A Nodally Bound-Preserving Finite Element Method for Hyperbolic Convection-Reaction Problems

Ben Ashby

With:

A. Hamdan, Bath

T. Pryer, Bath

April 2025

Institute for
Mathematical Innovation

UNIVERSITY OF
BATH

Outline

Motivation

Have PDE with solution u, solved using some numerical method to obtain numerical solution u_h.

Desired properties of u_h

- Preservation of maximum principle
- Accuracy (violation often caused by spurious oscillation)
- Physical relevance of numerical solution (i.e. dose, concentration etc)
- Compatibility (i.e. if solution is input to another model)

Illustrative Example: Advection-Reaction

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_-, \end{aligned}$$

where Γ_- is the inflow boundary defined by the flow field $\boldsymbol{b}$. Suppose that we set a rotational flow field

$$\boldsymbol{b}(x, y) := \frac{1}{\sqrt{x^2 + y^2}}(-y, x),$$

and piecewise constant data g on the inflow boundary, with $0 \leq g(\boldsymbol{x}) \leq 1$ for almost every $\boldsymbol{x} \in \partial\Omega$.

Then u should also be bounded between 0 and 1.

Illustrative Example

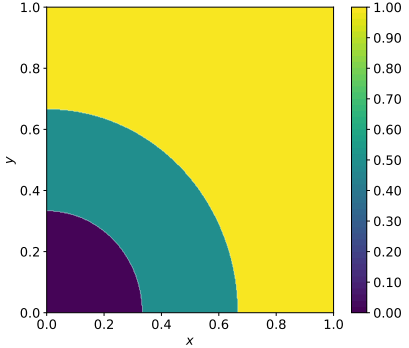


Figure: Contour plot of piecewise constant solution

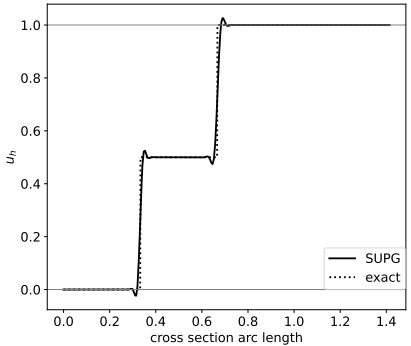


Figure: Diagonal cross-section of numerical solution

Illustrative Example

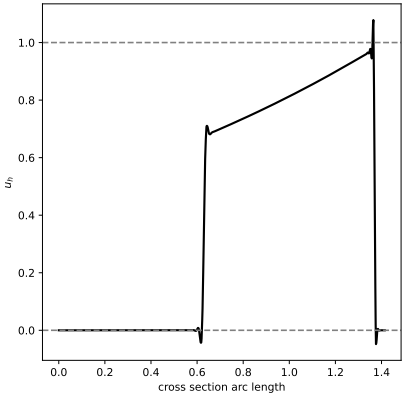


Figure: A less well-behaved case

Linear Advection-Reaction

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_-, \end{aligned} \tag{1}$$

Linear Advection-Reaction

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_-, \end{aligned} \tag{1}$$

The variational formulation of the advection reaction equation is to find $u \in \mathrm{H}_-(\Omega)$ such that

$$a(u, v) = l(v) \quad \forall v \in \mathrm{L}^2(\Omega),$$

Linear Advection-Reaction

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_-, \end{aligned} \tag{1}$$

The variational formulation of the advection reaction equation is to find $u \in \mathrm{H}_-(\Omega)$ such that

$$a(u, v) = l(v) \quad \forall v \in \mathrm{L}^2(\Omega),$$

where

$$a(w, v) = \int_\Omega (\boldsymbol{b} \cdot \nabla w + cw)\, v,$$

and

$$l(v) = \int_\Omega f v.$$

Linear Advection-Reaction

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_-, \end{aligned} \tag{1}$$

The variational formulation of the advection reaction equation is to find $u \in \mathrm{H}_-(\Omega)$ such that

$$a(u, v) = l(v) \quad \forall v \in \mathrm{L}^2(\Omega),$$

where

$$a(w, v) = \int_\Omega (\boldsymbol{b} \cdot \nabla w + cw)\, v,$$

and

$$l(v) = \int_\Omega fv.$$

We will assume a coercivity condition: $\exists \mu > 0$ such that $c(\boldsymbol{x}) - \frac{1}{2}\nabla \cdot \boldsymbol{b}(\boldsymbol{x}) \geq \mu$ for almost every $\boldsymbol{x}$.

