Available at www.MATHEMATICSweb.org
POWERED BY SCIENCE DIRECT
Journal of Computational and Applied Mathematics 154 (2003) 27–61
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

A finite element semi-Lagrangian explicit Runge–Kutta–Chebyshev method for convection dominated reaction–diffusion problems

Rodolfo Bermejo[a,*], Mofdi El Amrani[b]

[a] *Universidad de Castilla-La Mancha, Dpto. de Matemáticas, Facultad de Ciencias Ambientales Avda. Carlos III s/n, 45071 Toledo, Spain*

[b] *Universidad Complutense de Madrid, Facultad de CC Matemáticas, Departamento de Matemática Aplicada, 28040 Madrid, Spain*

Received 12 September 2001; received in revised form 8 August 2002

Abstract

Explicit Runge–Kutta–Chebyshev methods have proved to be efficient for reaction–diffusion problems of moderate stiffness. In this paper, we extend such an efficiency to convection-dominated-reaction–diffusion problems by giving a formulation of these methods in a semi-Lagrangian framework, using C^0-finite elements of degree $m \geqslant 2$ as the space discretization method. We also study the convergence in the L^2-norm of the methods proposed in this paper.

MSC: 65L0; 65M12; 65M25; 65M60

1. Introduction

In this paper, we introduce a numerical method to compute an approximate solution to convection–reaction–diffusion problems in which convection dominates the other terms of the equations. The method we propose consists of using a consistent space approximation method, which is C^0-finite element method in this paper (although one can also use finite differences or spectral methods), combined with a semi-Lagrangian formulation of a second order explicit Runge–Kutta–Chebyshev (hereafter RKC) scheme. A conventional (or Eulerian) formulation of RKC schemes in the context of the integration of parabolic problems has been developed and analyzed in a number of papers

* Corresponding author.
E-mail address: rbermejo@amb-to.ucml.es (R. Bermejo).

0377-0427/03/$ - see front matter
PII: S0377-0427(02)00746-X

such as [6,13,14]. Thus, using the method of lines approach, a parabolic problem is converted into a system of ordinary differential equations

$$\frac{dU}{dt} = G(t, U(t)), \quad 0 < t \leqslant T,$$
$$U(0) = U_0,$$

in $\mathbb{R}^M$ say, by means of a space discretization method. If (i) the eigenvalues of the Jacobian matrix $G'(t, U) = \partial G/\partial U$ lie in a narrow strip along the negative real axis in the complex plane and (ii) $G'(t, U)$ is close to normal, then RKC schemes can be an interesting choice to integrate this ODE system for a number of reasons such as: (a) they possess an extended stability interval, which can be enlarged as much as needed by adding more stages to the schemes, (b) they can be of second order and (c) they are explicit. The latter property is very attractive from a computational viewpoint. However, in convection dominated problems properties (i) and (ii) of the Jacobian matrix $G'(t, U)$ deteriorate, so that RKC schemes do not work well in these problems. Nevertheless, if one uses an operator splitting approach one can still apply RKC schemes for the parabolic step of such an approach. In this spirit, we propose the semi-Lagrangian formulation of the RKC schemes to overcome the difficulties brought about by the strong convection terms. In a way that we will see below, this formulation can be considered as a splitting method that is applied along the characteristics of the total derivative operator $D/Dt := \partial/\partial t + a \cdot \nabla$, where $a(x, t)$ is a flow velocity. The model equations are the following:

$$\frac{Du}{Dt} = \nabla \cdot (K\nabla u) + f(u), \quad (x, t) \in \Omega \times (0, T], \ \Omega \in \mathbb{R}^d, \ d = 1, 2 \text{ or } 3,$$
$$u(x, 0) = u_0(x), \quad x \in \Omega,$$
$$K\frac{\partial u}{\partial n} = 0, \quad (x, t) \in \partial\Omega \times (0, T], \tag{1}$$

