SpringerPlus
a SpringerOpen Journal

RESEARCH Open Access

Sparse tensor phase space Galerkin approximation for radiative transport

Konstantin Grella

Abstract

We develop, analyze, and test a sparse tensor product phase space Galerkin discretization framework for the stationary monochromatic radiative transfer problem with scattering. The mathematical model describes the transport of radiation on a phase space of the Cartesian product of a typically three-dimensional physical domain and two-dimensional angular domain. Known solution methods such as the discrete ordinates method and a spherical harmonics method are derived from the presented Galerkin framework. We construct sparse versions of these well-established methods from the framework and prove that these sparse tensor discretizations break the "curse of dimensionality": essentially (up to logarithmic factors in the total number of degrees of freedom) the solution complexity increases only as in a problem posed in the physical domain alone, while asymptotic convergence orders in terms of the discretization parameters remain essentially equal to those of a full tensor phase space Galerkin discretization. Algorithmically we compute the sparse tensor approximations by the combination technique. In numerical experiments on $2+1$ and $3+2$ dimensional phase spaces we demonstrate that the advantages of sparse tensorization can be leveraged in applications.

2010 Mathematics subject classification: 35Q79; 65N12; 65N30; 65N35

Keywords: Radiative transfer; Sparse grids; Discrete ordinates method; Spherical harmonics method; Combination technique

Introduction

In this paper, we consider the numerical solution of the radiative transfer problem (RTP). This transport problem is stated on the phase space $\Omega = D \times \mathcal{S}$ as the Cartesian product of a bounded physical domain $D \subset \mathbb{R}^d$, where $d = 2, 3$, and the unit $d_{\mathcal{S}}$-sphere as the parameter domain $\mathcal{S}$ of dimension $d_{\mathcal{S}} = d - 1 = 1, 2$. The RTP (see e.g. Modest 2003) is then given as the task of finding the unknown *radiative intensity* $u : \Omega \to \mathbb{R}$, a real function over the phase space satisfying

$$\boldsymbol{s} \cdot \nabla_x u(\boldsymbol{x}, \boldsymbol{s}) + (\kappa(\boldsymbol{x}) + \sigma(\boldsymbol{x})) u(\boldsymbol{x}, \boldsymbol{s}) = \kappa(\boldsymbol{x}) I_b(\boldsymbol{x}) + \sigma(\boldsymbol{x}) \int_{\mathcal{S}} \Phi(\boldsymbol{s}, \boldsymbol{s}') u(\boldsymbol{x}, \boldsymbol{s}') \, \mathrm{d}\boldsymbol{s}', \quad (1a)$$

$$u(\boldsymbol{x}, \boldsymbol{s}) = g(\boldsymbol{x}, \boldsymbol{s}), \qquad \boldsymbol{x} \in \partial D, \ \boldsymbol{s} \cdot \boldsymbol{n}(\boldsymbol{x}) < 0 \, . \quad (1b)$$

We refer to Eq. (1a) as the *stationary monochromatic radiative transfer equation* (RTE), while Eq. (1b) constitutes *inflow boundary conditions*. A ray of light of direction $\boldsymbol{s}$ is attenuated by absorption and scattering with the medium. In (1a), $\kappa \geq 0$ is the *absorption coefficient*, $\sigma \geq 0$ the *scattering coefficient*, and $\Phi > 0$ the *scattering kernel* or *scattering phase function*. The scattering phase function is normalized to $\int_{\mathcal{S}} \Phi(\boldsymbol{s}, \boldsymbol{s}') \, \mathrm{d}\boldsymbol{s}' = 1$ for each direction $\boldsymbol{s}$. Sources inside the domain D are modeled by the *blackbody intensity* $I_b \geq 0$, radiation from sources outside of the domain or from its enclosings is prescribed by the *boundary data* $g \geq 0$. The vector $\boldsymbol{n}(\boldsymbol{x})$ denotes the *outward unit normal* vector which is defined in (almost every) point $\boldsymbol{x}$ on the boundary ∂D of the physical domain.

Due to the high dimension of the phase space, the nonlocality of the scattering operator, and the hyperbolic nature of the PDE, the efficient numerical simulation of radiative transfer is a challenging computational task even today. Still, radiative transfer as such or as a means of energy transfer among others is of interest in many applications, e.g. in the fields of heat transfer (Modest 2003), neutron transport (Hébert 2010), atmospheric sciences (Evans 1998), medical imaging (e.g. Peng et al. 2011), or

Correspondence: konstantin.grella@sam.math.ethz.ch
Seminar for Applied Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland

other areas where transported particles interact with a background medium, but only negligibly with each other.

In this paper, we extend the range of sparse tensor product discretization methods for the RTP investigated before (Grella and Schwab 2011a; Grella and Schwab 2011b; Grella 2013; Widmer et al. 2008) by a new phase space Galerkin framework.

Apart from Monte Carlo methods for raytracing, the most popular deterministic approaches for the radiative transfer problem are the discrete ordinates method and the spherical harmonics approximation. We quote a brief overview of their properties from (Grella 2013).

In the discrete ordinate method (DOM) or S_N-approximation, the angular domain is discretized by a number of fixed directions, which are inserted into Eq. (1) so that a system of spatial PDEs results. Without scattering the equations for single directions are independent of each other, with scattering, however, they are coupled through the scattering integral. After the straightforward discretization of the angular domain, the spatial PDEs are typically solved using finite differences, finite elements, or finite volume methods.

The DOM is popular as it is simple to implement, offers straightforward parallelization, and can capture directed radiation relatively well as some of the ordinates can usually be chosen freely.

On the downside, the method can suffer from so-called "ray effects" (Lathrop 1968): due to the point evaluation in the angular domain, the scalar flux or incident radiation from small isotropic sources may appear star-like with rays emanating from the source into the chosen angular directions (Stone 2007, p. 2 and following). These effects occur especially pronounced in settings with low scattering and absorption, i.e. in optically thin media.

An example for truncated series expansion is the method of spherical harmonics or P_N-approximation. The solution of Eq. (1) is replaced by a series of spherical harmonics up to some order N with spatially dependent coefficients. Due to orthogonality relations, the scattering part often decouples or couples only few terms depending on the scattering kernel. However, the system of PDEs for the spatial coefficient functions is always coupled by the transport part $\boldsymbol{s} \cdot \nabla_x u$.

As low order series expansions in spherical harmonics do not permit a very localized resolution of the angular variable, the method performs best when the solutions are nearly isotropic in angle, which is the case in diffusive, so-called "optically thick" media. Then, rather low order spherical harmonics approximations might suffice for a good approximation. Indeed, the P_1 method can be formulated as a diffusion equation for the incident radiation (Modest 2003, Sec. 15.4). For smooth solutions, the spherical harmonics method exhibits spectral convergence in angle (Grella and Schwab 2011a).

On the other hand, beam-like solutions require a high spectral order to be resolved appropriately, leading to high computational complexity. In general, higher spectral orders also lead to a sharp increase in computational complexity when boundary conditions are to be satisfied (Modest and Yang 2008).

When combined with a standard finite element or finite volume discretization in the physical domain D, the deterministic, numerical S_N- and P_N-approximations exhibit the so-called "curse of dimensionality": the error (typically the L^2-error of the solution) with respect to the numbers of degrees of freedom (DoF) M_D and $M_{\mathcal{S}}$ on the component domains D and $\mathcal{S}$ scales with the dimension d and $d_{\mathcal{S}}$ of the application problem as $O\left(M_D^{-s/d} + M_{\mathcal{S}}^{-t/d_{\mathcal{S}}}\right)$ with constants s and t.

The first sparse finite element approximation method was proposed in (Zenger 1991) for the solution of Laplace equation in the unit square and cube. In this paper, Zenger developed the (direct) sparse grid approximation method which alleviates this curse of dimensionality: the computational complexity is reduced, up to logarithmic terms, to that of a one-dimensional problem.

The idea of sparse tensorization of finite element and finite difference methods was generalized by (Bungartz and Griebel 2004; Hegland 2003; Garcke 2007), and others, for the numerical solution of PDEs as well as for other applications where standard numerical methods are obstructed by the curse of dimensionality.

Sparse tensor methods were first applied to radiative transfer by (Widmer et al. 2008). In that paper, the authors formulated a least squares phase space Galerkin sparse tensor approximation with hierarchical finite elements as discretization of the physical domain and wavelets in the angular domain. They proved that sufficient regularity of the solution provided, their method breaks the curse of dimensionality: the problem complexity reduces to log-linear in the number of degrees of freedom, while convergence rates deteriorate only by a logarithmic factor. However, the discretization of the scattering operator had not been addressed in that work.

In earlier work (Grella and Schwab 2011a), we showed that the sparse tensor product method of (Widmer et al. 2008) can also be combined with a spectral discretization involving spherical harmonics, resulting in a sparse P_N-method which also treats scattering. Boundary conditions were satisfied in a strong sense by introducing piecewise spectral functions in angle.

Secondly we presented a sparse tensor version of the DOM as a sparse collocation method with a Galerkin ansatz in the physical domain and strong enforcement of the boundary conditions, while not yet accounting for scattering (Grella and Schwab 2011b). This sparse tensor S_N-method was realized computationally with the sparse grid combination technique (Griebel et al. 1992) to

construct a sparse approximation to the radiative transfer solution.

The sparse DOM was subsequently reformulated as a phase space Galerkin method with quadrature in angle (Grella 2013) in order to treat sparse P_N- and sparse S_N-method in a more uniform manner. In this reformulation, we employed streamline upwind Petrov Galerkin (SUPG) stabilization and weak satisfacion of boundary conditions. Sparse S_N-methods were derived as a direct sparse tensor method and implemented algorithmically via the combination technique.

In the present paper, we derive a sparse P_N- and sparse S_N-method from the same phase space Galerkin framework with transport stabilization and scattering. Boundary conditions are satisfied in a weak sense. In doing so we close a gap in the list of conceivable combinations of formulations regarding stabilization and type of angular approximation. In contrast to the previous approach (Grella 2013), we stabilize the formulation in a different way and the analytical focus will be on the direct sparse approach. With transport stabilization and direct sparse approach we follow (Widmer et al. 2008) more closely, extending their work by treatment of scattering and weak satisfaction of the boundary conditions.

Similar savings in computational effort are realized with other variational formulations, such as Petrov-Galerkin saddle point formulations (see e.g. Dahmen et al. (2012) and the references there).

The outline of this paper is as follows. In Section 'Phase space Galerkin method' we formulate the phase space Galerkin framework in operator form and outline how P_N and S_N-methods can be derived from it. Then we develop full tensor and sparse tensor discretizations based on the framework and analyze and compare their convergence properties.

Section 'Numerical experiments' presents several basic numerical experiments designed with the purpose of validating and illustrating the theoretical convergence results.

Finally we conclude this work in Section 'Conclusion' by summarizing and reviewing the results.

Phase space Galerkin method

We begin by introducing the radiative transfer problem in operator form. Using this compact notation we then state the variational formulation of our phase space Galerkin method and proceed to discretizations of the method.