Discretisation with SUPG

A common approach is to use a stabilised, conforming finite element approximation, that is, we find $u_h \in \mathbb{V}$ such that

$$a_h(u_h, v_h) = l_h(v_h) \quad \forall v_h \in \mathbb{V},$$

Discretisation with SUPG

A common approach is to use a stabilised, conforming finite element approximation, that is, we find $u_h \in \mathbb{V}$ such that

$$a_h(u_h, v_h) = l_h(v_h) \quad \forall v_h \in \mathbb{V},$$

where

$$a_h(w_h, v_h) = \int_\Omega \left(\boldsymbol{b} \cdot \nabla w_h + c v_h\right) v_h + \sum_{K \in \mathscr{T}} \delta_K \int_K \left(\boldsymbol{b} \cdot \nabla w_h + c w_h\right) \boldsymbol{b} \cdot \nabla v_h.$$

$$l_h(v_h) := \int_\Omega f v_h + \sum_{K \in \mathscr{T}} \delta_K \int_K f \left(\boldsymbol{b} \cdot \nabla v_h\right),$$

and

$$\mathbb{V} := \{v_h \in C^0(\Omega) : v_h|_K \in \mathcal{R}(K)\ \forall K \in \mathscr{T},\ v_h = 0 \text{ on } \Gamma_-\}.$$

Main Idea

The SUPG method is provably stable and convergent as $h \to 0$, however oscillations are still present and a discrete maximum principle does not hold.

Main Idea

The SUPG method is provably stable and convergent as $h \to 0$, however oscillations are still present and a discrete maximum principle does not hold.

Solution

Seek u_h in the convex subset K_h of $\mathbb{V}$ by restricting the nodal values of function in $\mathbb{V}$, that is

$$K_h := \{v_h \in \mathbb{V} : v_h(\boldsymbol{x}_i) \in [0,1],\ i = 1, \ldots, N\}. \tag{2}$$

Main Idea

The SUPG method is provably stable and convergent as $h \to 0$, however oscillations are still present and a discrete maximum principle does not hold.

Solution

Seek u_h in the convex subset K_h of $\mathbb{V}$ by restricting the nodal values of function in $\mathbb{V}$, that is

$$K_h := \{v_h \in \mathbb{V} : v_h(\boldsymbol{x}_i) \in [0,1],\ i = 1, \dots, N\}. \tag{2}$$

We find u_h via projection onto this convex set, that is, u_h is the element of K_h such that

$$a(u_h, v_h - u_h) \geq l(v_h - u_h) \quad \forall v_h \in K_h. \tag{3}$$

Main Idea

The SUPG method is provably stable and convergent as $h \to 0$, however oscillations are still present and a discrete maximum principle does not hold.

Solution

Seek u_h in the convex subset K_h of $\mathbb{V}$ by restricting the nodal values of function in $\mathbb{V}$, that is

$$K_h := \{v_h \in \mathbb{V} : v_h(\boldsymbol{x}_i) \in [0,1],\ i = 1, \dots, N\}. \tag{2}$$

We find u_h via projection onto this convex set, that is, u_h is the element of K_h such that

$$a(u_h, v_h - u_h) \geq l(v_h - u_h) \quad \forall v_h \in K_h. \tag{3}$$

Remark

We essentially solve a discrete 'obstacle' problem where we constrain the value of the discrete solution at degrees of freedom.

The choice of K_h is *consistent* in the sense that the bounds are the same as those satisfied by the PDE solution.

Nodally Bound-Preserving FEM

- Use of discrete variational inequalities suggested by Chang & Nakshatrala 2017 and Kirby & Shapero 2024.
- Bound-preserving stabilised FEM introduced by Barrenechea et al 2024 for elliptic problems. Solution shown to satisfy a variational inequality.

Nodally Bound-Preserving FEM

- Use of discrete variational inequalities suggested by Chang & Nakshatrala 2017 and Kirby & Shapero 2024.
- Bound-preserving stabilised FEM introduced by Barrenechea et al 2024 for elliptic problems. Solution shown to satisfy a variational inequality.

Desired Properties

- Satisfies bounds (nodally at least, see below)
- Satisfies the same optimal approximation properties as the regular FE solution

Nodally Bound-Preserving FEM

- Use of discrete variational inequalities suggested by Chang & Nakshatrala 2017 and Kirby & Shapero 2024.
- Bound-preserving stabilised FEM introduced by Barrenechea et al 2024 for elliptic problems. Solution shown to satisfy a variational inequality.

