Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 285 (2015) 730–751
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

A posteriori error analysis of IMEX multi-step time integration methods for advection–diffusion–reaction equations

Jehanzeb H. Chaudhry[a], Donald Estep[b], Victor Ginting[c], John N. Shadid[d,*], Simon Tavener[a]

[a] *Department of Mathematics, Colorado State University, Fort Collins, CO 80523, United States*
[b] *Department of Statistics, Colorado State University, Fort Collins, CO 80523, United States*
[c] *Department of Mathematics, University of Wyoming, Laramie, WY 82071, United States*
[d] *Computational Mathematics Department, Sandia National Laboratories, Albuquerque, NM 87123, United States*

Received 19 May 2014; received in revised form 9 November 2014; accepted 13 November 2014
Available online 25 November 2014

Abstract

Implicit–Explicit (IMEX) schemes are an important and widely used class of time integration methods for both parabolic and hyperbolic partial differential equations. We develop accurate *a posteriori* error estimates for a user-defined quantity of interest for two classes of multi-step IMEX schemes for advection–diffusion–reaction problems. The analysis proceeds by recasting the IMEX schemes into a variational form suitable for *a posteriori* error analysis employing adjoint problems and computable residuals. The *a posteriori* estimates quantify distinct contributions from various aspects of the spatial and temporal discretizations, and can be used to evaluate discretization choices. Numerical results are presented that demonstrate the accuracy of the estimates for a representative set of problems.
Published by Elsevier B.V.

Keywords: Error estimation; Adjoint operator; Implicit–explicit schemes

1. Introduction

We derive goal-oriented *a posteriori* error estimates for Implicit–Explicit (IMEX) multi-step numerical methods for scalar-valued advection–diffusion–reaction partial differential equations (PDEs) of the form,

$$\begin{cases} \dot{u}(x,t) - \nabla \cdot \epsilon(x)\nabla u(x,t) + \mathbf{b}(x) \cdot \nabla u(x,t) = R(u,x,t), & (x,t) \in \Omega \times (0,T], \\ u(x,0) = u_0(x), & x \in \Omega, \\ u(x,t) = u_d(t), & (x,t) \in \partial\Omega \times (0,T], \end{cases} \tag{1.1}$$

* Corresponding author.
E-mail addresses: jehanzeb@colostate.edu (J.H. Chaudhry), estep@stat.colostate.edu (D. Estep), vginting@uwyo.edu (V. Ginting), jnshadi@sandia.gov (J.N. Shadid), tavener@math.colostate.edu (S. Tavener).

http://dx.doi.org/10.1016/j.cma.2014.11.015
0045-7825/Published by Elsevier B.V.

where $\dot{u}(x,t) = \frac{\partial u(x,t)}{\partial t}$, $\epsilon(x) > 0$ is the diffusion coefficient, $\mathbf{b}(x)$ represents the advective vector field, $R(u,x,t)$ is a reaction term (possibly nonlinear), and Ω is a convex polygonal domain. We assume that ϵ, $\mathbf{b}$, and u_0 are smooth on Ω, u_d is smooth on $[0,T]$, R is smooth on $\mathbb{R}\times\Omega\times(0,T]$, $R(\cdot,x,t)$ is uniformly Lipschitz continuous on $\Omega\times(0,T]$, and $\epsilon(x) \geq \epsilon_0 > 0$ for some constant ϵ_0. Under these assumptions, (1.1) admits smooth solutions for some time T [1].

IMEX methods are a widely used class of time integration techniques for complex partial differential equations. While there are many forms of IMEX discretization, all IMEX schemes share the basic idea of decomposing the differential operator into two components in which one component is treated implicitly in the discretization and the other component explicitly. This flexibility of mixing explicit and implicit discretization allows the application of specialized numerical solution methods for systems composed of operators with differing time-scales. Consider the transient solution of a system of equations that includes coupled convection, diffusion and reaction mechanisms (e.g. simple convection–diffusion–reaction PDEs, Navier–Stokes with chemical reactions and/or radiation-diffusion models). In this context, IMEX methods can be used to treat the convection and reaction operators either explicitly or implicitly based on the time-step-size stability restrictions of these terms, while most commonly the diffusion is treated implicitly. A set of representative references from the scientific literature indicates both the recent interest in these methods, as well as the complexity of the computational physics/mathematical models to which these methods have been applied (see e.g. [2–12]). Specific references that consider multi-step IMEX approaches for complex applications include [13–16]. The stability, order-of-accuracy, and order-reduction results for these methods have been studied for a number of prototype systems.

Accurate numerical solution of advection–diffusion–reaction problems presents a significant computational challenge in general, and, consequently, numerical error is generally significant in practical applications [17,18]. Thus, it is important to accurately quantify the error in computed quantities of interest obtained from numerical solutions. For an important class of IMEX multi-step methods, we develop an *a posteriori* error analysis using variational analysis, computable residuals, and adjoint problems to derive accurate error estimates for a given quantity-of-interest. Such *a posteriori* error estimates are widely used for finite element methods [19,20,17,21–25]. The resulting estimates have the useful feature that the total error is decomposed as a sum of contributions from various aspects of discretization and therefore can provide insight into the effect of different choices for the parameters controlling the discretization (e.g. time step size and mesh spacing). Our development of the *a posteriori* error analysis is demonstrated in the context of convection–diffusion–reaction systems. IMEX approaches raise additional reasons for quantitative error estimation, since IMEX methods fall in the general category of operator decomposition/finite iteration methods [18,26–30], and thus give rise to additional sources of instability and discretization errors compared to fully implicit methods. Further, error estimates are required to construct adaptive algorithms and this is an active area of current research (see e.g. [22,31]).

Specifically, we derive *a posteriori* error estimates for two classes of one-step first-order and two-step second-order IMEX schemes. We begin the development by considering IMEX methods applied to ordinary differential equations obtained after semi-discretization of (1.1) in space. The analysis is then generalized by considering the full space–time discretization of (1.1). One difficulty in the analysis for deriving the *a posteriori* error estimates is that IMEX methods are usually presented as finite difference schemes. In order to apply adjoint-based techniques, we reformulate the IMEX schemes as finite element methods by employing certain quadrature formulas and considering the difference schemes over appropriate intervals. Finally we demonstrate the accuracy of the estimates for numerical solutions of a range of scalar-valued PDEs using both finite-difference and finite-element discretizations in space. The former provides estimates for temporal errors only, the latter provides estimates of both spatial and temporal errors.

The paper is organized as follows. In Section 2, we discuss variational formulations and discretization schemes for (1.1). We describe multi-step IMEX schemes in Section 3. We perform an *a posteriori* error analysis for one-step first-order and two-step second-order IMEX schemes in Section 4. Finally, we then present numerical examples in Section 5.

2. PDE variational formulation and discretizations

We present two *a posteriori* analyses. First, we treat the large dimension ODE in time that results from semi-discretization in space by employing the method of lines. This discretization is presented in Section 2.1 and the analysis is presented in Section 4.1. This analysis makes it relatively easy to focus on the effects of IMEX discretization

on the time solution. After that, we consider the full space–time discretization of the PDE. The discretization is presented in Section 2.2 and the analysis in Section 4.2. We then present numerical comparisons in Section 5.

2.1. Semidiscrete formulation

Semi-discretization in space

Without loss of generality we assume that $u_d = 0$ in (1.1). The weak formulation of (1.1) is: Find $u(t) \in H_0^1(\Omega)$ for $t \in (0, T]$ such that

$$(\dot{u}, v) - (\epsilon \nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v) = (R(u), v) \quad \forall v \in H_0^1(\Omega),\ t \in (0, T], \tag{2.1}$$

where $(\cdot, \cdot)$ represents the L_2 inner product on the spatial domain Ω. We discretize Ω into a quasi-uniform triangulation $\mathcal{T}_h$, where h denotes the maximum diameter of the elements. This triangulation is chosen so the union of the elements of $\mathcal{T}_h$ is Ω and the intersection of any two elements is either a common edge, node, or is empty. The finite-element approximation is a continuous piecewise linear polynomial with respect to $\mathcal{T}_h$. We let $V_h \subset H_0^1(\Omega)$ denote the space of piecewise linear continuous functions $v(x) \in \mathbb{R}$ defined on $\mathcal{T}_h$. The finite element semi-discretization reads: Find $u^h \in V^h$ such that,

$$(\dot{u}^h, v) - (\epsilon \nabla u^h, \nabla v) + (\mathbf{b} \cdot \nabla u^h, v) = (R(u^h), v), \quad \forall v \in V^h,\ t \in (0, T]. \tag{2.2}$$

This is a nonlinear system of ordinary differential equations. Fixing a basis for V^h, (2.2) can be written as

$$\dot{\widetilde{u}} = f(\widetilde{u}(t)) + g(\widetilde{u}(t)), \quad t \in (0, T], \tag{2.3}$$

where $\widetilde{u}$ is a m dimensional vector, $f(\widetilde{u}(t)) = -M^{-1}B\widetilde{u}(t) + M^{-1}N(\widetilde{u}(t))$ and $g(\widetilde{u}(t)) = M^{-1}A\widetilde{u}(t)$. Here M, A and B are matrices arising from the expressions $(\dot{u}^h, v)$, $(\epsilon \nabla u^h, \nabla v)$ and $(\mathbf{b} \cdot \nabla u^h, v)$ respectively, while $N(\widetilde{u}(t))$ is a vector arising from the non-linear term $(R(u^h), v)$. We note that this choice of f and g may not be the optimal choice, and in fact may lead to an ineffective method depending on the nature of the nonlinear reaction term. The *a posteriori analysis* presented in this paper applies to (2.3) for different choices of f and g as well, and may in fact guide the choice of f and g, as we illustrate with a numerical example in Section 5.

We have written the right-hand side as a sum of two terms for the purpose of IMEX discretization, where $f : \mathbb{R}^m \to \mathbb{R}^m$ represents the part of the equation that is treated explicitly and $g : \mathbb{R}^m \to \mathbb{R}^m$ is the term that is treated implicitly.

Time discretization of the semi-discretization

We now discretize (2.3) using a finite element method. The variational form of (2.3) is: Find $\widetilde{u}(t) \in H^1((0, T])^m$ such that,

$$\int_0^T (\dot{\widetilde{u}}, \widetilde{v})\, \mathrm{d}t = \int_0^T (f(\widetilde{u}(t)) + g(\widetilde{u}(t)), \widetilde{v})\, \mathrm{d}t, \quad \forall \widetilde{v} \in L_2((0, T])^m, \tag{2.4}$$

where we have abused notation to let (a, b) denote the standard inner product in $\mathbb{R}^m$. We consider the continuous Galerkin finite element method [32,20]. Let τ_N be a partition of $[0, T]$ where $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n < \cdots < t_N = T$ and $I_n = [t_{n-1}, t_n]$, $n = 1, \ldots, N$. On τ_N, let $\widetilde{W}^q[0, T]$ be the space of *continuous* piecewise polynomial vector valued functions of degree q, that is, if $\widetilde{v} \in \widetilde{W}^q[0, T]$, then $\widetilde{v}|_{I_n}$ is a vector-valued polynomial function of degree q for each interval I_n. The continuous Galerkin method of order q (cG(q)) for (2.4) is: Find $\widetilde{U} \in \widetilde{W}^q[0, T]$ such that

$$\sum_{n=1}^{N} \langle \dot{\widetilde{U}} - f(\widetilde{U}) - g(\widetilde{U}), \widetilde{v} \rangle_{I_n} = 0 \quad \forall \widetilde{v} \in \widetilde{W}^{q-1}(I_n), \tag{2.5}$$

where $\langle a, b \rangle_{I_n} = \int_{t_{n-1}}^{t_n} (a, b)\, \mathrm{d}t$. In a practical setting, the integral $\langle a, b \rangle_{I_n}$ is approximated by a quadrature rule.