Operator formulation

Problem (1) reads in operator form: Find the intensity $u(\boldsymbol{x},\boldsymbol{s}) : D \times \mathcal{S} \to \mathbb{R}$ such that

$$\mathrm{A}u = f, \quad u|_{\partial\Omega_-} = g. \tag{2}$$

In this, $\partial\Omega_-$ represents the inflow part of the boundary $\partial\Omega = \partial D \times \mathcal{S}$ of the computational domain or *phase space* $\Omega = D \times \mathcal{S}$. The *inflow boundary* is defined by

$$\partial\Omega_- := \{(\boldsymbol{x},\boldsymbol{s}) \in \Omega : \boldsymbol{x} \in \Gamma_-(\boldsymbol{s})\} \tag{3}$$

with the *physical part of the inflow boundary*

$$\Gamma_-(\boldsymbol{s}) := \{\boldsymbol{x} \in \partial D : \boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x}) < 0\}. \tag{4}$$

Correspondingly we define the *physical part of the outflow boundary* as

$$\Gamma_+(\boldsymbol{s}) := \{\boldsymbol{x} \in \partial D : \boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x}) > 0\}. \tag{5}$$

The *radiative transfer operator* $\mathrm{A} = \mathrm{T} + \mathrm{Q}$ consists of the *transport operator* T,

$$\mathrm{T}u := (\boldsymbol{s}\cdot\nabla_x + \kappa)u, \tag{6}$$

and the *scattering operator* Q,

$$\begin{aligned}\mathrm{Q}u &:= \sigma\mathrm{Q}_1 u := \sigma(\mathrm{Id}-\Sigma)u := \sigma(\boldsymbol{x})u(\boldsymbol{x},\boldsymbol{s}) \\ &\quad - \sigma(\boldsymbol{x})\int_{\mathcal{S}} \Phi(\boldsymbol{s},\boldsymbol{s}')u(\boldsymbol{x},\boldsymbol{s}')\,\mathrm{d}\boldsymbol{s}'\,.\end{aligned} \tag{7}$$

Here, $\mathrm{Q}_1 = \mathrm{Id} - \Sigma$ is the *unity scattering operator*, and Σ is the *scattering integral operator*, the integral of Φ and u. The *source function* f contains the sources of radiation in the domain,

$$f := \kappa I_b, \tag{8}$$

and g is the *incoming radiation* on the boundary $\partial\Omega_-$, as in Sec. 'Introduction'.

Properties of the scattering operator

Aside from the positivity and normalization requirements already mentioned in Sec. 'Introduction', we assume an isotropic medium, i.e. Φ does not depend on $\boldsymbol{x}$. As Φ models the type of scattering, this assumption can safely be made for most applications (cf. Modest 2003, p. 268). Variations in the strength of scattering due to e.g. varying spatial density of the medium are modeled by the scattering coefficient σ. As long as the following properties hold for almost every $\boldsymbol{x}$, the complexity and convergence analysis later on could also be conducted without this assumption.

Furthermore, if spherical scatterers are assumed, the scattering phase function does not vary with the azimuthal angle so that Φ only depends on the inner product of $\boldsymbol{s}$ and $\boldsymbol{s}'$. From this it follows immediately that $\Phi(\boldsymbol{s},\boldsymbol{s}') = \Phi(\boldsymbol{s}\cdot\boldsymbol{s}') = \Phi(\boldsymbol{s}',\boldsymbol{s})$.

From here on, we shall take Φ to be *forward dominant* (cf. Kanschat 2008, Def. 1) if $\Phi(\boldsymbol{s},\boldsymbol{s}') = \sum_{k=0}^{\infty} a_k \cos(k \arccos(\boldsymbol{s}\cdot\boldsymbol{s}'))$ with all $a_k \geq 0$. Then, one can show that Σ is positive semi-definite (Kanschat 2008, Lemmata 2 and 3), i.e.

$$(v, \Sigma v)_{L^2(\mathcal{S})} \geq 0 \quad \forall v \in L^2(\mathcal{S}). \tag{9}$$

Normalization and symmetry of Φ with respect to its arguments leads to normalization of the operator norm $||\Sigma||_{L^2(\mathcal{S})\to L^2(\mathcal{S})} = 1$ (Kanschat 2008, Lemma 5).

From these properties and a Hilbert-Schmidt theorem for integral operators (e. g. Knapp 2005, Thm. 2.4), one can derive that the spectrum of Q_1 lies in $[0, 1]$ with an isolated eigenvalue $\lambda_0 = 0$, from which the next largest eigenvalue λ_1 differs by a positive constant (Ávila et al. 2011, Sec. 2.2).

With the previous considerations, one arrives at the following properties of Q:

Lemma 1. *For any $u \in L^2(\Omega)$, the scattering operator* Q *as defined by Eq. (7) satisfies (cf. Ávila et al. 2011, Eq. (11))*

$$\lambda_1 \left\| \sigma \mathrm{P}^{\perp} u \right\|_{L^2(\Omega)} \leq \|\mathrm{Q}u\|_{L^2(\Omega)} \leq \|\sigma\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)}, \tag{10}$$

$$(u, \mathrm{Q}u)_{L^2(\Omega)} \geq \|\mathrm{Q}u\|^2_{L^2(\Omega)} \geq 0, \tag{11}$$

in which the projector $\mathrm{P}^{\perp}$ maps $u(\boldsymbol{x}, \cdot) \in L^2(\mathcal{S})$ to $(\ker \mathrm{Q})^{\perp}$, the space orthogonal to the kernel of Q*, and $\lambda_1 \in (0, 1]$ is the smallest nonzero eigenvalue of Q_1.*

For a proof of (11) we refer to (Grella 2013).

Variational formulation

Our variational formulation will be based on a Galerkin finite element framework over the phase space Ω with stabilization applied to the operator RTP (2).

A generic stabilized phase space variational formulation

To begin with, we define the Hilbert space

$$\mathcal{V} := \left\{ u \in L^2(\Omega) : \boldsymbol{s} \cdot \nabla_x u \in L^2(\Omega) \right\} \tag{12}$$

with the usual $L^2(\Omega)$ inner product

$$(u, v)_{L^2(\Omega)} := \int_{\mathcal{S}} \int_D u(\boldsymbol{x}, \boldsymbol{s}) v(\boldsymbol{x}, \boldsymbol{s}) \mathrm{d}x \, \mathrm{d}\boldsymbol{s}, \tag{13}$$

and the *triple bar norm*

$$|||v|||^2 := \|v\|^2_{L^2(\Omega)} + \|\boldsymbol{s} \cdot \nabla_x v\|^2_{L^2(\Omega)} + \|\mathrm{Q}_1 v\|^2_{L^2(\Omega)}, \quad v \in \mathcal{V}. \tag{14}$$

For the weak enforcement of boundary conditions, we define the boundary form

$$b(u, v) := (v, |\boldsymbol{s} \cdot \boldsymbol{n}| u)_{L^2(\partial\Omega_-)} = \int_{\mathcal{S}} \int_{\Gamma_-(\boldsymbol{s})} |\boldsymbol{s} \cdot \boldsymbol{n}| u v \mathrm{d}x \, \mathrm{d}\boldsymbol{s}, \tag{15}$$

in which we have omitted the dependence of the outward unit normal $\boldsymbol{n}$ on the position $\boldsymbol{x}$. This boundary form was introduced by Manteuffel et al. (2000, Eq. (2.16)). It is well-defined for functions $v \in L^2(\Omega)$ with finite *inflow norm*

$$\|v\|_- := b(v, v)^{1/2}. \tag{16}$$

Combining (14) and (16) yields the new norm

$$\|v\|_1 := \left(|||v|||^2 + \|v\|^2_- \right)^{1/2}, \tag{17}$$

which gives rise to the closed, linear subspace of $\mathcal{V}$,

$$\mathcal{V}_1 := \{ v \in \mathcal{V} : \|v\|_1 < \infty \} \tag{18}$$

which, with the inner product related to $\|v\|_1$, is a Hilbert space. For functions $u, v \in \mathcal{V}_1$, we define the bilinear form

$$a(u, v) := (\mathrm{R}v, \mathrm{A}u)_{L^2(\Omega)} + 2b(u, v), \tag{19}$$

where R is a stabilization operator on the test function side yet to be specified. Together with the linear form

$$l(v) := \left(\mathrm{R}v, f\right)_{L^2(\Omega)} + 2b(g, v), \tag{20}$$

the bilinear form constitutes the following variational problem: Find $u \in \mathcal{V}_1$ such that

$$a(u, v) = l(v) \quad \forall v \in \mathcal{V}_1. \tag{21}$$

Different ways of stabilization are conceivable and have been used in the literature, e. g. the least squares approach by (Manteuffel et al. 2000), or SUPG introduced by (Brooks and Hughes 1982). For our purposes here, we will choose the T-stabilized formulation (Grella and Schwab 2011a) to avoid mesh-dependent quantities and the square of the scattering operator. More precisely, we set $\mathrm{R} = \varepsilon \mathrm{T}$ with a stabilization parameter ε that depends on the absorption coefficient κ.

Properties of the variational formulation

At this point, we introduce the *anisotropic* or *mixed Sobolev spaces* $H^{s,t}(\Omega) = H^s(D) \otimes H^t(\mathcal{S})$ as

$$H^{s,t}(\Omega) := \Big\{ v \in L^2(\Omega) : \mathrm{D}^{\boldsymbol{\beta}}_{\boldsymbol{s}} \mathrm{D}^{\boldsymbol{\alpha}}_{\boldsymbol{x}} v \in L^2(\Omega), \; 0 \leq |\boldsymbol{\alpha}| \leq s, \; 0 \leq |\boldsymbol{\beta}| \leq t \Big\} \tag{22}$$

with the corresponding *mixed Sobolev norms* $\|\cdot\|_{H^{s,t}(\Omega)}$, given by

$$\|v\|^2_{H^{s,t}(\Omega)} := \sum_{0 \leq |\boldsymbol{\alpha}| \leq s} \sum_{0 \leq |\boldsymbol{\beta}| \leq t} \left\| \mathrm{D}^{\boldsymbol{\beta}}_{\boldsymbol{s}} \mathrm{D}^{\boldsymbol{\alpha}}_{\boldsymbol{x}} v \right\|^2_{L^2(\Omega)}. \tag{23}$$

Here, $\mathrm{D}^{\boldsymbol{\beta}}_{\boldsymbol{s}} \mathrm{D}^{\boldsymbol{\alpha}}_{\boldsymbol{x}} v$ denotes the weak derivative of $v : D \times \mathcal{S} \to \mathbb{R}$ of order $|\boldsymbol{\alpha}|$ w. r. t. $\boldsymbol{x} \in D$ and order $|\boldsymbol{\beta}|$ w. r. t. $\boldsymbol{s} \in \mathcal{S}$, with the multi-indices $\boldsymbol{\alpha} \in \mathbb{N}^d_0$ and $\boldsymbol{\beta} \in \mathbb{N}^{d_{\mathcal{S}}+1}_0$.

The following lemma collects auxiliary results which will become helpful later.