where Ω is a bounded domain with boundary $\partial\Omega$ and n is the unit outward normal to $\partial\Omega$. In these equations, K denotes a prescribed positive diffusion coefficient, which in a general case may be a second order tensor that might depend on (x, t) and u, but for the sake of simplicity we take it as constant in this paper. $f(u)$ is the reaction term. There are many environmental, chemical and fluid dynamics problems modelled by this equation, in which the convective term $a \cdot \nabla u$ is the dominant one, whereas $K \ll 1$ and f is a moderately stiff reaction term. In such cases, numerical methods that combine schemes based on the methods of characteristics, such as semi-Lagrangian and Characteristic-Galerkin methods, to deal with the convective term, and conventional implicit schemes, such as Crank–Nicolson and backward Euler, to manage the diffusive and reaction terms have proved to be accurate and efficient schemes. See for instance [1,7,9]. However, the implicitness of both diffusion and nonlinear reaction terms yields, every time step, a non linear system of equations, which requires the use of costly iterative methods to compute its numerical solution. So that, the idea of the numerical methods proposed in this paper is motivated by the excellent results obtained by the application of explicit RKC schemes to find the approximate solution of reaction–diffusion problems of moderate stiffness, and the fact that the numerical schemes, which integrate the term Du/Dt backward in time along the characteristics, reduce the procedure to find the numerical solution of (1) to the numerical integration of a reaction–diffusion problem. As for the characteristics based methods, we propose semi-Lagrangian ones for the reason that they offer computational advantages as compared with Characteristic-Galerkin methods and are as accurate as the latter.

The layout of the paper is the following. In Section 2 we introduce some well known results and useful definitions which are needed in the development of the paper. Sections 3 and 4 are devoted to the formulation of the finite element semi-Lagrangian RKC methods. The numerical analysis, some numerical experiments illustrating the performance of such methods and the conclusions are presented in Sections 5, 6 and 7, respectively. For the sake of completeness, and for those readers who may be interested in knowing the technical details of the proof of convergence in Section 5, we end the paper with two appendices which contain such material.

2. Preliminaries

We recast our model problem (1) in a weak form. To do so, we consider the real Sobolev spaces $W^{m,p}(\Omega)$, m and p integers, $0 \leqslant m < \infty$, $0 \leqslant p \leqslant \infty$, with norm $\|\cdot\|_{m,p}$ and semi-norm $|\cdot|_{m,p}$. We denote by $C^m(\bar{\Omega})$ the class of functions on $\bar{\Omega}$ that can be extended to be m-times differentiable in $\mathbb{R}^d$. $C_0^m(\Omega)$ is the set of m-times differentiable functions having compact support in Ω. The closure with respect to the norm $\|\cdot\|_{m,p}$ of the space $C_0^\infty(\Omega)$ is denoted by $W_0^{m,p}(\Omega)$. For $p=2, W^{m,p}(\Omega)$ (resp. $W_0^{m,p}(\Omega)$) coincides with the real Hilbert space $H^m(\Omega)$ (resp. $H_0^m(\Omega)$) with norm $\|\cdot\|_m$ and semi-norm $|\cdot|_m$. The dual space of $W_0^{m,p}(\Omega)$ (resp. $H_0^m(\Omega)$) is denoted by $W^{-m,q}(\Omega)$ (resp. $H^{-m}(\Omega)$) with $1/p+1/q=1$. For $m=0$, $H^m(\Omega)$ is the Hilbert space $L^2(\Omega)$ whose inner product and norm are denoted by $(\cdot,\cdot)$ and $\|\cdot\|$, respectively. The dual space of $L^2(\Omega)$ is itself. We require spaces that incorporate time dependency. Let X be any of the spaces introduced above and $[0,T]$ a time interval, if $u(x,t)$ represents a function defined on $\Omega\times[0,T]$ the following norms are used:

$$\|u\|_{L^p(0,T;X)} = \left(\int_0^T \|u(t)\|_X^p\right)^{1/p}, \quad 1 \leqslant p < \infty,$$

$$\|u\|_{L^\infty(0,T:X)} = \operatorname*{ess\,sup}_{0\leqslant t\leqslant T} \|u(t)\|_X.$$

The space $L^p(0,T;X)$ is the set of u such that the above norm is finite

We need the following assumptions:

A1. The function f is $C^2(a,b;\mathbb{R})$ for some interval (a,b) in $\mathbb{R}$.