Desired Properties

- Satisfies bounds (nodally at least, see below)
- Satisfies the same optimal approximation properties as the regular FE solution

Remark

For polynomial degree 1, K_h consists of precisely the finite element functions which satisfy upper and lower bounds pointwise due to the bound-preserving properties of linear interpolation. For polynomial degree 2 or higher, this is NOT the case, and functions are nodally bound-preserving only.

Error Analysis

It remains to show that it satisfies optimal error estimates.

Lemma (consistency)

For any $v_h \in \mathbb{V}$,

$$l_h(v_h) - a_h(u, v_h) = 0. \tag{4}$$

We note that, in general, this is not true in the finite element approximation of variational inequalities. The natural norm for this problem is

$$|||w^2||| := \mu \|w\|^2_{\mathrm{L}^2(\Omega)} + \sum_{K \in \mathscr{T}} \left\| \delta_K^{\frac{1}{2}} \boldsymbol{b} \cdot \nabla w \right\|^2_{\mathrm{L}^2(K)} + |w|^2_{\Gamma^+} \tag{5}$$

Lemma (continuity & coercivity)

$$a_h(w, v) \leq C |||w|||_* |||v|||, \quad a_h(w, w) \geq \frac{1}{2} |||w|||^2. \tag{6}$$

Error Analysis

Theorem (Best Approximation)

There exists a constant $C > 0$ independent of h such that

$$|||u - u_h||| \leq C \inf_{v_h \in K_h} |||u - v_h|||_*. \tag{7}$$

Error Analysis

Theorem (Best Approximation)

There exists a constant $C > 0$ independent of h such that

$$|||u - u_h||| \leq C \inf_{v_h \in K_h} |||u - v_h|||_*. \tag{7}$$

- Instead of the usual Galerkin orthogonality, we use consistency to show

$$a_h(u - u_h, u_h - v_h) \geq 0. \tag{8}$$

Error Analysis

Theorem (Best Approximation)

There exists a constant $C > 0$ independent of h such that

$$|||u - u_h||| \leq C \inf_{v_h \in K_h} |||u - v_h|||_*. \tag{7}$$

- Instead of the usual Galerkin orthogonality, we use consistency to show

$$a_h(u - u_h, u_h - v_h) \geq 0. \tag{8}$$

- Combining the above with continuity and coercivity in the usual way gives the result.

Error Analysis

Theorem (Best Approximation)

There exists a constant $C > 0$ independent of h such that

$$|||u - u_h||| \leq C \inf_{v_h \in K_h} |||u - v_h|||_*. \tag{7}$$

- Instead of the usual Galerkin orthogonality, we use consistency to show

$$a_h(u - u_h, u_h - v_h) \geq 0. \tag{8}$$

- Combining the above with continuity and coercivity in the usual way gives the result.
- The result is somewhere between Céa's lemma and the best approximation result in Falk 1974, where extra terms appear relating to the approximability of the Ritz projection in the convex set.

Error Analysis

Since the Lagrange interpolant of u is in the set K_h, we can invoke interpolation estimates and obtain convergence rates.

Theorem (Convergence Rates)

Let $\boldsymbol{b} \in \mathrm{W}^{1,\infty}(\Omega)$, $c \in \mathrm{L}^\infty(\Omega)$, and let u be the unique solution of (1), with $u_h \in K_h$ the finite element solution. Let $k \geq 1$, and assume that $u \in \mathrm{H}^r(\Omega)$, where $r > \frac{d}{2}$ is sufficiently large so that u is regular enough to belong to the domain of the Lagrange interpolation operator. Suppose that $C_\delta > 0$ is sufficiently small so that

$$\delta_K := C_\delta h_K \leq \frac{\mu}{\|c\|^2_{\mathrm{L}^\infty(K)}}.$$

Then there exists a constant $C > 0$ independent of h such that

$$|||u - u_h||| \leq C h^{\min\{k+1,r\}-\frac{1}{2}} |u|_{\mathrm{H}^r(\Omega)}. \tag{9}$$

So we get the usual SUPG rate of $k + \frac{1}{2}$.