Lemma 2 (Auxiliary results)**.**

1. *Let $v \in \mathcal{V}$. Then $(v, \boldsymbol{s} \cdot \nabla_x v)_{L^2(\Omega)} \geq \frac{1}{2} \int_{\mathcal{S}} \int_{\Gamma_{-(\mathcal{S})}} v^2 \boldsymbol{s} \cdot \boldsymbol{n}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{s}$. If furthermore $v \in \mathcal{V}_0$, then $(v, \boldsymbol{s} \cdot \nabla_x v)_{L^2(\Omega)} \geq 0$.*
2. *For $v \in H^{1,0}(\Omega)$, it holds $\|\boldsymbol{s} \cdot \nabla_x v\| \leq \sqrt{d} \|v\|_{H^{1,0}(\Omega)}$.*

Proof. 1. A proof is given by (Manteuffel et al. 2000, Thm. 2.1). It uses the divergence theorem and exploits the fact that $v|_{\partial\Omega_-} = 0$ for $\boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x}) < 0$ if $v \in \mathcal{V}_0$, where $\boldsymbol{n}(\boldsymbol{x})$ is the outward unit normal on the boundary ∂D:

$$\begin{aligned}(v, \boldsymbol{s}\cdot\nabla_x v)_{L^2(\Omega)} &= \frac{1}{2}\int_{\mathcal{S}}\int_D \nabla_x\cdot(\boldsymbol{s}v^2)\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{s}\\ &= \frac{1}{2}\int_{\mathcal{S}}\int_{\partial D} v^2\boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{s}\\ &= \frac{1}{2}\int_{\mathcal{S}}\int_{\Gamma_-(\boldsymbol{s})} v^2\boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{s}\\ &\quad+\frac{1}{2}\int_{\mathcal{S}}\int_{\Gamma+(\boldsymbol{s})} v^2\boldsymbol{s}\cdot\boldsymbol{n}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{s}.\end{aligned}$$

As $\boldsymbol{s}\cdot\boldsymbol{n} \geq 0$ in the second integral, we obtain the first assertion. If additionally $v \in \mathcal{V}_0$, then the first integral vanishes, and the second assertion follows.

2. We again quote Manteuffel et al. (2000, Lemma 4.1 (i)):

$$\begin{aligned}\|\boldsymbol{s}\cdot\nabla_x v\|^2_{L^2(\Omega)} &\leq \int_D\int_{\mathcal{S}}\left(\sum_{i=1}^d s_i D_{x_i} v\right)^2 \mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{x}\\ &\leq d\int_D\int_{\mathcal{S}}\sum_{i=1}^d \left(s_i D_{x_i} v\right)^2 \mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{x}\\ &\leq d\sum_{i=1}^d \left\|D_{x_i} v\right\|^2 \leq d\|v\|^2_{H^{1,0}(\Omega)}.\end{aligned}$$

□

In order to establish well-posedness of the variational formulation (21), we prove continuity and coercivity of the bilinear form (19) and continuity of the linear form (20) in the following.

Lemma 3 (Continuity of bilinear form)**.** *Let σ, κ, $\varepsilon \in L^\infty(D)$ with $\|\sigma\|_{L^\infty(D)} =: \sigma_{\max}$, $\|\kappa\|_{L^\infty(D)} =: \kappa_{\max}$, $\|\varepsilon\|_{L^\infty(D)} =: \varepsilon_{\max}$, then there is a constant $0 < c_c < \infty$ such that for all $u, v \in \mathcal{V}_1$*

$$|a(u,v)| \leq c_c\,\|u\|_1\,\|v\|_1.$$

Proof. We proceed analogously to (Manteuffel et al. 2000, Thm. 3.3). To begin with, we estimate for all $u, v \in \mathcal{V}$

$$\begin{aligned}\|\mathrm{R}v\| &= \|\varepsilon\kappa v + \varepsilon\boldsymbol{s}\cdot\nabla_x v\| \leq \varepsilon_{\max}\kappa_{\max}\|v\| + \varepsilon_{\max}\,\|\boldsymbol{s}\cdot\nabla_x v\|\\ &\leq \max\{\varepsilon_{\max}\kappa_{\max}, 1\}\,\|v\|_1\,, \qquad (24)\\ \|\mathrm{A}u\| &\leq \kappa_{\max}\,\|u\| + \|\boldsymbol{s}\cdot\nabla_x u\| + \sigma_{\max}\,\|\mathrm{Q}_1 u\|\\ &\leq \max\{\kappa_{\max}, 1, \sigma_{\max}\}|||u|||. \qquad (25)\end{aligned}$$

Using the Cauchy-Schwarz inequality as well as estimates (24) and (25) it holds

$$\begin{aligned}|a(u,v)| &\leq |\,\|\mathrm{R}v\|\,\|\mathrm{A}u\| + 2\|v\|_-\,\|u\|_-\,|\\ &\leq 2\left(\|\mathrm{R}v\|^2 + \|v\|^2_-\right)^{1/2}\left(\|\mathrm{A}u\|^2 + \|u\|^2_-\right)^{1/2}\\ &\leq 2\max\{1, \varepsilon_{\max}\kappa_{\max}\}\max\{\kappa_{\max}, 1, \sigma_{\max}\}\,\|u\|_1\,\|v\|_1\,.\end{aligned}$$

□

Lemma 4 (Continuity of linear form)**.** *Given the assumptions of Lemma 3 on κ, σ, ε, and additionally $f \in L^2(\Omega)$, $g : \partial\Omega_- \to \mathbb{R}$ with $\|g\|_- < \infty$, there is a constant $0 < c_l < \infty$ such that for $v \in \mathcal{V}_1$ it holds*

$$|l(v)| \leq c_l\|v\|_1\,.$$

Proof. The proof is analogous to that of Lemma 3:

$$\begin{aligned}|l(v)| &\leq |\,\|\mathrm{R}v\|\,\|f\| + 2\|v\|_-\,\|g\|_-\,|\\ &\leq 2\left(\|\mathrm{R}v\|^2 + \|v\|^2_-\right)^{1/2}\left(\|f\|^2 + \|g\|^2_-\right)^{1/2}\\ &\leq 2\max\{1, \varepsilon_{\max}\kappa_{\max}\}(\|f\| + \|g\|_-)\,\|v\|_1\,.\end{aligned}$$

□

Next, we show coercivity of the bilinear form. For ease of exposition we shall assume ε and κ to be constant on the physical domain. Coercivity can also be obtained for non-constant coefficients (see Widmer 2009, Thm. 2.2, for an example). Coercivity of the SUPG variational formulation for the RTP has also been proved by (Ávila et al. 2011, Lemma 2), although in a different norm. Previously, we had proved coercivity of the T-stabilized variational formulation without the boundary form $b(\cdot,\cdot)$ (Grella 2013, Lemma 4.1), here we include this boundary form in the formulation, which will motivate the choice of the stabilization parameter ε.

Lemma 5 (Coercivity of bilinear form)**.** *Let κ, ε be positive functions which are constant on the physical domain D. Assume $\min_{\boldsymbol{x}\in D}\sigma =: \sigma_{\min} > 0$ and $\sigma_{\max} = \|\sigma\|_{L^\infty(D)}$, and additionally that*

$$\sigma^2_{\max} < 4\kappa\sigma^2_{\min}, \quad \varepsilon < \frac{2}{\kappa}. \qquad (26)$$

Then the bilinear form $a(\cdot,\cdot)$ from (19) is coercive on $\mathcal{V}_1\times\mathcal{V}_1$: there is a constant $c_e > 0$ such that for all $v \in \mathcal{V}_1$ it holds

$$a(v,v) \geq c_e\,\|v\|^2_1\,.$$

Proof. For an overview of the involved terms we split the bilinear form into separate inner products:

$$\begin{aligned} a(\nu,\nu) &= (\varepsilon\kappa\nu + \varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \mathrm{A}\nu) + 2b(\nu,\nu) \\ &= (\varepsilon\kappa\nu + \varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \boldsymbol{s}\cdot\nabla_x\nu + \kappa\nu + \mathrm{Q}\nu) + 2b(\nu,\nu) \\ &= (\varepsilon\kappa\nu, \boldsymbol{s}\cdot\nabla_x\nu) + (\varepsilon\kappa\nu, \kappa\nu) + (\varepsilon\kappa\nu, \mathrm{Q}\nu) \\ &\quad + (\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \boldsymbol{s}\cdot\nabla_x\nu) + (\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \kappa\nu) \\ &\quad + (\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \mathrm{Q}\nu) + 2b(\nu,\nu) \end{aligned} \tag{27}$$

As we assumed ε and κ to be constant, we can factor these coefficients out of the inner products. We begin by analyzing the sum of first and fifth inner product.

Applying statement 1 of Lemma 2 yields

$$2\varepsilon\kappa\,(\nu, \boldsymbol{s}\cdot\nabla_x\nu) \geq -\varepsilon\kappa\|\nu\|_-^2.$$

Together with the boundary term, we obtain

$$(\varepsilon\kappa\nu, \boldsymbol{s}\cdot\nabla_x\nu) + (\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \kappa\nu) + 2b(\nu,\nu) \geq (2-\varepsilon\kappa)\|\nu\|_-^2.$$

The second inner product is bounded from below by

$$(\varepsilon\kappa\nu, \kappa\nu) \geq \varepsilon\kappa^2\|\nu\|^2.$$

To estimate the third inner product, we use property (11) of the scattering operator:

$$\varepsilon\kappa\,(\nu, \mathrm{Q}\nu) \geq \varepsilon\kappa\,\|\mathrm{Q}\nu\|^2 \geq \varepsilon\kappa\sigma_{\min}^2\,\|\mathrm{Q}_1\nu\|^2\,.$$

The fourth inner product in Eq. (27) is

$$(\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \boldsymbol{s}\cdot\nabla_x\nu) = \varepsilon\,\|\boldsymbol{s}\cdot\nabla_x\nu\|^2\,.$$

For the sixth inner product we apply Cauchy-Schwarz inequality and Young's inequality with a parameter $\theta > 0$:

$$\begin{aligned} (\varepsilon\boldsymbol{s}\cdot\nabla_x\nu, \mathrm{Q}\nu) &\geq -\varepsilon\sigma_{\max}\,\|\boldsymbol{s}\cdot\nabla_x\nu\|\,\|\mathrm{Q}_1\nu\| \\ &\geq -\varepsilon\sigma_{\max}\left(\frac{\theta}{2}\,\|\boldsymbol{s}\cdot\nabla_x\nu\|^2 + \frac{1}{2\theta}\,\|\mathrm{Q}_1\nu\|^2\right) \end{aligned}$$

Combining all estimates yields the result:

$$\begin{aligned} a(\nu,\nu) &\geq \varepsilon\kappa^2\,\|\nu\|^2 + \varepsilon\left(1 - \frac{\theta}{2}\sigma_{\max}\right)\|\boldsymbol{s}\cdot\nabla_x\nu\|^2 \\ &\quad + \varepsilon\left(\kappa\sigma_{\min}^2 - \frac{1}{2\theta}\sigma_{\max}\right)\|\mathrm{Q}_1\nu\|^2 \\ &\quad + (2-\varepsilon\kappa)\,\|\nu\|_-^2 \\ &\geq \min\left\{\varepsilon\kappa^2, \varepsilon\left(1 - \frac{\theta}{2}\sigma_{\max}\right),\right. \\ &\qquad \left.\varepsilon\left(\kappa\sigma_{\min}^2 - \frac{1}{2\theta}\sigma_{\max}\right), 2-\varepsilon\kappa\right\}\|\nu\|_1^2\,. \end{aligned}$$

By eliminating θ we obtain the condition $\sigma_{\max}^2 < 4\kappa\sigma_{\min}^2$. The condition $\varepsilon < 2/\kappa$ results from the last of the terms over which the minimum is taken. □

The previous condition on the stabilization parameter leads to a choice of $\varepsilon = 1/\kappa$. Well-posedness of the variational formulation now follows directly.