A2. For any $(x,t)\in\bar{\Omega}\times[0,T], a(x,t)\in L^\infty(0,T;W^{1,\infty}(\bar{\Omega}))$ and $a\cdot n=0$ on the solid boundaries of Ω.

Hypothesis A2 guarantees for all time the existence and uniqueness of the trajectories of the points of Ω, the computation of which is an important ingredient of the semi-Lagrangian methods.

If u, f and a are sufficiently smooth, then u satisfies the weak form of (1)

$$\left(\frac{Du}{Dt}, v\right) + (K\nabla u, \nabla v) = (f(u), v), \quad \forall v\in H^1(\Omega),\ 0<t\leqslant T,$$

$$u(\cdot,t)\in H^1(\Omega), \quad 0<t\leqslant T,$$

$$u(x,0)=u_0(x), \quad x\in\Omega. \tag{2}$$

Next, given h_0, $0<h_0<1$, let h be a space discretization parameter such that $0<h<h_0$. To compute the numerical solution of (2) we generate a quasi-uniform partition D_h in $\bar{\Omega}$ of elements T_j that satisfy the following conditions.

(i) Let NE be the number of elements of D_h and let $J=\{1,2,\dots,NE\}$ be an index set, then $\bar{\Omega}=\bigcup_{j\in J} T_j$.

(ii) For $j,l\in J, j\neq l$,

$$T_j\cap T_l=\begin{cases} P_i, & \text{a mesh point, or}\\ \Gamma_{jl} & \text{a common side, or}\\ \emptyset & \text{empty set.}\end{cases}$$

(iii) There exists a positive constant μ such that for all $j\in J$, $d_j/h_j>\mu$, where d_j is the diameter of the circle inscribed in T_j and $h_j(h_j\leqslant h)$ is the largest side of T_j. Associated with the partition D_h there exists a family of finite dimensional subspaces S_h (further details on these spaces are given below) with the following approximation property.

Given integers s, m and r such that $0\leqslant s\leqslant r\leqslant m+1$, if $S_h\subset H^1(\Omega)$ then $\forall u\in W^{r,p}(\Omega)$ there exists a constant $K>0$ such that

$$\inf_{v_h\in V_h}\|u-v_h\|_{s,p}\leqslant Kh^{r-s}|u|_{r,p}\quad \text{with } 2\leqslant p\leqslant\infty.$$

Given $V_h\subseteq S_h$, we define some discrete operators which are needed below.

The orthogonal projection $P_0: H^{-1}(\Omega)\to V_h$

$$(P_0 v,\varphi)=(v,\varphi),\quad \forall\varphi\in V_h.$$

The polynomial interpolant of degree m I_m: $C^0(\bar{\Omega})\to S_h$

$$I_m u(x_i)=u(x_i),\quad 1\leqslant i\leqslant M,$$

where $\{x_i\}$ is the set of mesh points in the partition D_h.

The linear continuous operator $A: H^1(\Omega\to H^{-1}(\Omega))$ defined as

$$\langle Au,w\rangle=(K\nabla u,\nabla w),\quad \forall w\in H^1(\Omega),$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing. A is also a symmetric positive definite operator on $V=H^1(\Omega)/R$.

The discrete operator $A_h: V_h\to V_h$

$$(A_h v,\varphi)=\langle Av,\varphi\rangle,\quad \forall\varphi\in V_h.$$

The Ritz projection operator $R: V\to V_h$

$$(K\nabla Rv,\nabla\varphi)=\langle Av,\varphi\rangle,\quad \forall\varphi\in V_h.$$

The discrete operator $f_h: V\to V_h$

$$(f_h(v),\varphi)=(f(v),\varphi),\quad \forall\varphi\in V_h.$$

It is easy to see that

$$A_h R=P_0 A\quad \text{and}\quad f_h=P_0 f.$$

In our error analysis we shall need standard estimates of $\eta=u-Ru$ which are encapsulated in the following lemma [12]:

Lemma 1. *If u belongs to $L^\infty(0,T;H^r(\Omega))$, then there exist constants C_1 and C_2 independent of h such that for $1 \leqslant r \leqslant m+1$*

$$\|\eta\|_{L^\infty(0,T;L^2)} + h\|\eta\|_{L^\infty(0,T;H^1)} \leqslant C_1 h^r \|u\|_{L^\infty(0,T;H^r)}$$

and

$$\|u - P_0 u\|_{L^\infty(0,T;L^2)} + h\|u - P_0 u\|_{L^\infty(0,T;H^1)} \leqslant C_2 h^r \|u\|_{L^\infty(0,T;H^r)}.$$

We also assume that there exists a positive bounded $C(u)$ independent of h such that the following estimate holds:

$$\|f(u) - f(Ru)\|_{L^\infty(0,T;L^2)} \leqslant C h^r \|u\|_{L^\infty(0,T;H^r)}. \tag{3}$$

3. The finite element semi-Lagrangian RKC methods

To formulate the numerical methods we divide the interval $[0,T]$ into N subintervals $[t_n, t_{n+1}]$ such that $[0,T] = \bigcup_{n=0}^{N-1} [t_n, t_{n+1}]$. For the sake of simplicity, we take subintervals of equal length k, although this is not essential for the method to work well. Next, we define the family of finite element subspaces V_h. To this end, we consider an element of reference $\hat{T} \subset \mathbb{R}^d$ such that for each element T_j of the partition D_h we can define a one-to-one mapping $F_j : \hat{T} \to T_j$. Let $\hat{R}_m(\hat{T})$ be the set of polynomials $\hat{p}(\hat{x})$ of degree $\leqslant m$ defined on $\hat{T}$, then for each T_j we define the set

$$R_m(T_j) = \{P(x), x \in T_j\colon\ p(x) = \hat{p}(F_j^{-1}(x))\}.$$

Note that in the conventional literature of finite elements, $\hat{R}_m(\hat{T})$ is $P_m(\hat{T})$ if the elements of D_h are d-simplexes, whereas $\hat{R}_m(\hat{T})$ is $Q_m(\hat{T})$ if the elements of D_h are d-quadrilaterals. Defining the subspace S_h as

$$S_h = \{v_h \in W^{1,\infty}(\bar{\Omega})\colon\ v_h|_{T_j} \in R_m(T_j)\ \forall T_j \in D_h\},$$

we have that

$$V_h = \left\{ v_h \in S_h\colon\ K \left.\frac{\partial v_h}{\partial n}\right|_{\partial\Omega} = 0. \right\}$$

Hereafter, we say that S_h is a *high order subspace* if $m \geqslant 2$, whereas if $m = 1$, we say that S_h is a *low order subspace* denoted by L_h. So that, unless otherwise stated, we shall assume that S_h is a high order subspace.

Furthermore, we also need the finite dimensional spaces Z_h where the flow velocity is approximated,

$$Z_h = \{z_h \in S_h \times S_h\colon\ z_h \cdot n|_{\partial\Omega_s} = 0\},$$

where $\partial\Omega_s$ denotes the set of solid boundaries of Ω.

Let M be the number of mesh points of the partition D_h, then any element of V_h is expressed as

$$v_h = \sum_{j=1}^{M} V_j \phi_j(x),$$

where $V_j = v_h(x_j), x_j$ being the jth mesh point, and $\{\phi_j\}$ is the set of global nodal basis functions of V_h characterized by the property $\phi_j(x_i) = \delta_{ji}$. Similarly for the elements of Z_h. However, the elements of L_h are expressed as

$$l_h = \sum_{k=1}^{ML} L_k \psi_k,$$

where ML is the number of mesh points in the partition D_h corresponding to the space L_h, such points being the vertices of the elements T_l of D_h, so that $ML < M$. $\{\psi_k\}$ is the set of global nodal basis functions of L_h characterized also by the property $.\psi_k(x_l) = \delta_{kl}$, with x_l being the kth mesh point corresponding to the space L_h.