2.2. Space–time formulation

Eq. (2.5) represents a full space–time discretization of (1.1). For the sake of exposition however, we write out the space–time discretization directly. Letting $\mathcal{V} = H^1((0, T]); H_0^1(\Omega)$ and $\mathcal{W} = L_2((0, T]); H_0^1(\Omega)$, the weak

space–time variational form of (1.1) is: Find $u(x,t) \in \mathcal{V}$ such that,

$$\int_0^T (\dot{u}, v) + (\epsilon \nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v)\, \mathrm{d}t = \int_0^T (R(u), v)\, \mathrm{d}t, \quad \forall v \in \mathcal{W}. \tag{2.6}$$

On each space–time slab $S_n = \Omega \times I_n$, we choose finite-element approximations that are polynomials in time and continuous piecewise linear polynomials in space with respect to $\mathcal{T}_h$ and define

$$W_n^q = \left\{ w(x,t) : w(x,t) = \sum_{j=0}^{q} t^j v_j(x),\ v_j \in V_h,\ (x,t) \in S_n \right\}.$$

The finite element solution is sought in the space W^q where if $v \in W^q$, then $v|_{I_n} \in W_n^q$. The continuous Galerkin method of order 1, cG(1), for (1.1) is to find $U \in W^1$ such that,

$$\int_{t_{n-1}}^{t_n} (\dot{U}, v) + (\epsilon \nabla U, \nabla v) + (\mathbf{b} \cdot \nabla U, v)\, \mathrm{d}t = \int_{t_{n-1}}^{t_n} (R(U), v)\, \mathrm{d}t \quad \forall v \in W_n^0. \tag{2.7}$$

Evaluation of the time integrals in the weak formulation (2.7) using quadrature leads to IMEX schemes, as we show in the next section.

In the development that follows we assume that full space–time discretizations of the PDE systems have sufficient physical dissipation that the convection–diffusion–reaction system integrated at the time-step of interest with the IMEX multi-step methods are stable (see e.g. [3] for detailed linear stability analysis of these methods).

3. IMEX schemes and their representation as finite element methods

Following the pattern established in Section 2, we consider IMEX schemes for the semidiscrete formulation in Section 3.1, and construct IMEX schemes for the space–time finite-element formulation in Section 3.2.

3.1. IMEX time integration for the semidiscrete formulation

A generic r-step implicit–explicit (IMEX) scheme for (2.3) has the form,

$$\widetilde{U}_n = \sum_{j=1}^{r} a_j \widetilde{U}_{n-j} + \sum_{j=1}^{r} b_j \Delta t f(\widetilde{U}_{n-j}) + \sum_{j=0}^{r} c_j \Delta t g(\widetilde{U}_{n-j}), \tag{3.1}$$

where the parameters a_j, b_j and c_j are chosen to obtain a rth order scheme. We consider the one-step first-order and two-step second-order IMEX schemes presented in [10]. Throughout this article, we use the notation $f_n = f(\widetilde{U}(t_n))$ and $g_n = g(\widetilde{U}(t_n))$.

First-order IMEX schemes

The single-step, first-order IMEX schemes are,

$$\widetilde{U}_n - \widetilde{U}_{n-1} = \Delta t f_{n-1} + \Delta t[(1-\gamma) g_{n-1} + \gamma g_n], \tag{3.2}$$

where $0 \le \gamma \le 1$. Different choices of γ lead to different schemes. For example, $\gamma = 0$ leads to the fully explicit Forward Euler scheme, whereas $\gamma = 1$ leads to a semi-implicit BDF (SBDF) scheme [10],

$$\widetilde{U}_n - \widetilde{U}_{n-1} = \Delta t f_{n-1} + \Delta t g_n.$$

Second-order IMEX schemes

The two-step, second-order IMEX schemes [10] are,

$$\begin{aligned} &\left(\gamma + \frac{1}{2}\right) \widetilde{U}_n - 2\gamma \widetilde{U}_{n-1} + \left(\gamma - \frac{1}{2}\right) \widetilde{U}_{n-2} \\ &\quad = \Delta t \left[(\gamma+1) f_{n-1} - \gamma f_{n-2} + \left(\gamma + \frac{c}{2}\right) g_n + (1 - \gamma - c) g_{n-1} + \frac{c}{2} g_{n-2} \right]. \end{aligned} \tag{3.3}$$

Different choices of γ and c lead to some popular schemes [10]. For example $(\gamma, c) = (\frac{1}{2}, 0)$ yields the CNAB (Crank–Nicolson, Adams Bashforth) scheme, $(\gamma, c) = (0, 1)$ the CNLF (Crank–Nicolson, Leap Frog) scheme, and $(\gamma, c) = (1, 0)$ the SBDF (or Gear) scheme [11].

Representation of the IMEX schemes as finite element methods

To represent the IMEX schemes as a type of cG(1) finite element method (2.5), we evaluate the integrals in the weak formulation via quadrature,

$$\begin{aligned}\langle \dot{\widetilde{U}}, \widetilde{v}\rangle_{I_n} &= \langle f(\widetilde{U}) + g(\widetilde{U}), \widetilde{v}\rangle_{I_n} = \langle f(\widetilde{U}), \widetilde{v}\rangle_{I_n} + \langle g(\widetilde{U}), \widetilde{v}\rangle_{I_n} \\ &\approx \langle f(\widetilde{U}), \widetilde{v}\rangle_{I_{n,m_1}} + \langle g(\widetilde{U}), \widetilde{v}\rangle_{I_{n,m_2}} = I^{(1)} + I^{(2)}. \end{aligned} \tag{3.4}$$

Here

$$I^{(1)} = \sum_{j=1}^{m_1} w_j^{(1)}\ f(\widetilde{U}(t_j^{(1)}))\ \widetilde{v}(t_j^{(1)}), \qquad I^{(2)} = \sum_{j=1}^{m_2} w_j^{(2)}\ g(\widetilde{U}(t_j^{(2)}))\ \widetilde{v}(t_j^{(2)}), \tag{3.5}$$

for pairs $(t_j^{(1)}, w_j^{(1)})$, $j = 1 \cdots m_1$ and $(t_j^{(2)}, w_j^{(2)})$, $j = 1 \cdots m_2$. We note that the use of notation $\langle \cdot, \cdot \rangle_{I_{n,m}}$ for some integer m does not fully specify a m point quadrature rule. However, whenever we use this notation, we also specify the m locations t_j and weights w_j immediately afterwards and hence avoid any ambiguity.

Equivalency of first-order IMEX schemes

Through appropriate choice of the quadrature rules $I^{(1)}$ and $I^{(2)}$, we show nodal equivalency of the first-order IMEX scheme with a cG(1) method, that is, we choose the quadrature locations and quadrature weights for the cG(1) method so that the two schemes have the same nodal values.

Theorem 3.1 (*Equivalency of First-Order Schemes*)**.** *The following choices for quadrature weights and quadrature points in Eq. (3.4) ensure nodal equivalency between the* cG(1) *finite element method with a particular quadrature and the single-step first-order IMEX scheme. Let* $m_1 = 1$ *and* $m_2 = 2$*,*

$$w_1^{(1)} = \Delta t, \qquad w_2^{(1)} = \Delta t(1-\gamma), \qquad w_2^{(2)} = \Delta t(\gamma),$$

and

$$t_1^{(1)} = t_{n-1}, \qquad t_2^{(1)} = t_{n-1}, \qquad t_2^{(2)} = t_n.$$

Proof. For the cG(1) scheme, $\widetilde{v} \equiv 1$. The terms in Eq. (3.4) are then

$$\langle \dot{\widetilde{U}}, 1\rangle_{I_n} = \widetilde{U}_n - \widetilde{U}_{n-1}, \qquad \langle f(\widetilde{U}), 1\rangle_{I_{n,m_1}} = \Delta t f_{n-1}, \qquad \langle g(\widetilde{U}), 1\rangle_{I_{n,m_2}} = \Delta t(1-\gamma) g_{n-1} + \Delta t \gamma g_n$$

which gives Eq. (3.2). □

Equivalency of second-order IMEX schemes

To analyze the two-step second-order scheme, we consider how it arises from a variational principle. To this end, we consider the integrals in Eq. (2.5) in pairs, i.e.,

$$\begin{cases} \langle \dot{\widetilde{U}} - f(\widetilde{U}) - g(\widetilde{U}), \widetilde{v}_{n-1}\rangle_{I_{n-1}} = 0, \quad \forall \widetilde{v}_{n-1} \in \widetilde{W}^{q-1}(I_{n-1}), \\ \langle \dot{\widetilde{U}} - f(\widetilde{U}) - g(\widetilde{U}), \widetilde{v}_n\rangle_{I_n} = 0, \quad \forall \widetilde{v}_n \in \widetilde{W}^{q-1}(I_n). \end{cases} \tag{3.6}$$

We multiply the first equation by $(\gamma + \frac{1}{2})$ and the second equation by $(-\gamma + \frac{1}{2})$ and sum. We then evaluate the integrals via quadrature, namely

$$\begin{aligned} &\left(\gamma + \frac{1}{2}\right) \langle \dot{\widetilde{U}}, \widetilde{v}_n\rangle_{I_n} + \left(-\gamma + \frac{1}{2}\right) \langle \dot{\widetilde{U}}, \widetilde{v}_{n-1}\rangle_{I_{n-1}} \\ &\quad = \left(\gamma + \frac{1}{2}\right) \langle f(\widetilde{U}) + g(\widetilde{U}), \widetilde{v}_n\rangle_{I_n} + \left(-\gamma + \frac{1}{2}\right) \langle f(\widetilde{U}) + g(\widetilde{U}), \widetilde{v}_{n-1}\rangle_{I_{n-1}}. \end{aligned}$$

For cG(1), $v(t) \equiv 1$ and this simplifies to

$$\begin{aligned}\left(\gamma+\frac{1}{2}\right)\langle\dot{\widetilde{U}},1\rangle_{I_n}+\left(-\gamma+\frac{1}{2}\right)\langle\dot{\widetilde{U}},1\rangle_{I_{n-1}} &\approx \frac{1}{2}\langle f(\widetilde{U}),1\rangle_{\{I_{n-1}\cup I_n\},m_1}+\frac{1}{2}\langle g(\widetilde{U}),1\rangle_{\{I_{n-1}\cup I_n\},m_2}\\ &\quad+\gamma\langle f(\widetilde{U}),1\rangle_{I_n,m_3}+\gamma\langle g(\widetilde{U}),1\rangle_{I_n,m_4}\\ &\quad-\gamma\langle f(\widetilde{U}),1\rangle_{I_{n-1},m_5}-\gamma\langle g(\widetilde{U}),1\rangle_{I_{n-1},m_6},\\ &= I^{(1)}+I^{(2)}+I^{(3)}+I^{(4)}+I^{(5)}+I^{(6)}. \end{aligned}\tag{3.7}$$

Here,

$$I^{(1)}=\frac{1}{2}\sum_{j=1}^{m_1}w_j^{(1)}\ f(\widetilde{U}(t_j^{(1)})),\qquad I^{(2)}=\frac{1}{2}\sum_{j=1}^{m_2}w_j^{(2)}\ g(\widetilde{U}(t_j^{(2)})),$$

$$I^{(3)}=\gamma\sum_{j=1}^{m_3}w_j^{(3)}\ f(\widetilde{U}(t_j^{(3)})),\qquad I^{(4)}=\gamma\sum_{j=1}^{m_4}w_j^{(4)}\ g(\widetilde{U}(t_j^{(4)})),$$

$$I^{(5)}=-\gamma\sum_{j=1}^{m_5}w_j^{(5)}\ f(\widetilde{U}(t_j^{(5)})),\qquad I^{(6)}=-\gamma\sum_{j=1}^{m_6}w_j^{(6)}\ g(\widetilde{U}(t_j^{(6)})).$$