Theorem 6 (Existence and uniqueness of solution to variational formulation)**.** *Provided that $f \in L^2(\Omega)$ and $\|g\|_- < \infty$ there exists a unique solution $u \in \mathcal{V}_1$ to the variational formulation (21).*

Proof. Since $(\mathcal{V}_1, \|\cdot\|_1)$ is a Hilbert space and Lemmata 3–5 guarantee continuity of the augmented SUPG bilinear form and linear form as well as coercivity of the bilinear form, the Lax-Milgram theorem (Brenner and Scott 2008, Thm. 2.7.7) ensures existence and uniqueness of the solution to (21). □

Discretization

For the discretization of the variational problem (21), we restrict the space $\mathcal{V}_1$ in the variational formulation (21) to tensor products of hierarchic, finite dimensional approximation spaces over the component domains D and $\mathcal{S}$.

Full tensor discretization

In the standard full tensor approximation, we choose a full tensor product space $V^{L,N}$ to approximate $\mathcal{V}_1$:

$$\mathcal{V}_1 \approx V^{L,N} := V_D^L \otimes V_{\mathcal{S}}^N. \tag{28}$$

As $H^{1,0}(\Omega) \cong H^1(D) \otimes L^2(\mathcal{S})$ is a dense subspace of $\mathcal{V}_1$, we define the family of physical approximation spaces as

$$V_D^{l_D} := S^{0,1}\left(D, \mathcal{T}_D^{l_D}\right) \subset H^1(D), \quad l_D = 1, \ldots, L, \tag{29}$$

the spaces of continuous, piecewise linear functions on a dyadically refined mesh $\mathcal{T}_D^{l_D}$ over D. Here, the parameter l_D stands for the physical resolution. It is related to the mesh width h in $\mathcal{T}_D^{l_D}$ by $h = O\left(2^{-l_D}\right)$. With respect to the resolution $l_D = 0, \ldots, L$, the spaces $V_D^{l_D}$ form a nested sequence

$$V_D^0 \subset V_D^1 \subset \ldots \subset V_D^L \subset H^1(D).$$

Let $M_D := \dim V_D^L$ denote the number of degrees of freedom for the FE space V_D^L in the physical domain D. Then

$$M_D = O\left(2^{dL}\right) \tag{30}$$

with the dimension d of the physical domain. The exact number will depend on the geometry of the domain. For a square or cube $D = [0,1]^d$, respectively, we obtain

$$M_D = \left(2^L + 1\right)^d. \tag{31}$$

In the angular domain, we distinguish between the P_N-method and the S_N-method.

S_N-method Here, the family of approximation spaces is given by

$$V_{\mathcal{S}}^{l_{\mathcal{S}}} := S^{-1,0}\left(\mathcal{S}, \mathcal{T}_{\mathcal{S}}^{l_{\mathcal{S}}}\right) \subset L^2(\mathcal{S}), \quad l_{\mathcal{S}} = 1, \ldots, N, \tag{32}$$

the spaces of piecewise constant functions on a dyadically refined mesh $\mathcal{T}_{\mathcal{S}}^{l_{\mathcal{S}}}$. As the physical spaces, these spaces exhibit a nested structure. The angular resolution N and the dimension of $V_{\mathcal{S}}^N$ are related by

$$M_{\mathcal{S}} := \dim V_{\mathcal{S}}^N = O\left(2^{d_{\mathcal{S}} N}\right). \tag{33}$$

P_N-method To define the angular approximation spaces of the P_N-method, we first introduce the spaces of spectral functions of the $d_{\mathcal{S}}$-sphere,

$$\mathbb{P}_{\tilde{N}}^{d_{\mathcal{S}}} = \operatorname{span}\left\{ Y_{n,m}^{(d_{\mathcal{S}})} : n = 0, \ldots, \tilde{N}; \; m = 1, \ldots, m_{n,d_{\mathcal{S}}} \right\} \subset L^2(\mathcal{S}), \tag{34}$$

where $Y_{n,m}^{(d_{\mathcal{S}})}$ are the spherical harmonics of the $d_{\mathcal{S}}$-sphere, and $m_{n,d_{\mathcal{S}}}$ is the largest value of the secondary index m depending on the value of the primary index n and the dimension. These spaces offer an inherent nested structure. To obtain the same relation (33) between resolution level and degrees of freedom as in the S_N-method, we connect the resolution level N and $\tilde{N}$ by $\tilde{N} = 2^N - 1$.

Then, the angular approximation spaces are

$$V_{\mathcal{S}}^{l_{\mathcal{S}}} := \mathbb{P}_{2^{l_{\mathcal{S}}}-1}^{d_{\mathcal{S}}}, \quad l_{\mathcal{S}} = 1, \ldots, N, \tag{35}$$

and relation (33) also holds here. Up to the index relabeling and the additional boundary form, we obtain the spherical harmonics method already analyzed by (Grella and Schwab 2011a).

In both methods, the full tensor approximation space consequently has the dimension

$$M_{L,N} := \dim V^{L,N} = M_D \cdot M_{\mathcal{S}} = O\left(2^{dL + d_{\mathcal{S}} N}\right). \tag{36}$$

The full tensor approximate solution can be expressed by means of a physical basis $\{\alpha_i(\boldsymbol{x})\}_{i=1}^{M_D}$ of V_D^L and an angular basis $\{\beta_j\}_{j=1}^{M_{\mathcal{S}}}$ of $V_{\mathcal{S}}^N$ as

$$u_{L,N}(\boldsymbol{x}, \boldsymbol{s}) := \sum_{i=1}^{M_D} \sum_{j=1}^{M_{\mathcal{S}}} u_{ij} \alpha_i(\boldsymbol{x}) \beta_j(\boldsymbol{s}) \tag{37}$$

with solution coefficients $u_{ij} \in \mathbb{R}$. The *discrete variational formulation* finally reads: Find $u_{L,N} \in V^{L,N}$ such that

$$a(u_{L,N}, v_{L,N}) = l(v_{L,N}) \quad \forall v_{L,N} \in V^{L,N}, \tag{38}$$

with the bilinear form $a(\cdot,\cdot)$ from (19) and the linear form $l(\cdot)$ from (20). As $V^{L,N}$ is a subspace of $\mathcal{V}_1$ well-posedness ensured by Thm. 6 for the continuous problem follows also for this discrete problem.

By choosing a subset of $H^1(D) \otimes L^2(\mathcal{S})$ as trial space we effectively assume a slightly higher regularity on the solution than what is guaranteed by the definition (12) of $\mathcal{V}_1$. For instance, solutions with line discontinuities due to the transport of discontinuous boundary data into the domain are not included in $V^{L,N}$. However, since $V^{L,N}$ is dense in $\mathcal{V}_1$, even discontinuous solutions will be approximated with increasing resolution. Furthermore, in order to leverage the advantages of a sparse tensor approximation, a higher regularity of the solution will be required in any case.

Equivalence of collocation DOM and phase space Galerkin DOM with quadrature

Ordinarily the discrete ordinates method is presented as a collocation method in angle: Fixed directions $\boldsymbol{s}_j \in \mathcal{S}$, $j = 1, \ldots, M_{\mathcal{S}}$, are inserted into the RTE (1a), and for each direction, the intensity $u_j(\boldsymbol{x}) := u(\boldsymbol{x}, \boldsymbol{s}_j) \in V_D^L$ is sought as the solution to a purely spatial PDE. In these PDEs, the scattering integral is replaced by a quadrature rule

$$\int_{\mathcal{S}} \Phi(\boldsymbol{s}_j, \boldsymbol{s}') u(\boldsymbol{x}, \boldsymbol{s}') \, \mathrm{d}\boldsymbol{s}' \approx \sum_{m=1}^{M_{\mathcal{S}}} w_m \Phi(\boldsymbol{s}_j, \boldsymbol{s}_m) u_m \tag{39}$$

with weights $w_m > 0$. By applying a Galerkin ansatz with stabilization in the physical domain to the PDEs, a system of coupled variational formulations

$$\begin{aligned} &\left(\mathrm{R}_j v, \; \mathrm{T}_j u_j + \sigma u_j - \sum_{m=1}^{M_{\mathcal{S}}} w_m \Phi(\boldsymbol{s}_j, \boldsymbol{s}_m) u_m \right)_{L^2(D)} \\ &\qquad + 2\left(v, |\boldsymbol{s}_j \cdot \boldsymbol{n}| u_j\right)_{L^2(\Gamma_-(\boldsymbol{s}_j))} \\ &\quad = \left(\mathrm{R}_j v f\right)_{L^2(D)} + 2\left(v, |\boldsymbol{s}_j \cdot \boldsymbol{n}| g_j\right)_{L^2(\Gamma_-(\boldsymbol{s}_j))} \quad \forall v \in V_D^L \end{aligned} \tag{40}$$

results with directional stabilization and transport operators

$$\mathrm{R}_j := \mathrm{R}|_{\boldsymbol{s}=\boldsymbol{s}_j}, \quad \mathrm{T}_j := \mathrm{T}|_{\boldsymbol{s}=\boldsymbol{s}_j}, \quad j = 1, \ldots, M_{\mathcal{S}}. \tag{41}$$

In the phase space Galerkin approach, variational formulation (38) is discretized further by substituting the angular quadrature rule (39) for all angular integrals so that the bilinear form (19) is approximated by

$$\begin{aligned} a(u, v) \approx \tilde{a}(u, v) = \sum_{j=1}^{M_{\mathcal{S}}} w_j \Bigg(&\mathrm{R}_j v_j, \; \mathrm{T}_j u_j + \sigma u_j \\ &- \sum_{m=1}^{M_{\mathcal{S}}} w_m \Phi(\boldsymbol{s}_j, \boldsymbol{s}_m) u_m \Bigg)_{L^2(D)} \\ &+ 2 \sum_{j=1}^{M_{\mathcal{S}}} w_j \left(v_j, |\boldsymbol{s}_j \cdot \boldsymbol{n}| u_j\right)_{L^2(\Gamma_-(\boldsymbol{s}_j))}. \end{aligned}$$

Let the linear functional $l(\cdot)$ from (20) be approximated by a functional $\tilde{l}(\cdot)$ with angular quadrature correspondingly, then the directional solutions u_j are determined from the variational formulation with angular quadrature

$$\tilde{a}(u,v) = \tilde{l}(v) \quad \forall v \in V^{L,N},\ j = 1,\dots,M_{\mathcal{S}}. \tag{42}$$

Since this formulation has to hold for all $v \in V^{L,N}$, it follows that for test functions which vanish at every angular quadrature node $\boldsymbol{s}_i$, $i = 1,\dots,M_{\mathcal{S}}$ except one s_j, formulation (42) can be reduced to the variational formulation (40) from the collocation discretization. This condition on the test functions is satisfied e. g. for a basis of the test space of characteristic functions on the angular mesh if each mesh cell contains exactly one angular quadrature node. With such a one-point quadrature rule and characteristic basis functions of $V_{\mathcal{S}}^N$, the phase space Galerkin DOM is therefore equivalent to the collocation DOM after discretization.