Linear Examples

Let $\Omega = (0,1) \times (0,1)$, and let $\boldsymbol{b}_1 = (1, \sqrt{2})$, so that the inflow boundary is

$$\Gamma^+ = \{(x,y) \in \partial\Omega : x = 0\} \cup \{(x,y) \in \partial\Omega : y = 0\}.$$

Smooth inflow boundary data is prescribed:

$$g_1(x,y) := \begin{cases} \exp\left(1 - \frac{1}{1-5(x-\frac{1}{2})^2}\right) & \text{if } |x - \frac{1}{2}| < \frac{1}{\sqrt{5}}, \\ 0 & \text{otherwise.} \end{cases}$$

And non-smooth:

$$g_2(x,y) := \begin{cases} 1 & \text{if } |x - \frac{1}{2}| < \frac{1}{\sqrt{5}}, \\ 0 & \text{otherwise.} \end{cases}$$

Convergence to Smooth and Non-Smooth solutions

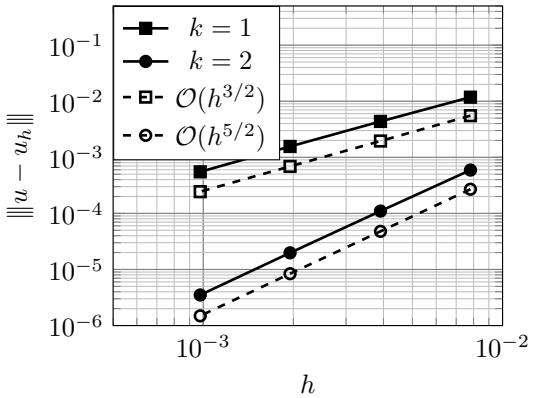


Figure: Optimal rates for $k = 1, 2$.

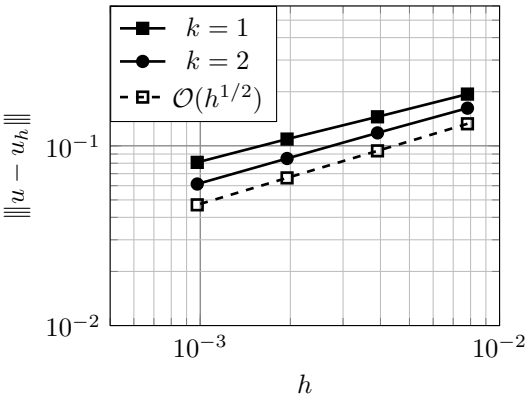


Figure: Suboptimal Convergence

Pure Advection Problem

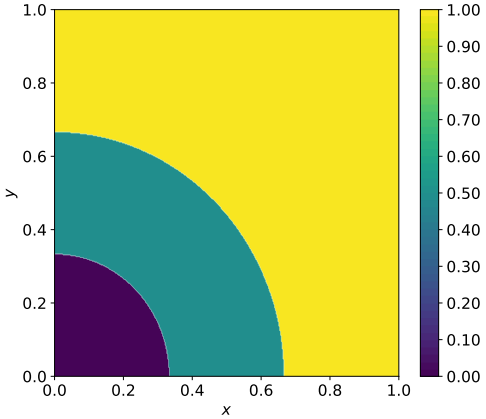


Figure: Contour plot of piecewise constant solution

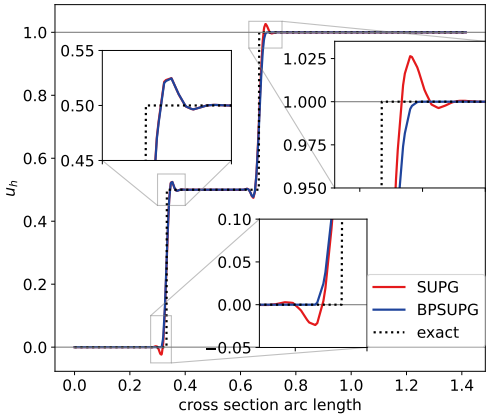


Figure: Diagonal cross-section of numerical solution

A Nonlinear Example

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + |u|^{p-2} u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma_-, \end{aligned} \tag{10}$$

where $1 < p \leq 2$.