Theorem 3.2 (*Equivalency of Second-Order Schemes*)**.** *The choices* $m_1=1$, $m_2=3$ *and* $m_3=m_4=m_5=m_6=1$,

$$w_1^{(1)}=2\Delta t,\quad w_1^{(2)}=c\Delta t,\quad w_2^{(2)}=(2-2c)\Delta t,\quad w_3^{(2)}=c\Delta t,\quad w_1^{(3)}=\Delta t,$$
$$w_1^{(4)}=\Delta t,\quad w_1^{(5)}=\Delta t,\quad w_1^{(6)}=\Delta t,$$

and

$$t_1^{(1)}=t_{n-1},\quad t_1^{(2)}=t_n,\quad t_2^{(2)}=t_{n-1},\quad t_3^{(2)}=t_{n-2},\quad t_1^{(3)}=t_{n-1},$$
$$t_1^{(4)}=t_n,\quad t_1^{(5)}=t_{n-2},\quad t_1^{(6)}=t_{n-1}$$

ensure nodal equivalency between the second-order IMEX scheme and cG(1) *finite element method with a particular quadrature.*

Proof. Observing that $\langle\dot{\widetilde{U}},1\rangle_{I_n}=\widetilde{U}_n-\widetilde{U}_{n-1}$ and $\langle\dot{\widetilde{U}},1\rangle_{I_{n-1}}=\widetilde{U}_{n-1}-\widetilde{U}_{n-2}$ and approximating terms on the right hand side of (3.7) with the quadrature rules

$$\langle f(\widetilde{U}),1\rangle_{\{I_{n-1}\cup I_n\},m_1}=2\Delta t f_{n-1},\qquad \langle g(\widetilde{U}),1\rangle_{\{I_{n-1}\cup I_n\},m_2}=\Delta t[cg_n+(2-2c)g_{n-1}+cg_{n-2}],$$
$$\langle f(\widetilde{U}),1\rangle_{I_n,m_3}=\Delta t f_{n-1},\qquad \langle g(\widetilde{U}),1\rangle_{I_n,m_4}=\Delta t g_n,$$
$$\langle f(\widetilde{U}),1\rangle_{I_{n-1},m_5}=\Delta t f_{n-2},\qquad \langle g(\widetilde{U}),1\rangle_{I_{n-1},m_6}=\Delta t g_{n-1},$$

leads to Eq. (3.3). These quadrature rules are easily recognizable as either the left-hand or right-hand rule. Note that $\langle g(\widetilde{U}),1\rangle_{\{I_{n-1}\cup I_n\},m_2}=\Delta t[cg_n+(2-2c)g_{n-1}+cg_{n-2}]$ is Simpson's Rule when $c=1/6$. □

3.2. IMEX time integration for the space–time formulation

The first-order IMEX scheme for (2.7) is,

$$\begin{aligned}(U_n-U_{n-1},v) &= \Delta t(R(U_{n-1}),v)+\Delta t\left[(1-\gamma)((-\mathbf{b}\cdot\nabla U_{n-1},v)+(-\epsilon\nabla U_{n-1},\nabla v))\right.\\ &\quad\left.+\gamma((-\mathbf{b}\cdot\nabla U_n,v)+(-\epsilon\nabla U_n,\nabla v))\right],\end{aligned}\tag{3.8}$$

for all $v\in V^h$.

The second-order IMEX scheme for (2.7) is of the form,

$$\begin{aligned}&\left(\left(\gamma+\frac{1}{2}\right)U_n-2\gamma U_{n-1}+\left(\gamma-\frac{1}{2}\right)U_{n-2},v\right)\\&=\Delta t\Bigg[(\gamma+1)(R(U_{n-1}),v)-\gamma(R(U_{n-2}),v)+\left(\gamma+\frac{c}{2}\right)((-\mathbf{b}\cdot\nabla U_n,v)+(-\epsilon\nabla U_n,\nabla v))\\&\quad+(1-\gamma-c)((-\mathbf{b}\cdot\nabla U_{n-1},v)+(-\epsilon\nabla U_{n-1},\nabla v))+\frac{c}{2}((-\mathbf{b}\cdot\nabla U_{n-2},v)+(-\epsilon\nabla U_{n-2},\nabla v))\Bigg].\end{aligned}\tag{3.9}$$

It should be noted that a specific choice of the assignment of convection, diffusion and reaction operators has been made in this development. Other choices would of course be possible with a re-interpretation of the terms above. In the examples that follow alternate choices for the convection operator to the explicit or implicit terms are demonstrated in the case of the space–time discretizations. The equivalency of the space–time formulations with cG(1) finite element method with a particular quadrature mirrors that of semidiscrete formulation. However, for concreteness a brief description given below.

Equivalency of first-order IMEX schemes

As before, we evaluate the time integrals in the weak formulation (2.7) by quadrature,

$$\begin{aligned}\langle(\dot{U},v)\rangle_{I_n}&=\langle(R(U),v)\rangle_{I_n}-\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_n}\\&\approx\langle(R(U),v)\rangle_{I_{n,m_1}}-\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_{n,m_2}}=I^{(1)}+I^{(2)},\end{aligned}\tag{3.10}$$

where

$$\begin{aligned}I^{(1)}&=\sum_{j=1}^{m_1}w_j^{(1)}\,(R(U(t_j^{(1)})),v(t_j^{(1)})),\\I^{(2)}&=-\sum_{j=1}^{m_2}w_j^{(2)}\,\left[(\epsilon\nabla U(t_j^{(2)}),\nabla v(t_j^{(2)}))+(\mathbf{b}\cdot\nabla U(t_j^{(2)}),v(t_j^{(2)}))\right].\end{aligned}\tag{3.11}$$

Theorem 3.1 defines the weights which ensure the equivalency of the cG(1) finite element method with a particular quadrature and the first-order IMEX scheme.

Equivalency of second-order IMEX schemes

To show how the two-step second-order schemes arise from the variational formulation, we consider the integrals in Eq. (2.7) in pairs, i.e.,

$$\begin{cases}\langle(\dot{U},v)+(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)-(R(U),v)\rangle_{I_{n-1}}=0,\\\langle(\dot{U},v)+(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)-(R(U),v)\rangle_{I_n}=0.\end{cases}\tag{3.12}$$

Let $\langle(a,b)\rangle_{I_n}=\int_{I_n}(a,b)\,dt$ where (a,b) denotes the L_2 inner product in space. As before, we multiply the first equation by $(\gamma+\frac{1}{2})$ and the second equation by $(-\gamma+\frac{1}{2})$ and sum, giving

$$\begin{aligned}&\left(\gamma+\frac{1}{2}\right)\langle(\dot{U},v)\rangle_{I_n}+\left(-\gamma+\frac{1}{2}\right)\langle(\dot{U},v)\rangle_{I_{n-1}}\\&=\left(\gamma+\frac{1}{2}\right)\langle(-\epsilon\nabla U,\nabla v_n)-(\mathbf{b}\cdot\nabla U,v)+(R(U),v_n)\rangle_{I_n}\\&\quad+\left(-\gamma+\frac{1}{2}\right)\langle(-\epsilon\nabla U,\nabla v_{n-1})-(\mathbf{b}\cdot\nabla U,v)+(R(U),v_{n-1})\rangle_{I_{n-1}}.\end{aligned}$$

We now evaluate the integrals via quadrature,

$$\begin{aligned}
&\left(\gamma+\frac{1}{2}\right)\langle(\dot{U},v)\rangle_{I_n}+\left(-\gamma+\frac{1}{2}\right)\langle(\dot{U},v)\rangle_{I_{n-1}}\\
&\quad\approx\frac{1}{2}\langle(R(U),v)\rangle_{\{I_{n-1}\cup I_n\},m_1}-\frac{1}{2}\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{\{I_{n-1}\cup I_n\},m_2}\\
&\qquad+\gamma\langle(R(U),v)\rangle_{I_n,m_3}-\gamma\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_n,m_4}\\
&\qquad-\gamma\langle(R(U),v)\rangle_{I_{n-1},m_5}+\gamma\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_{n-1},m_6},\\
&\quad=I^{(1)}+I^{(2)}+I^{(3)}+I^{(4)}+I^{(5)}+I^{(6)}
\end{aligned}\tag{3.13}$$

where

$$I^{(1)}=\frac{1}{2}\sum_{j=1}^{m_1}w_j^{(1)}\,(R(U(t_j^{(1)})),v),\qquad I^{(2)}=\frac{1}{2}\sum_{j=1}^{m_2}w_j^{(2)}\,(\epsilon\nabla U(t_j^{(2)}),\nabla v)+(\mathbf{b}\cdot\nabla U(t_j^{(2)}),v),$$

$$I^{(3)}=\gamma\sum_{j=1}^{m_3}w_j^{(3)}\,(R(U(t_j^{(3)})),v),\qquad I^{(4)}=-\gamma\sum_{j=1}^{m_4}w_j^{(4)}\,(\epsilon\nabla U(t_j^{(4)}),\nabla v)+(\mathbf{b}\cdot\nabla U(t_j^{(4)}),v),$$

$$I^{(5)}=-\gamma\sum_{j=1}^{m_5}w_j^{(5)}\,(R(U(t_j^{(5)})),v),\qquad I^{(6)}=-\gamma\sum_{j=1}^{m_6}w_j^{(6)}\,(\epsilon\nabla U(t_j^{(6)}),\nabla v)+(\mathbf{b}\cdot\nabla U(t_j^{(6)}),v).$$

Theorem 3.2 gives the weights which ensure that the second-order IMEX scheme and cG(1) finite element method with a particular quadrature are equivalent.

4. *A posteriori* analysis of IMEX schemes

In this section, we derive *a posteriori* error estimates. Note that there is no unique definition for adjoint operators corresponding to nonlinear operators. We employ a definition based on linearization which is useful for error analysis.

An *a posteriori* error analysis for multi-stage methods has been developed in [33]. *A posteriori* analysis of the multi-step BDF schemes based on a probabilistic estimate is presented in [34]. Another recent approach based on Petrov–Galerkin type finite element methods is analyzed in [35]. Our approach for the multi-step IMEX schemes is based on employing quadrature and weighted sum of finite element schemes over adjacent intervals as presented in Section 3. The idea of employing quadrature for *a posteriori* analysis has been explored earlier in adjoint based analysis, e.g. see [36].

In this section it is important to note that in the case of spatially semi-discretized PDEs, we quantify the error in a Quantity of Interest (QoI) for the resulting system of ODEs, not the error in the original PDE. The error analysis for this case is still significant since a method of lines approach using a fixed spatial discretization and an ODE solver is a commonly used method for solving PDEs. For the space–time discretization, the estimates quantify the error in the discrete solution of the PDE (that is, the error in this case is the difference between the true solution of the PDE and the discrete solution of the PDE).

4.1. A posteriori analysis for the IMEX time integration of the semidiscrete formulation

Consider the (temporal) finite element method in (2.5)

$$\begin{cases}\displaystyle\sum_{n=1}^{N}\langle\dot{\widetilde{U}}-f(\widetilde{U})-g(\widetilde{U}),v\rangle_{I_n}=0\quad\forall v\in\widetilde{W}^{q-1}(I_n),\\ \widetilde{U}(0)=u_0.\end{cases}\tag{4.1}$$

Let $h(\widetilde{u})=f(\widetilde{u})+g(\widetilde{u})$, $\widetilde{e}=\widetilde{u}-\widetilde{U}$, and $z=s\widetilde{u}+(1-s)\widetilde{U}$, and define the linearized operator $\overline{H(\widetilde{u},\widetilde{U})}$ such that

$$\overline{H(\widetilde{u},\widetilde{U})}\widetilde{e}=\int_0^1 h'(z)\,\mathrm{d}s=\int_0^1 f'(z)+g'(z)\,\mathrm{d}s=(f(\widetilde{u})-f(\widetilde{U}))+(g(\widetilde{u})-g(\widetilde{U})).\tag{4.2}$$

Assuming that we are interested in a QoI that is a linear functional of the solution at some time T, i.e., QoI = $(\widetilde{u}(T), \psi)$, we define the adjoint problem as,

$$\begin{cases} -\dot{\widetilde{\phi}} = \overline{H(\widetilde{u}, \widetilde{U})}^{\top} \widetilde{\phi}, & t \in [0, T), \\ \widetilde{\phi}(T) = \psi. \end{cases} \tag{4.3}$$

Nonlinear QoIs require special treatment, and are often dealt by linearization of the QoI [37,38]. The rest of the ideas are the same, and hence we limit ourselves to linear QoIs in this paper. We have the following error representation formula, based on (4.3).