Sparse tensor discretization

The full tensor approach presented before shows the typical complexity for full tensor approximations: The number of degrees of freedoms increases exponentially with the dimension and the resolution levels in a dyadically refined scheme.

A way to counter this exponential increase is found in sparse tensorization. Using the same approximation spaces on the component domains $V_D^{l_D}$ and $V_{\mathcal{S}}^{l_{\mathcal{S}}}$ as for the full tensor approximation we define a sparse tensor approximation space $\hat{V}^{L,N}$ by

$$\mathcal{V}_1 \approx \hat{V}^{L,N} := \sum_{0 \le f(l_D,l_{\mathcal{S}}) \le L} V_D^{l_D} \otimes V_{\mathcal{S}}^{l_{\mathcal{S}}}, \tag{43}$$

where the *sparsity profile* $f : \{0,\dots,L\} \times \{0,\dots,N\} \to \mathbb{R}$ determines which tensor product subspaces $V_D^{l_D} \otimes V_{\mathcal{S}}^{l_{\mathcal{S}}}$ are to be included in the approximation. The sparsity profile usually depends on N as well. Here, we employ a linear profile

$$f(l_D, l_{\mathcal{S}}) = l_D + L l_{\mathcal{S}}/N, \tag{44}$$

which is normally chosen if the component complexities M_D and $M_{\mathcal{S}}$ depend on the resolution parameters L and N in the same way and identical order of approximation is sought over both component domains (cf. Zenger 1991; Bungartz and Griebel 2004; Griebel and Harbrecht 2013a).

If direct sum decompositions of the component approximation spaces $V_D^{l_D}$ and $V_{\mathcal{S}}^{l_{\mathcal{S}}}$ into detail spaces $W_D^{l_D}$ and $W_{\mathcal{S}}^{l_{\mathcal{S}}}$, i. e.

$$V_D^{l_D} = V_D^{l_D-1} \oplus W_D^{l_D}, \quad l_D = 1,\dots,L$$

are available (correspondingly in the angular domain), then the sparse tensor approximation space $\hat{V}^{L,N}$ can also be written as

$$\hat{V}^{L,N} = \sum_{0 \le f(l_D,l_{\mathcal{S}}) \le L} W_D^{l_D} \otimes W_{\mathcal{S}}^{l_{\mathcal{S}}}. \tag{45}$$

By choosing hierarchical bases for $V_D^{l_D}$ and $V_{\mathcal{S}}^{l_{\mathcal{S}}}$, each degree of freedom u_{ij} can directly be associated with a tensor product detail space $W_D^{l_D} \otimes W_{\mathcal{S}}^{l_{\mathcal{S}}}$. The sparse solution is then given by

$$\hat{u}_{L,N} = \sum_{0 \le f(l_D,l_{\mathcal{S}}) \le L} u_{l_D,l_{\mathcal{S}}},$$

$$u_{l_D,l_{\mathcal{S}}} = \sum_{i=1}^{\dim W_D^{l_D}} \sum_{j=1}^{\dim W_{\mathcal{S}}^{l_{\mathcal{S}}}} u_{ij} \alpha_i^{l_D}(\boldsymbol{x}) \beta_j^{l_{\mathcal{S}}}(\boldsymbol{s}) \in W_D^{l_D} \otimes W_{\mathcal{S}}^{l_{\mathcal{S}}}.$$

Thus, the *sparse discrete variational problem* reads: Find $\hat{u}_{L,N} \in \hat{V}^{L,N}$ such that

$$a(\hat{u}_{L,N}, \hat{v}_{L,N}) = l(\hat{v}_{L,N}) \quad \forall \hat{v}_{L,N} \in \hat{V}^{L,N}. \tag{46}$$

The dimension of the sparse tensor product space $\hat{V}^{L,N}$ depends on the sparsity profile $f(l_D, l_{\mathcal{S}})$. For a linear sparsity profile as in (44), the following complexity estimate is known (e. g. Bungartz and Griebel 2004, Lemma 3.6), or Griebel and Harbrecht (2013a, Thm. 4.1)).

Lemma 7. *Assuming the dimensions of the detail spaces $W_D^{l_D}$ and $W_{\mathcal{S}}^{l_{\mathcal{S}}}$ scale as $\dim(W_i^{l_i}) \le c_i 2^{d_i l_i}$ with constants $c_i > 0$ and dimensions d_i, $i = D, \mathcal{S}$, with $d_D = d$, and given a linear sparsity profile $f(l_D, l_{\mathcal{S}})$ as in (44), the dimension of the sparse tensor product approximation space $\hat{V}^{L,N}$ as defined by (45) is*

$$\hat{M}_{L,N} \lesssim L^{\theta} 2^{\max\{dL, d_{\mathcal{S}}N\}} \lesssim (\log M_D)^{\theta} \max\{M_D, M_{\mathcal{S}}\}, \tag{47}$$

where $\theta = 1$ if $dL = d_{\mathcal{S}}N$ and $\theta = 0$ otherwise. Relation "$\lesssim$" defines an order up to constants with respect to the relevant scaling parameters L, N: $a \lesssim b$ iff $a \le Cb$ with constant C independent of L and N.

Error analysis

In this section, we shall show that the convergence rates of the full tensor and sparse tensor Galerkin methods differ only by a logarithmic factor in the degrees of freedom, provided that somewhat stronger regularity requirements are met for the exact solution.

The analysis will proceed along the usual fashion, cp. (Bungartz and Griebel 2004). We define the *Galerkin projector* $\mathrm{P}^{L,N} : \mathcal{V}_1 \to V^{L,N}$ into the full tensor product approximation space

$$a\left(\mathrm{P}^{L,N} u, v\right) = a(u,v) \quad \forall v \in V^{L,N}. \tag{48}$$

Letting $L \to \infty$ ($N \to \infty$) the fact that the subspaces are closed and dense implies that in the respective limits we obtain *semidiscrete Galerkin projectors* $\mathrm{P}_S^N := \lim_{L\to\infty} \mathrm{P}^{L,N}$ $\left(\mathrm{P}_D^L := \lim_{N\to\infty} \mathrm{P}^{L,N}\right)$ on the physical (angular) domain, as the Galerkin projector is stable in the $\|\cdot\|_1$-norm:

Lemma 8 (Stability of the Galerkin projector)**.** *Let $v \in \mathcal{V}_1$. Then there is a constant $c_P > 0$ independent of L and N so that*

$$\left\|\mathrm{P}^{L,N} v\right\|_1 \leq c_P \|v\|_1.$$

Proof. With continuity (Lemma 3) of the bilinear form we obtain

$$\begin{aligned} |a\left(\mathrm{P}^{L,N} v, v_{L,N}\right)| &= |a(v, v_{L,N})| \\ &\leq c_c \|v\|_1 \left\|v_{L,N}\right\|_1 \quad \forall v_{L,N} \in V^{L,N}. \end{aligned}$$

Since this holds for all $v_{L,N} \in V^{L,N}$, we can set $v_{L,N} = \mathrm{P}^{L,N} v$ and exploit coercivity of the bilinear form (Lemma 5):

$$\begin{aligned} c_e \left\|\mathrm{P}^{L,N} v\right\|_1^2 &\leq |a\left(\mathrm{P}^{L,N} v, \mathrm{P}^{L,N} v\right)| \\ &= |a\left(v, \mathrm{P}^{L,N} v\right)| \leq c_c \|v\|_1 \left\|\mathrm{P}^{L,N} v\right\|_1. \end{aligned}$$

If $\mathrm{P}^{L,N} v \neq 0$ we obtain the result with $c_P = c_c / c_e$. □

Error estimates on the physical domain

To begin with, we require some approximation results in the $H^1(D)$-norm on the physical domain. With a Clément-type quasi-interpolation operator P_I^L (Scott and Zhang, 1990, Thm. 4.1 and Cor. 4.1) we obtain

Lemma 9 (Approximation of quasi-interpolation)**.** *For polyhedral $D \subset \mathbb{R}^d$ and a shape-regular triangulation $\mathcal{T}_D^L$ on D with mesh width $h = 2^{-L}$, the quasi-interpolation $\mathrm{P}_\mathrm{I}^L v$ of a function $v \in H^{s+1}(D)$, $s \in [0,1]$, to the space $V_D^L = S^{0,1}\left(D, \mathcal{T}_D^L\right)$ of piecewise affine functions on $\mathcal{T}_D^L$ satisfies the error estimate*

$$\|v - \mathrm{P}_\mathrm{I}^L v\|_{H^1(D)} \leq c_H 2^{-sL} \|v\|_{H^{s+1}(D)},$$

where $c_H > 0$ is a constant independent of L.

Lemma 10 (Stability of quasi-interpolation)**.** *Under the assumptions of Lemma 9, quasi-interpolation is H^1-stable, i. e. there exists a constant $c_B > 0$ independent of L such that for all $v \in H^1(D)$ it holds*

$$\|\mathrm{P}_\mathrm{I}^L v\|_{H^1(D)} \leq c_B \|v\|_{H^1(D)}.$$

Next we derive an error estimate for the Galerkin approximation on the physical domain. At this point, the approximation is semidiscrete.

