ELSEVIER

Comput. Methods Appl. Mech. Engrg. 191 (2002) 4641–4659

Computer methods in applied mechanics and engineering

www.elsevier.com/locate/cma

# On fully discrete schemes for the Fermi pencil-beam equation

Mohammad Asadzadeh [a,*], Alexandros Sopasakis [b,1]

[a] *Department of Mathematics, Chalmers University of Technology and Göteborg University, SE-412 96 Göteborg, Sweden*
[b] *Department of Mathematics, University of California Berkeley, Berkeley, CA 94720, USA*

Received 25 May 2001; received in revised form 27 March 2002; accepted 24 May 2002

**Abstract**

We consider a Fermi pencil-beam model in two-space dimensions $(x, y)$, where $x$ is aligned with the beam's penetration direction and $y$ together with the scaled angular variable $z$ correspond to a, bounded symmetric, transversal cross-section. The model corresponds to a forward–backward degenerate, convection dominated, convection–diffusion problem. For this problem we study some fully discrete numerical schemes using the standard- and Petrov–Galerkin finite element methods, for discretizations of the transversal domain, combined with the backward Euler, Crank–Nicolson, and discontinuous Galerkin methods for discretizations in the penetration variable. We derive stability estimates for the semi-discrete problems. Further, assuming sufficiently smooth exact solution, we obtain optimal a priori error bounds in a triple norm. These estimates give rise to a priori error estimates in the $L^2$-norm. Numerical implementations presented for some examples with the data approximating Dirac $\delta$ function, confirm the expected performance of the combined schemes.


*Keywords:* Fermi equation; Pencil beam; Standard Galerkin; Semi-streamline diffusion; Fully discrete schemes; Convergence rate

## 1. Introduction

The Fermi pencil-beam equation is derived from the Fokker–Planck equation through an asymptotic expansion. The Fokker–Planck equation itself is yet another asymptotic limit of the linear Boltzmann equation, see [6]. Both asymptotic derivations relay on the assumption of *forward-peaked scattering* in a transport process.

In this work we focus on approximation techniques and study some fully discrete schemes for the numerical solution of a pencil-beam model in two-space dimensions. Introducing a scaled angular variable $z$ our model problem would correspond to a forward–backward, degenerate type, convection dominated, convection–diffusion problem in a slab of thickness $L$, $x \in I_x := [0, L]$, with a symmetric cross-section

* Corresponding author. Fax: +49-31-16-1973.
*E-mail addresses:* mohammad@math.chalmers.se (M. Asadzadeh), sopasak@math.berkeley.edu (A. Sopasakis).
[1] Supported by TMR, EU contract no. ERB FMRX-CT97-0157.

0045-7825/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0045-7825(02)00397-3

$I_\perp := I_y \times I_z := [-y_0, y_0] \times [-z_0, z_0]$, where $y_0,\ z_0 \in \mathbb{R}^+$. Thus the physical domain, $I_x \times I_\perp$, is now three-dimensional and the corresponding Fermi equation is modelling the penetration (in the direction of the $x$-axis) of narrowly focused pencil-beam particles, incident at the transversal boundary of an *isotropic* slab, entering into the domain at the origin $(x, y, z) = (0, 0, 0)$. Because of the forward-peakedness assumption the radiation scattering from the beam particles would cover bounded transversal intersections which, for simplicity, are assumed to be convex. Further, for the isotropic background media we may assume that all involved functions are symmetric, i.e., they are *even* in $y$ and $z$.

In this setting, our model problem is thus formulated as follows: given the incident source intensity $f$ at $x = 0$, find the current $u$ defined on the domain $\Omega := I_x \times I_\perp$ satisfying the Fermi pencil-beam equation

$$\begin{cases} u_x + z u_y = \varepsilon u_{zz} & \text{in } \Omega = I_x \times I_\perp, \\ u_z(x, y, \pm z_0) = 0 & \text{for } (x, y) \in I_x \times I_y, \\ u(0, x_\perp) = f(x_\perp) & \text{for } x_\perp \in I_\perp, \\ u(x, x_\perp) = 0 & \text{on } \Gamma^-_{\tilde\beta} \setminus \{(0, x_\perp)\}, \end{cases} \tag{1.1}$$

where $\Gamma^-_{\tilde\beta} = \{(x, x_\perp) \in \Gamma := \partial\Omega,\ \tilde{\mathbf{n}} \cdot \tilde{\boldsymbol\beta} < 0\}$ is the inflow boundary with respect to $\tilde{\boldsymbol\beta} := (1, z, 0)$, ($z = \tan\theta$, $-\pi/2 < \theta < \pi/2$, corresponds to scaled angular variable), $x_\perp := (y, z)$ and $2\varepsilon = \sigma_{\mathrm{tr}}(x, y)$. Here $\sigma_{\mathrm{tr}}(x, y)$ is called the transport cross-section which is a positive small and decreasing function of $(x, y)$ and corresponds to the deposit of energy due to particle collisions. Finally $\tilde{\mathbf{n}} := \tilde{\mathbf{n}}(x, x_\perp)$ is the outward unit normal to $\Gamma$ at $(x, x_\perp) \in \Gamma$. This problem corresponds to a forward–backward (depending on sign of $z$ in (1.1)) convection dominated (small $\varepsilon$) convection–diffusion problem which can be interpreted as a time-dependent (with $x$ corresponding to the time variable) degenerate type (convection in $y$, diffusion in $z$) problem.

For the convection dominated problems having hyperbolic nature, (assuming that the exact solution in the Sobolev space $H^{k+1}$), the standard finite element schemes with a quasi-uniform triangulation and a mesh size $h$, would have a convergence of order $\mathcal{O}(h^k)$, versus $\mathcal{O}(h^{k+1})$ for elliptic and parabolic problems, see [10]. These estimates are in some modified $L_2$-norms associated with the weak formulation of the problem. The idea of including artificial viscosity term, e.g., by adding some amount of diffusion in the equation, is to create smoothing effects improving the poor behaviour of the standard Galerkin (SG) method for hyperbolic type problems, see, e.g., [8]. Here, using a semi-streamline diffusion (SSD) method through a modified form of the test functions, we can automatically add a proper amount of *spatial viscosity* in the $y$-direction. If we could add the same amount of viscosity in the $z$-direction then we would have an improved convergence rate by $\mathcal{O}(h^{1/2})$. However, because of the assumption of forward peaked scattering in angle and energy, creating more amount of diffusion in the $z$-direction is unphysical. Therefore $\varepsilon$ will be kept in the range $h^2 \leqslant \varepsilon \leqslant h$, and the optimality of our final $L_2$-error estimates are stated in this context. Note that the SSD method is performed only on $x_\perp$, whereas the usual streamline diffusion (SD) finite element method is applied also on the $x$ variable. This improves the convergence rate in the triple norm by $\mathcal{O}(h^{1/2})$, see, e.g., [2,8,10]. However, in the $L^2$-estimates, because of the absence of an absorption term in the equation, the relation between $\varepsilon$ and the mesh size $h$: $\varepsilon \geqslant h^2$, would cause to a reduced convergence rate by $\mathcal{O}(\varepsilon^{1/2}) \sim \mathcal{O}(h)$.

Some related studies of the Fokker–Planck and Fermi pencil-beam models can be found in [2–4]. In [2] a priori error estimates are derived for a fully discrete problem using the usual SD and discontinuous Galerkin (DG) finite element methods, while [3] is devoted to the a posteriori error estimates in the same setting. The *characteristic* methods, based on the technique of exact transport + projections are considered in [4], see also [11].

Below first we study the semi-discrete schemes where we discretize in the transversal variable $x_\perp =: (y, z)$, using the SG and SSD finite element methods with *weakly imposed boundary conditions*, and derive some stability estimates. The convergence results in the semi-discrete part are based on Galerkin orthogonality and strong stability estimates derived for certain bilinear forms. As for the fully discrete problem: because

of the structure of the equation the penetrating variable $x$ is interpreted as a time variable and treated by usual time discretization techniques such as: DG, backward Euler or Crank–Nicolson methods.

One of the basic applications of the Fermi pencil-beam models is in the *dose* calculations of the radiative cancer therapy, see [9]. Our computational results, in concrete examples, indicate the reliability of different numerical schemes proposed in this context. We present fast and efficient, deterministic, schemes competitive with the commonly used stochastic algorithms and derive convergence rates and stability estimates.

An outline of this note is as follows: we start by presenting the semi-discrete approximation by the SG finite element method in Section 2, prove a stability result for this type of discretization in Section 2.1 and derive the error estimates in Section 2.2. The corresponding investigations using the SSD approximation are presented in Section 3. Section 4 is devoted to fully discrete algorithms. Numerical simulations, for some relevant examples, together with the study of the behaviour of either discretization algorithms are introduced in Section 5. Finally in Section 6 we comment the numerical results. Throughout the paper $C$ will denote an absolute constant unless otherwise explicitly stated.

## 2. The standard Galerkin method

In this section we discretize in $x_\perp = (y,z)$ using a finite element approximation based on quasi-uniform triangulation of the rectangular domain $I_\perp = I_y \times I_z$ with a mesh size $h$. To this approach we let $\beta = (z,0)$ and define the inflow (outflow) boundary as

$$\Gamma_\beta^{-(+)} := \{x_\perp \in \Gamma := \partial I_\perp : \mathbf{n}(x_\perp)\cdot\beta < 0 \ (>0)\}, \tag{2.1}$$

where $\mathbf{n}(x_\perp)$ is the outward unit normal to the boundary $\Gamma$ at $x_\perp \in \Gamma$. Now we introduce a discrete finite dimensional function space $V_{h,\beta} \subset H^1_\beta(I_\perp)$ with,

$$H^1_\beta(I_\perp) = \left\{ v \in H^1(I_\perp) : v = 0 \text{ on } \Gamma_\beta^- \setminus \{(0,x_\perp)\} \right\}, \tag{2.2}$$

such that, $\forall v \in H^1_\beta(I_\perp) \cap H^r(I_\perp)$,

$$\inf_{\chi\in V_{h,\beta}} \|v-\chi\|_j \leqslant Ch^{\alpha-j}\|v\|_\alpha, \quad j = 0,1 \text{ and } 1 \leqslant \alpha \leqslant r, \tag{2.3}$$

where for positive integer $s$, $\|\cdot\|_s$ denotes the $L_2$-based Sobolev norm of functions with all their partial derivatives of order $\leqslant s$ in $L_2$, see Adams [1]. An example of such $V_{h,\beta}$ is the set of sufficiently smooth piecewise polynomials $P(x_\perp)$ of degree $\leqslant r$, satisfying the boundary condition given in (2.2).