Theorem 4.1 (*Error Representation*)**.** *Let $\widetilde{e}_n = \widetilde{u}_n - \widetilde{U}_n$ and $\widetilde{\phi}_n$ denote the error and adjoint solution at time t_n respectively, then*

$$(\widetilde{e}_n, \widetilde{\phi}_n) = (\widetilde{e}_{n-1}, \widetilde{\phi}_{n-1}) + \langle f(\widetilde{U}) + g(\widetilde{U}) - \dot{\widetilde{U}}, \widetilde{\phi} \rangle_{I_n}. \tag{4.4}$$

Proof. The proof is standard, e.g. see [32]. □

Lemma 4.2 (*Error Representation for Quadrature for the Interval I_n*)**.**

$$(\widetilde{e}_n, \widetilde{\phi}_n) = (\widetilde{e}_{n-1}, \widetilde{\phi}_{n-1}) + DE_n + QE_{f,n} + QE_{g,n}, \tag{4.5}$$

where

$$\begin{aligned} DE_n &= \langle f(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_1}} + \langle g(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_2}} - \langle \dot{\widetilde{U}}, \widetilde{\phi} \rangle_{I_n}, \\ QE_{f,n} &= \langle f(\widetilde{U}), \widetilde{\phi} \rangle_{I_n} - \langle f(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_1}}, \\ QE_{g,n} &= \langle g(\widetilde{U}), \widetilde{\phi} \rangle_{I_n} - \langle g(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_2}}. \end{aligned}$$

Proof. From (4.4),

$$(\widetilde{e}_n, \widetilde{\phi}_n) = (\widetilde{e}_{n-1}, \widetilde{\phi}_{n-1}) + \langle f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi} \rangle_{I_n} - \langle \dot{\widetilde{U}}, \widetilde{\phi} \rangle_{I_n}.$$

Adding and subtracting $\langle f(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_1}}$ and $\langle g(\widetilde{U}), \widetilde{\phi} \rangle_{I_{n,m_2}}$ proves the result. □

The term DE_n in Eq. (4.5) describes the contribution to the total error due to discretization, whereas the terms $QE_{f,n}$ and $QE_{g,n}$ describe the error contributions arising from the approximation of integrals involving f and g respectively.

Remark 4.1. We may apply Galerkin orthogonality to Eq. (4.5) to modify DE_n as

$$DE_n = \langle f(\widetilde{U}), \widetilde{\phi} - \pi\widetilde{\phi} \rangle_{I_{n,m_1}} + \langle g(\widetilde{U}), \widetilde{\phi} - \pi\widetilde{\phi} \rangle_{I_{n,m_2}} - \langle \dot{\widetilde{U}}, \widetilde{\phi} - \pi\widetilde{\phi} \rangle_{I_n}, \tag{4.6}$$

where π is a projection operator from $H^1(0, T)$ to $\widetilde{W}^q(0, T)$. This form is potentially more useful to localization of error and adaptive refinement.

We sum the errors on each time interval to construct the following error representation for first-order IMEX schemes.

Theorem 4.3 (*Error Representation for First-Order IMEX Schemes (3.2)*)**.** *The error, $\widetilde{e} = \widetilde{u} - \widetilde{U}$, in the quantity of interest ψ, at the final time $t_N = T$, for first-order IMEX schemes is given by,*

$$(\widetilde{e}_N, \widetilde{\phi}_N) = \underbrace{DE}_{\text{Discretization error}} + \underbrace{QE_f}_{\text{Quadrature Error for } f} + \underbrace{QE_g}_{\text{Quadrature Error for } g}, \tag{4.7}$$

where

$$DE = \sum_{n=1}^{N} DE_n, \qquad QE_f = \sum_{n=1}^{N} QE_{f,n}, \qquad QE_g = \sum_{n=1}^{N} QE_{g,n}. \tag{4.8}$$

Proof. From Lemma 4.2 the error over time step (t_{n-1}, t_n) is,

$$(\widetilde{e}_n, \widetilde{\phi}_n) = (\widetilde{e}_{n-1}, \widetilde{\phi}_{n-1}) + \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n,m_1}} + \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n,m_2}} - \langle \dot{\widetilde{U}}, \widetilde{\phi}\rangle_{I_n} + \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} - \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n,m_1}} + \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} - \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n,m_2}}. \tag{4.9}$$

Summing $(\widetilde{e}_n, \widetilde{\phi}_n)$ for $n = 1, 2, \ldots, N$, and assuming that $\widetilde{e}_0 = 0$, proves the result. □

Theorem 4.4 (*Error Representation for Second-Order IMEX Schemes (3.3)*)**.** *The error, $\widetilde{e} = \widetilde{u} - \widetilde{U}$, in the quantity of interest ψ, at the final time $t_N = T$, for second-order IMEX schemes is given by,*

$$(\widetilde{e}_n, \widetilde{\phi}_N) = E_0 + E_N + \underbrace{DE}_{\text{Discretization error}} + \underbrace{QE_f}_{\text{Quadrature Error for } f} + \underbrace{QE_g}_{\text{Quadrature Error for } g}, \tag{4.10}$$

where

$$\begin{aligned}
DE \;=\; & \sum_{n=1}^{N}\Bigg[\left(\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}}, \widetilde{\phi}\rangle_{I_n} + \left(-\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}}, \widetilde{\phi}\rangle_{I_{n-1}} + \frac{1}{2}\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_1} \\
& + \frac{1}{2}\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_2} + \gamma\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_n,m_3} + \gamma\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n,m_4} \\
& - \gamma\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1},m_5} - \gamma\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1},m_6}\Bigg], \\
QE_f \;=\; & \sum_{n=2}^{N}\Bigg[\frac{1}{2}(\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\}} - \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_1}) \\
& + \gamma(\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} - \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_n,m_3}) - \gamma(\langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1}} - \langle f(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1},m_5})\Bigg], \\
QE_g \;=\; & \sum_{n=2}^{N}\Bigg[\frac{1}{2}(\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\}} - \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_2}) \\
& + \gamma(\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} - \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n,m_4}) - \gamma(\langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1}} - \langle g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1},m_6})\Bigg], \\
E_0 = & \left(\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}} + f(\widetilde{u}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_1}, \\
E_N = & \left(-\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}} + f(\widetilde{u}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_N}.
\end{aligned} \tag{4.11}$$

Proof. Summing (4.4) over the N intervals leads to,

$$\begin{aligned}
(\widetilde{e}_N, \widetilde{\phi}_N) &= \sum_{n=1}^{N}\langle -\dot{\widetilde{U}} + f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} \\
&= \left(\gamma+\frac{1}{2}\right)\sum_{n=1}^{N}\langle -\dot{\widetilde{U}} + f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} + \left(-\gamma+\frac{1}{2}\right)\sum_{n=1}^{N}\langle -\dot{\widetilde{U}} + f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} \\
&= E_0 + \sum_{n=2}^{N}\Bigg[\left(\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}} + f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_n} \\
&\quad + \left(-\gamma+\frac{1}{2}\right)\langle -\dot{\widetilde{U}} + f(\widetilde{U}) + g(\widetilde{U}), \widetilde{\phi}\rangle_{I_{n-1}}\Bigg] + E_N.
\end{aligned} \tag{4.12}$$

Adding and subtracting

$$\sum_{n=2}^{N}\Big[\frac{1}{2}\langle f(\widetilde{U}),\widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_1}+\frac{1}{2}\langle g(\widetilde{U}),\widetilde{\phi}\rangle_{\{I_{n-1}\cup I_n\},m_2}$$
$$+\gamma\langle f(\widetilde{U}),\widetilde{\phi}\rangle_{I_n,m_3}+\gamma\langle g(\widetilde{U}),\widetilde{\phi}\rangle_{I_n,m_4}-\gamma\langle f(\widetilde{U}),\widetilde{\phi}\rangle_{I_{n-1},m_5}-\gamma\langle g(\widetilde{U}),\widetilde{\phi}\rangle_{I_{n-1},m_6}\Big], \tag{4.13}$$

proves the theorem. □

4.2. A posteriori analysis for the space–time formulation with IMEX time integration

Recall that the space–time finite element method for (1.1) is,

$$\int_{t_{n-1}}^{t_n}(\dot{U},v)+(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\,\mathrm{d}t=\int_{t_{n-1}}^{t_n}(R(U),v)\,\mathrm{d}t\quad\forall v\in W_n^0,\ n=1\cdots N. \tag{4.14}$$

We now define a linearized form of the nonlinear reaction term to form an adjoint. Let

$$\overline{R(u,U)}=\int_0^1\frac{\partial R}{\partial u}(z)\,\mathrm{d}s, \tag{4.15}$$

where $z=su+(1-s)U$. Consequently, in an argument similar to (4.2), we arrive at

$$R(u)-R(U)=\overline{R(u,U)}(u-U). \tag{4.16}$$

With this linearized form, we define the adjoint problem as,

$$\begin{cases}-\dot{\phi}(x,t)-\nabla\cdot\epsilon\nabla\phi-\nabla\cdot(\mathbf{b}\phi)=\overline{R(u,U)}\phi, & (x,t)\in\Omega\times(0,T],\\ \phi(x,T)=\psi, & x\in\Omega.\end{cases} \tag{4.17}$$

This leads to the following error representation for the cG(1) method.

Theorem 4.5 (*Error Representation for the Space–Time cG(1) Solution to* (1.1))**.**

$$(e_n,\phi_n)=(e_{n-1},\phi_{n-1})+\int_{t_{n-1}}^{t_n}(-\dot{U},\phi)-(\epsilon\nabla U,\nabla\phi)-(\mathbf{b}\cdot\nabla U,\phi)+(R(U),\phi)\,\mathrm{d}t. \tag{4.18}$$

Proof. In view of (4.16), this is standard. □

Lemma 4.6 (*Error Representation for Quadrature for the Interval* I_n)**.**

$$(e_n,\phi_n)=(e_{n-1},\phi_{n-1})+\mathcal{DE}_n+\mathcal{QE}_{f,n}+\mathcal{QE}_{g,n}, \tag{4.19}$$

where,

$$\mathcal{DE}_n=\langle(R(U),\phi)\rangle_{I_{n,m_1}}-\langle(\epsilon\nabla U,\nabla\phi)+(\mathbf{b}\cdot\nabla U,\phi)\rangle_{I_{n,m_2}}-\langle(\dot{U},\phi)\rangle_{I_n},$$
$$\mathcal{QE}_{f,n}=\langle(R(U),\phi)\rangle_{I_n}-\langle(R(U),\phi)\rangle_{I_{n,m_1}},$$
$$\mathcal{QE}_{g,n}=-\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_n}+\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_{n,m_2}}.$$

Proof. The proof is similar to Lemma 4.2 and follows by adding and subtracting $\langle(R(U),\phi)\rangle_{I_{n,m_1}}$ and $\langle(\epsilon\nabla U,\nabla v)+(\mathbf{b}\cdot\nabla U,v)\rangle_{I_{n,m_2}}$ to (4.18). □

The term $\mathcal{DE}_n$ measures the discretization error, while the terms $\mathcal{QE}_{f,n}$ and $\mathcal{QE}_{g,n}$ describe the numerical quadrature errors.