A Nonlinear Example

$$\begin{aligned} \boldsymbol{b} \cdot \nabla u + |u|^{p-2} u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma_-, \end{aligned} \tag{10}$$

where $1 < p \leq 2$. The weak form is to find $u \in \mathrm{H}_{-,p}(\Omega)$ such that

$$a(u, v) + b(u; u, v) = l(v) \quad \forall v \in \mathrm{L}^2(\Omega), \tag{11}$$

where

$$b(u; v, w) = \int_\Omega |u|^{p-2} vw \, \mathrm{d}x. \tag{12}$$

Finite Element Discretisation

The bound-preserving finite element method is to find $u_h \in K_h$ such that

$$a_h(u_h, v_h - u_h) + b(u_h; u_h, v_h - u_h) \geq l_h(v_h - u_h) \quad \forall v_h \in K_h. \tag{13}$$

Note that we treat the advection term only with SUPG stabilisation, giving an inconsistent method. This time we have:

Lemma (lack of consistency)

$$l_h(w_h) - a_h(u, w_h) - b(u; u, w_h) = \sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \left(\boldsymbol{b} \cdot \nabla w_h\right). \tag{14}$$

Quasi-Norm

Problems such as this are often analysed using quasi-norms to achieve optimal convergence rates.
Introduce the notation, for fixed $w \in \mathrm{L}^p(\Omega)$,

$$\|v\|^2_{(w,p)} := \int_\Omega |v|^2 \left(|v| + |w|\right)^{p-2} \mathrm{d}x, \tag{15}$$

for all $v \in \mathrm{L}^p(\Omega)$.
In this quasi-norm, the form $b(\cdot;\cdot,\cdot)$ satisfies monotonicity and boundedness properties that take the place of the usual continuity and coercivity in the error analysis.

Main Result

Theorem

Assume that $\operatorname{div} \boldsymbol{b} = 0$, that u is the exact solution, and that u_h is the finite element solution. Then

$$|||u - u_h|||^2 + \|u - u_h\|_{(u,p)}^2 \leq C \inf_{v_h \in K_h} \left(\|u - v_h\|_{(u,p)}^2 + |||u - v_h|||_*^2 \right) + \sup_{0 \neq w_h \in \mathbb{V}} \frac{\sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \, (\boldsymbol{b} \cdot \nabla w_h)}{|||w_h|||}. \quad (16)$$

Main Result

Theorem

Assume that $\operatorname{div} \boldsymbol{b} = 0$, that u is the exact solution, and that u_h is the finite element solution. Then

$$|||u - u_h|||^2 + \|u - u_h\|_{(u,p)}^2 \leq C \inf_{v_h \in K_h} \left(\|u - v_h\|_{(u,p)}^2 + |||u - v_h|||_*^2 \right) + \sup_{0 \neq w_h \in \mathbb{V}} \frac{\sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \, (\boldsymbol{b} \cdot \nabla w_h)}{|||w_h|||}. \quad (16)$$

- The extra term results from the inconsistency of the finite element method.

Main Result

Theorem

Assume that $\operatorname{div} \boldsymbol{b} = 0$, that u is the exact solution, and that u_h is the finite element solution. Then

$$|||u - u_h|||^2 + \|u - u_h\|_{(u,p)}^2 \leq C \inf_{v_h \in K_h} \left(\|u - v_h\|_{(u,p)}^2 + |||u - v_h|||_*^2 \right) + \sup_{0 \neq w_h \in \mathbb{V}} \frac{\sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \, (\boldsymbol{b} \cdot \nabla w_h)}{|||w_h|||}. \quad (16)$$

- The extra term results from the inconsistency of the finite element method.
- The achievable convergence rate now requires specific choices of the SUPG weights δ_K due to the inconsistency.

Main Result

Theorem

Assume that $\operatorname{div} \boldsymbol{b} = 0$, that u is the exact solution, and that u_h is the finite element solution. Then

$$|||u - u_h|||^2 + \|u - u_h\|_{(u,p)}^2 \leq C \inf_{v_h \in K_h} \left(\|u - v_h\|_{(u,p)}^2 + |||u - v_h|||_*^2 \right) + \sup_{0 \neq w_h \in \mathbb{V}} \frac{\sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \, (\boldsymbol{b} \cdot \nabla w_h)}{|||w_h|||}. \quad (16)$$

- The extra term results from the inconsistency of the finite element method.
- The achievable convergence rate now requires specific choices of the SUPG weights δ_K due to the inconsistency.
- Nonlinearity poses challenges for the analysis if the consistent analogue is used.