To proceed we introduce a bilinear form, $A : H^1_\beta(I_\perp) \times H^1_\beta(I_\perp)$, defined by

$$A(u,v) = (u_x, v)_\perp + (zu_y, v)_\perp, \quad \forall u, v \in H^1_\beta(I_\perp), \tag{2.4}$$

then the continuous variational problem is: find a solution $u$ to (1.1) such that

$$\begin{cases} A(u,\chi)_\perp + (\varepsilon u_z, \chi_z)_\perp = 0 \quad \forall \chi \in H^1_\beta(I_\perp), \\ u(0,x_\perp) = f(x_\perp). \end{cases} \tag{2.5}$$

Let $\tilde{u} \in V_{h,\beta}$ be an auxiliary interpolant of the solution $u$ of (1.1) defined by

$$A(u-\tilde{u},\chi) = 0 \quad \forall \chi \in V_{h,\beta}. \tag{2.6}$$

Now the objective is to solve the following discrete variational problem: find $u_h \in V_{h,\beta}$, such that

$$\begin{cases} (u_{h,x},\chi)_\perp + (zu_{h,y},\chi)_\perp + (\varepsilon u_{h,z},\chi_z)_\perp = 0 \quad \forall \chi \in V_{h,\beta}, \\ u_h(0,x_\perp) = f_h(x_\perp), \end{cases} \tag{2.7}$$

where $f_h$ is assumed to be a finite element approximation of $f$ which coincides with the interpolant $\tilde{u}(0,x_\perp)$ of $u(0,x_\perp)$. Here, $(u,v)_\perp = \int_{I_\perp} u(x_\perp)v(x_\perp)\,\mathrm{d}x_\perp$ and $\|u\|_{L_2(I_\perp)} = (u,u)_\perp^{1/2}$. To distinguish, we use the following inner products notations: $(\cdot,\cdot)_\perp$ and $(\cdot,\cdot)_\Omega$, where $\Omega = [0,L]\times I_\perp := I_x \times I_\perp$, for integrations over $I_\perp$ and $I_x \times I_\perp$, respectively. Finally, we assume that the mesh size $h$ is related to $\varepsilon$ according to:

$$h^2 \leqslant \varepsilon \leqslant h.$$

### 2.1. Stability

In this part we prove a stability lemma in both inner products, $(\cdot,\cdot)_\perp$ and $(\cdot,\cdot)_\Omega$, to guarantee the control of both continuous and discrete solutions by the data. For simplicity we introduce the triple norm,

$$|||v|||_{\tilde{\beta}}^2 = \frac{1}{2}\int_{\Gamma_{\tilde{\beta}}^+} v^2(\mathbf{n}\cdot\tilde{\beta})\,\mathrm{d}\Gamma + \|\varepsilon^{1/2}v_z\|_{L_2(\Omega)}^2, \tag{2.8}$$

where $\tilde{\beta} = (1,\beta)$ and $\Gamma_{\tilde{\beta}}^+ := \Gamma \setminus \Gamma_{\tilde{\beta}}^- = [0,L]\times\Gamma_\beta^+ \cup \{\{L\}\times I_\perp\}$.

**Lemma 2.1.** *For u satisfying* (1.1) *we have that*,

$$\sup_{x\in I_x}\|u(x,\cdot)\|_{L_2(I_\perp)} \leqslant \|f\|_{L_2(I_\perp)}, \tag{2.9}$$

$$|||u|||_{\tilde{\beta}}^2 = \tfrac{1}{2}\|u(0,\cdot)\|_{L_2(I_\perp)}^2. \tag{2.10}$$

**Proof.** We let $\chi = u$, in the first equation, in (2.5). Using (2.4) we obtain,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u\|_{L_2(I_\perp)}^2 + (zu_y,u)_\perp + \|\varepsilon^{1/2}u_z\|_{L_2(I_\perp)}^2 = 0. \tag{2.11}$$

Using integration by parts, in $y$, we may write

$$(zu_y,u)_\perp = \frac{1}{2}\int_{I_z} z\big(u^2(y_0) - u^2(-y_0)\big)\,\mathrm{d}z = \frac{1}{2}\int_{\Gamma_\beta^+}(\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma, \tag{2.12}$$

which, inserting in (2.11), gives that,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u\|_{L_2(I_\perp)}^2 + \frac{1}{2}\int_{\Gamma_\beta^+}(\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma + \|\varepsilon^{1/2}u_z\|_{L_2(I_\perp)}^2 = 0. \tag{2.13}$$

Now, since $\int_{\Gamma_\beta^+}(\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma \geqslant 0$, by (2.13), $(\mathrm{d}/\mathrm{d}x)\|u\|_{L_2(I_\perp)}^2 \leqslant 0$, i.e., $\|u\|_{L_2(I_\perp)}^2$ is decreasing in $x$ and hence,

$$\|u(x,\cdot)\|_{L_2(I_\perp)} \leqslant \|u(x',\cdot)\|_{L_2(I_\perp)}, \quad 0\leqslant x'\leqslant x\leqslant L \tag{2.14}$$

and thus

$$\|u(x,\cdot)\|_{L_2(I_\perp)} \leqslant \|f\|_{L_2(I_\perp)}, \quad \forall x\in[0,L]. \tag{2.15}$$

This gives the first statement of the lemma. Integrating (2.13) over $x\in[0,L]$ we get,

$$\frac{1}{2}\|u(L,\cdot)\|_{L_2(I_\perp)}^2 + \frac{1}{2}\int_0^L\int_{\Gamma_\beta^+}(\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma + \|\varepsilon^{1/2}u_z\|_{L_2(\Omega)}^2 = \tfrac{1}{2}\|u(0,\cdot)\|_{L_2(I_\perp)}^2. \tag{2.16}$$

Observe that the first two terms above add up to $(1/2)\int_{\Gamma_{\tilde{\beta}}^+}(\mathbf{n}\cdot\tilde{\beta})u^2\,\mathrm{d}\Gamma$, so that we obtain the second assertion of the lemma and the proof is complete. □

By the same argument as in Lemma 2.1 we obtain the semi-discrete version of the stability estimate for SG problem:

**Corollary 2.1.** *The solution $u_h$ of problem* (2.7) *satisfies the stability relations*,

$$\sup_{x\in I_x}\|u_h(x,\cdot)\|_{L_2(I_\perp)}\leqslant\|u_h(0,\cdot)\|_{L_2(I_\perp)}, \tag{2.17}$$

$$|||u_h|||_{\tilde\beta}^2=\tfrac{1}{2}\|u_h(0,\cdot)\|_{L_2(I_\perp)}^2. \tag{2.18}$$

### 2.2. Convergence

In this part we state and prove convergence rates, both in the $L_2$-norm and in the triple norm, for the SG-method for the semi-discrete problem with weakly imposed boundary conditions. Our main results are Lemma 2.2 and Theorem 2.1 below. For the hyperbolic problems with an absorption term of $\mathcal{O}(1)$, and $u\in H^r(\Omega)$, the optimal convergence rate for the SG in the $L_2$-norm is $\mathcal{O}(h^{r-1})$. Our equation, although degenerate, is not purely hyperbolic: the diffusive term in $z$ on the right hand side corresponds to add of artificial viscosity, of order $\mathcal{O}(\varepsilon)$, in the $z$-direction. This improves the triple norm estimate by $\mathcal{O}(\sqrt{\varepsilon})\sim\mathcal{O}(\sqrt{h})$. However, turning to the $L_2$-norm estimate because of the lack of absorption term, using Poincare inequality, and due to $h^2\leqslant\varepsilon\leqslant h$, we get a factor of $\mathcal{O}(\varepsilon^{-1/2})\sim\mathcal{O}(h^{-1})$ on the right hand side of the $L_2$-estimate, which corresponds to lose of an accuracy of $\varepsilon^{-1/2}\sim h^{-1}$. We have shown these phenomena in Lemma 2.2 and Theorem 2.1 below.

**Lemma 2.2** (error estimate in the triple norm). *Assume that u and $u_h$ satisfy* (1.1) *and* (2.7), *respectively. Let $u\in H^r(\Omega)$, $r\geqslant 2$, then there is a constant C such that*,

$$|||u_h-u|||_{\tilde\beta}\leqslant Ch^{r-1/2}\|u\|_r. \tag{2.19}$$

**Proof.** By adding first equation in (2.7) and (2.6) we get, using (2.5), that

$$\begin{aligned}((u_h-\tilde u)_x,\chi)_\perp+(z(u_h-\tilde u)_y,\chi)_\perp+(\varepsilon(u_h-\tilde u)_z,\chi_z)_\perp&=-(u_x,\chi)_\perp-(zu_y,\chi)_\perp-(\varepsilon u_z,\chi_z)_\perp+(\varepsilon(u-\tilde u)_z,\chi_z)_\perp\\&=0+(\varepsilon(u-\tilde u)_z,\chi_z)_\perp.\end{aligned}$$