Theorem 4.7 (*Error Representation for First-Order IMEX Schemes (3.8)*)**.** *The error, $e = u - U$, in the quantity of interest ψ, at the final time $t_N = T$ is estimated as,*

$$(e_n, \phi_N) = \underbrace{\mathcal{DE}}_{\text{Discretization error}} + \underbrace{\mathcal{QE}_f}_{\text{QE for reaction}} + \underbrace{\mathcal{QE}_g}_{\text{QE for convection/diffusion}} \tag{4.20}$$

where

$$\mathcal{DE} = \sum_{n=1}^{N} \mathcal{DE}_n, \qquad \mathcal{QE}_f = \sum_{n=1}^{N} \mathcal{QE}_{f,n}, \qquad \mathcal{QE}_g = \sum_{n=1}^{N} \mathcal{QE}_{g,n}. \tag{4.21}$$

Proof. From Lemma 4.6, the error over time step (t_{n-1}, t_n) is

$$\begin{aligned}(e_n, \phi_n) = {} & (e_{n-1}, \phi_{n-1}) + \langle (R(U), \phi) \rangle_{I_{n,m_1}} - \langle (\epsilon \nabla U, \nabla \phi) + (\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_{n,m_2}} - \langle (\dot{U}, \phi) \rangle_{I_n} \\ & + \langle (R(U), \phi) \rangle_{I_n} - \langle (R(U), \phi) \rangle_{I_{n,m_1}} - \langle (\epsilon \nabla U, \nabla v) + (\mathbf{b} \cdot \nabla U, v) \rangle_{I_n} \\ & - \langle (\epsilon \nabla U, \nabla v) + (\mathbf{b} \cdot \nabla U, v) \rangle_{I_{n,m_2}}. \end{aligned} \tag{4.22}$$

Summing (e_n, ϕ_n) for $n = 1, 2, \ldots, N$, and assuming that $e_0 = 0$, proves the result. □

Theorem 4.8 (*Error Representation for Second-Order IMEX Schemes (3.9)*)**.** *The error in the QoI at the final time is,*

$$(e_n, \phi_N) = \mathcal{E}_0 + \mathcal{E}_N + \underbrace{\mathcal{DE}}_{\text{Discretization error}} + \underbrace{\mathcal{QE}_f}_{\text{QE for reaction}} + \underbrace{\mathcal{QE}_g}_{\text{QE for convection/diffusion}}, \tag{4.23}$$

where

$$\begin{aligned}\mathcal{DE} = {} & \sum_{n=2}^{N} \Bigg[\left(\gamma + \frac{1}{2}\right) \langle (-\dot{U}, \phi) \rangle_{I_n} + \left(-\gamma + \frac{1}{2}\right) \langle (-\dot{U}, \phi) \rangle_{I_{n-1}} + \frac{1}{2} \langle (R(U), \phi) \rangle_{\{I_{n-1} \cup I_n\}, k_1} \\ & + \frac{1}{2} \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{\{I_{n-1} \cup I_n\}, m_1} \\ & + \gamma \langle (R(U), \phi) \rangle_{I_n, k_2} + \gamma \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_n, m_2} \\ & - \gamma \langle R(U), \phi \rangle_{I_{n-1}, k_2} - \gamma \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_{n-1}, m_2} \Bigg], \\ \mathcal{QE}_f = {} & \sum_{n=2}^{N} \Bigg[\frac{1}{2} \Big(\langle (R(U), \phi) \rangle_{\{I_{n-1} \cup I_n\}} - \langle (R(U), \phi) \rangle_{\{I_{n-1} \cup I_n\}, k_1} \Big) \\ & + \gamma \Big(\langle R(U), \phi \rangle_{I_n} - \langle R(U), \phi \rangle_{I_n, k_2} \Big) - \gamma \Big(\langle R(U), \phi \rangle_{I_{n-1}} - \langle R(U), \phi \rangle_{I_{n-1}, k_2} \Big) \Bigg], \\ \mathcal{QE}_g = {} & \sum_{n=2}^{N} \Bigg[\frac{1}{2} (\langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{\{I_{n-1} \cup I_n\}} \\ & - \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{\{I_{n-1} \cup I_n\}, m_1}) \\ & + \gamma \Big(\langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_n} - \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_n, m_2} \Big) \\ & - \gamma \Big(\langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_{n-1}} - \langle (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_{n-1}, m_2} \Big) \Bigg], \\ \mathcal{E}_0 = {} & \left(\gamma + \frac{1}{2}\right) \langle (-\dot{U} + R(U), \phi) + (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_1}, \\ \mathcal{E}_N = {} & \left(-\gamma + \frac{1}{2}\right) \langle (-\dot{U} + R(U), \phi) + (-\epsilon \nabla U, \nabla \phi) + (-\mathbf{b} \cdot \nabla U, \phi) \rangle_{I_N}. \end{aligned} \tag{4.24}$$

Proof. The proof is similar to the one for Theorem 4.4. □

Remark 4.2. We may further decompose the $\mathcal{DE}$ term into separate error contributions from spatial and temporal discretizations. This involves adding and subtracting suitable projections of the adjoint solution to the expression involving $\mathcal{DE}$, see [31] for details. Such a decomposition is useful when considering adaptive refinement of the spatial or temporal mesh.

5. Numerical examples

We present numerical examples exploring the accuracy and behavior of the error estimates for the IMEX schemes.

5.1. Computational details

The computational examples share the following details.

Spatial discretization

We discretize in space using continuous piecewise linear finite elements with respect to a uniform triangulation of Ω. For the semidiscrete problem, we employ the trapezoidal rule quadrature (lumped mass quadrature) for evaluation of the integrals $(\dot{u}^h, v)$ and $(R(u^h), v)$, and evaluating the integrals $(\epsilon\nabla u^h, \nabla v)$ and $(\mathbf{b}\cdot\nabla u^h, v)$ exactly. This is equivalent to using a second order central finite difference scheme in space. For the space–time finite element method we evaluate the spatial integrals using a Gauss quadrature rule.

Implementation of the error estimates

The error representation formulas require formulating and solving the adjoint (4.3). In theory, the adjoint is obtained by linearizing around a combination of the discrete solution and the true solution. In practice it is common to linearize around the discrete solution only [17] and we follow this approach, that is, we approximate $\tilde{u}$ by $\tilde{U}$ and u by U in (4.3) and (4.17) respectively. This "linearization error" can negatively affect the accuracy of the error estimates in some cases, see example in Section 5.2.4. Since we effectively require time derivatives of the adjoint solution, we solve the adjoints in time using the cG(1) method but employing time steps 4 times smaller than the steps used in the forward problem. In the case of the full space–time discretization, the adjoint is solved using a continuous piecewise quadratic finite element method in space.

We note that computing the error estimate involves the cost of solving the adjoint problem in addition to computing the original approximation. The computational cost depends on how the numerical adjoint problem is discretized in space and time, however the adjoint problem is at least linear, and hence avoids a nonlinear solution technique which is often required for the original problem. We also note that while we use the fully implicit cG(1) method to solve the adjoint problem, alternate methods such as an IMEX scheme may also be employed. Moreover, the computation of estimates involves evaluation of the original approximation at all time steps and not simply at the final time. For large scale problems, storing the solution at each time step may be a concern and large scale implementations may require efficient checkpointing/recomputational type algorithm (see e.g. [39]) or some type of data compression/reconstruction methods to approximate the forward solution [40,41] and control storage.

Verification of the accuracy of the estimates

In the numerical examples, we plot the different error contributions, as well the effectivity ratios. The effectivity ratio measures the accuracy of the estimator and is defined as,

$$\rho_{\text{eff}} = \frac{\text{Estimated error}}{\text{True error}}.$$

An accurate error estimator has an effectivity ratio close to one. In cases in which the true solution is unknown, we compute a more accurate reference numerical solution using finer spatial and temporal meshes and a second-order IMEX scheme for the time integration.

5.2. Examples using IMEX time integration for semidiscrete problems

In this section we present examples of the schemes considered in (3.2) and (3.3). For illustrative purposes and to compare the results with previous studies we apply the error estimation technique to a number of scalar-valued PDEs in one-dimensional spatial domains.

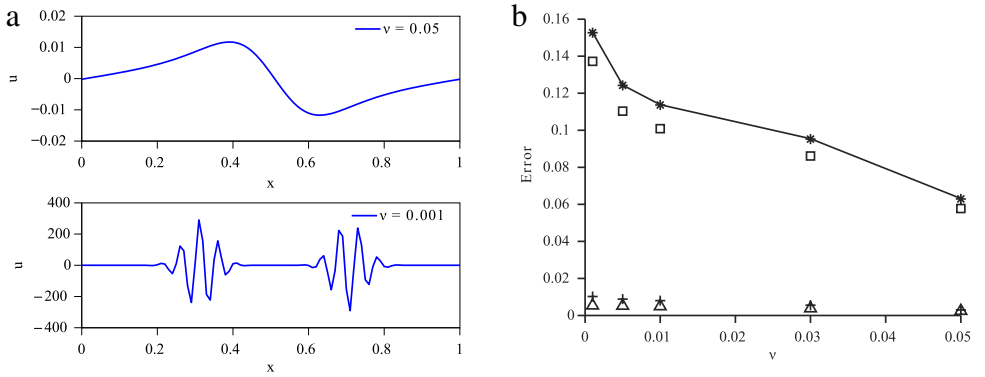


Fig. 5.1. (a) First-order SBDF solutions for two different ν values at $T = 1.0$. The solution develops oscillations as ν decreases. (b) Error contributions for the example in Section 5.2.1 using a first-order scheme, $h = 0.01$, $\Delta t = 2h$. Here $*$ is the true error, – (the connected line) is the error estimated using the error representation formula, $\square$ is the discretization error (*DE*), $+$ is the quadrature error for f (QE_f) and $\triangle$ is the quadrature error for g (QE_g).

For all examples except in Section 5.2.4, the QoI is the average value of the solution over the first half of the one-dimensional spatial domain at the final time. That is, if the spatial domain is $[a, b]$, then the QoI was given by

$$\text{QoI} = (u, \psi) = \int_a^{\frac{a+b}{2}} u(x, T) \quad \mathrm{d}x. \tag{5.1}$$

In the numerical experiments below, the spatial discretization parameter is chosen to be $h = 1/100$. For second-order IMEX schemes, the "true" solution is computed to a high degree of accuracy using MATLAB's ODE solver (ode23s). We note that in these examples the "true" solution was the solution to the ODE after spatial discretization, not the solution to the original PDE and the error we seek to estimate is the error with respect to the true solution of the ODE problem. Apart from the total error, we also indicate different contributions due to discretization, quadrature for f and quadrature for g. These terms, *DE*, QE_f and QE_g are defined in Theorems 4.3 and 4.4.

