 0, & \text{otherwise.} \end{array} \right.$$

The blackbody intensity is equal to 1 everywhere in the domain. For $x_1 < -\frac{1}{15}$, the intensity is smooth thanks to the large absorption coefficient, whereas for $x_1 > -\frac{1}{15}$ there is only radiation in directions with positive x_1-components. Due to the discontinuity at $x_1 = -\frac{1}{15}$, the solution does not satisfy the regularity requirements for the sparse tensor product approximation. However, as the solution is smooth with respect to a large part of $(D \times S^2)$, most of the largest wavelet coefficients are still contained in the sparse tensor product space (see Fig. 24, 25). This is also confirmed by the convergence rates in Figs. 28 to 31. Apart from the error arising from the discontinuity at $x_1 = -\frac{1}{15}$ and the non-absorbing medium for $x_1 > -\frac{1}{15}$, a considerable contribution to the discretization error occurs at the boundary, as inflow and outflow directions are not well resolved by the angular discretization (Figs. 26 and 27).

As the solution is either smooth or directed into the positive x_1-direction, many degrees of freedom within the sparse tensor product space can be neglected. As can be seen in Figs. 28 to 31, the adaptive sparse tensor product method significantly reduces the number of degrees of freedom without any loss of accuracy with respect to the sparse tensor method. When looking at Fig. 26, it becomes clear that all methods fail to produce an accurate solution and that a much finer discretization level is required. Due to limitations of the current implementation, we could
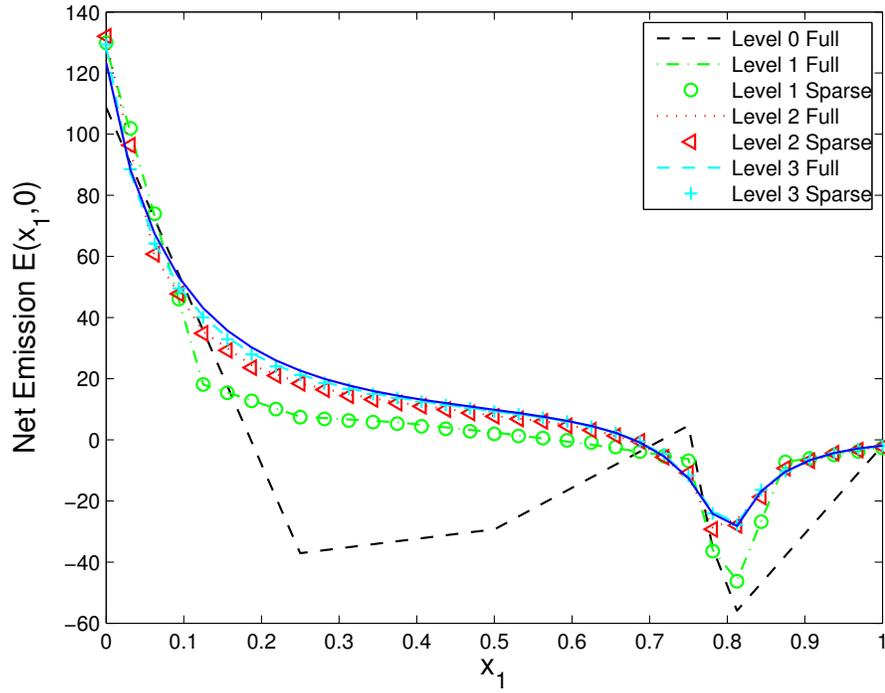
Figure 19: Net emission for example 1.
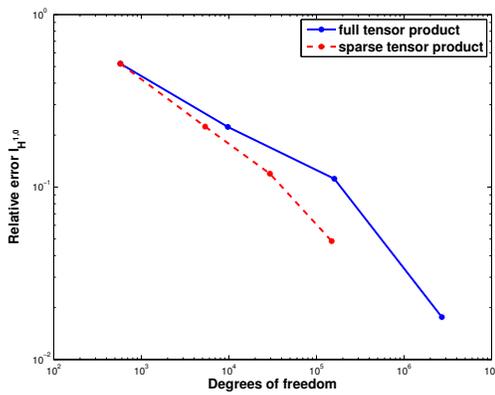
Figure 20: Example 1: Relative intensity er-Figure 21: Example 1: Relative intensity error in the $H^{1,0}(D \times S^2)$-norm. ror in the $A(D \times S^2)$-norm.
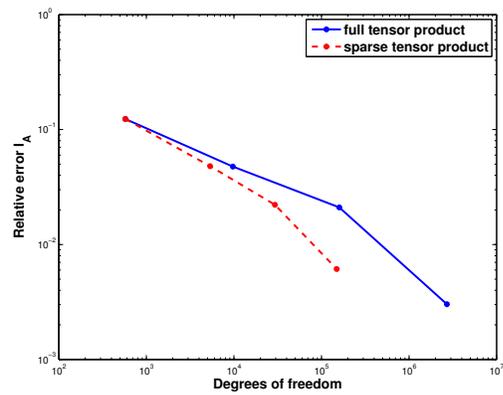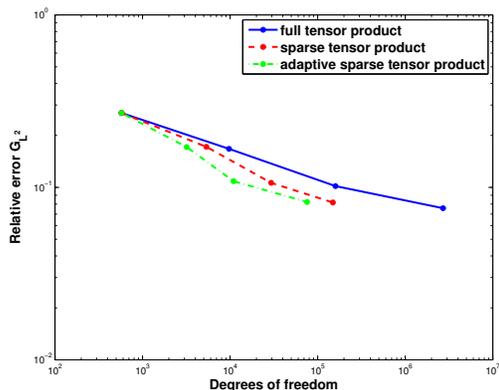
28
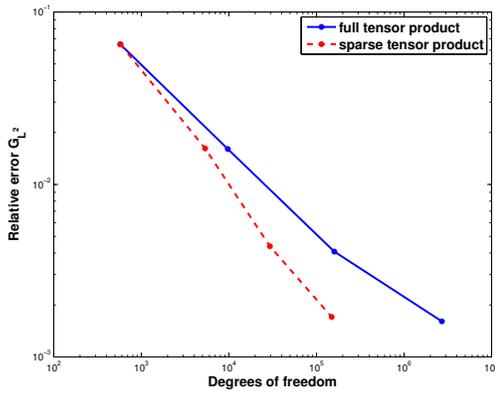
Figure 22: Example 1: Relative error of the incident radiation in the $L^2(D)$-norm.
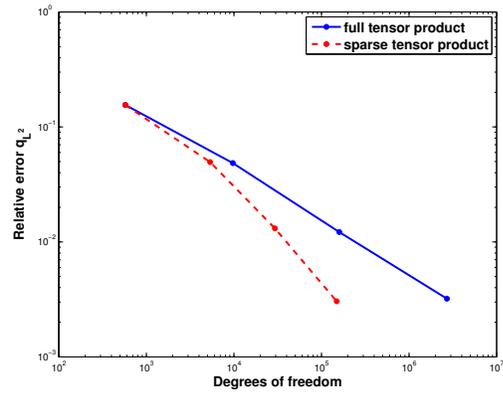
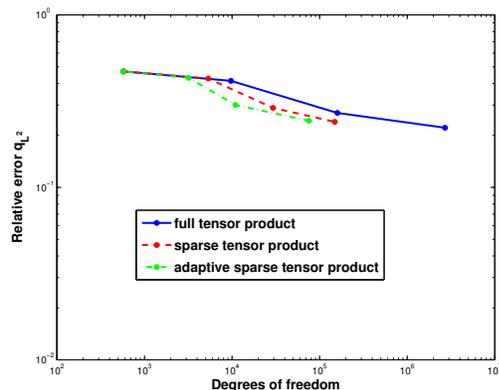Figure 23: Example 1: Relative error of the heat flux in the $L^2(D)$-norm.
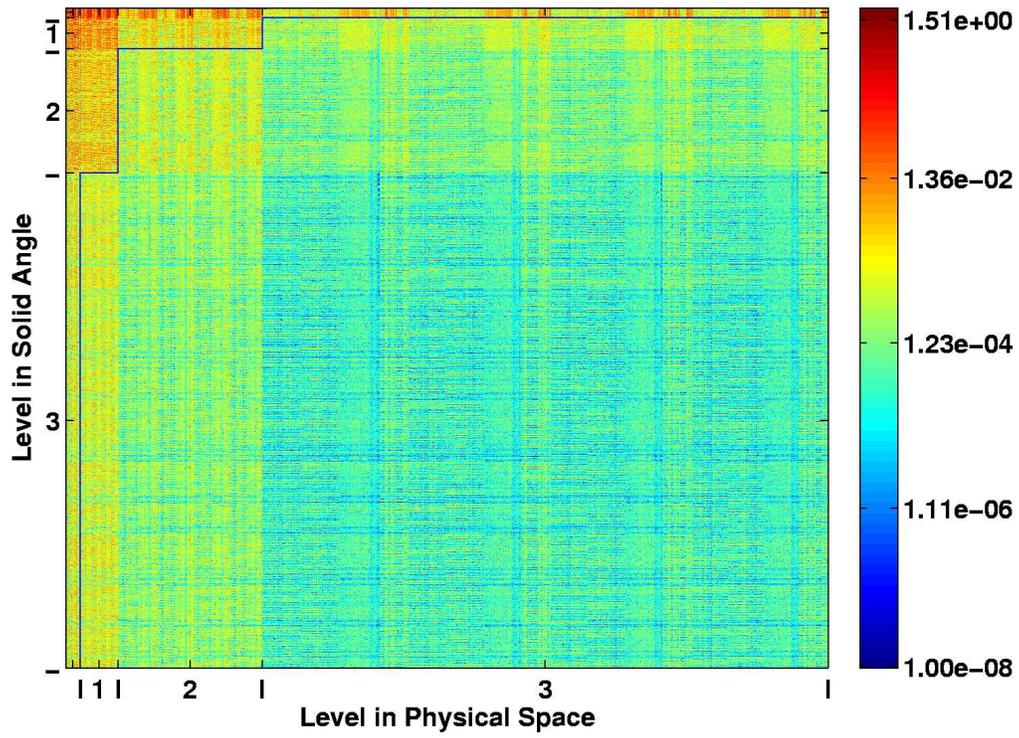