Let now $\chi=u_h-\tilde u$, then using the same argument as in the stability estimate we may write,

$$\begin{aligned}&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u_h-\tilde u\|_{L_2(I_\perp)}^2+\frac{1}{2}\int_{\Gamma_{\tilde\beta}^+}(\mathbf{n}\cdot\beta)(u_h-\tilde u)^2\,\mathrm{d}\Gamma+\|\varepsilon^{1/2}(u_h-\tilde u)_z\|_{L_2(I_\perp)}^2\\&\quad\leqslant\tfrac{1}{2}\|\varepsilon^{1/2}(u_h-\tilde u)_z\|_{L_2(I_\perp)}^2+\tfrac{1}{2}\|\varepsilon^{1/2}(u-\tilde u)_z\|_{L_2(I_\perp)}^2,\end{aligned}$$

or equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}x}\|u_h-\tilde u\|_{L_2(I_\perp)}^2+\int_{\Gamma_{\tilde\beta}^+}(\mathbf{n}\cdot\beta)(u_h-\tilde u)^2\,\mathrm{d}\Gamma+\|\varepsilon^{1/2}(u_h-\tilde u)_z\|_{L_2(I_\perp)}^2\leqslant\|\varepsilon^{1/2}(u-\tilde u)_z\|_{L_2(I_\perp)}^2.$$

Now integrating over $x\in[0,L]$, implies that

$$\begin{aligned}&\|(u_h-\tilde u)(L,\cdot)\|_{L_2(I_\perp)}^2+\int_{\Gamma_{\tilde\beta}^+\setminus\Gamma_L}(\tilde{\mathbf{n}}\cdot\tilde\beta)(u_h-\tilde u)^2\,\mathrm{d}\Gamma+\|\varepsilon^{1/2}(u_h-\tilde u)_z\|_{L_2(\Omega)}^2\\&\quad\leqslant\|\varepsilon^{1/2}(u-\tilde u)_z\|_{L_2(\Omega)}^2+\|(u_h-\tilde u)(0,\cdot)\|_{L_2(I_\perp)}^2,\end{aligned}$$

where $\Gamma_L = \{\{L\} \times I_\perp\}$. Thus recalling $u_h(0,\cdot) = \tilde{u}(0,\cdot) = f_h$, and the definition of the $|||\cdot|||_{\tilde{\beta}}$ norm we have

$$|||u_h - \tilde{u}|||^2_{\tilde{\beta}} \leqslant \|\varepsilon^{1/2}(u-\tilde{u})_z\|^2_{L_2(\Omega)}.$$

Writing $u_h - u = (u_h - \tilde{u}) + (\tilde{u} - u)$, the desired result follows from the following interpolation estimate: □

**Proposition 2.1.** *Let $h^2 \leqslant \varepsilon(x,y) \leqslant h$, then there is a constant $\tilde{C}$ such that,*

$$|||u - \tilde{u}|||_{\tilde{\beta}} \leqslant \tilde{C} h^{r-1/2} \|u\|_r. \tag{2.20}$$

**Proof.** The proof is based on classical interpolation error estimates, see [5], [7] and [10]: Let $u \in H^r(\Omega)$, then there exists an interpolant $\tilde{u} \in V_{h,\beta}$, of $u$ and interpolation constants $C_1$ and $C_2$ such that

$$\|u - \tilde{u}\|_s \leqslant C_1 h^{r-s}\|u\|_r, \quad s = 0, 1,$$
$$|u - \tilde{u}|_{\tilde{\beta}} \leqslant C_2 h^{r-1/2}\|u\|_r,$$

where

$$|\varphi|_{\tilde{\beta}} = \left( \int_{\Gamma^+_{\tilde{\beta}}} \varphi^2 (\tilde{\mathbf{n}} \cdot \tilde{\beta})\, \mathrm{d}\Gamma \right)^{1/2}.$$

Now recalling the definition of $|||\cdot|||_{\tilde{\beta}}$ we have,

$$\begin{aligned} |||u - \tilde{u}|||^2_{\tilde{\beta}} &= \tfrac{1}{2}|u - \tilde{u}|^2_{\tilde{\beta}} + \|\varepsilon^{1/2}(u-\tilde{u})_z\|^2_{L_2(\Omega)} \\ &\leqslant \tfrac{1}{2}|u - \tilde{u}|^2_{\tilde{\beta}} + \|\varepsilon^{1/2}\|^2_{L_\infty(\Omega)} \|(u-\tilde{u})_z\|^2_{L_2(\Omega)} \|u - \tilde{u}\|^2_{H^1(\Omega)} \leqslant \tfrac{1}{2} C_2^2 h^{2r-1}\|u\|^2_r \\ &\leqslant \tfrac{1}{2}|u - \tilde{u}|^2_{\tilde{\beta}} + (\sup_{I_x \times I_y} \varepsilon) + C_1^2 \tilde{\varepsilon} h^{2r-2}\|u\|^2_r \leqslant C h^{2r-1}\|u\|^2_r, \end{aligned}$$

where in the last step we used $\tilde{\varepsilon} := \sup \varepsilon \leqslant h$ and $C = \max(C_1^2, C_2^2/2)$. Letting now $\tilde{C} = C^{1/2}$ the proof is complete. □

From this result we now obtain the desired estimate in the $L_2$-norm:

**Theorem 2.1.** *For $u \in H^r(\Omega)$, satisfying* (2.7) *and with $u_h$ being the solution of* (2.7), *there is a constant $C = C(\Omega, f)$ such that*

$$\|u - u_h\|_{L_2(\Omega)} \leqslant C h^{r-3/2}\|u\|_r. \tag{2.21}$$

**Proof.** By a simple application of the Poincare inequality

$$\|u - u_h\|_{L_2} \leqslant C\|(u-u_h)_z\|_{L_2} \tag{2.22}$$

and using Lemma 2.2 we have

$$\|\varepsilon^{1/2}(u-u_h)_z\|_{L_2} \leqslant C h^{r-1/2}\|u\|_r. \tag{2.23}$$

Thus, since $\varepsilon \geqslant h^2$, we obtain (2.21). □

Observe that in $C = C(\Omega, f)$, the $\Omega$ dependence is because of $\varepsilon = \varepsilon(x, y)$ and the Poincare inequality, while the $f$ dependence comes from the assumed identity $u_h(0) := \tilde{u}(0) = f$.

## 3. A smoothing Petrov–Galerkin method

Below we introduce a SSD approach which includes a diffusion generating test function in the $y$-direction over the usual SG procedure. We show the strong stability of this scheme. Its smoothing properties are obvious, from the diffusivity (in $y$ and $z$), as seen by the numerical implementations in Section 5. However, the admissible size of, $\varepsilon(x, y)$ is very small, this implies unrealistically fine degree of numerical resolution. In the numerical examples we have chosen $\varepsilon$ in the borderline of a numerically realistic and physically admissible value. So at the end the dominant smallness parameter is $\varepsilon$ not the (comparably) large diffusion that the method adds in the $y$-direction. We shall not carry out convergence analysis of this scheme. Instead we refer to [2] for the complete analysis in the general SD case.

The degenerate character of the problem (1.1) contributes to the anisotropic nature of the diffusion. Using the SSD scheme we obtain an equation with somewhat improved regularity in the $y$-direction. More precisely the SSD test functions having the form $v + \delta v_\beta$ automatically add an extra diffusion term, $\delta(v_\beta, v_\beta)$, to the variational formulation, which combined with $(v, -\varepsilon v_{zz}) = (\varepsilon v_z, v_z)$ term gives a, non-degenerate, *weakly*, diffusive equation, ($x$ is interpreted as the time variable), with a full diffusion of order $\mathrm{O}(\varepsilon)$, $h^2 \leqslant \varepsilon \leqslant h$, (while we have assumed $\delta \sim h$).

Below we derive stability estimates for the continuous problem based on the SSD variational formulation. The corresponding discrete version is obtained in a similar way and therefore omitted. As we mentioned above the modified test function has the form: $v + \delta v_\beta$ with $\delta \geqslant \varepsilon$, $\beta = (z, 0)$, $v_\beta = \beta \cdot \nabla_\perp v$ and $\nabla_\perp = (\partial/\partial y, \partial/\partial z)$, and $v$ satisfying the boundary conditions in (1.1). Multiplying the differential equation in (1.1) by $v + \delta v_\beta$ and integrating over $I_\perp$ yields,

$$\begin{aligned}(u_x + u_\beta - \varepsilon u_{zz}, v + \delta v_\beta)_\perp &= (u_x, v)_\perp + \delta(u_x, v_\beta)_\perp + (u_\beta, v)_\perp + \delta(u_\beta, v_\beta)_\perp \\ &\quad + (\varepsilon u_z, v_z)_\perp + \delta(\varepsilon u_z, (v_\beta)_z)_\perp = 0. \end{aligned} \tag{3.1}$$

To derive the basic stability estimate we let $v = u$ in (3.1) and get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u\|_\perp^2 + \delta(u_x, u_\beta)_\perp + \frac{1}{2}\int_{\Gamma_\beta^+} (\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma + \delta\|u_\beta\|_\perp^2 + \|\varepsilon^{1/2}u_z\|_\perp^2 + \delta(\varepsilon u_z, (u_\beta)_z)_\perp = 0. \tag{3.2}$$

The inner product in the last term can be written as,

$$(\varepsilon u_z, (zu_y)_z)_\perp = (\varepsilon u_z, zu_{yz})_\perp + (\varepsilon u_z, u_y)_\perp = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}y}\left(\int_{I_\perp} \varepsilon z u_z^2\,\mathrm{d}y\,\mathrm{d}z\right) - \frac{1}{2}\int_{I_\perp} \varepsilon_y z u_z^2\,\mathrm{d}y\,\mathrm{d}z + (\varepsilon u_z, u_y)_\perp.$$

Now since by the symmetry assumption $u$ is even in $y$ and $z$, the integrands above are odd functions in $z$, and their integrals over the symmetric interval $I_z$ are identically zero. Hence (3.2) can be written as:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u\|_\perp^2 + \delta(u_x, u_\beta)_\perp + \frac{1}{2}\int_{\Gamma_\beta^+} (\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma + \delta\|u_\beta\|_\perp^2 + \|\varepsilon^{1/2}u_z\|_\perp^2 + \delta(\varepsilon u_z, u_y)_\perp = 0. \tag{3.3}$$

Now we multiply the differential equation in (1.1) by $\delta u_x$, integrate over $I_\perp$ and perform an integration by parts to obtain

$$\delta\|u_x\|_\perp^2 + \delta(u_x, u_\beta)_\perp + \delta(\varepsilon u_z, u_{xz})_\perp = 0. \tag{3.4}$$