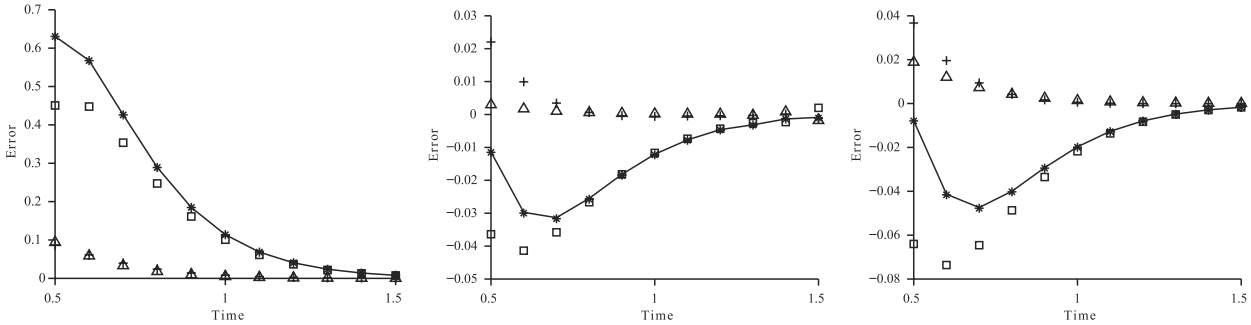
The first three examples arise from the finite difference discretization of scalar-valued PDEs, all of which were previously considered in [10,11]. Consistent with previous authors, we chose f to be the term arising from the first-order spatial derivatives, while g represented the diffusive term (that is, the term arising from second-order spatial derivatives). The error estimates are quite accurate for all three. The final two examples concern cases when the true error in the discrete solution is quite large. This may be due to the choice of the problem (example in Section 5.2.4), or due to the choice of the functions f and g (example in Section 5.2.5).

5.2.1. Linear PDE

Consider the scalar valued linear PDE

$$\begin{cases} \dot{u} + \sin(2\pi x) u_x = \nu u_{xx}, & (x, t) \in [0, 1] \times (0, T], \\ u(x, 0) = \sin(2\pi x), & x \in [0, 1], \end{cases} \tag{5.2}$$

with periodic boundary conditions. We choose $\Delta t = 2h$, where we recall that $h = 1/100$. It should be noted as the dissipation (ν) goes to zero the centered difference approximation to the first order convection operator results in the discretized solution to the PDE being unstable as shown in Fig. 5.1(a). However, as discussed earlier, once the spatial discretization is fixed, we are analyzing the error in the ODE system, not the error in the PDE. Fig. 5.1(b) shows the error estimates for the first-order SBDF scheme as ν was decreased. These estimates remain accurate even for small ν and show that the discretization error, *DE* is the dominant error. Fig. 5.2(a) shows the results for different final times T for the first-order SBDF scheme at fixed $\nu = 0.01$ and the estimator captures the error quite accurately in all cases. For small T, all three components of error are significant. Fig. 5.2(b) and (c) show the error components arising from the second-order CNAB and SBDF schemes. We observe that the error due to the second order schemes is considerably less than for the first-order scheme. The effectivity ratios in Table 5.1 quantify the accuracy of the error estimates for the first-order scheme and for both second-order schemes.



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

Fig. 5.2. Error contributions for the example in Section 5.2.1 for $\nu = 0.01$. Here $(*)$ true error, $(-)$ error estimate, $(\square)$ discretization error (DE), $(+)$ quadrature error for f (QE_f) and $(\triangle)$ quadrature error for g (QE_g).

Table 5.1
Effectivity ratios for the example in Section 5.2.1.

(a) ρ_{eff} for Fig. 5.1(b)		(b) ρ_{eff} for Fig. 5.2(a)		(c) ρ_{eff} for Fig. 5.2(b)		(d) ρ_{eff} for Fig. 5.2(c)	
ν	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.001	1.0007	0.5	1.0003	0.5	0.9888	0.5	0.9833
0.005	1.0012	0.7	1.0008	0.7	0.9904	0.7	0.9937
0.01	1.0015	0.9	1.0012	0.9	0.9873	0.9	0.9921
0.03	1.0022	1.1	1.0019	1.1	0.9821	1.1	0.9888
0.05	1.0033	1.3	1.0033	1.3	0.9712	1.3	0.9829
		1.5	1.0061	1.5	0.8295	1.5	0.9715

Table 5.2
Effectivity ratios for the example in Section 5.2.2.

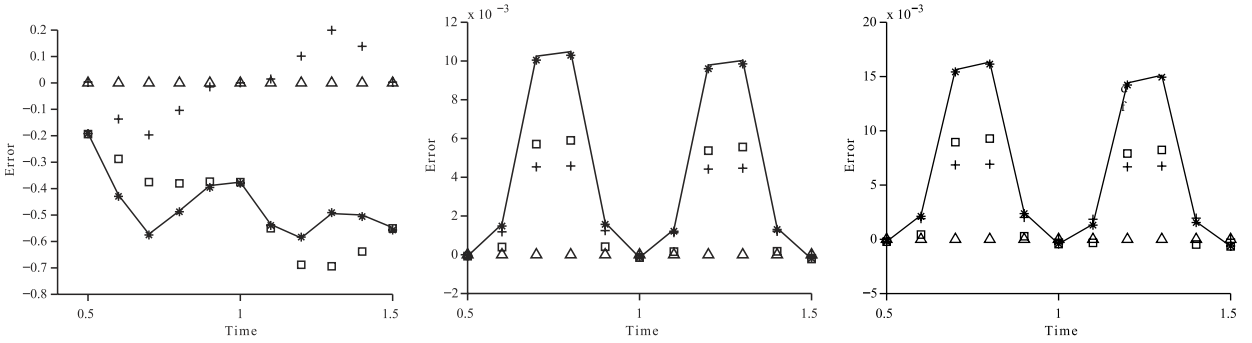
(a) ρ_{eff} for Fig. 5.3(a)		(b) ρ_{eff} for Fig. 5.3(b)		(c) ρ_{eff} for Fig. 5.3(c)	
T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.5	0.9941	0.5	1.1782	0.5	1.0544
0.7	0.9941	0.7	1.0201	0.7	1.0132
0.9	0.9822	0.9	1.0570	0.9	1.0378
1.1	0.9929	1.1	1.0736	1.1	1.0683
1.3	1.0070	1.3	1.0179	1.3	1.0119
1.5	0.9836	1.5	1.168	1.5	1.0512

5.2.2. *Non-linear PDE with explicit terms only*

Consider

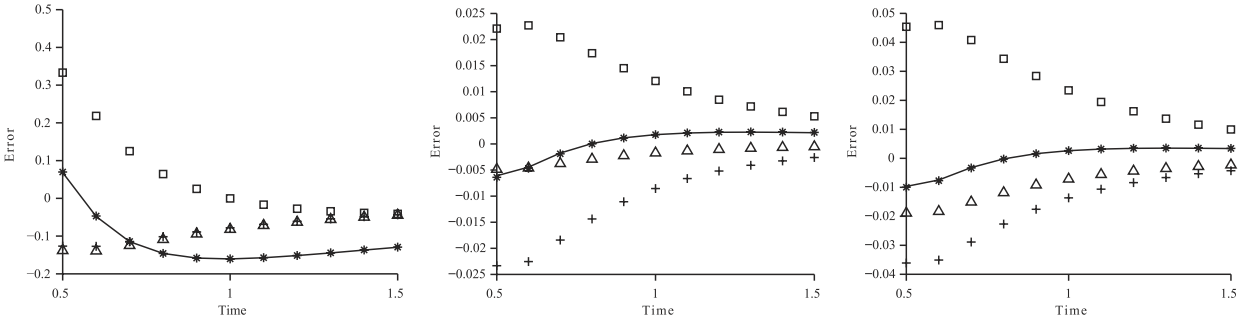
$$\begin{cases} \dot{u} + \dfrac{1}{2}\cos(2\pi t)(1+u)u_x = 0, & (x,t) \in [0,1] \times (0,T], \\ u(x,0) = \sin(2\pi x), & x \in [0,1], \end{cases} \tag{5.3}$$

with periodic boundary conditions. Here $g(t) \equiv 0$. We chose $\Delta t = 0.5h$, where we recall that $h = 1/100$. The results for different final times T for the first-order and second-order schemes are shown in Fig. 5.3(a)–(c). We observe that the component QE_g is zero due to the absence of the stiff term. There is significant cancellation of error components for $T > 1.1$ in the first-order SBDF scheme, however the discretization error is the dominant term. The second-order schemes are again more accurate than the first-order scheme. Moreover, both discretization error and quadrature error for f are comparable in magnitude for the second-order schemes. The effectivity ratios, shown in Table 5.2, highlight the accuracy of the estimator.



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

Fig. 5.3. Error components for the example in Section 5.2.2. Here (∗) true error, (–) error estimate, (□) discretization error (*DE*), (+) quadrature error for f (QE_f) and (△) quadrature error for g (QE_g).



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

Fig. 5.4. Error components for Burgers' equation example in Section 5.2.3 with $\nu = 0.01$. Here (∗) true error, (–) error estimate, (□) discretization error (*DE*), (+) quadrature error for f (QE_f) and (△) quadrature error for g (QE_g).

Table 5.3
Effectivity ratios for Burgers' equation example in Section 5.2.3.

(a) ρ_{eff} for Fig. 5.4(a)		(b) ρ_{eff} for Fig. 5.4(b)		(c) ρ_{eff} for Fig. 5.4(c)	
T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.5	0.9789	0.5	0.9538	0.5	0.9707
0.7	1.0036	0.7	0.9250	0.7	0.9580
0.9	1.0004	0.9	1.0205	0.9	1.0141
1.1	1.0010	1.1	0.9884	1.1	0.9922
1.3	1.0019	1.3	0.9805	1.3	0.9873
1.5	1.0027	1.5	0.9760	1.5	0.9845

5.2.3. *Damped non-linear Burgers' equation*

The damped non-linear Burgers' equation is

$$\begin{cases} \dot{u} + uu_x = \nu u_{xx}, & (x,t) \in [-1,1] \times (0,T], \\ u(x,0) = \sin(\pi x), & x \in [-1,1], \end{cases} \tag{5.4}$$

which we consider with periodic boundary conditions. We chose $\Delta t = 1/160$ and recall that we chose $h = 1/100$. The computational results appear in Fig. 5.4(a)–(c), and show similar characteristics to the first two examples. The error for second-order is predictably less than the error for first-order schemes. Moreover, discretization error dominates for first-order schemes. For second-order schemes, both discretization error and quadrature error for f are large, whereas the quadrature error for g is relatively small. Finally, the effectivity ratios are given in Table 5.3 and demonstrate the accuracy of the estimator.

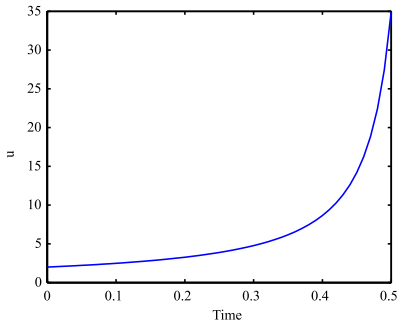


Fig. 5.5. IMEX (first-order SBDF) solution for the example in Section 5.2.4.

5.2.4. Blow up in finite time

Now we consider a scalar-valued ODE which has a blow up in finite time,

$$\begin{cases} u_t + \lambda u = u^2, \quad t \in (0, T], \\ u(0) = u_0. \end{cases} \tag{5.5}$$

This equation has finite time blow up when $\lambda < u_0$. We chose $\lambda = 0.01$, $u_0 = 2$, $f(u) = u^2$ and $g(u) = \lambda u$, and $\Delta t = 0.01$ and the QoI to be the value of the solution at the final time. The IMEX solution for $t = [0, 0.5]$ is shown in Fig. 5.5. Table 5.4 provides the results for different final times for the first-order SBDF scheme. The estimator is accurate for small T when the discrete and the true solutions are close to each other, but becomes inaccurate near the blow up when the discrete solution has significant error. This inaccuracy is due to the linearization of the computed adjoint using only the discrete solution. When the discrete solution is highly inaccurate, this approximation is no longer valid. However, we note that the estimator is reliable in the sense that even though it does not reflect the true value of the error, it does indicate the error is quite large.

5.2.5. Linear PDE—effect of choice of f and g

We revisit the linear PDE in Section 5.2.1 to investigate the effects on error based on the choice of f and g. The choice made in Section 5.2.1 was quite obvious, however, this choice may not be obvious for complicated systems. Accordingly, we perform two experiments with different choices of f and g. In both cases, we set $\nu = 1$, $h = 0.1$ and $\Delta t = 0.01$ and use the first-order SBDF scheme.