Main Result

Theorem

Assume that $\operatorname{div} \boldsymbol{b} = 0$, that u is the exact solution, and that u_h is the finite element solution. Then

$$\|\!|u - u_h|\!\|^2 + \|u - u_h\|_{(u,p)}^2 \leq C \inf_{v_h \in K_h} \left(\|u - v_h\|_{(u,p)}^2 + \|\!|u - v_h|\!\|_*^2 \right) + \sup_{0 \neq w_h \in \mathbb{V}} \frac{\sum_{K \in \mathscr{T}} \delta_K \int_K |u|^{p-2} u \, (\boldsymbol{b} \cdot \nabla w_h)}{\|\!|w_h|\!\|}. \quad (16)$$

- The extra term results from the inconsistency of the finite element method.
- The achievable convergence rate now requires specific choices of the SUPG weights δ_K due to the inconsistency.
- Nonlinearity poses challenges for the analysis if the consistent analogue is used.
- Note that the norm $\|\!|\cdot|\!\|$ is now weaker (it no longer has the L^2 component), resulting in slightly weaker error control.

Nonlinear Reaction

- We set $p = \frac{3}{2}$
- Smooth boundary data is prescribed
- An exact solution can be found using the method of characteristics, which is piecewise smooth, and can be shown to be in $\mathrm{H}^2(\Omega)$ (in general the maximum regularity one can expect from this problem, even for smooth boundary data).
- In the numerical solver, the Jacobian (which is singular around $u = 0$) had a small amount of regularisation added to improve numerical performance.