Note that,

$$(\varepsilon u_z, u_{xz})_\perp = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\int_{I_\perp} \varepsilon u_z^2 \,\mathrm{d}x_\perp - \frac{1}{2}\int_{I_\perp} \varepsilon_x u_z^2 \,\mathrm{d}x_\perp. \tag{3.5}$$

Adding (3.3) and (3.4) and using (3.5) we have,

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\|u\|_\perp^2 + \delta\|u_x + u_\beta\|_\perp^2 + \frac{1}{2}\int_{\Gamma_\beta^+} (\mathbf{n}\cdot\beta)u^2 \,\mathrm{d}\Gamma + \|\varepsilon^{1/2}u_z\|_\perp^2 + \delta(\varepsilon u_z, u_y)_\perp \\ + \frac{\delta}{2}\frac{\mathrm{d}}{\mathrm{d}x}\int_{I_\perp} \varepsilon u_z^2 \,\mathrm{d}x_\perp - \frac{\delta}{2}\int_{I_\perp} \varepsilon_x u_z^2 \,\mathrm{d}x_\perp = 0.\end{aligned} \tag{3.6}$$

We shall also use the following trivial inequality,

$$|(\varepsilon u_z, u_y)_\perp| \leqslant \tfrac{1}{2}\|\varepsilon^{1/2}u_z\|_\perp^2 + \tfrac{1}{2}\|\varepsilon^{1/2}u_y\|_\perp^2. \tag{3.7}$$

Now we make an additional symmetry assumption on the transversal plane viz.,

$$\|\varepsilon^{1/2}u_y\|_\perp \sim \|\varepsilon^{1/2}u_z\|_\perp. \tag{3.8}$$

Observe that $\varepsilon = (1/2)\sigma_{\mathrm{tr}}(x,y) \sim 1/l$, where $l$, the mean distance between two successive collisions, is an increasing function of $x$ and $y$. The justification of this phenomenon lies on the fact that we have a model starting with dense collisions which gradually in the penetration direction $x$, towards to the end, i.e., on leaving the physical domain, transfers to a particle distribution with rarefied character. (Note that, because of the lack of absorption term in the equation, a problem with an increasing $\varepsilon$ would be unstable.) Thus, as we have assumed, $\varepsilon$ is decreasing and $\varepsilon_x \leqslant 0$, hence

$$\int_{I_\perp} \varepsilon_x u_z^2 \,\mathrm{d}x_\perp \leqslant 0. \tag{3.9}$$

Inserting (3.7)–(3.9) in (3.6) we get,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\left(\|u\|_\perp^2 + \delta\int_{I_\perp} \varepsilon u_z^2 \,\mathrm{d}x_\perp\right) + \frac{1}{2}\int_{\Gamma_\beta^+} u^2(\mathbf{n}\cdot\beta)\,\mathrm{d}\Gamma + \delta\|u_x + u_\beta\|_\perp^2 + (1-\delta)\|\varepsilon^{1/2}u_z\|_\perp^2 \leqslant 0. \tag{3.10}$$

Thus for sufficiently small $\delta \sim \sqrt{\varepsilon} \ll 1$, (actually $\delta < 1$ would suffice)

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\|u\|_\perp^2 + \delta\int_{I_\perp} \varepsilon u_z^2 \,\mathrm{d}x_\perp\right) < 0 \tag{3.11}$$

and hence, $(\|u\|_\perp^2 + \delta\int_{I_\perp} \varepsilon u_z^2 \,\mathrm{d}x_\perp)$ is strictly decreasing in $x$. As a consequence, we have $\forall x' \in [0, L]$,

$$\|u(x',\cdot)\|_{L_2(I_\perp)}^2 + \delta\|\varepsilon^{1/2}u_z(x',\cdot)\|_{L_2(I_\perp)}^2 \leqslant \|u(0,\cdot)\|_{L_2(I_\perp)}^2 + \delta\|\varepsilon^{1/2}u_z(0,\cdot)\|_{L_2(I_\perp)}^2, \tag{3.12}$$

which gives the first stability estimate for the continuous SSD method:

$$\|u(L,\cdot)\|_{L_2(I_\perp)}^2 + \delta\|\varepsilon^{1/2}u_z(L,\cdot)\|_{L_2(I_\perp)}^2 \leqslant \|f\|_{L_2(I_\perp)}^2 + \delta\|\varepsilon^{1/2}f_z\|_{L_2(I_\perp)}^2 \tag{3.13}$$

and also, integrating over $x' \in [0, L]$, we get the second stability estimate:

**Lemma 3.1.** *Assuming* (3.8) and with $\delta < 1$ *we have the stability estimate*

$$|||u|||_{\tilde{\beta}}^2 + \delta\|u_x + u_\beta\|_{L_2(\Omega)}^2 \leqslant \overline{C}\left(\|f\|_{L_2(I_\perp)}^2 + \delta\|\varepsilon^{1/2}f_z\|_{L_2(I_\perp)}^2\right). \tag{3.14}$$

**Remark.** Note that in deriving (3.14) from (3.10)–(3.13) we get the $\delta$-terms: $\delta\int_{I_\perp}\varepsilon u_z^2(L,x_\perp)\,\mathrm{d}x_\perp = \delta\|\varepsilon^{1/2}u_z(L,\cdot)\|^2_{L_2(I_\perp)}$ and $(1-\delta)\|\varepsilon^{1/2}u_z\|^2_{L_2(I_x\times I_\perp)}$ adding up to $\sim\|\varepsilon^{1/2}u_z\|^2_{L_2(\Omega)}$, which is included in $|||u|||^2_{\tilde\beta}$. Further,

$$\frac{1}{2}\|u(L,\cdot)\|^2_{L_2(\Omega)}+\frac{1}{2}\int_0^L\int_{\Gamma^+_\beta}(\mathbf{n}\cdot\beta)u^2\,\mathrm{d}\Gamma=\frac{1}{2}\int_{\Gamma^+_{\tilde\beta}}(\mathbf{n}\cdot\tilde\beta)u^2\,\mathrm{d}\Gamma, \tag{3.15}$$

is also included in $|||u|||^2_{\tilde\beta}$. Thus the assertion of Lemma 3.1, is simply,

$$\delta\|u_x+u_\beta\|^2_{L_2(\Omega)}+\|\varepsilon^{1/2}u_z\|^2_{L_2(\Omega)}+\frac{1}{2}\int_{\Gamma^+_{\tilde\beta}}(\mathbf{n}\cdot\tilde\beta)u^2\,\mathrm{d}\Gamma\leqslant\overline{C}\Big(\|f\|^2_{L_2(I_\perp)}+\delta\|\varepsilon^{1/2}f_z\|^2_{L_2(I_\perp)}\Big), \tag{3.16}$$

where we may take $\overline{C}\sim(1/2(1-\delta))\sim(1/2(1-h))<1/2$, for $h<1/4$. Comparing this estimate with the second assertion of Lemma 2.1, i.e., (2.10), we get,

$$\|u_x+u_\beta\|_{L_2(\Omega)}\leqslant\overline{C}\|\varepsilon^{1/2}f_z\|_{L_2(I_\perp)},\quad \overline{C}<1/2. \tag{3.17}$$

Using the equation this yields,

$$\|\varepsilon u_{zz}\|_{L_2(\Omega)}\leqslant\overline{C}\|\varepsilon^{1/2}f_z\|_{L_2(I_\perp)},\quad \overline{C}<1/2. \tag{3.18}$$

The estimate (3.18) states that if $\varepsilon=\mathcal{O}(1)$ then the solution is regularized in the sense that $f\in H^r_{\sqrt{\varepsilon}}(I_\perp)$ implies $u\in H^{r+1}_\varepsilon(\Omega)$. However, this is obviously affected by small $\varepsilon$ values and, in particular, distorted when $\varepsilon=(1/2)\sigma_{\mathrm{tr}}(x,y)\to 0$.

**Remark.** The discrete version is now obtained by replacing $u$ by a suitable $u_h$, having the desired approximation properties. The corresponding semi-discrete convergence analysis would slightly improve the results of Section 2 (Lemma 2.2, Proposition 2.1 and Theorem 2.1) by $\mathcal{O}(h^r)$ where $0\leqslant r<1/2$. We emphasise that, the general convergence studies for the SD method show that the suitable *compatibility relations* are $\delta\sim h$ and $\varepsilon\sim h$, whereas, as we mentioned earlier, the physical parameter $\varepsilon\sim h^2$. To be concise we skip deriving SSD convergence rates and refer the reader to related estimates in [2] and [10].

## 4. The fully discrete problem

In this section, we derive the algorithms corresponding to the SG and SSD schemes for $I_\perp$ combined with DG, backward-Euler (BE) and Crank–Nicolson (CN) methods for the penetration interval $I_x$. The approximation techniques in Sections 2 and 3 are designed for discretizations in the transversal variable $x_\perp=(y,z)$. We could include the penetration variable $x$, in this procedure as an additional space variable, as it is, see the analysis in [2], where full discretizations are made in all three variables using both the usual SD and the DG methods, see also [10,12]. However, in order to efficiently determine the beam intensity at different transversal cross-sections, discretization procedures for the penetration variable $x$ is treated separately and as a time variable, in similar time dependent problems. Thus, in extending our semi-discrete algorithms to a higher dimensional case containing also discretizations in $x$, we consider the time discretization schemes for $I_x$, such as DG, BE and CN.