In the first case, we deliberately make the unstable choice and set $f(u) = \nu u_{xx}$ and $g(u) = \sin(2\pi x)u_x$. Since $0.5h/\Delta t^2 > 1$, we know from standard finite difference theory that the numerical solution exhibits unbounded growth. We show the results for $T = 0.91$ in Table 5.5. The estimator captures the true error quite well, even though the error is quite large. Further, we see that the contribution to the explicit quadrature, $\mathcal{QE}_f$, is the significant term in this example. Since in all previous examples for first-order IMEX time integration the discretization error has been dominant, this suggests a different choice of f and g may lead to a more accurate solution. We verify this by switching the choices of f and g. The results are again shown in Table 5.5, which shows that now we have a significant decrease in the error.

5.3. Examples using IMEX time integration for the space–time formulation

Now we present numerical examples for the analysis in Section 4.2, which takes into account the effects of spatial discretization. Our results indicate the total estimate as well as contribution due to discretization, quadrature for f and quadrature for g. These terms, $\mathcal{DE}$, $\mathcal{QE}_f$ and $\mathcal{QE}_g$, are defined in Theorems 4.7 and 4.8.

5.3.1. Linear PDE

Consider

$$\begin{cases} \dot{u} - \nabla^2 u = \pi^2 u, & (x, y, t) \in \Omega \times (0, T], \\ u(x, y, t) = 0, & (x, y, t) \in \partial\Omega \times (0, T], \\ u(x, y, 0) = \sin(2\pi x)\sin(2\pi y), & (x, y) \in \Omega. \end{cases} \tag{5.6}$$

Table 5.4
Errors and effectivity ratios for the example in Section 5.2.4.

T	True error	Estimated error	ρ_{eff}
0.1	0.013341	0.013326	0.9989
0.2	0.053041	0.05288	0.9970
0.3	0.20381	0.2016	0.9893
0.4	1.2296	1.1680	0.9499
0.5	765.76	53.13	0.0694

Table 5.5
Errors for different choices of f and g for the example in Section 5.2.5 for first-order SBDF scheme.

f	g	True err.	Err est.	ρ_{eff}	DE	QE_g	QE_f
νu_{xx}	$\sin(2\pi x)u_x$	175.97	178.40	1.01	9.916	52.946	115.54
$\sin(2\pi x)u_x$	νu_{xx}	−0.0043	−0.0043	1.00	−0.005	5.6e−4	−6.1e−6

Table 5.6
Effectivity ratios for the example in Section 5.3.1.

(a) ρ_{eff} for Fig. 5.6(a)		(b) ρ_{eff} for Fig. 5.6(b)		(c) ρ_{eff} for Fig. 5.6(c)	
T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.5	0.9979	0.5	0.9956	0.5	0.9992
0.7	0.9970	0.7	1.0765	0.7	0.9932
0.9	0.9966	0.9	1.0059	0.9	1.0126
1.1	0.9965	1.1	1.0693	1.1	1.0012
1.3	0.9964	1.3	1.0024	1.3	0.9952
1.5	0.9964	1.5	0.9759	1.5	1.0155

The true solution is given by

$$u = e^{-\pi^2 t}\sin(2\pi x)\sin(2\pi y).$$

The QoI is $(u(T), \psi)$ with $\psi = \sin(2\pi x)\sin(2\pi y)$. As in our space–time analysis above we choose the source term to be represented in the explicit term, f, and the diffusion term to be represented in the implicit term, g. The spatial domain is the unit square which is discretized using a uniform triangular mesh of 30 elements in each direction. The computational results are shown in Fig. 5.6(a)–(c). The different components of the error sum to accurately capture the error in the solution. The error in the second-order schemes is considerably less than the first-order scheme. All components, $\mathcal{DE}$, $\mathcal{QE}_f$ and $\mathcal{QE}_g$, contribute relatively equally to the error. The effectivity ratios are given in Table 5.6 and highlight the accuracy of the estimator.

5.3.2. Thermal wave

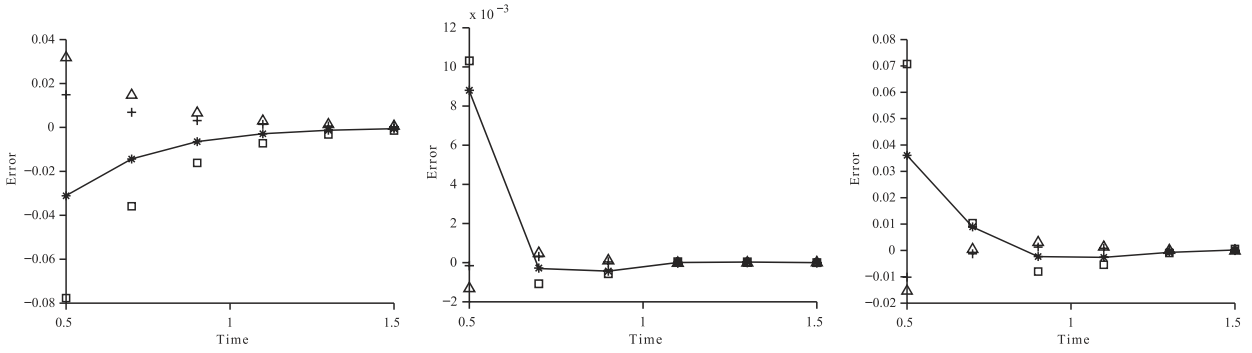
We consider the thermal wave problem [42],

$$\begin{cases} \dot{u} - \nu\nabla^2 u + \mathbf{b}(x)\cdot\nabla u(x,t) = 8\dfrac{\nu}{\delta^2}u^2(1-u) & (x,y,t)\in\Omega\times(0,T], \\ u(x,y,0) = \dfrac{1}{2}\left(1-\tanh\left[\dfrac{x}{\delta}\right]\right), & (x,y)\in\Omega \end{cases} \tag{5.7}$$

where $\mathbf{b}(x) = [\epsilon/\delta\,,\ 0]^\top$. The true solution is,

$$u(x,y,t) = \frac{1}{2}\left(1-\tanh\left[\frac{x-(2\nu-\epsilon)t/\delta}{\delta}\right]\right). \tag{5.8}$$

The spatial domain is the rectangle $[-10\,,\ 10]\times[-1\,,\ 1]$. We set Dirichlet boundary conditions (given by the true solution) along the x boundary, and Neumann boundary conditions along the y boundary. Moreover, we set



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

Fig. 5.6. Error components for the example in Section 5.3.1. Here (∗) true error, (–) error estimate, (□) discretization error ($\mathcal{DE}$), (+) quadrature error for f ($\mathcal{QE}_f$) and (△) quadrature error for g ($\mathcal{QE}_g$).

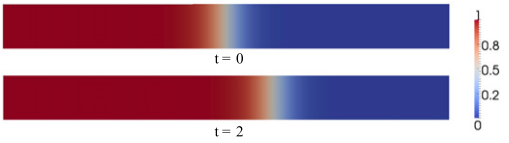
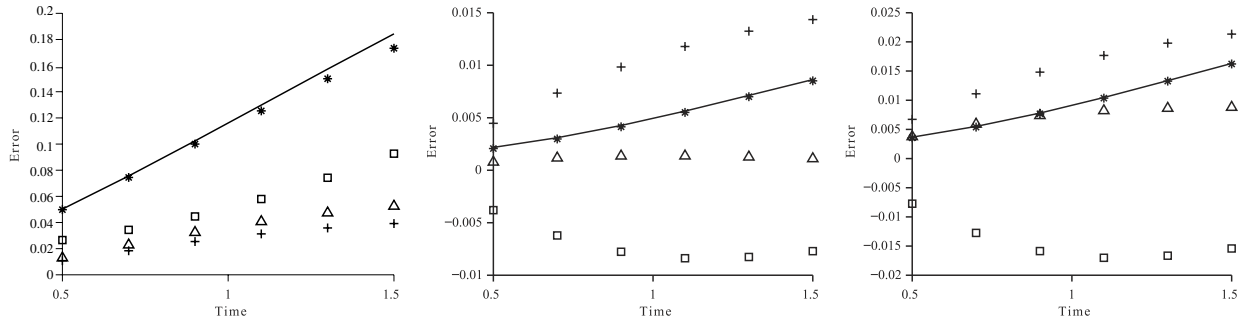


Fig. 5.7. Solution at different times for the example in Section 5.3.2.



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

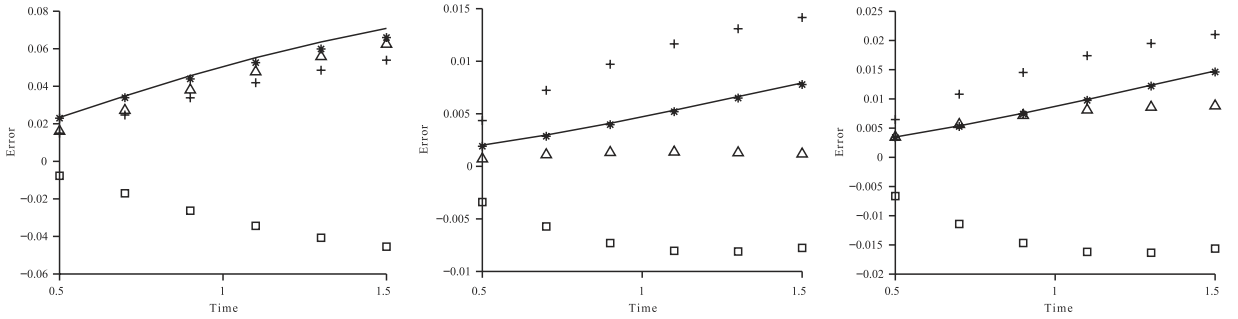
Fig. 5.8. Error components for the example in Section 5.3.2 with convection solved implicitly. Here (∗) true error, (–) error estimate, (□) discretization error ($\mathcal{DE}$), (+) quadrature error for f ($\mathcal{QE}_f$) and (△) quadrature error for g ($\mathcal{QE}_g$).

$\epsilon = \delta = \nu = 1$. The solution has a sharp gradient which moves in the direction of the field $\mathbf{b}(x)$, as shown in Fig. 5.7 for $t = 0$ and $t = 2$. The spatial domain is discretized using 500 elements along the x direction and 8 elements along the y direction. The QoI is $(u(T), \psi)$ where ψ is a mesh dependent function, equal to one on the patch $[0\,,\,3] \times [-0.5\,,\,0.5]$, and then decreasing linearly to zero in the adjacent elements, and zero everywhere else. This choice of QoI implies that the spatial integration, $(u(T), \psi)$, involves the sharp gradient in the solution for $0 < T < 2$.

The computational results solve diffusion implicitly. Results for convection solved implicitly and explicitly are shown in Figs. 5.8 and 5.9 respectively. For the first-order SBDF scheme with convection solved implicitly, all error components have the same sign, and sum to give the total error. For rest of the results, the discretization error and quadrature errors have different signs, and cancel each other to give an accurate estimate of the error. The effectivity ratios are given in Tables 5.7 and 5.8 demonstrate the accuracy of the estimator.

6. Conclusions

We have developed an adjoint-based *a posteriori* analysis to estimate the error in numerical solutions of PDEs solved using IMEX schemes. We derived error estimates for both finite difference and finite element discretizations



(a) First-order SBDF. (b) Second-order CNAB. (c) Second-order SBDF.

Fig. 5.9. Error components for the example in Section 5.3.2 with convection solved explicitly. Here (∗) true error, (–) error estimate, (□) discretization error ($\mathcal{DE}$), (+) quadrature error for f ($\mathcal{QE}_f$) and (△) quadrature error for g ($\mathcal{QE}_g$).