Nonlinear Reaction

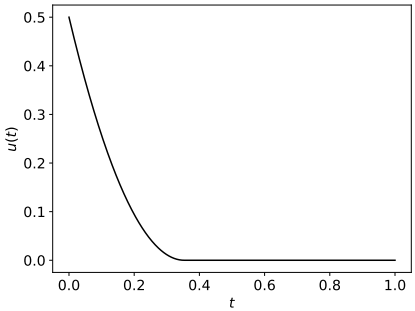


Figure: Plot of exact solution along a characteristic

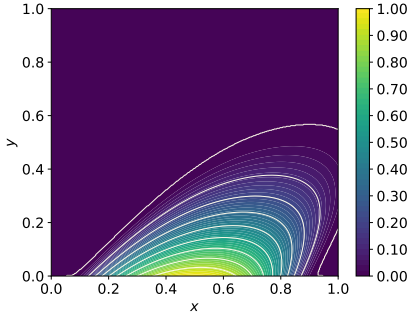


Figure: Contour plot of numerical solution

Problem with Nonlinear Reaction Term

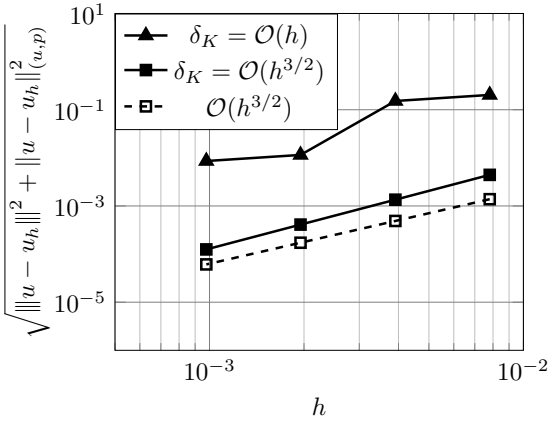


Figure: Approximation error, solution in $\mathrm{H}^2(\Omega)$

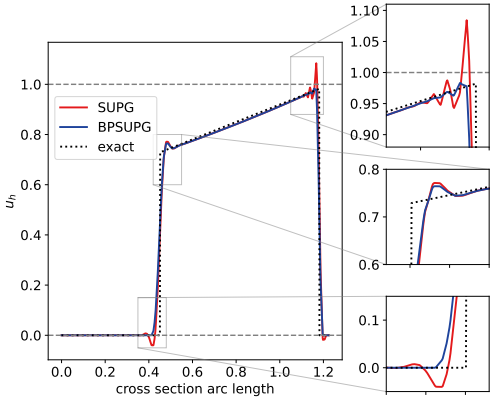


Figure: Diagonal cross-section of numerical solution

Application: Simple Model of Proton Transport

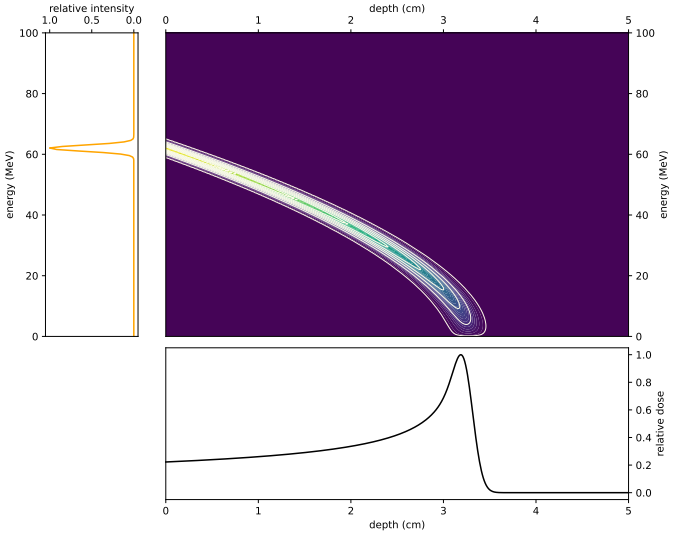
A simplified model of proton transport is given by

$$\boldsymbol{\omega} \cdot \nabla_{\boldsymbol{x}} \psi(\boldsymbol{x}, E) - \frac{\partial}{\partial E}(S(E)\psi(\boldsymbol{x}, E)) - \epsilon \Delta_{\boldsymbol{\omega}} \psi(\boldsymbol{x}, E) = f \tag{17}$$

This is solved with the nodally bound-preserving method presented here.

- ψ is the proton fluence
- Physical dose (the quantity of interest for practitioners) is calculated from fluence by integrating over energy.
- Efficient proton transport and dose computations are an integral part of treatment planning in Proton Beam Therapy (PBT).
- It is crucial that the physics of the problem are appropriately represented in numerical models (e.g. can't have negative absorbed dose).
- The angular Laplacian approximates Coulomb scattering, while the stopping power term approximates energy loss from ionisation.

Application: Proton Transport Computations



Application: Proton Transport Computations

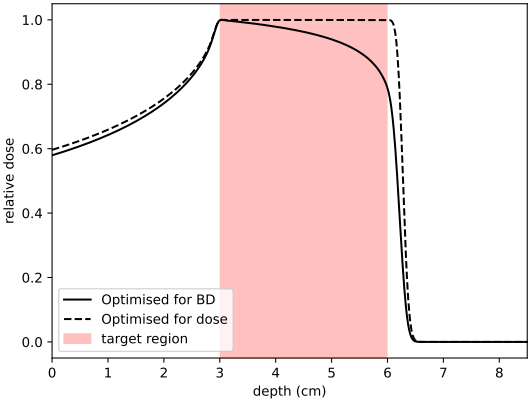


Figure: Aim to produce conforming dose profiles using combinations of Bragg peaks. Method must be efficient enough to evaluate many different solutions as part of an optimisation routine.

Application: Proton Transport Computations

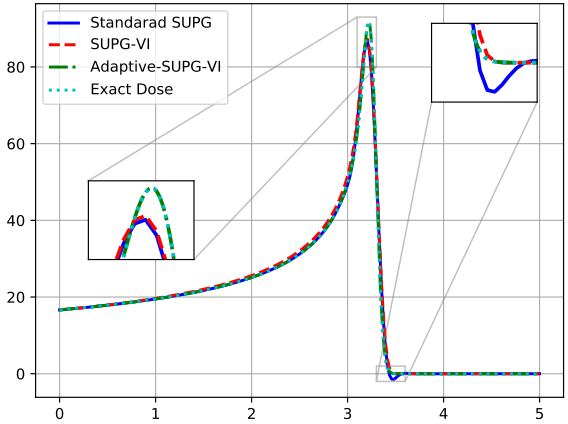


Figure: Comparison of dose curves solved with conventional SUPG and the bound-preserving version.

Conclusions

- The framework presented here is flexible, and applicable to a broad range of problems satisfying a variational formulation.
- For Piecewise (bi)linear finite element methods, bounds on the solution are satisfied. For higher order methods, bounds are preserved at the degrees of freedom.
- In the linear case, the same convergence rates are achieved as the standard SUPG method.
- Similar rates are achievable in the nonlinear case as long as the SUPG weights are chosen correctly.
- Efficient iterative algorithms are available in many computational packages (e.g. the simulations here used firedrake with PETSc). The cost of solving the variational inequality (i.e. replacing a linear problem with a nonlinear one) is smaller than one might expect