Figure 24: Size of wavelet coefficients of the full tensor product solution for example 2.

Figure 25: $\hat{\mathcal{N}}_L = 149120$ largest wavelet coefficients of the full tensor product solution for example 2 at level $L = 3$.
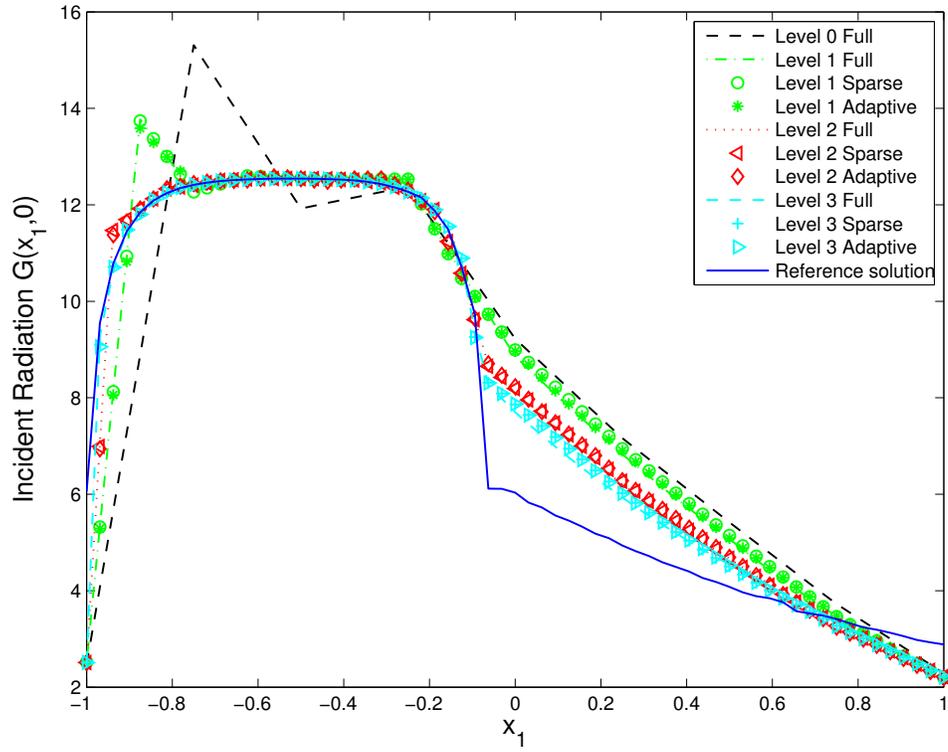
Figure 26: Incident radiation for example 2.
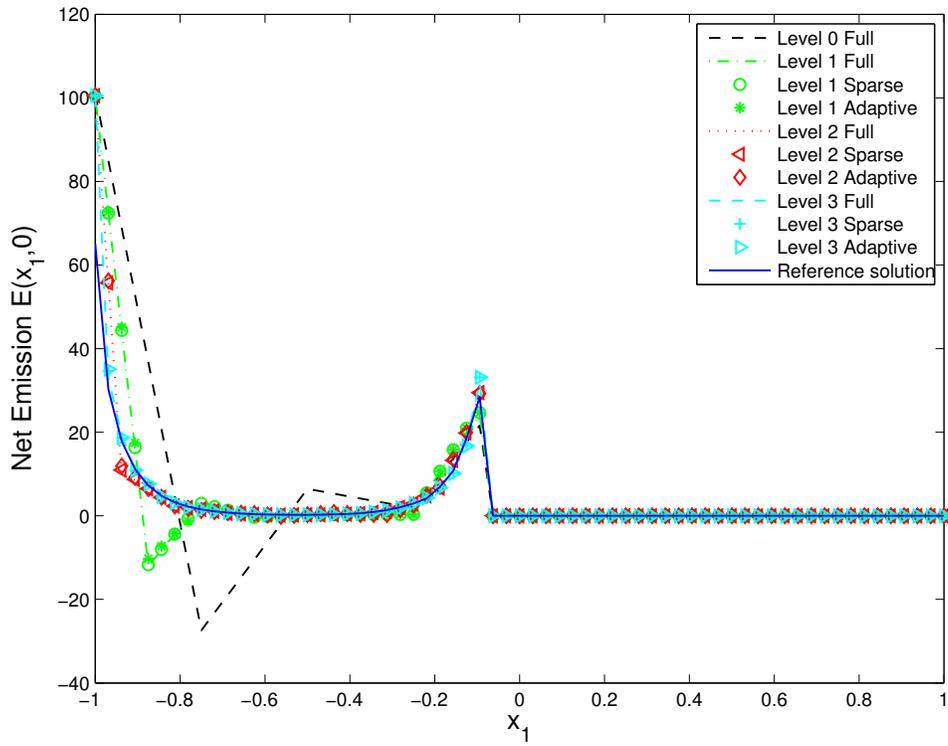
Figure 27: Net emission for example 2.

not increase the level any further. However, the (adaptive) sparse tensor product approach has potential for further refinement and therefore for providing a more accurate discretization.
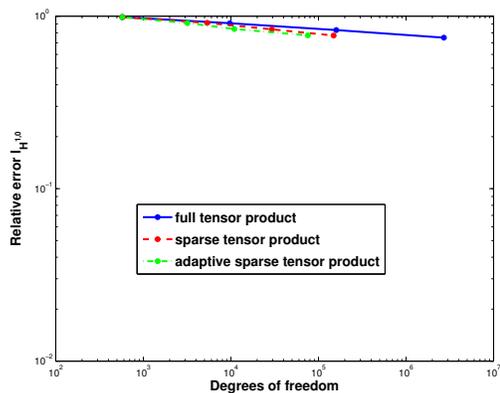
Figure 28: Example 2: Relative intensity error in the $H^{1,0}(D \times S^2)$-norm.
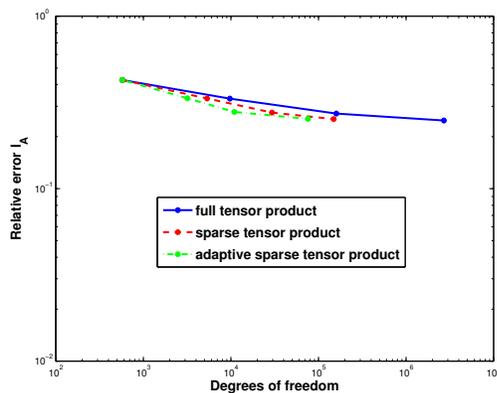
Figure 29: Example 2: Relative intensity error in the $A(D \times S^2)$-norm.

Figure 30: Example 2: Relative error of the incident radiation in the $L^2(D)$-norm.

Figure 31: Example 2: Relative error of the heat flux in the $L^2(D)$-norm.

In **example (3)** (see Figs. 32, 33), there is a radiating zone in the center of the domain with an exponential decay of the blackbody intensity to zero between $|\mathbf{x}| = 0.2$ and roughly $|\mathbf{x}| = 0.7$. As the absorption coefficient has the same shape, energy is emitted in the center and transported to the boundary of the domain.

Again, the structure of the wavelet coefficients indicates that the solution can be well approximated in the (adaptive) sparse tensor product space (Figs. 34, 35).

The results of the incident radiation and net emission confirm that. With the exception of the adaptive sparse tensor approximation on level 3, the error increases only slightly when the (adaptive) sparse tensor product method is applied. Figs. 38 to 41 show that the sparse tensor product method is clearly superior to the full tensor product approximation. Introducing adaptivity, we further improve the efficiency. However, as in particular the results in the heat flux in Figs. 37 and 41 indicate, the selection of the degrees of freedom in the adaptive algorithm could most likely be improved.

In **example (4)**, the radiating source in the center is not of radial symmetry. The absorption coefficient is given in Fig. 42, while $f(\mathbf{x}) = 1$.