We introduce the bilinear forms:

$$a(u,v):=(u_\beta,v)_\perp+\delta(u_\beta,v_\beta)_\perp+(\varepsilon u_z,v_z)_\perp+\delta(\varepsilon u_z,(v_\beta)_z)_\perp, \tag{4.1}$$

$$b(u,v):=\delta(u,v_\beta)_\perp+(u,v)_\perp \tag{4.2}$$

Table 1
Backward Euler

| | Dirac $e^*_{4h}$–$e^*_{2h}$ | Hyperbolic $e^*_{4h}$–$e^*_{2h}$ | Maxwellian $e^*_{4h}$–$e^*_{2h}$ | Cone $e^*_{4h}$–$e^*_{2h}$ |
|---|---|---|---|---|
| *Galerkin elements* | | | | |
| $L_2$ | 26.79–8.44 | 0.155–0.062 | 0.333–0.184 | 0.344–0.207 |
| $L_1$ | 27.92–10.28 | 0.182–0.091 | 0.434–0.267 | 0.442–0.293 |
| $L_\infty$ | 48.69–16.99 | 0.224–0.090 | 0.407–0.263 | 0.476–0.313 |
| $\tilde{L}_2$ | 13.63–1.806 | 0.064–0.013 | 0.123–0.042 | 0.115–0.047 |
| *Semi-streamline diffusion elements* | | | | |
| $L_2$ | 27.16–8.522 | 0.151–0.064 | 0.322–0.179 | 0.320–0.203 |
| $L_1$ | 27.75–9.676 | 0.184–0.095 | 0.420–0.261 | 0.441–0.287 |
| $L_\infty$ | 50.93–15.22 | 0.208–0.083 | 0.418–0.246 | 0.433–0.287 |
| $\tilde{L}_2$ | 13.33–1.801 | 0.063–0.014 | 0.118–0.041 | 0.110–0.045 |

Table 2
Crank–Nicolson

| | Dirac $e^*_{4h}$–$e^*_{2h}$ | Hyperbolic $e^*_{4h}$–$e^*_{2h}$ | Maxwellian $e^*_{4h}$–$e^*_{2h}$ | Cone $e^*_{4h}$–$e^*_{2h}$ |
|---|---|---|---|---|
| *Galerkin elements* | | | | |
| $L_2$ | 26.93–8.473 | 0.156–0.063 | 0.356–0.180 | 0.345–0.209 |
| $L_1$ | 27.99–10.32 | 0.183–0.091 | 0.454–0.257 | 0.442–0.295 |
| $L_\infty$ | 49.12–17.27 | 0.225–0.091 | 0.469–0.270 | 0.478–0.315 |
| $\tilde{L}_2$ | 13.73–1.814 | 0.065–0.014 | 0.122–0.041 | 0.115–0.047 |
| *Semi-streamline diffusion elements* | | | | |
| $L_2$ | 27.33–8.527 | 0.152–0.064 | 0.332–0.178 | 0.321–0.204 |
| $L_1$ | 27.80–9.697 | 0.185–0.095 | 0.426–0.254 | 0.415–0.288 |
| $L_\infty$ | 51.50–15.24 | 0.208–0.084 | 0.426–0.245 | 0.434–0.286 |
| $\tilde{L}_2$ | 13.44–1.806 | 0.063–0.015 | 0.117–0.040 | 0.110–0.045 |

Table 3
Discontinuous Galerkin

| | Dirac $e^*_{4h}$–$e^*_{2h}$ | Hyperbolic $e^*_{4h}$–$e^*_{2h}$ | Maxwellian $e^*_{4h}$–$e^*_{2h}$ | Cone $e^*_{4h}$–$e^*_{2h}$ |
|---|---|---|---|---|
| *Galerkin elements* | | | | |
| $L_2$ | 29.04–9.809 | 0.161–0.061 | 0.309–0.200 | 0.303–0.237 |
| $L_1$ | 29.75–11.56 | 0.202–0.088 | 0.428–0.275 | 0.420–0.311 |
| $L_\infty$ | 52.54–21.13 | 0.224–0.089 | 0.429–0.417 | 0.404–0.498 |
| $\tilde{L}_2$ | 13.40–2.065 | 0.064–0.012 | 0.117–0.043 | 0.110–0.051 |
| *Semi-streamline diffusion elements* | | | | |
| $L_2$ | 29.24–10.17 | 0.152–0.067 | 0.304–0.193 | 0.296–0.220 |
| $L_1$ | 30.19–10.51 | 0.188–0.090 | 0.437–0.247 | 0.428–0.264 |
| $L_\infty$ | 53.34–19.71 | 0.221–0.100 | 0.428–0.382 | 0.401–0.455 |
| $\tilde{L}_2$ | 13.28–2.068 | 0.063–0.014 | 0.117–0.042 | 0.110–0.049 |

and rewrite the problem (3.1) as finding a solution $u \in H^1_\beta(I_\perp)$ such that,

$$b(u_x, v) + a(u, v) = 0, \quad \forall v \in H^1_\beta(I_\perp). \tag{4.3}$$

We subsequently use the finite dimensional subspace $V_{h,\beta} \subset H^1_\beta(I_\perp)$ and represent the discrete solution $u_h$ by a separation of variables viz:

$$u_h(x, y, z) = \sum_{i=1}^{M} \xi_i(x)\phi_i(y, z), \tag{4.4}$$

where $M \sim 1/h$. Now we let $v = \phi_j$ for $j = 1, \ldots, M$, and insert (4.4) into the semi-discrete counterpart of (4.3) to obtain,

$$\sum_{i=1}^{M} \xi_i'(x) b(\phi_i, \phi_j) + \sum_{i=1}^{M} \xi_i(x) a(\phi_i, \phi_j) = 0, \quad j = 1, \ldots, M. \tag{4.5}$$

In the matrix form (4.5) may be represented by $B\xi'(x) + A\xi(x) = 0$, where $B = (b_{ij})$ with $b_{ij} = b(\phi_j, \phi_i)$ and $A = (a_{ij})$ with $a_{ij} = a(\phi_j, \phi_i)$. For small $\delta$ the matrix $B$, being positive definite, is invertible and therefore we can reformulate (4.5) as,

$$\xi'(x) + \overline{A}\xi(x) = 0, \tag{4.6}$$

where $\overline{A} = B^{-1}A$. However inverting $B$ is among other things expensive. Therefore we instead consider a Choleski decomposition of $B = E^{\mathrm{T}}E$, which leads to

$$\eta'(x) + \overline{A}\eta(x) = 0, \quad \eta(0) = \eta_0, \tag{4.7}$$

where now $\overline{A} = (E^{-1})^{\mathrm{T}} A E^{-1}$ and $\eta = E\xi$. The (stiff) solution of (4.7) is:

$$\eta(x) = \eta_0 \exp(-\overline{A}x). \tag{4.8}$$

The matrix equations presented in this section can now be easily implemented for usual finite element test functions (with $\delta = 0$). In this manner our algorithm can be used to compare the SG and SSD methods.

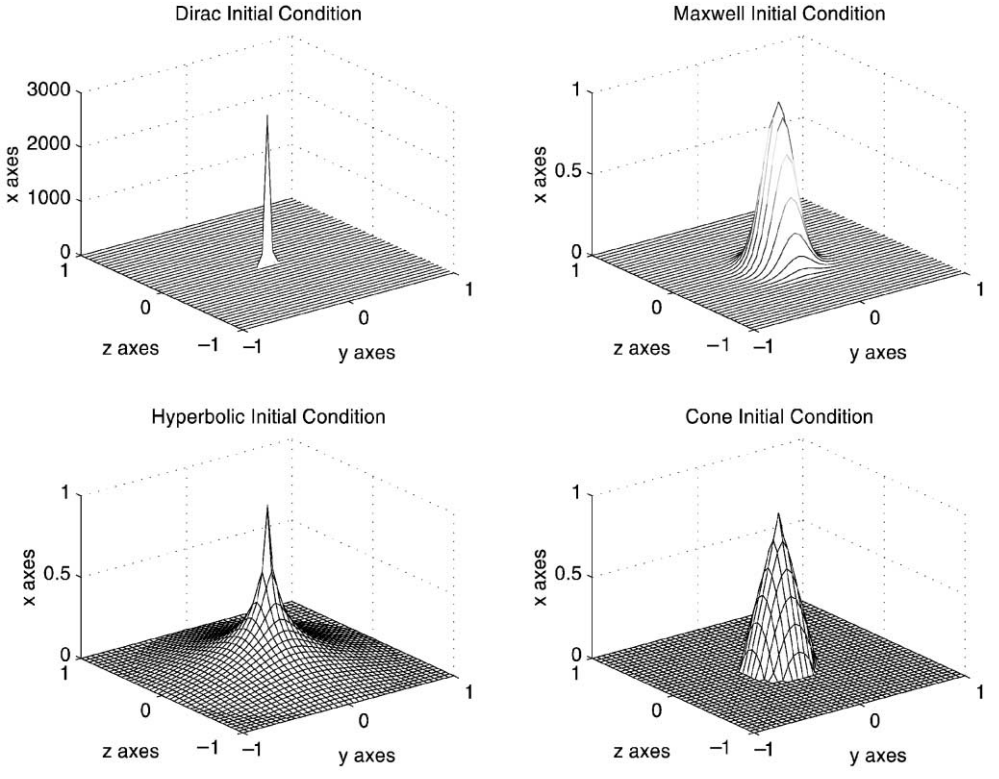


Fig. 1. Some of the initial conditions used.

A fully discrete scheme is obtained by also discretizing (4.6) in the $x$ variable. Below we combine both SG and SSD schemes, for discretization in $x_\perp$, with the most common time discretization techniques applied to our $x$ variable. To achieve the most general schemes for the $x$ discretization we extract them from Pade approximations of the form, $U^{n+1} = E_{\mu\nu}U^n$ for $n \geqslant 0$, where $E_{\mu\nu} = r_{\mu\nu}(\overline{A})$. Here, $r_{\mu\nu}(x) = \eta_{\mu\nu}(x)/d_{\mu\nu}(x)$ with,

$$\eta_{\mu\nu}(x) = \sum_{j=0}^{\nu} \frac{(\mu+\nu-j)!\nu!}{(\mu+\nu)!j!(\nu-j)!}(-x)^j!, \tag{4.9}$$

$$d_{\mu\nu}(x) = \sum_{j=0}^{\mu} \frac{(\mu+\nu-j)!\mu!}{(\mu+\nu)!j!(\mu-j)!}x^j. \tag{4.10}$$

For instance $r_{01}(x) = 1 - x$ corresponds to forward Euler, $r_{10}(x) = 1/(1+x)$ to backward Euler and $r_{11}$ to Crank–Nicolson scheme:

$$\left(\frac{U_h^n - U_h^{n-1}}{k_n}\right) + \overline{A}U_h^n = 0, \quad \text{backward Euler}, \tag{4.11}$$

$$\left(\frac{U_h^n - U_h^{n-1}}{k_n}\right) + \overline{A}\left(\frac{U_h^n + U_h^{n-1}}{2}\right) = 0, \quad \text{Crank–Nicolson}. \tag{4.12}$$

Other such choices will easily provide comparisons for alternative methods.