Lemma 11 (Error estimate for Galerkin projection on physical domain)**.** *Let $u \in H^{s+1,0}(\Omega)$, $s \in \{0,1\}$, be the exact solution to problem (21) and $u_L := \mathrm{P}_D^L u \in V_D^L \otimes L^2(\mathcal{S})$ the Galerkin projected solution to*

$$a(u_L, v_L) = l(v_L) \quad \forall v_L \in V_D^L \otimes L^2(\mathcal{S}) \tag{49}$$

with $a(\cdot,\cdot)$ from (19) and $l(\cdot)$ from (20). Then, there is a constant $c_p > 0$ independent of L such that

$$\|u - u_L\|_1 \leq c_p 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)}.$$

Proof. The proof is standard, and is based on coercivity and Galerkin orthogonality. We proceed analogous to (Ávila et al. 2011, Lemma 3 and Theorem 1). After inserting the quasi-interpolated solution $\hat{u}_L := (\mathrm{P}_\mathrm{I}^L \otimes \mathrm{Id}_\mathcal{S}) u$ with P_I^L from Lemma 9 the triangle inequality permits us to write

$$\|u - u_L\|_1 \leq \left\|u - \hat{u}_L\right\|_1 + \left\|\hat{u}_L - u_L\right\|_1. \tag{50}$$

For the first part, we use the fact that there is a constant $c_n > 0$ for all $v \in H^1(D) \otimes L^2(\mathcal{S})$ such that

$$\|v\|_1 \leq c_n \|v\|_{H^{1,0}(\Omega)}.$$

Thus, we can apply Lemma 9:

$$\left\|u - \hat{u}_L\right\|_1 \leq c_n \left\|u - \hat{u}_L\right\|_{H^{1,0}(\Omega)} \leq c_n c_H 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)}.$$

For the second part in (50), we use coercivity of the bilinear form, then in a second step Galerkin orthogonality, and finally continuity of the bilinear form to write

$$\begin{aligned} \left\|u_L - \hat{u}_L\right\|_1^2 &\leq c_e^{-1} a(u_L - \hat{u}_L, u_L - \hat{u}_L) \\ &\leq c_e^{-1} a(u - \hat{u}_L, u_L - \hat{u}_L) \\ &\leq c_c c_e^{-1} \left\|u - \hat{u}_L\right\|_1 \left\|u_L - \hat{u}_L\right\|_1 \\ &\leq c_c c_e^{-1} c_n \left\|u - \hat{u}_L\right\|_{H^{1,0}(\Omega)} \left\|u_L - \hat{u}_L\right\|_1, \end{aligned}$$

and therefore with Lemma 9

$$\left\|u_L - \hat{u}_L\right\|_1 \leq c_c c_e^{-1} c_n c_H 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)}.$$

By inserting into (50) we arrive at the result

$$\|u - u_L\|_1 \leq c_n c_H \left(1 + c_c c_e^{-1}\right) 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)}.$$

□

Error estimates on the angular domain

On the angular domain, the considerations in the following require an approximation result for L^2-projections.

Lemma 12. *For functions $v \in H^t(\mathcal{S})$, $t \in \{0,1\}$, the L^2-projection to the space $V_\mathcal{S}^N$ satisfies the error estimate*

$$\left\|v - \mathrm{P}_{L^2(\mathcal{S})}^N v\right\|_{L^2(\mathcal{S})} \leq c_l 2^{-tN} \|v\|_{H^t(\mathcal{S})}, \tag{51}$$

where the constant $c_l > 0$ is independent of N.

This result can be obtained for approximation by spectral functions as in the spherical harmonics method (in which case $t \geq 0$ is arbitrary), for instance, as well as for

approximation by piecewise constants as in the discrete ordinates method (in which case $0 \leq t \leq 1$). It allows the derivation of the same approximation rate for the semidiscrete Galerkin projection on the angular domain.

Lemma 13 (Error estimate for angular Galerkin projection)**.** *Let $u \in H^{1,t}(\Omega)$, $t \in \{0,1\}$, be the exact solution to problem (21) and $u_N := \mathrm{P}_{\mathcal{S}}^N u \in H^1(D) \otimes V_{\mathcal{S}}^N$ the Galerkin projected solution with angular part from the subspace $V_{\mathcal{S}}^N$ of $L^2(\mathcal{S})$. Then there is a constant $c_a > 0$ independent of N such that*

$$\|u - u_N\|_1 \leq c_a N^{-t} \|u\|_{H^{1,t}(\Omega)}.$$

Proof. The proof proceeds analogously to the one of Lemma 11 while substituting the L^2-projected solution with Lemma 12 for the quasi-interpolated solution, the details are therefore omitted here. □

Error estimate for the full tensor phase space Galerkin method

The following theorem gives an error estimate for the full tensor approximation.

Theorem 14 (Error estimate full tensor Galerkin method)**.** *The full tensor Galerkin approximation $u_{L,N} = \mathrm{P}^{L,N} u$ of a solution $u \in H^{s+1,0}(\Omega) \cap H^{1,t}(\Omega)$, $s \in \{0,1\}$, $t \in \{0,1\}$, to the variational problem (21) satisfies the asymptotic error estimate*

$$\left\|u - u_{L,N}\right\|_1 \lesssim 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)} + 2^{-tN} \|u\|_{H^{1,t}(\Omega)}, \tag{52}$$

with relation "$\lesssim$" as in Lemma 7.

Proof. By Céa's Lemma (Brenner and Scott 2008, Thm. 2.8.1) the Galerkin approximation is quasi-optimal in $V^{L,N}$, its error can therefore be bounded (up to constants) by the error of any other approximation to u in $V^{L,N}$, for example the quasi-interpolated and L^2-projected approximation $\mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{P}_{L^2}^N u$:

$$\begin{aligned}
\left\|u - \mathrm{P}^{L,N} u\right\|_1 &\lesssim \left\|u - \mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{P}_{L^2}^N u\right\|_1 \leq \left\|u - \mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{Id} u\right\|_1 \\
&\quad + \left\|\mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{Id} u - \mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{P}_{L^2}^N u\right\|_1 \\
&\lesssim 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)} \\
&\quad + \left\|(\mathrm{Id} - \mathrm{Id} \otimes \mathrm{P}_{L^2}^N)\mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{Id} u\right\|_1 \\
&\lesssim 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)} \\
&\quad + 2^{-tN} \left\|\mathrm{P}_{\mathrm{I}}^L \otimes \mathrm{Id} u\right\|_{H^{1,t}(\Omega)} \\
&\lesssim 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)} + 2^{-tN} \|u\|_{H^{1,t}(\Omega)}.
\end{aligned}$$

Here, we used the approximation properties of the quasi-interpolant from Lemma 9 and of the angular L^2-projection from Lemma 12. The last step is a consequence of the H^1-stability asserted in Lemma 10 of the quasi-interpolation. □

Error estimate for the sparse tensor phase space Galerkin method

After the full tensor approximation properties, we consider the convergence properties of a direct sparse tensor approximation on the sparse tensor product space $\hat{V}^{L,N}$ as defined in (45).

In analogy to the full tensor Galerkin projector $\mathrm{P}^{L,N}$, we can define a *sparse tensor Galerkin projector* $\hat{\mathrm{P}}^{L,N}$ by the orthogonality relation

$$a\left(\hat{\mathrm{P}}^{L,N} u, v\right) = a(u,v) \quad \forall v \in \hat{V}^{L,N}.$$

The error of the sparse tensor solution $\hat{u}_{L,N} = \hat{\mathrm{P}}^{L,N} u$ is estimated in the following theorem (see also Widmer 2009, Thm. 2.6) and (Griebel and Harbrecht 2013a, Thms. 4.3 and 7.1)).

Theorem 15 (Error estimate of direct sparse tensor solution)**.** *Let the linear sparsity profile as in (44) be given. Assume further that L and N vary such that $-sL + tN = \zeta = \text{const}$, then the direct sparse tensor approximation $\hat{u}_{L,N}$ of a function $u \in H^{s+1,t}(\Omega)$, $s,t \in \{0,1\}$, satisfies the error estimate*

$$\left\|u - \hat{u}_{L,N}\right\|_1 \lesssim L\left(2^{-sL} + 2^{-tN}\right) \|u\|_{H^{s+1,t}(\Omega)},$$

where relation "$\lesssim$" is defined as in Lemma 7.

Proof. We follow the proof of Thm. 2.6 by (Widmer 2009). First we introduce so-called difference projectors $\Delta_{\mathrm{I}}^{l_D} := \mathrm{P}_{\mathrm{I}}^{l_D} - \mathrm{P}_{\mathrm{I}}^{l_D - 1}$ and $\Delta_{L^2}^{l_S} := \mathrm{P}_{L^2}^{l_S} - \mathrm{P}_{L^2}^{l_S - 1}$ as the difference between projections to two consecutive resolution levels with the convention $\mathrm{P}_{\mathrm{I}}^{-1} = 0 = \mathrm{P}_{L^2}^{-1}$. They project onto the detail spaces $W_D^{l_D}$ and $W_{\mathcal{S}}^{l_S}$, respectively.

With these difference projectors, a sparse quasi-interpolated and L^2-projected approximation $\bar{u}_{L,N} \in \hat{V}^{L,N}$ to u can be expressed as

$$\bar{u}_{L,N} = \sum_{l_D=0}^{L} \sum_{l_S=0}^{l_S^{\max}(l_D)} \Delta_{\mathrm{I}}^{l_D} \otimes \Delta_{L^2}^{l_S} u,$$

where $l_S^{\max}(l_D)$ is the largest feasible angular resolution index which results from solving $f(l_D, l_S) \leq L$ with respect to l_S.

Now we exploit quasi-optimality of the Galerkin approximation on the sparse tensor product space to replace the Galerkin approximation error by the error of the quasi-interpolated and L^2-projected approximation.

Additionally applying the norm estimate $\|v\|_1 \lesssim \|v\|_{H^{1,0}(\Omega)}$ yields

$$\left\|u - \hat{u}_{L,N}\right\|_1 \lesssim \left\|u - \sum_{l_D=0}^{L} \sum_{l_S=0}^{l_S^{\max}(l_D)} \Delta_{\mathrm{I}}^{l_D} \otimes \Delta_{L^2}^{l_S} u\right\|_{H^{1,0}(\Omega)}. \tag{53}$$

The error is split into two terms:

$$\left\|u - \bar{u}_{L,N}\right\|_{H^{1,0}(\Omega)} \leq \underbrace{\left\|\sum_{l_D=0}^{L} \sum_{l_S=l_S^{\max}(l_D)+1}^{\infty} \Delta_{\mathrm{I}}^{l_D} \otimes \Delta_{L^2}^{l_S} u\right\|_{H^{1,0}(\Omega)}}_{=:I} + \underbrace{\left\|\sum_{l_D=L+1}^{\infty} \sum_{l_S=0}^{\infty} \Delta_{\mathrm{I}}^{l_D} \otimes \Delta_{S}^{l_S} u\right\|_{H^{1,0}(\Omega)}}_{=:II}. \tag{54}$$

The second term on the right hand side can be estimated by Lemma 9:

$$II = \left\|\left(\mathrm{Id} - \mathrm{P}_{\mathrm{I}}^{L}\right) \otimes \mathrm{Id}\, u\right\|_{H^{1,0}(\Omega)} \leq c_H 2^{-sL} \|u\|_{H^{s+1,0}(\Omega)}. \tag{55}$$

This term will not contribute to the asymptotic terms.