The results are displayed in Figs. 43 to 50 and confirm that radial symmetry is not required for the approximation properties.
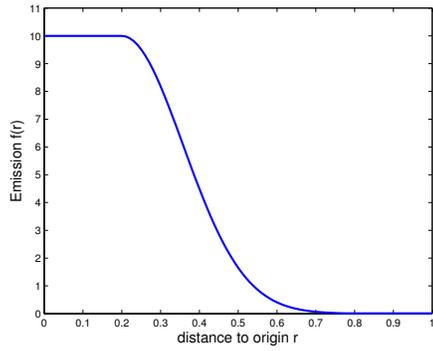
Figure 32: Blackbody intensity for example 3.
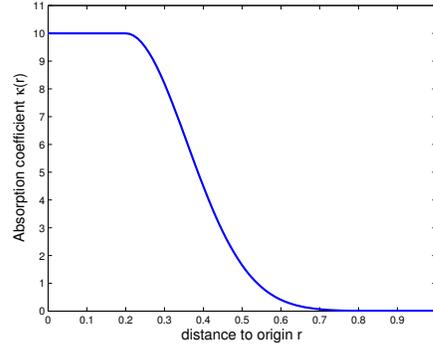
Figure 33: Absorption coefficient for example 3.

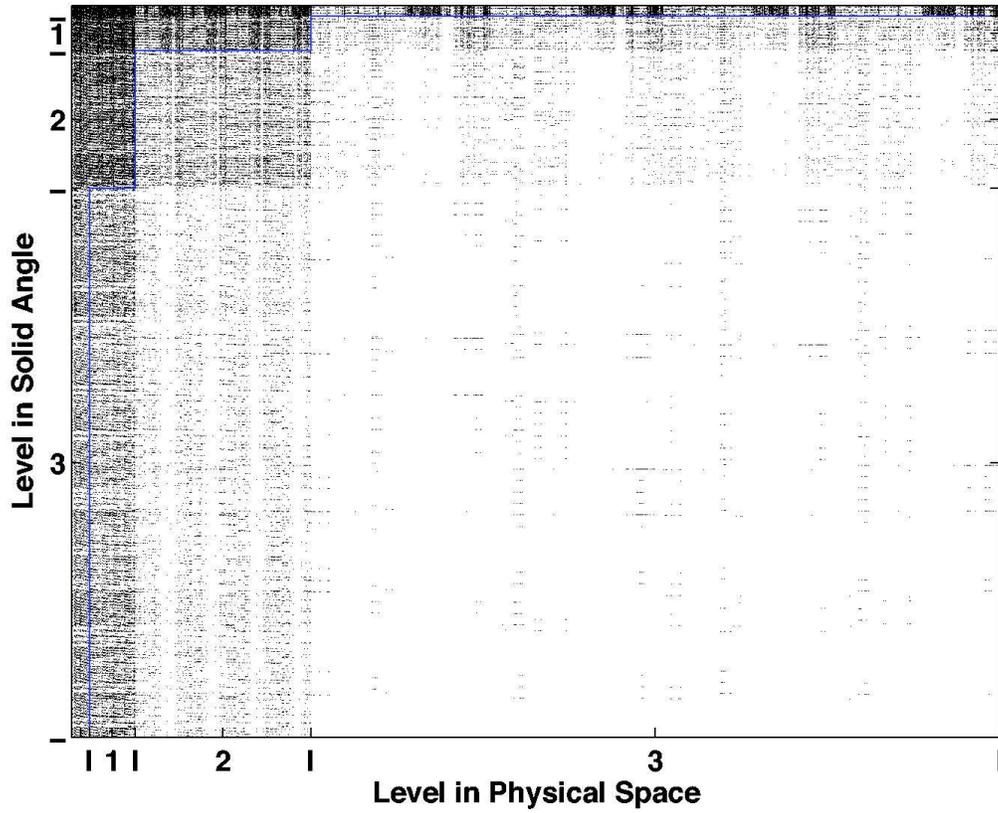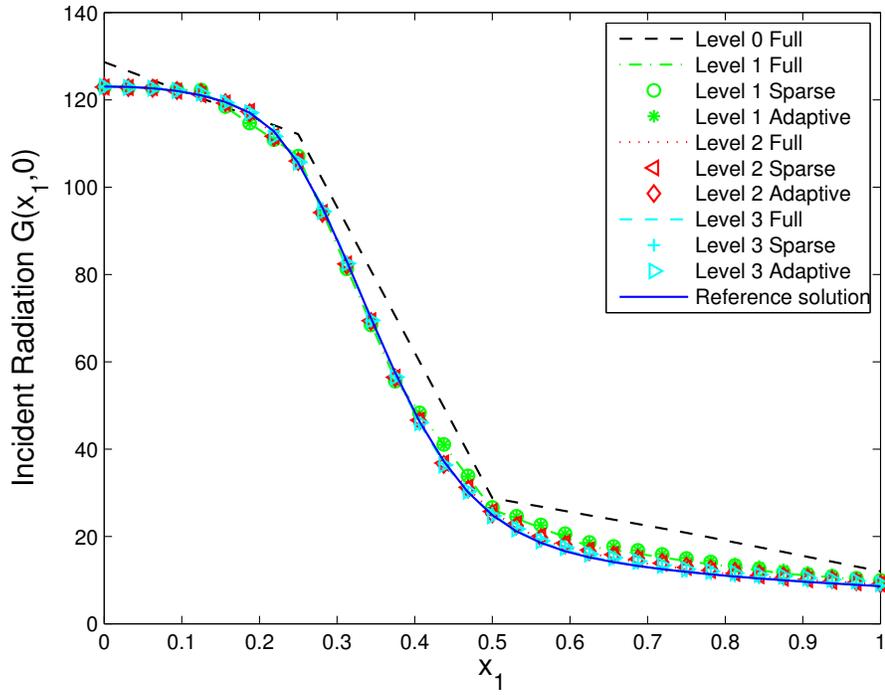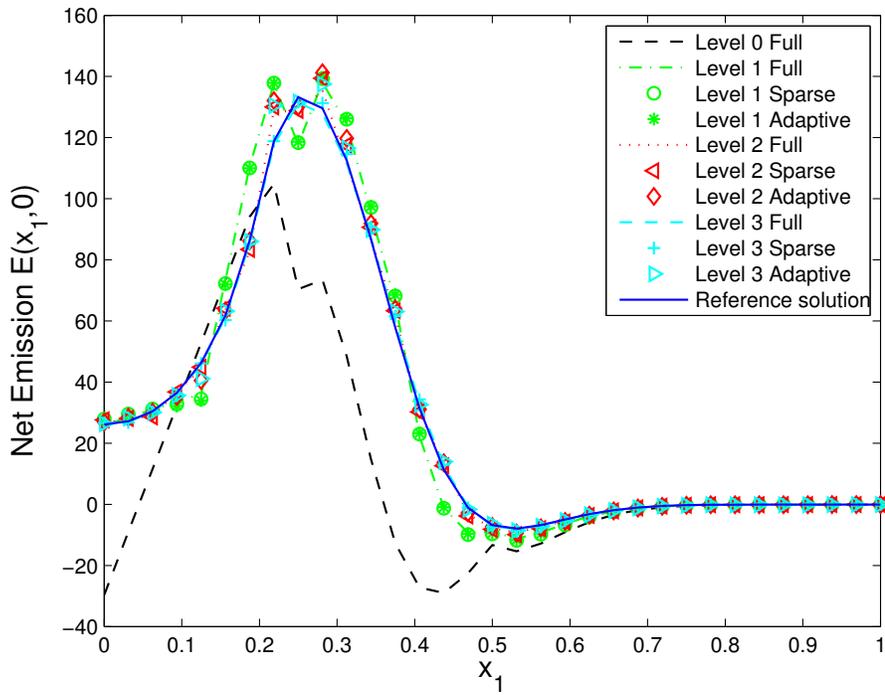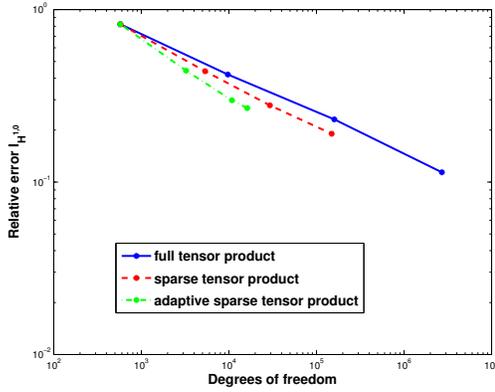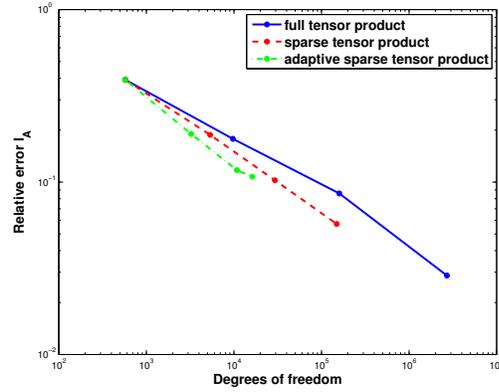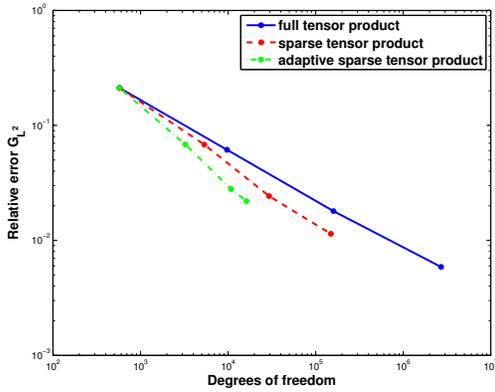