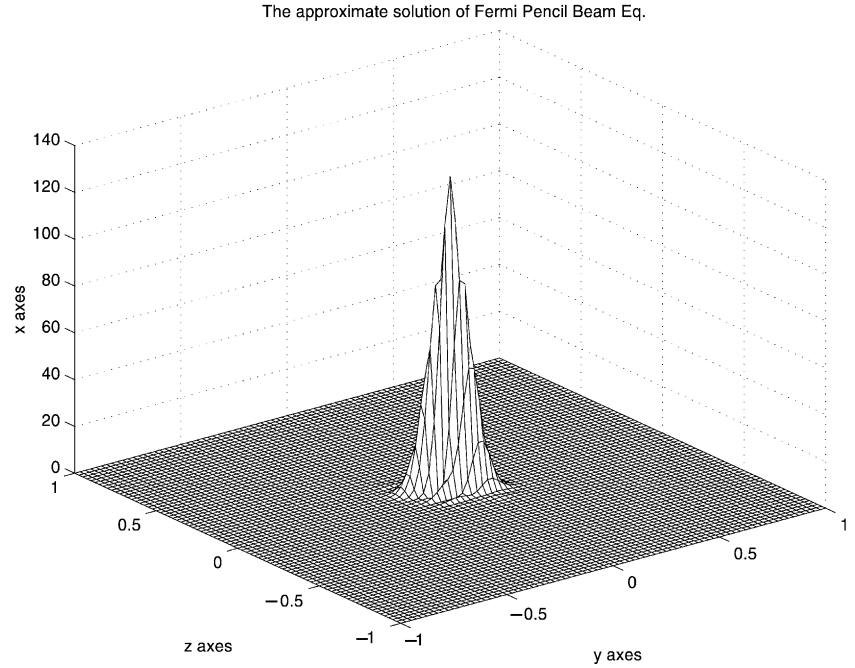


Fig. 2. An example of a solution. The Dirac initial condition is used with $\varepsilon = 0.002$, $h = 0.025$ and $k = 0.0005$.

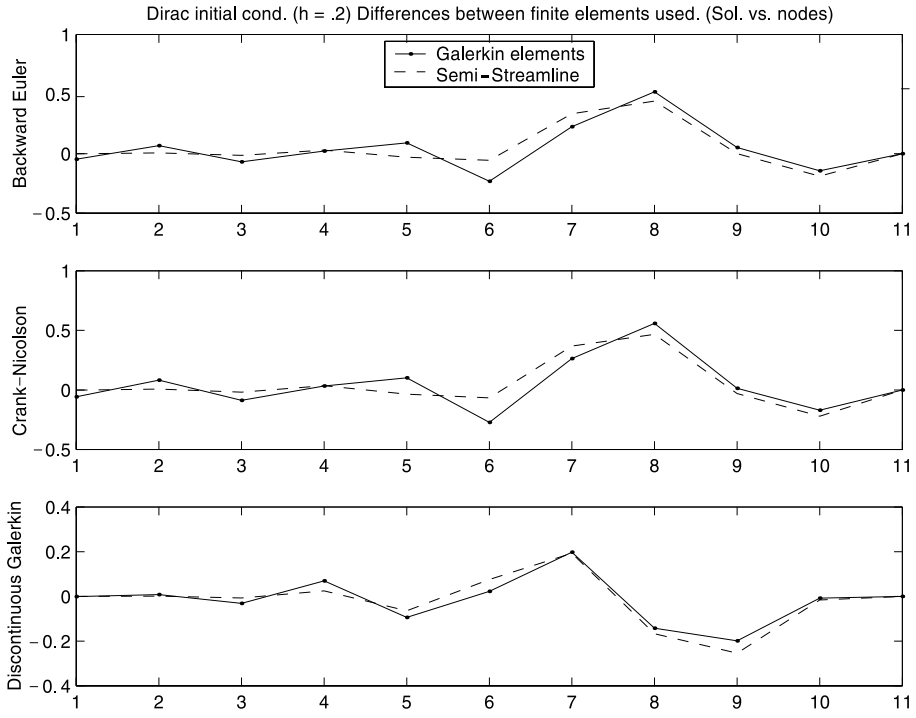


Fig. 3. Galerkin vs. semi-streamline elements for Dirac initial condition at $h = 0.2$ for the slice, $-1 \leqslant y \leqslant 1$ at $z = -0.9$.

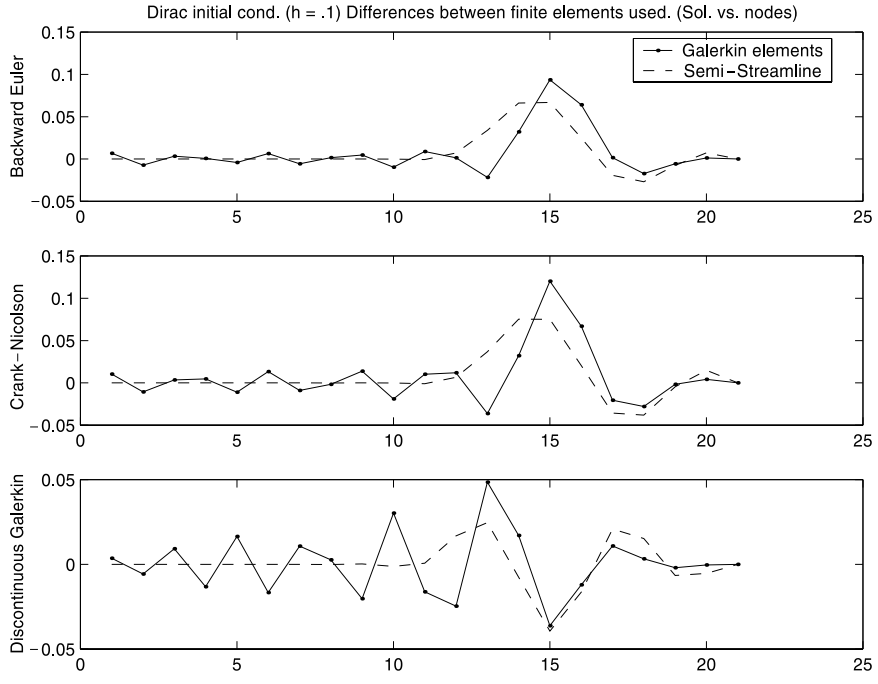


Fig. 4. Galerkin vs. semi-streamline elements for Dirac initial condition at $h = 0.1$ for the slice, $-1 \leqslant y \leqslant 1$ at $z = -0.9$.

## 5. Numerical examples

To justify the theoretical estimates of Sections 2 and 3 we present numerical examples testing the convergence rates of both the SG and SSD. The implementations are performed over four different initial

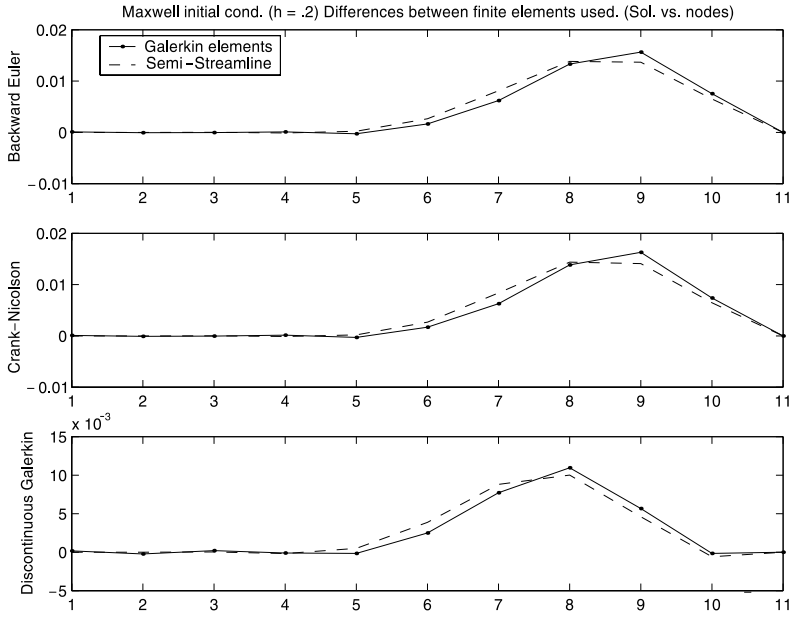


Fig. 5. Galerkin vs. semi-streamline elements for Maxwell initial condition at $h = 0.2$ for the slice, $-1 \leqslant y \leqslant 1$ at $z = -0.9$.