The first term on the right hand side of (54) is split up further:

$$\begin{aligned} I &= \left\|\sum_{l_D=0}^{L} \left(\mathrm{P}_{\mathrm{I}}^{l_D} - \mathrm{P}_{\mathrm{I}}^{l_D-1}\right) \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{1,0}(\Omega)} \\ &= \left\|\sum_{l_D=0}^{L} \left(\mathrm{P}_{\mathrm{I}}^{l_D} - \mathrm{Id} + \mathrm{Id} - \mathrm{P}_{\mathrm{I}}^{l_D-1}\right) \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{1,0}(\Omega)} \\ &\leq \sum_{l_D=0}^{L} \left(\left\|\left(\mathrm{Id} - \mathrm{P}_{\mathrm{I}}^{l_D}\right) \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{1,0}(\Omega)} \right. \\ &\quad \left. + \left\|\left(\mathrm{Id} - \mathrm{P}_{\mathrm{I}}^{l_D-1}\right) \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{1,0}(\Omega)} \right). \end{aligned} \tag{56}$$

Both norms on the right hand side of (56) can be estimated by Lemma 9 and Lemma 12:

$$\begin{aligned} &\left\|\left(\mathrm{Id} - \mathrm{P}_{\mathrm{I}}^{l_D}\right) \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{1,0}(\Omega)} \\ &\leq c_H 2^{-sl_D} \left\|\mathrm{Id} \otimes \left(\mathrm{Id} - \mathrm{P}_{L^2}^{l_S^{\max}(l_D)}\right) u\right\|_{H^{s+1,0}(\Omega)} \\ &\leq c_H c_l 2^{-sl_D - t l_S^{\max}(l_D)} \|u\|_{H^{s+1,t}(\Omega)}. \end{aligned}$$

Inserting back into (56) yields

$$I \leq 2 c_H c_l \|u\|_{H^{s+1,t}(\Omega)} \sum_{l_D=0}^{L} 2^{-sl_D - t l_S^{\max}(l_D)}. \tag{57}$$

The task is now to estimate the series. Using the assumption $\zeta = -s + tN/L$:

$$\begin{aligned} \sum_{l_D=0}^{L} 2^{-sl_D - tN/L(L-l_D)} &= 2^{-tN} \sum_{l_D=0}^{L} 2^{(-s+tN/L)l_D} \\ &= 2^{-tN} \sum_{l_D=0}^{L} 2^{\zeta l_D}. \end{aligned} \tag{58}$$

We estimate the sum on the right hand side of (58) by its largest summand. Two cases can be distinguished here:

1. If $\zeta \leq 0$, the largest summand occurs for $l_D = 0$:

$$2^{-tN} \sum_{l_D=0}^{L} 2^{\zeta l_D} \leq L 2^{-tN}.$$

2. If $\zeta > 0$, the largest summand occurs for $l_D = L$:

$$2^{-tN} \sum_{l_D=0}^{L} 2^{\zeta l_D} \leq 2^{-tN} L 2^{-sL+tN} = L 2^{-sL}.$$

In summary, we may write

$$\sum_{l_D=0}^{L} 2^{-sl_D - t l_S^{\max}(l_D)} \leq L 2^{-sL-tN}.$$

By combining this estimate with relations (53) to (57), we finally arrive at

$$\left\|u - \hat{u}_{L,N}\right\|_{H^{1,0}(\Omega)} \lesssim L 2^{-sL-tN} \|u\|_{H^{s+1,t}(\Omega)}.$$

□

In conclusion, we find that the convergence rate of $O(2^{-sL-tN})$ of the full tensor approximation is maintained up to an additional factor L, which by $M_D = O(2^{dL})$ is logarithmic in the number of degrees of freedom. This result in conjunction with the greatly reduced complexity of the sparse tensor method (Lemma 7) shows its superior efficiency provided that the function u to be approximated is at least in $H^{s+1,t}(\Omega)$, with $s, t \in \{0, 1\}$.

Numerical experiments

Algorithms

For the numerical experiments we compute a sparse tensor solution with the help of the combination technique. The sparse solution is constructed according to the formula

$$\check{u}_{L,N} = \sum_{\ell_D=0}^{L} \left(u_{\ell_D, \ell_S^{\max}(\ell_D)} - u_{\ell_D, \ell_S^{\max}(\ell_D+1)}\right)$$

from a number of solutions $u_{\ell_D,\ell_S} \in V^{\ell_D,\ell_S}$ to the full tensor discrete variational formulation (38) of reduced physical resolution ℓ_D and angular resolution ℓ_S.

Clearly $\check{u}_{L,N}$ is in the space $\hat{V}^{L,N} = \sum_{l_D=0}^{L} V^{\ell_D,\ell_S^{\max}(\ell_D)}$, which is identical to the sparse tensor approximation space from (43). However, in general the combination approximation differs from a direct sparse approximation $\hat{u}_{L,N}$ (see also Grella 2013, Sec. 2.3.1). Due to the quasi-optimality of the direct sparse solution as an approximation in $\hat{V}^{L,N}$, the error of the combination approximation can serve as an upper bound (up to factors) for the error $\|u - \hat{u}_{L,N}\|_1$ of the direct sparse approximation.

Note that the convergence of the combination technique for the radiative transfer problem has not been shown formally yet. A recent proof for elliptic operators by (Griebel and Harbrecht 2013b) would be applicable under certain stability assumptions on the semidiscrete Galerkin projectors (for details we refer to (Grella 2013, Sec. 5.3.7)). However, the use of the combination technique approximation has practical advantages over the direct sparse approximation. First, to construct the subproblem solutions of lower resolution, an existing full tensor solver with standard nonhierarchical FEM bases can be reused, no direct sparse solver needs to be implemented. Second, the splitting into subproblems entails a natural level for parallelism in the algorithm, which can still be combined with parallel solution procedures at the level of each subproblem (an implementation is described in (Grella 2013, Chap. 7)).

Each of the full tensor subproblems is solved by a phase space Galerkin finite element method with nonhierarchical affine hat functions as physical basis and piecewise constants as angular basis. In the experiment of Sec. 'Experiment 2', the midpoint rule is used for angular quadrature which corresponds to the S_N-method. However, in situations where ray effects (Lathrop 1968) pollute the results, adaptive quadrature may help (Stone 2007). As a simple adaptive rule we link the number of quadrature points n_q per dimension and per mesh element to the resolution levels l_D, l_S of the subproblem by $n_q = \max\{l_D/l_S, 1\}$ in the experiment of Sec. 'Experiment 1'. Even though the overall computational effort is then not bounded by Lemma 7, the total number of degrees of freedom still is. As the iterative, approximate solution of the linear system constitutes the most time consuming part, the sparse tensor method is, in practice, more efficient than the full tensor method.

Quantities of interest

In applications, the radiative intensity is often coupled to other modes of energy transport via the net emission (e. g. Larsen et al. 2002, Eq. (1.1a)). The net emission can be computed in turn from the *incident radiation*

$$G(\boldsymbol{x}) = \int_{\mathcal{S}} u(\boldsymbol{x}, \boldsymbol{s}) \, \mathrm{d}\boldsymbol{s}. \tag{59}$$

For this reason, we choose the incident radiation as a lower-dimensional variable to visualize results and to analyze errors. The relative L^2- or H^1-error of the incident radiation is given by

$$err(G_{L,N})_X = \|G - G_{L,N}\|_X / \|G\|_X, \quad X = L^2(D), H^1(D).$$

Numerical experiments

All experiments are set on the domains $D = [0,1]^d$, $\mathcal{S} = \mathcal{S}^{d_{\mathcal{S}}}$, with $d = d_{\mathcal{S}} + 1$. We solve the RTP with isotropic scattering $\Phi(\boldsymbol{s}, \boldsymbol{s}') = 1/|\mathcal{S}|$ and zero inflow boundary conditions $g = 0$.

Experiment 1

We search the solution to the Gaussian blackbody radiation

$$I_b(\boldsymbol{x}) = 2 \exp\left(-3^2(\boldsymbol{x} - \boldsymbol{c})^2\right), \quad \boldsymbol{c} = (0.5, 0.5)^\top,$$

with absorption and scattering coefficient $\kappa = \sigma = 1$.

The H^1-error of the incident radiation indeed converges faster in the sparse approximation than the full approximation (Figure 1). Note that the L^2-error of the sparse approximation can be larger than the error of the full approximation because the sparsity profile $f(l_D, l_S)$ has been optimized for essentially undeteriorated convergence in the $\|\cdot\|_1$-norm of the error in the radiative intensity, which is more closely represented by the H^1-error than the L^2-error of the incident radiation.

Experiment 2

A blackbody radiation $I_b(\boldsymbol{x}, \boldsymbol{s})$ corresponding to the exact solution

$$u(\boldsymbol{x}, \boldsymbol{s}) = \frac{3}{16\pi}(1 + (\boldsymbol{s} \cdot \boldsymbol{s}')^2) \prod_{i=1}^{3} (-4x_i(x_i - 1)),$$

with fixed $\boldsymbol{s}' = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^\top$ is inserted into the right hand side functional in (38) (Grella 2013, Sec. 8.2, Exp. 1). The absorption is set to $\kappa = 1$, the scattering coefficient to $\sigma = 0.5$.

For this experiment we employed a discrete ordinates solver in which the angular resolution N' is related to the angular degrees of freedom by $M_{\mathcal{S}} = (N' + 1)^2$ so that $N \approx \lfloor \log_2(N' + 1) \rfloor$, where N is the angular resolution used otherwise in this paper.

Figure 2 shows the superior efficiency of the sparse approach with respect to number of degrees of freedom vs. achieved error. The convergence rates indicate that the curse of dimensionality is mitigated by the sparse discrete ordinates method.

For a comparison to other sparse tensor approaches we refer to the numerical experiment section of (Grella 2013), which features a sparse tensor spherical harmonics approximation and a sparse collocation discrete ordinates method realized via the combination technique. We

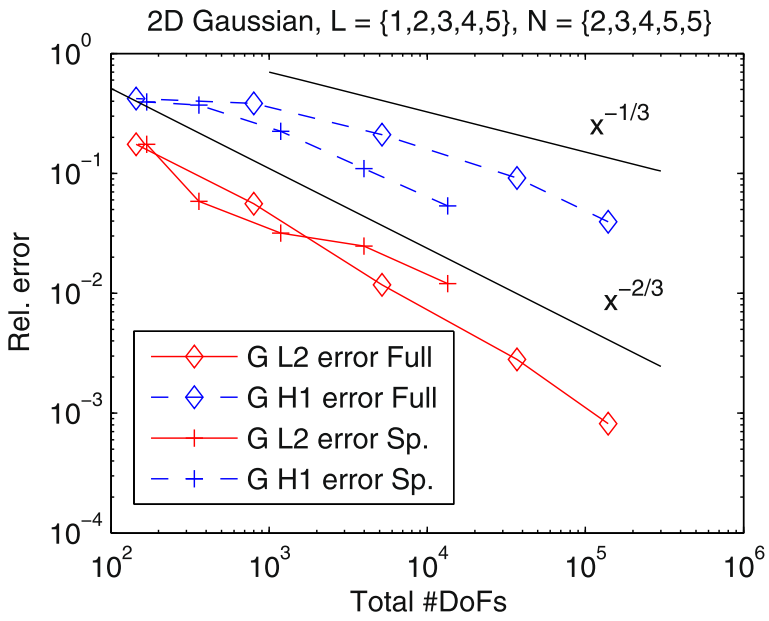

Figure 1 Experiment 1: Convergence in incident radiation with full and sparse phase space Galerkin approximation. Reference resolution was $L_{\text{ref}} = 6/N_{\text{ref}} = 6$. Reference slopes provided as visual aids only. Even with the lowest order sparse tensor phase space Galerkin discretization, the savings in DoFs to reach engineering accuracy of 1%–10% in the H^1 error is about an order of magnitude.

observed that the approach presented here performs similarly to the sparse collocation DOM combination technique as the methods are similar from the point of view of implementation, even though their theoretical derivation is different. The presented approach is somewhat less susceptible to ray effects at the expense of slightly longer computational times as the angular quadrature is adapted to the resolution of the angular mesh. The spherical harmonics method is most effective for solutions with highly regular angular part because of its regularity requirements for spectral convergence. In general, at the same resolution levels L and N, the combination technique approach