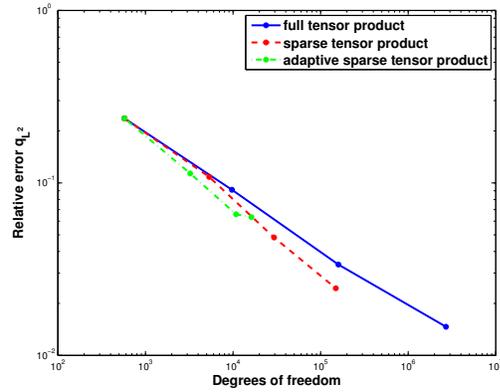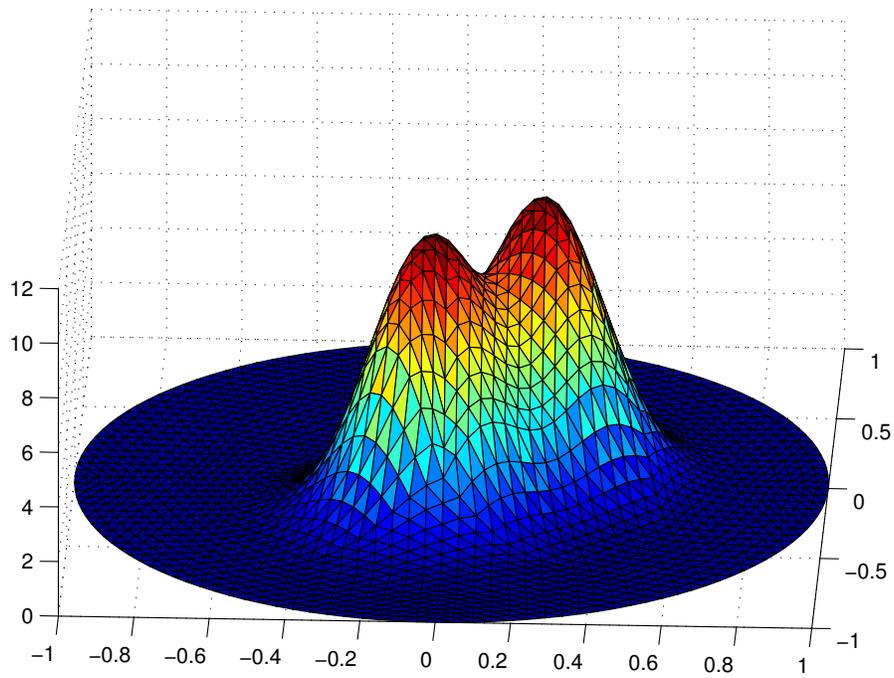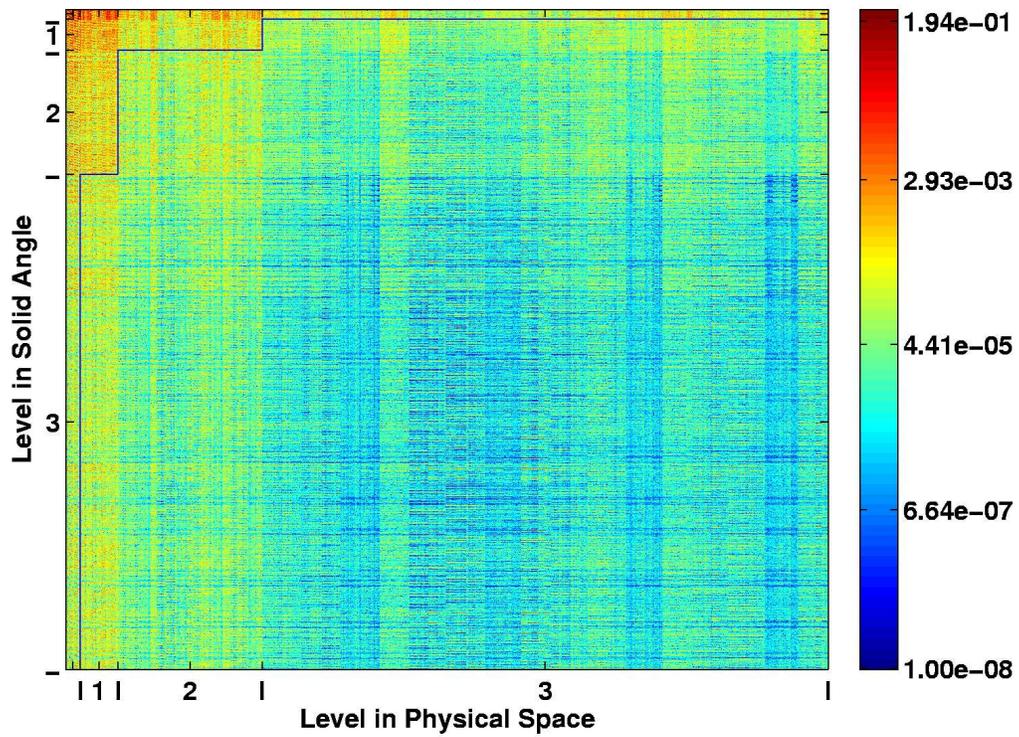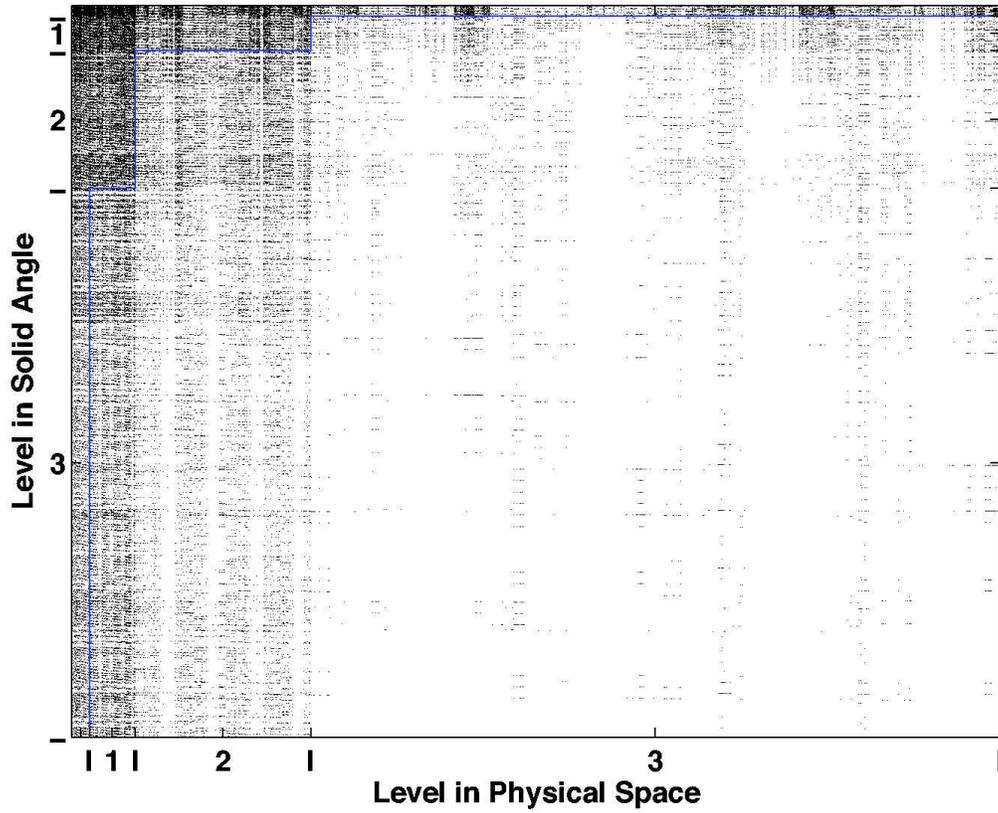Figure 34: Size of wavelet coefficients of the full tensor product solution for example 3.