Table 5.7
Effectivity ratios for the example in Section 5.3.2 with convection solved implicitly.

(a) ρ_{eff} for Fig. 5.8(a)		(b) ρ_{eff} for Fig. 5.8(b)		(c) ρ_{eff} for Fig. 5.8(c)	
T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.5	1.0082	0.5	1.0509	0.5	1.0291
0.7	1.0148	0.7	1.0392	0.7	1.0210
0.9	1.0242	0.9	1.03	0.9	1.0149
1.1	1.0358	1.1	1.0230	1.1	1.0104
1.3	1.0490	1.3	1.0179	1.3	1.0071
1.5	1.0638	1.5	1.0141	1.5	1.0049

Table 5.8
Effectivity ratios for the example in Section 5.3.2 with convection solved explicitly.

(a) ρ_{eff} for Fig. 5.9(a)		(b) ρ_{eff} for Fig. 5.9(b)		(c) ρ_{eff} for Fig. 5.9(c)	
T	ρ_{eff}	T	ρ_{eff}	T	ρ_{eff}
0.5	1.0167	0.5	1.0554	0.5	1.0311
0.7	1.0263	0.7	1.0414	0.7	1.0223
0.9	1.038	0.9	1.0325	0.9	1.0175
1.1	1.0503	1.1	1.0263	1.1	1.0144
1.3	1.0626	1.3	1.0219	1.3	1.0126
1.5	1.0745	1.5	1.0189	1.5	1.0115

in space. In the former case, we classify only the error due to temporal discretization, while the effects of spatial discretization are also included in the latter case. In both cases, we show the nodal equivalence of the IMEX scheme with a variational formulation. This equivalence is necessary for forming error estimates using adjoint analysis. Analysis of first-order schemes is based on recognizing that the IMEX scheme is equivalent to the continuous Galerkin finite element scheme using special quadrature rules. The key insight in showing the equivalence of a second-order IMEX scheme to a finite element method is to recognize that the IMEX scheme arises by taking a *weighted* sum of finite element solutions over multiple time steps. The error estimates quantify various sources of error in a quantity of interest and include terms arising from discretization error and quadrature errors. We illustrate the accuracy of the estimator with a wide range of examples. The examples demonstrate how these different sources of error sum to give an accurate estimate of the error in the solution. We believe that the techniques presented in this article can be used to analyze higher order multi-step IMEX methods as well. However, the techniques do not encompass the analysis of multi-stage IMEX schemes (such as the popular IMEX Runge–Kutta methods). An *a posteriori* error analysis for multi-stage methods has been developed in [33], and the techniques employed there should prove useful in an analysis of IMEX multi-stage methods.

Acknowledgments

J.H. Chaudhry's work is supported in part by the Department of Energy (DE-SC0005304). D. Estep's work is supported in part by the Defense Threat Reduction Agency (HDTRA1-09-1-0036), Department of Energy (DE-FG02-04ER25620, DE-FG02-05ER25699, DE-FC02-07ER54909, DE-SC0001724, DE-SC0005304, INL0012-0133, DE0000000SC9279), Dynamics Research Corporation PO672TO001, Idaho National Laboratory (00069249, 00115474), Lawrence Livermore National Laboratory (B573139, B584647, B590495), National Science Foundation (DMS-0107832, DMS-0715135, DGE-0221595003, MSPA-CSE-0434354, ECCS-0700559, DMS-1065046, DMS-1016268, DMS-FRG-1065046, DMS-1228206), National Institutes of Health (#R01GM096192). V. Ginting's work is supported in part by the National Science Foundation (DMS-1016283), the Department of Energy (DE-SC0004982). S. Tavener's work is supported in part by the Department of Energy (DE-FG02-04ER25620, INL00120133) and National Science Foundation (DMS-1016268). J.N. Shadid's work is partially supported by the DOE Office of Science ASCR Applied Math Program at Sandia National Laboratory under contract DE-AC04-94AL85000.

References

[1] J. Smoller, Shock Waves and Reaction–Diffusion Equations, Springer-Verlag, New York, 1994.
[2] L. Pareschi, G. Russo, Implicit–explicit Runge–Kutta schemes and applications to hyperbolic systems with relaxation, J. Sci. Comput. 25 (2005) 129–154.
[3] W. Hundsdorfer, S.J. Ruuth, IMEX extensions of linear multistep methods with general monotonicity and boundedness properties, J. Comput. Phys. 225 (2007) 2016–2042.
[4] R. Donat, I. Higueras, A. Martinez-Gavara, On stability issues for IMEX schemes applied to 1D scalar hyperbolic equations with stiff reaction terms, Math. Comp. 276 (2011) 2097–2126.
[5] Y. Kadioglu, D.A. Knoll, R.B. Lowrie, R.M. Rauenzhan, A second order self-consistent IMEX method for radiation hydrodynamics, J. Comput. Phys. 229 (2010).
[6] S.Y. Kadioglu, D.A. Knoll, A fully second order implicit/explicit time integration technique for hydrodynamics plus nonlinear heat conduction problems, J. Comput. Phys. 229 (2010) 3237–3249.
[7] M. Svard, S. Mishra, Implicit–explicit schemes for flow equations with stiff source terms, J. Comput. Appl. Math. 235 (2011) 1564–1577.
[8] S.R. Lau, G. Lovelace, H.P. Pfeiffer, Implicit–explicit evolution of single black holes, Phys. Rev. D 84 (2011) 084023.
[9] C. Roedig, O. Zanotti, D. Alic, General relativistic radiation hydrodynamics of accretion flows—II. Treating stiff source terms and exploring physical limitations, Mon. Not. R. Astron. Soc. 426 (2012) 1613–1631.
[10] U.M. Ascher, S.J. Ruuth, B.T.R. Wetton, Implicit–explicit methods for time-dependent partial differential equations, SIAM J. Numer. Anal. 32 (1995) 797–823.
[11] U.M. Ascher, S.J. Ruuth, R.J. Spiteri, Implicit–explicit Runge–Kutta methods for time-dependent partial differential equations, Appl. Numer. Math. 25 (1997) 151–167.
[12] M.H. Carpenter, C.A. Kennedy, H. Bijl, S.A. Viken, V.N. Vatsa, Fourth-order Runge–Kutta schemes for fluid mechanics applications, J. Sci. Comput. 25 (2005) 157–194.
[13] H.D. Ceniceros, G.O. Mohler, A practical splitting method for stiff sdes with applications to problems with small noise, Multiscale Model. Simul. 6 (2007) 212–227.
[14] I. Grooms, K. Julien, Linearly implicit methods for nonlinear PDEs with linear dispersion and dissipation, J. Comput. Phys. 230 (2012) 1307–1325.
[15] D.R. Durran, P.N. Blossey, Implicit–explicit multistep methods for fast-wave-slow-wave problems, Mon. Weather Rev. 140 (2011) 3630–3650.
[16] F. Garcia, M. Net, J. Sanchez, A Comparison of High-order Time Integrators for Highly Supercritical Thermal Convection in Rotating Spherical Shells, in: Lecture Notes in Computational Science and Engineering, vol. 95, Springer International Publishing, 2014.
[17] D.J. Estep, M.G. Larson, R.D. Williams, A.M. Society, Estimating the Error of Numerical Solutions of Systems of Reaction–Diffusion Equations, American Mathematical Society, 2000.
[18] D. Estep, Error estimates for multiscale operator decomposition for multiphysics models, in: J. Fish (Ed.), Multiscale Methods: Bridging the Scales in Science and Engineering, Oxford University Press, USA, 2009.
[19] D. Estep, *A posteriori* error bounds and global error control for approximation of ordinary differential equations, SIAM J. Numer. Anal. 32 (1995) 1–48.
[20] K. Eriksson, D. Estep, P. Hansbo, C. Johnson, Computational Differential Equations, Cambridge University Press, Cambridge, 1996.
[21] M. Ainsworth, T. Oden, *A Posteriori* Error Estimation in Finite Element Analysis, John Wiley-Teubner, 2000.
[22] W. Bangerth, R. Rannacher, Adaptive Finite Element Methods for Differential Equations, Birkhäuser Verlag, 2003.
[23] T.J. Barth, *A posteriori* Error Estimation and Mesh Adaptivity for Finite Volume and Finite Element Methods, in: Lecture Notes in Computational Science and Engineering, vol. 41, Springer, New York, 2004.
[24] R. Becker, R. Rannacher, An optimal control approach to *a posteriori* error estimation in finite element methods, Acta Numer. (2001) 1–102.
[25] M.B. Giles, E. Süli, Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality, Acta Numer. 11 (2002) 145–236.

[26] D. Estep, V. Ginting, D. Ropp, J. Shadid, S.J. Tavener, An *a posteriori-a priori* analysis of multiscale operator splitting, SIAM J. Numer. Anal. 46 (2008) 1116–1146.
[27] D. Estep, V. Carey, V. Ginting, S.J. Tavener, T. Wildey, *A posteriori error* analysis of multiscale operator decomposition methods for multiphysics models, J. Phys. Conf. Ser. 125 (2008) 012075.
[28] M.G. Larson, F. Bengzon, Adaptive finite element approximation of multiphysics problems, Comm. Numer. Methods Engrg. 24 (2008) 505–521.
[29] A. Logg, Multi-adaptive time integration, Appl. Numer. Math. 48 (2004) 339–354.
[30] J.H. Chaudhry, D. Estep, V. Ginting, S. Tavener, A posteriori analysis of an iterative multi-discretization method for reaction–diffusion systems, Comput. Methods Appl. Mech. Engrg. 267 (2013) 1–22.
[31] V. Carey, D. Estep, A. Johansson, M. Larson, S. Tavener, Blockwise adaptivity for time dependent problems based on coarse scale adjoint solutions, SIAM J. Sci. Comput. 32 (2010) 2121–2145.
[32] K. Eriksson, D. Estep, P. Hansbo, C. Johnson, Introduction to adaptive methods for differential equations, in: Acta Numerica, 1995, Acta Numerica, Cambridge Univ. Press, Cambridge, 1995, pp. 105–158.
[33] J. Collins, D. Estep, S.J. Tavener, *A posteriori* error estimates for explicit time integration methods, BIT (2014) in press.
[34] Y. Cao, L. Petzold, A posteriori error estimation and global error control for ordinary differential equations by the adjoint method, SIAM J. Sci. Comput. 26 (2004) 359–374.
[35] D. Beigel, Efficient goal-oriented global error estimation for BDF-type methods using discrete adjoints (Ph.D. thesis), Universitat Heidelberg, 2012.
[36] D. Estep, M. Pernice, D. Pham, S. Tavener, H. Wang, A posteriori error analysis of a cell-centered finite volume method for semilinear elliptic problems, J. Comput. Appl. Math. 233 (2009) 459–472.
[37] V. Heuveline, R. Rannacher, A posteriori error control for finite element approximations of elliptic eigenvalue problems, Adv. Comput. Math. 15 (2001).
[38] V. Carey, D. Estep, S. Tavener, A posteriori analysis and adaptive error control for multiscale operator decomposition solution of elliptic systems I: triangular systems, SIAM J. Numer. Anal. 47 (2009) 740–761.
[39] A. Griewank, A. Walther, Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation, ACM Trans. Math. Software 26 (2000) 19–45.
[40] D. Estep, B. Mckeown, D. Neckels, J. Sandelin, GAASP: globally accurate adaptive sensitivity package. 2006, write to estep@math.colostate.edu for information.
[41] E.C. Cyr, J.N. Shadid, T.M. Widely, Towards efficient backward-in-time adjoint computations using data compression techniques, Comput. Methods Appl. Mech. Engrg. (2014) in press.
[42] D.L. Ropp, J.N. Shadid, Stability of operator splitting methods for systems with indefinite operators: Advection–diffusion–reaction systems, J. Comput. Phys. 228 (2009) 3508–3516.