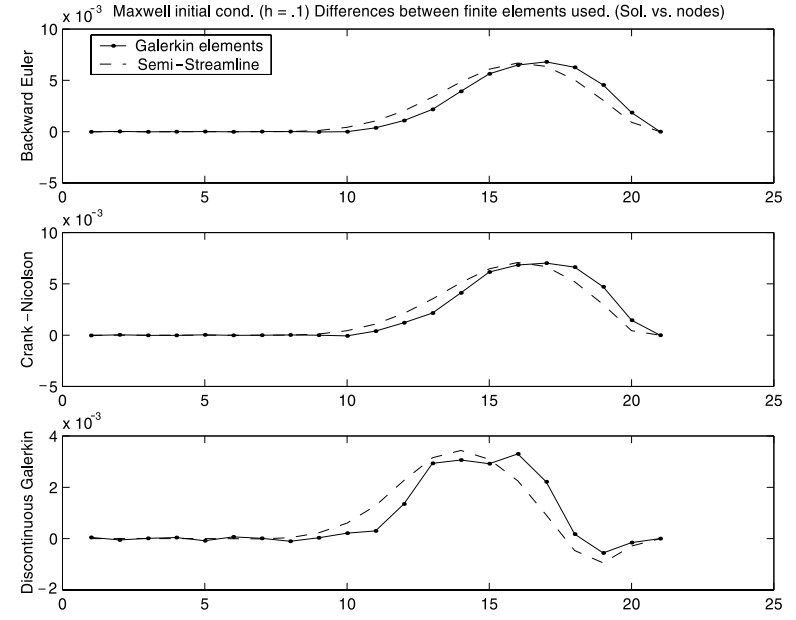


Fig. 6. Galerkin vs. semi-streamline elements for Maxwell initial condition at $h = 0.1$ for the slice, $-1 \leqslant y \leqslant 1$ at $z = -0.9$.

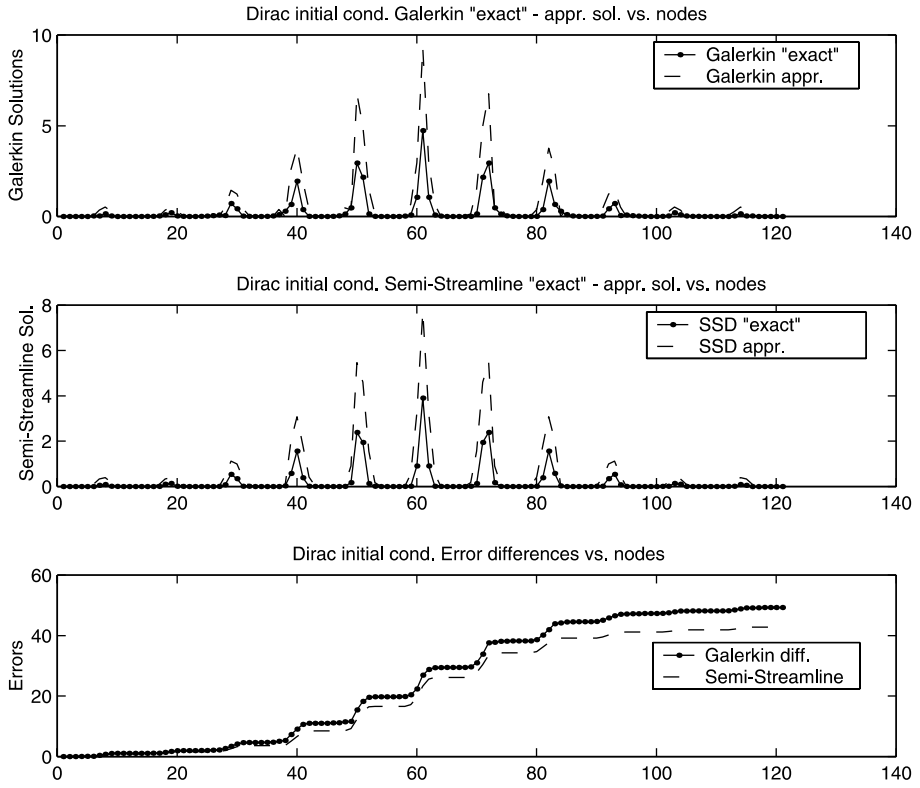


Fig. 7. $L_1$ cumulative error vs nodes for SG and SSD for Dirac. The solutions at $x = 1$ are found for $h = 0.2$.

conditions: cone-Maxwellian-hyperbolic- and modified Dirac-*impulses*, all approximating our data: the Dirac $\delta$ function.

We split the problem into two steps. First we discretize the two-dimensional domain $I_\perp = I_y \times I_z$ by means of continuous piecewise linear Galerkin approximation "cG(1)", and establish a mesh there in order to obtain a semi-discrete solution and subsequently we apply one of the three schemes, BE, CN or DG to step advance in the $x$-direction, (higher order elements could also be implemented in a similar way). Our cG(1) basis functions have the form, $\phi_i = a_1 y + a_2 z + a_3$.

In some special cases (for instance, for $\varepsilon = \varepsilon(x)$, see [9]) the closed form exact solution for (1.1) is available:

$$u(x,y,z) = \frac{\sqrt{3}}{\pi\varepsilon x^2} \exp(-2[3(y/x)^2 - 3(y/2)z + z^2]/(\varepsilon x)). \tag{5.1}$$

This allows us to draw some limited comparisons in terms of the actual error. In addition to being a limited case, (5.1) also displays singularities near the origin which, (although removable), makes it difficult to numerically implement as is. Obviously the final solution depends on initial conditions and therefore it is not correct to compare (5.1) with the solutions we obtain numerically since the underlying initial conditions were not the same to start with. For instance we can not numerically provide an initial data of the form of a Dirac $\delta$ function. We therefore use four different types of *computable initial conditions*, each approximating the Dirac $\delta$ function, in the $L_1$ sense, for comparison purposes. Through these examples we will also ascertain how strongly can differences in initial conditions affect our estimates on convergence established in

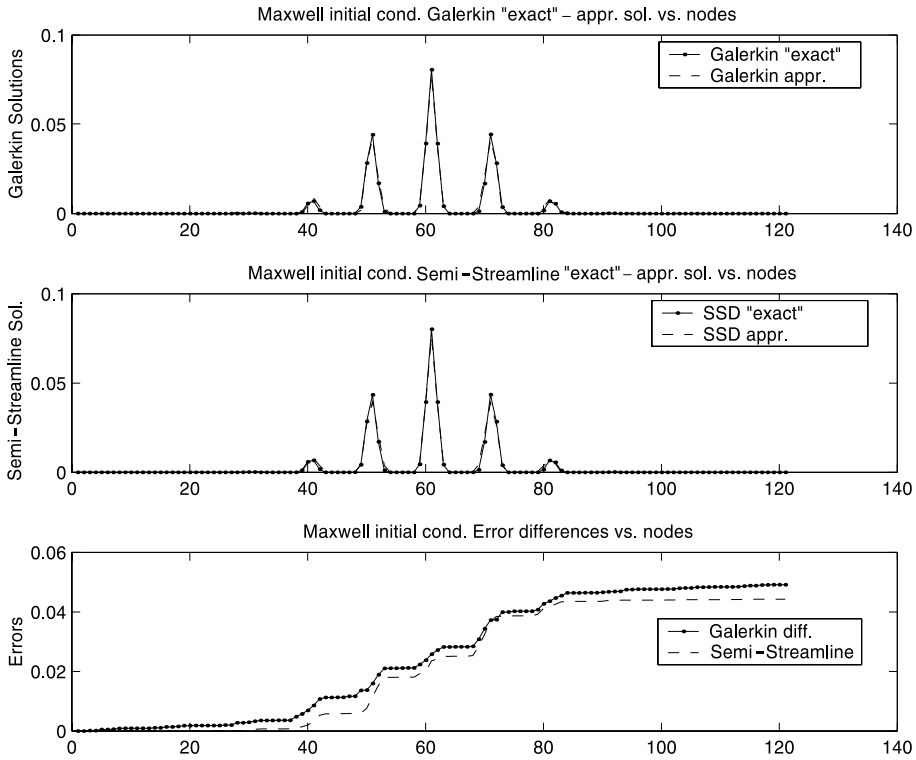


Fig. 8. $L_1$ cumulative error vs nodes for SG and SSD for Maxwell. The solutions at $x = 1$ are found for $h = 0.2$.

Sections 2 and 3. In the results that follow we see comparable convergence rates for solutions with corresponding initial data for both SG and SSD discretizations.

Tables 1–3 above summarise errors: the comparisons (differences) between solutions obtained from all the initial conditions used under both types of finite elements. We particularly calculate the error in $L_2$, $L_1$, $L_\infty$ and $\tilde{L}_2$-norms, where $\tilde{L}_2$ is a weighted $L_2$-norm defined by

$$\|\varphi\|_{\tilde{L}_2} = \left( \frac{1}{3} \sum_{\tau_i} |\tau_i| \sum_{j=1}^{3} (\varphi(\zeta_j^i))^2 \right)^{1/2}, \tag{5.2}$$

where $\tau_i$ are the mesh elements and $\zeta_j^i$ denote the midpoints of the edges of $\tau_i$.

The calculations are performed on an **Origin 2000** supercomputer with varying number of processors used at each running occasion. A total of almost 200 supercomputer hours were used for all the necessary computations. The finest mesh used (for at least some of the examples) for our domain $\Omega$ has a step size of $h = 0.025$ in the $y$ and $z$ variables and step size $k = 0.0005$ in the $x$ variable. This creates a total of 6,561,000 nodes (in three dimensions) and in particular 6561 ($= 81 \times 81$) nodes (in two dimensions) for each "exact" solution which will be used to calculate the norms in error estimates with solutions due to lesser number of nodes. The values in the tables are provided for $\varepsilon = 0.05$ and $\delta = 0.05$. These calculations are performed under all four types of considered initial conditions, for each case of finite elements, under all three types of time increment methods for three different mesh sizes. This gives a total of, $5 \times 2 \times 3 \times 3 = 90$ different solution evaluations. The accepted "exact" solution for these comparisons is denoted by $u^*$. We let $e_h^* = u^* - U_h$, where $U_h$ denotes the approximate solution for a mesh size $h$ on $(y, z)$. All solutions are provided for $x = 1$ and therefore the norms are calculated at this value of $x$.