Figure 35: $\hat{\mathcal{N}}_L = 149120$ largest wavelet coefficients of the full tensor product solution for example 3 at level $L = 3$.

Figure 36: Incident radiation for example 3.

Figure 37: Net emission for example 3.

Figure 38: Example 3: Relative intensity error in the $H^{1,0}(D \times S^2)$-norm.

Figure 39: Example 3: Relative intensity error in the $A(D \times S^2)$-norm.

Figure 40: Example 3: Relative error of the incident radiation in the $L^2(D)$-norm.

Figure 41: Example 3: Relative error of the heat flux in the $L^2(D)$-norm.

Figure 42: Absorption coefficient for example 4.

Figure 43: Size of wavelet coefficients of the full tensor product solution for example 4.

Figure 44: $\hat{\mathcal{N}}_L = 149120$ largest wavelet coefficients of the full tensor product solution for example 4 at level $L = 3$.
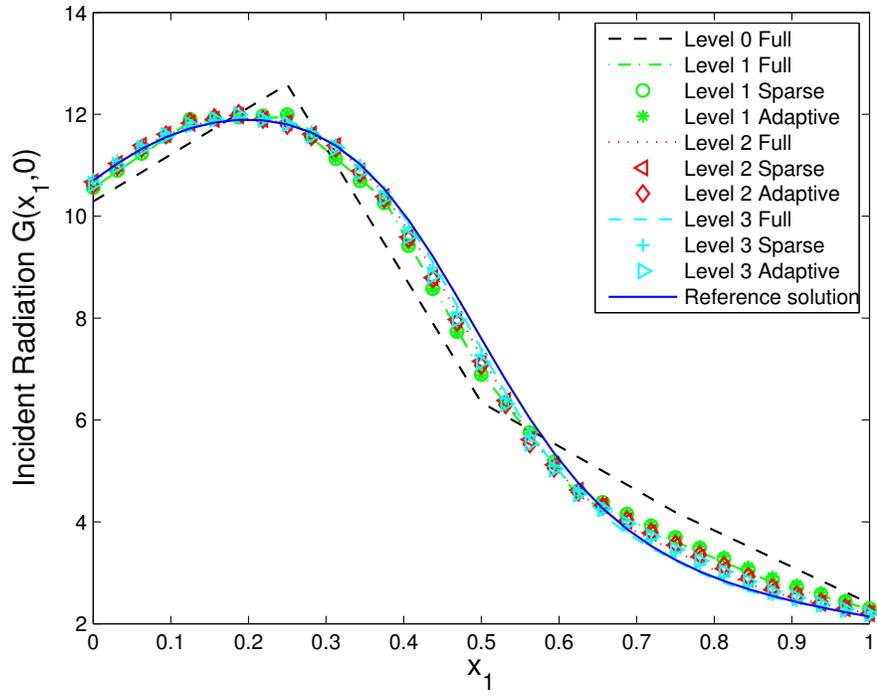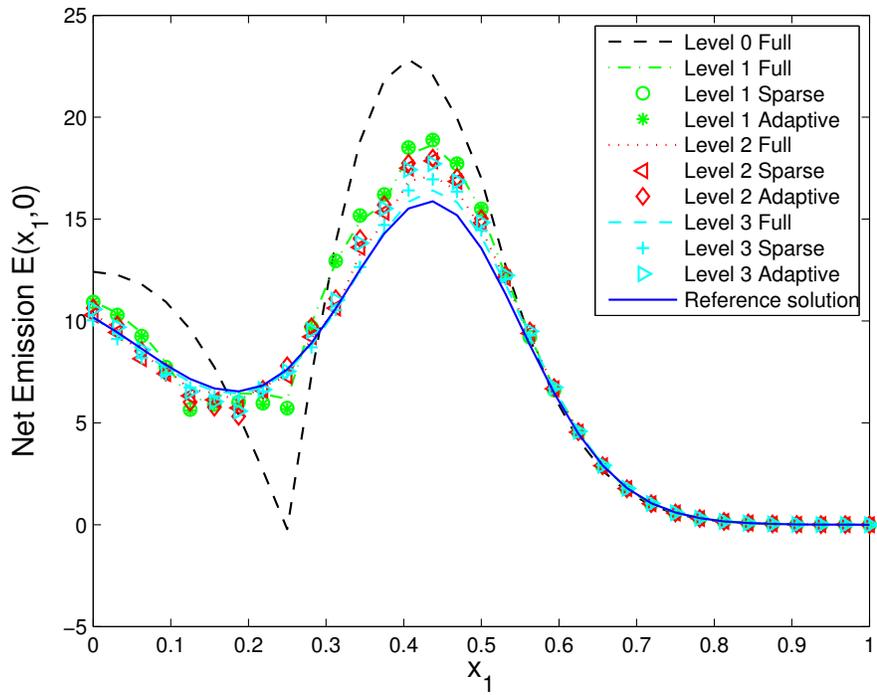
Figure 45: Incident radiation for example 4.
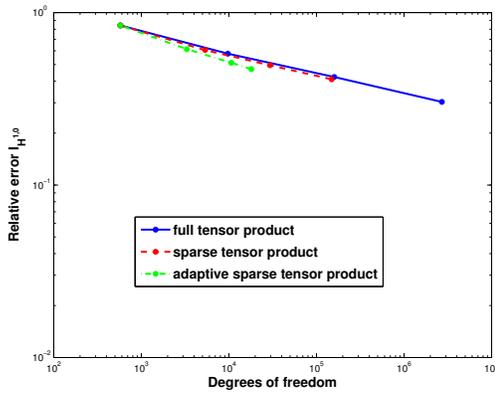

Figure 46: Net emission for example 4.

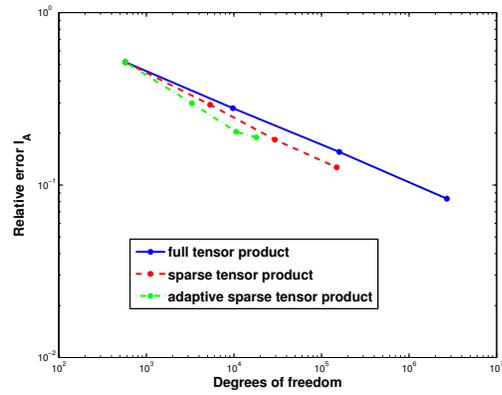Figure 47: Example 4:Relative intensity error in the $H^{1,0}(D \times S^2)$-norm.

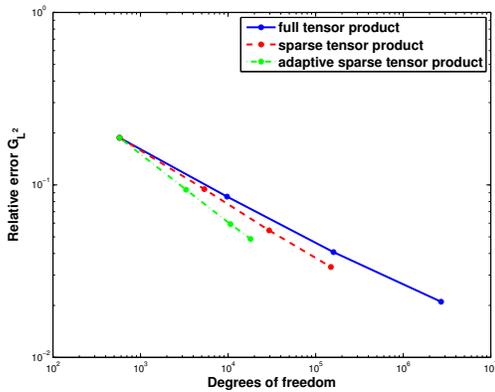Figure 48: Example 4:Relative intensity error in the $A(D \times S^2)$-norm.