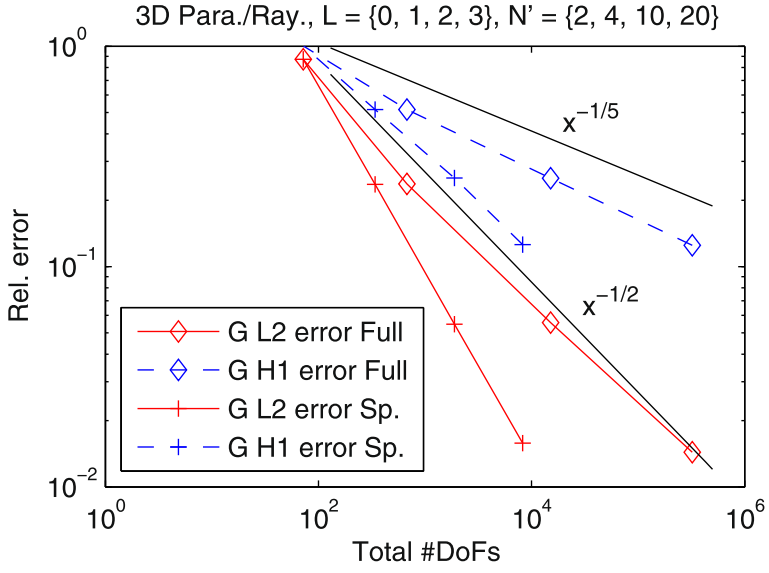

Figure 2 Experiment 2: Convergence in incident radiation with full and sparse DOM. Reference resolution was $L_{\text{ref}} = 4$. Angular resolution N' corresponds to $N \approx \{1, 2, 3, 4\}$. Reference slopes provided as visual aids only. The savings in DoFs to reach engineering accuracy of 1%–10% are about two orders of magnitude.

realizes approximately the same error as the direct sparse approach, while the number of degrees of freedom in the combination technique is larger than in the direct sparse approach because the approximation spaces of different subproblems in the combination technique overlap in the degrees of freedom. It is therefore slightly less efficient than the direct sparse approach, but considerably more efficient than the full tensor approach and advantageous in practice due to faster and simpler implementation and parallelization.

Conclusion

We have shown a direct sparse tensor phase space Galerkin approximation of the radiative intensity in the stationary monochromatic radiative transfer problem can be computed with only $O(\log M_D(M_D + M_{\mathcal{S}}))$ degrees of freedom as opposed to $O(M_D M_{\mathcal{S}})$ degrees of freedom for a standard full tensor approximation. Here, M_D is the number of physical degrees of freedom and $M_{\mathcal{S}}$ the number of angular degrees of freedom. At the same time, the error of the sparse approximation in the $\|\cdot\|_1$-norm still decreases essentially as the error of the full approximation, namely with the order $O\left(\log M_D \left(M_D^{-s/d} + M_{\mathcal{S}}^{-t/d_{\mathcal{S}}}\right)\right)$ as compared to $O\left(M_D^{-s/d} + M_{\mathcal{S}}^{-t/d_{\mathcal{S}}}\right)$ in the full tensor approximation. The parameters $s, t \in \{0, 1\}$ indicate the regularity of the exact solution which is required to be in the space of mixed smoothness $H^{s+1,t}(D \times \mathcal{S})$ to achieve the sparse convergence rate, whereas $H^{s+1,0}(D \times \mathcal{S}) \cap H^{1,t}(D \times \mathcal{S})$ is sufficient in the full tensor approximation.

To simplify implementation, we realized the sparse tensor approximation algorithmically via the combination technique. Together with suitable quadrature rules, we demonstrated in numerical experiments that this sparse tensor combination approximation retains the analyzed theoretical advantages of the direct sparse tensor method while allowing for straightforward parallelization also at the level of subproblems.

The proposed specialization of the phase space Galerkin framework investigated here has the advantage that both discrete ordinates and spherical harmonics method can be derived from it so that the sparse tensorization benefits hold for the sparse variants of both methods alike.

Therefore, for problems whose solutions exhibit so-called mixed regularity, the sparse tensor product phase space Galerkin approximations realize a significant increase in efficiency, i. e. achievable error per number of degrees of freedom. Even in applications where high numerical accuracy is the main objective, a sparse tensor product approximation might be of value as an initial value for an iterative solver or in a problem-adapted preconditioning scheme.

Competing interests
The author declares that he has no competing interests.

Acknowledgments
The author wishes to thank Prof. Dr. Ch. Schwab for helpful discussions and valuable suggestions to improvements of this article. Financial support for this work by Schweizerischer Nationalfonds (SNF) under project no. 121892, by Deutsche Forschungsgemeinschaft (DFG) within SPP1324, and by the European Research Council (ERC) under ERC Advanced Grant 247277 is gratefully acknowledged.

Received: 29 December 2013 Accepted: 15 April 2014
Published: 7 May 2014

References

Ávila M, Codina R, Principe J (2011) Spatial approximation of the radiation transport equation using a subgrid-scale finite element method. Comput Meth Appl Mech Eng 200: 425–438. doi:10.1016/j.cma.2010.11.003

Brenner SC, Scott LR (2008) The mathematical theory of finite element methods, volume 15 of Texts Applied in Mathematics. Springer, New York. doi:10.1007/978-0-387-75934-0

Brooks A, Hughes TJR (1982) Streamline upwind/Petrov-Galerkin formulation for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. Comput Methods Appl Mech Engrg 32(1–3): 199–259. doi:10.1016/0045-7825(82)90071-8

Bungartz H-J, Griebel M (2004) Sparse grids. In: Iserles A (ed). Acta numerica volume 13, Cambridge University Press, pp 147–269. www5.in.tum.de/pub/bungartz04sparse.pdf.

Dahmen W, Huang C, Schwab C, Welper G (2012) Adaptive Petrov-Galerkin methods for first order transport equations. SIAM J Numer Anal 50(5): 2420–2445. ISSN 0036-1429. doi:10.1137/110823158.

Evans KF (1998) The spherical harmonic discrete ordinate method for three-dimensional atmospheric radiative transfer. J Atmos Sci 55(3): 429–446. doi:10.1175/1520-0469(1998)055<0429:TSHDOM>2.0.CO;2.

Garcke J (2007) A dimension adaptive sparse grid combination technique for machine learning. In: Read W, Larson J W, Roberts A J (eds) Proceedings of the 13th Biennial Computational Techniques and Applications Conference, CTAC-2006, volume 48 of ANZIAM J, pp C725–C740. http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/70.

Grella K (2013) Sparse tensor approximation for radiative transport. PhD thesis ETH Zurich, No.21388. doi:10.3929/ethz-a-009970281.

Grella K, Schwab C (2011a) Sparse tensor spherical harmonics approximation in radiative transfer. J Comput Phys 230(23): 8452–8473. ISSN 0021-9991. doi:10.1016/j.jcp.2011.07.028.

Grella K, Schwab C (2011b) Sparse discrete ordinates method in radiative transfer. Comput Meth Appl Math 11(3): 305–326. ISSN 1609-9389. doi:10.2478/cmam-2011-0017.

Griebel M, Schneider M, Zenger C (1992) Iterative Methods in Linear Algebra, chapter A combination technique for the solution of sparse grid problems. Amsterdam, North-Holland

Griebel M, Harbrecht H (2013a) On the construction of sparse tensor product spaces. Math Comp 82: 975–994. doi:10.1090/S0025-5718-2012-02638-X.

Griebel M, Harbrecht H (2013b) On the convergence of the combination technique. Technical Report 1304, Institut für Numerische Simulation, Rheinische Friedrich-Wilhelms-Universität Bonn, March. http://wissrech.ins.uni-bonn.de/research/pub/griebel/CombiTechniqueConvergence.pdf.

Hegland M (2003) Adaptive sparse grids. In: Burrage K, Sidje RB (eds) Proc. of 10th Computational Techniques and Applications Conference CTAC-2001, volume 44 of ANZIAM J., pp C335–C353. http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/685.

Hébert A (2010) Handbook of nuclear engineering, chapter multigroup neutron transport and diffusion computations. Springer. doi:10.1007/978-0-387-98149-9_8.

Kanschat G (2008) Solution of radiative transfer problems with finite elements. In: Kanschat G, Meinköhn E, Rannacher R, Wehrse R (eds) Numerical methods in multidimensional radiative transfer. Springer, pp 49–98. doi:10.1007/978-3-540-85369-5.

Knapp AW (2005) Advanced Real Analysis. Cornerstones, Birkhäuser Boston. doi:10.1007/0-8176-4442-3. ISBN 978-0-8176-4382-9

Larsen EW, Thömmes G, Klar A, Seaïd M, Götz T (2002) Simplified P_N approximations to the equations of radiative heat transfer and

applications. J Comput Phys 183(2): 652–675. ISSN 0021-9991. doi:10.1006/jcph.2002.7210.
Lathrop KD (1968) Ray effects in discrete ordinates equations. Nucl Sci Eng 32(3): 357
Manteuffel TA, Ressel KJ, Starke G (2000) A boundary functional for the least-squares finite-element solution of neutron transport problems. SIAM J Numer Anal 37(2): 556–586. doi:10.1137/S0036142998344706.
Modest MF (2003) Radiative heat transfer, 2nd edition. Elsevier, Amsterdam
Modest MF, Yang J (2008) Elliptic PDE formulation and boundary conditions of the spherical harmonics method of arbitrary order for general three-dimensional geometries. J Quant Spectrosc Radiative Transf 109: 1641–1666. doi:10.1016/j.jqsrt.2007.12.018.
Peng K, Gao X, Qu X, Ren N, Chen X, He X, Wang X, Liang J, Tian J (2011) Graphics processing unit parallel accelerated solution of the discrete ordinates for photon transport in biological tissues. Appl Opt 50(21): 3808–3823. doi:10.1364/AO.50.003808.
Scott LR, Zhang S (1990) Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math Comp 54: 483–493. doi:10.1090/S0025-5718-1990-1011446-7.
Stone JC (2007) Adaptive discrete-ordinates algorithms and strategies. PhD thesis, Texas A&M University. http://repository.tamu.edu//handle/1969.1/85857.
Widmer G, Hiptmair R, Schwab C (2008) Sparse adaptive finite elements for radiative transfer. Comput Phys 227: 6071-6105. doi:10.1016/j.jcp.2008.02.025.
Widmer G (2009) Sparse finite elements for radiative transfer. PhD thesis, ETH Zürich. http://e-collection.ethbib.ethz.ch/view/eth:374. No. 18420. doi:10.3929/ethz-a-005916456.
Zenger C (1991) Sparse grids. In: Hackbusch W (ed) Parallel algorithms for partial differential equations, number 31 in notes on numerical fluid mechanics. Vieweg. http://www5.in.tum.de/pub/zenger91sg.pdf.

doi:10.1186/2193-1801-3-230
Cite this article as: Grella: **Sparse tensor phase space Galerkin approximation for radiative transport.** *SpringerPlus* 2014 **3**:230.