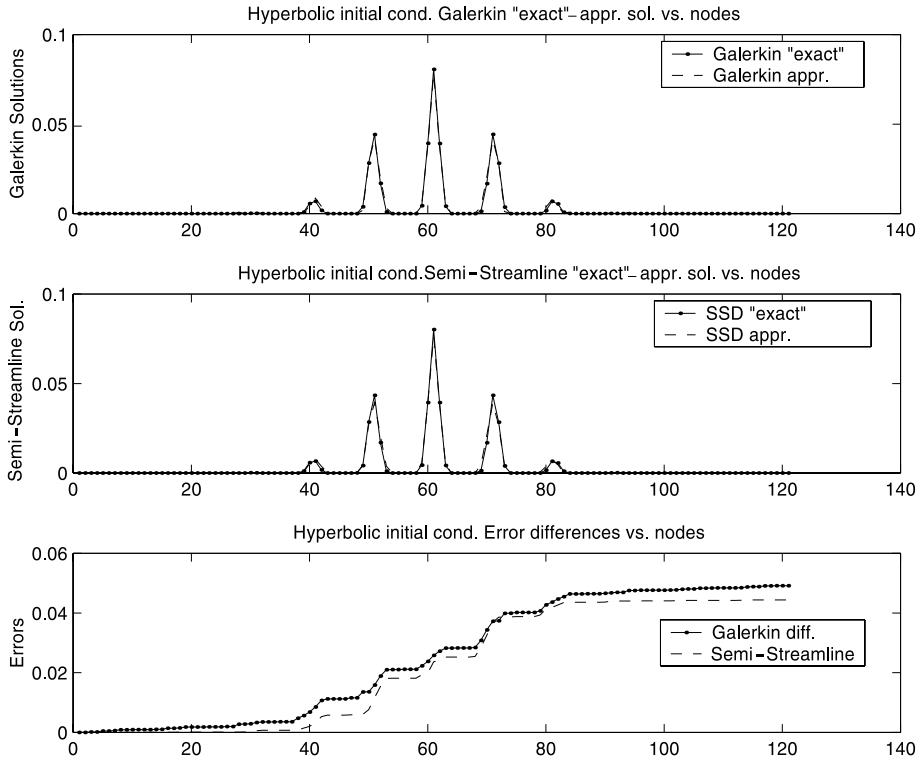


Fig. 9. $L_1$ cumulative error vs nodes for SG and SSD for hyperbolic. The solutions at $x = 1$ are found for $h = 0.2$.

The initial conditions used in calculating the values in the tables can be seen in Fig. 1. In Fig. 2 the approximate solution is displayed for a Dirac type of initial condition in the finest mesh used.

In the Tables 1–3 we see the convergence of each scheme as the step size is reduced. We also detect a slight improvement in using SSD over SG in terms of the consistent decrease in the respective error. Depending on the initial condition used the rates of this decay vary.

In figures studies on two of the initial data (Dirac and Maxwellian) are presented in some detail, while the remaining cases (because of space limitations) are shown rather briefly. More specifically in Figs. 3–6 we consider Dirac and Maxwellian data and look at slices of the domain $\Omega$ and the differences between the "exact" and approximate solutions over all three cases of time discretization schemes, thus providing us a *"local"* picture of the variation depicted in the tables. Further we provide an overall view of this variation, for all considered initial data cases, over the whole domain in Figs. 7–10. For each initial data in these figures we explore the effects of the SG and SSD solutions ONLY under the case of backward Euler time discretization where particular emphasis is given on plotting the cumulative $L_1$ error of the schemes. The computational parameters that are used depend on the theoretical results presented in Sections 2 and 3. For instance $\varepsilon$ must be chosen to be small and given such a choice we take $h^2 \sim \varepsilon$, also $\delta \sim h$. Specifically Figs. 7–10 were produced for values of $\varepsilon = 0.05$ and $h = 0.1$. In these examples the value of $\delta$ is taken as $\delta = h/2$, and the time increment was chosen as $k = h^2$.

In Fig. 7 we see the improvement due to SSD over SG for a Dirac type of initial condition. In particular we see that the error of using SSD is reduced by approximately $1/5$ (20%) over the error of SG. This behaviour is maintained for a Maxwellian type of initial condition as can be seen in Fig. 8. Here the error for using SSD is reduced by approximately $1/6$ (17%) over using SG. Similar such results are displayed in

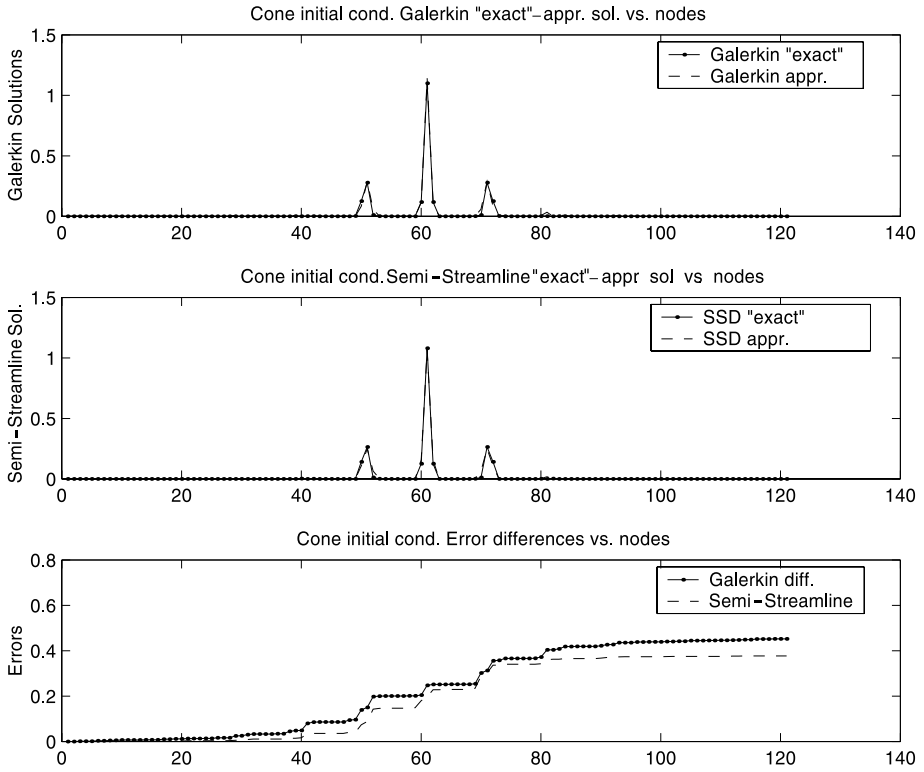


Fig. 10. $L_1$ cumulative error vs nodes for SG and SSD for cone data. The solutions at $x = 1$ are found for $h = 0.2$.

Figs. 9 and 10. For instance for a cone type of initial condition and a value of $\varepsilon = 0.04$ (for variety purposes), we observe a $1/6$ (17%) improvement of the error due to SSD over SG.

## 6. Conclusions

We have shown stability and convergence estimates for two finite element methods: the SG and a Petrov–Galerkin method here referred as SSD. We have proved strong stability of these methods and derived detailed convergence analysis of the SG method. The corresponding convergence study for the SSD method is similar, but lengthy, and therefore is omitted. Subsequent simple numerical examples were carried out to further illustrate the results under different cases of initial conditions. In theory, in the triple norm, the SSD converges by a factor of $h^{r-1/2}$ (compared with SG which converges with a factor of $h^{r-1}$) for functions in $H^r(\Omega)$. Our numerical examples correspond to use of linear basis functions. For such simple examples it is virtually impossible to calculate the exact estimates as required for instance in Proposition 2.1, since the $|||\cdot|||_{\tilde{\beta}}$ norm is not easily computable. Therefore, tables are constructed for the usual $L_p$, $p = 1, 2, \infty$ and $\tilde{L}_2$-norms, whereas in the figures we display a simpler cumulative $L_1$-norm of the errors for the SSD and SG schemes. Although we do not directly evaluate the $|||\cdot|||_{\tilde{\beta}}$ norm but only an $L_1$-norm, the speed up of SSD over SG is evident, and as we have shown it depends on the type of initial data used, as expected.

In general the theoretical estimates (Sections 2 and 3) are much harder to detect numerically since, as we previously remarked, we have no exact solution of the Fermi equation available but rather only an approximation of it in the finest mesh that we can produce and subsequently calculate the errors from it.

In summary we have considered a simple forward–backward degenerate type equation. Our objective is to extend the approximation techniques of the non-degenerate equations to a degenerate case. In a forthcoming paper we shall extend further our studies to a more realistic, three-dimensional, model problem.

## References

[1] R.A. Adams, Sobolev Spaces, Academic Press, New York, 1975.
[2] M. Asadzadeh, Streamline diffusion methods for Fermi and Fokker–Planck equations, Transport Theory Stat. Phys. 26 (3) (1997) 319–340.
[3] M. Asadzadeh, A posteriori error estimates for the Fokker–Planck and Fermi pencil beam equations, Math. Models Meth. Appl. Sci. 10 (5) (2000) 737–769.
[4] M. Asadzadeh, Characteristic methods for Fokker–Planck and Fermi pencil beam equations, in: R. Brun, R. Campargue, R. Gatignol, J.-C. Lengrand (Eds.), Rarefied Gas Dynamics Cépaduès Éditions, vol. 2, 1999, pp. 202–212.
[5] J. Bergh, J. Löfsröm, Interpolation Spaces, Springer, Berlin, 1976.
[6] C. Börgers, E.W. Larsen, Asymptotic derivation of the fermi pencil-beam approximation, Nucl. Sci. Engrg. 123 (1996) 343–357.
[7] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, Springer, Berlin, 1994.
[8] A. Brooks, T.J.R. Hughes, Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations, FENOMECH '81, Part I (Stuttgart, 1981), Comput. Meth. Appl. Mech. Engrg. 32 (1–3) (1982) 199–259.
[9] D. Jette, Electron dose calculations using multiple-scattering theory. A new theory of multiple-scattering, J. Med. Phys. 23 (1996) 459–476.
[10] C. Johnson, Numerical solution of partial differential equations by the finite element method, Studentlitteratur, 1991.
[11] C. Johnson, A new approach to algorithms for convection problems which are based on exact transport + projection, Comput. Meth. Appl. Mech. Engrg. 100 (1) (1992) 45–62.
[12] A.F.D. Loula, T.J.R. Hughes, L.P. Franca, Petrov–Galerkin formulations of the Timoshenko beam problem, Comput. Meth. Appl. Mech. Engrg. 63 (2) (1987) 115–132.