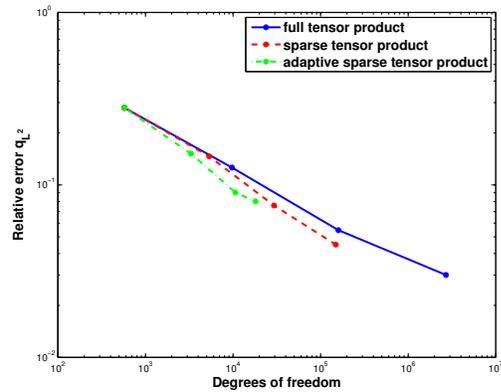Figure 49: Example 4: Relative error of the incident radiation in the $L^2(D)$-norm.

Figure 50: Example 4: Relative error of the heat flux in the $L^2(D)$-norm.

40

Remark 2. The simple diagonally preconditioned CG-solver combined with projection methods to impose the boundary conditions is not satisfactory, as the number of iterations increases with the number of levels and the algorithm even does not converge to a relative error of 10^{-6} on level 3 for some problems. However, as we focus on the approximation properties of the different spaces, we content ourselves here with that straightforward method for the moment.

7 Conclusions

We have presented an efficient method to discretize the radiative transfer equation for arbitrary absorption coefficients without scattering. For solutions of sufficient regularity, the sparse tensor product approximation is (almost) as accurate as the full tensor product approximation, while the number of degrees of freedom is reduced from $N_L \cdot M_L$ to $\mathcal{O}(N_L \log M_L + M_L \log N_L)$, where N_L is the number of degrees of freedom in physical space D and M_L the number of degrees of freedom in solid angle S^2. Here, we used only the lowest degree finite elements, namely $p = 1$ in D and $q = 0$ in S^2. Even with this lowest order method, in our numerical experiments we could achieve with our sparse tensor product method an accuracy comparable to that of the full tensor product method at about 1 to 2 order magnitude fewer degrees of freedom.

We obtained with a simple adaptive refinement strategy based on thresholding the solution's wavelet coefficients in various examples an additional reduction in the number of the degrees of freedom by a factor of 10 while still retaining the accuracy of the scheme. If the radiation intensity u is piecewise smooth, increasing the approximation order to $p > 1$ in D and to $q > 0$ in S^2 and applying wavelet coefficient thresholding allows a further reduction of degrees of freedom by selecting the most relevant contributions.

References

[1] M. F. Modest. *Radiative heat transfer*. Academic Press, second edition, 2003.

[2] R. Wehrse and W. Kalkofen. Advances in radiative transfer. *Astron. Astrophysics. Rev.*, 12(1/2):3–29, 2006.

[3] T. A. Manteuffel, K. J. Ressel, and G. Starke. A boundary functional for the least-squares finite-element solution of neutron transport problems. *SIAM J. Numer. Anal.*, 37(2):556–586, 2000.

[4] G. Kanschat. Parallel and adaptive Galerkin methods for radiative transfer problems. *Thesis University of Heidelberg*, 1996.

[5] R. Bellmann. *Adaptive control processes: A guided tour*. Princeton University Press, 1961.

[6] B. Jiang. Non-oscillatory and non-diffusive solution of convection problems by the iteratively reweighted least squares finite element method. *J. Comp. Phys.*, 105:108–121, 1993.

[7] T. A. Manteuffel and K. J. Ressel. Least-squares finite-element solution of the neutron transport equation in diffusive regimes. *SIAM J. Numer. Anal.*, 35(2):806–835, 1998.

[8] L. Bourhrara. New variational formulations for the neutron transport equation. *Transport theory and statistical physics*, 33(2):93–124, 2004.

[9] C. Zenger. Sparse grids. *In Parallel algorithms for partial differential equations: Proceedings of the Sixth GAMM-Seminar, Kiel*, January 1990.

[10] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18(3-4):209–232, 1998.

[11] H. Yserentant. Sparse grid spaces for the numerical solution of the electronic schrödinger equation. *Numerische Mathematik*, 101(2):381–389, 2005.

[12] M. Griebel. Adaptive sparse grid multilevel methods for elliptic pdes based on finite differences. *Computing*, 61(2):151–179, 1998.

[13] V. H. Hoang and C. Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul.*, 3(1):168–194, 2005.

[14] T Von Petersdorff and C. Schwab. Numerical solution of parabolic equations in high dimensions. *M2AN Math. Model. Numer. Anal.*, 38(1):93–127, 2004.

[15] H.-J. Bungartz. A multigrid algorithm for higher order finite elements on sparse grids. *Electronic Transactions on Numerical Analysis*, 6:63–77, December 1997.

[16] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numer. Math.*, 83(2):279–312, 1999.

[17] R. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[18] R. Balder and C. Zenger. The solution of multidimensional real Helmholtz equations on sparse grids. *SIAM J. Sci. Comput.*, 17(3):631–646, May 1996.

[19] J. Nordmans and P. W. Hemker. Application of an adaptive sparse-grid technique to a model singular perturbation problem. *Computing*, 65:357–378, 2000.

[20] A. Cohen, Dahmen W., and R. DeVore. Adaptive wavelet methods for elliptic operator equations : Convergence rates. *Math. Comp.*, 70(233):27–75, 2000.

[21] A. Cohen, Dahmen W., and R. DeVore. Adaptive wavelet methods ii- beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[22] T. Gantumur, H. Harbrecht, and R. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Preprint*, 1325, 2005.

[23] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet techniques in numerical simulation. *Encyclopedia of Computational Mechanics.*, 2004.

[24] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, Cambridge, 1987.

[25] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–123, 2004.

[26] H. Nguyen. Finite element wavelets for solving partial differential equations. *Dissertation University of Utrecht*, 2005.

[27] G. Schmidlin. Fast solutions for integral equations in $\mathbb{R}^3$. *Dissertation ETH*, 15016, 2003.

[28] P.-A. Nitsche. Best N term approximation spaces for tensor product wavelet bases. *Constructive Approximation*, 24(1):49–70, 2006.

[29] C Schwab and R. Stevenson. Adaptive wavelet algorithms for elliptic pde's on product domains. *Research Report SAM ETH Zurich (in review)*, 2006-10, 2006.

[30] A. Cohen. *Numerical analysis of wavelet methods*. Elsevier Amsterdam, 2003.