Anticipating things that we need below, it is convenient to introduce at this point the element nodal basis functions associated with the elements T_l of the partition D_h. Let $\{x_1^l, \dots, x_{\mathrm{NH}}^l\}$ be the set of nodes of the element T_l associated with the high order subspace $S_h \cdot \{\varphi_i^l\}_{i=1}^{\mathrm{NH}}$ denotes the set of element nodal basis functions of high order defined on T_l, which satisfy for all $i, \varphi_i^l \in R_m(T_l)$ and $\varphi_i^l(x_p^l) = \delta_{ip}$. Likewise, $\{\chi_i^l\}_{i=1}^{\mathrm{NL}}$ denotes the set of element nodal basis functions of low order defined on T_l, such that $\chi_i^l(\bar{x}_m^l) = \delta_{im}$ with $\{\bar{x}_1^l, \dots, \bar{x}_{\mathrm{NL}}^l\}$ being the set of vertices of T_l associated with L_h. Note that the sets $\{\varphi_i^l\}_{i=1}^{\mathrm{NH}}$ and $\{\chi_i^l\}_{i=1}^{\mathrm{NL}}$ are the restrictions on the element T_l of the sets of global nodal basis functions $\{\phi_j\}_{j=1}^{M}$ and $\{\psi_k\}_{k=1}^{ML}$, respectively. Hereafter, unless otherwise stated, we use the notations $b^n(x) \equiv b(x, t_n)$ and $b_i^n \equiv b(x_i, t_n)$, with $b_h^n(x)$ and b_{hi}^n denoting their corresponding space discretization, respectively.

The numerical solution to (2) is the map $u_h : (0, T] \to V_h$ such that for $t = t_n, n = 1, 2, \dots$.

$$u_h^n = \sum_{j=1}^{M} U_j^n \phi_j. \tag{4}$$

To compute such a solution we consider for each subinterval $[t_n, t_{n+1}]$ the map $\bar{u}_h : [t_n, t_{n+1}] \to V_h$ defined by

$$\bar{u}_h(x,t) = \sum_{j=1}^{M} \bar{U}_j(t)\phi_j(x), \quad t \in [t_n, t_{n+1}], \tag{5}$$

where

$$\bar{U}_j(t) = u_h(X_h(x_j, t_{n+1}; t), t), \tag{6}$$

and $X_h(x_j, t_{n+1}; t)$ is the trajectory of a point that departing from $X_{hj}^n = X_h(x_j, t_{n+1}; t_n)$ at time t_n will arrive at the mesh point x_j at time t_{n+1}. So that, for all n,

$$\bar{U}_j^n = u_h^n(X_{hj}^n), \tag{7}$$

and for $t \in (t_n, t_{n+1}]$

$$\left(\frac{\mathrm{d}\bar{u}_h}{\mathrm{d}t}, V\right) + (A_h \bar{u}_h, V) = (f_h(\bar{u}_h), V) \quad \text{for all } V \in V_h,$$

$$(\bar{u}_h^n, V) = \left(\sum_{j=1}^{M} \bar{U}_j^n \phi_j, V\right). \tag{8}$$

Few remarks are now in order. First, $\mathrm{d}\bar{u}_h/\mathrm{d}t$ denotes $\sum_l (\mathrm{d}\bar{U}_l/\mathrm{d}t)\phi_l$, with

$$\frac{\mathrm{d}\bar{U}_l}{\mathrm{d}t} = \frac{\partial \bar{U}_l}{\partial t} + a_h(X_h(x_l, t_{n+1}; t), t) \cdot \nabla u_h(X_h(x_l, t_{n+1}; t), t).$$

Second, (8) represents an initial value problem for $\bar{u}_h \in V_h$. To make this point clear, we write (8) as an ODE system for the components $\{\bar{U}_l(t)\}$. By virtue of definitions of the discrete operators given above, it follows that (8) becomes

$$\mathbf{M}\left[\frac{\mathrm{d}\bar{U}}{\mathrm{d}t}\right] = \mathbf{S}[\bar{U}] + \mathbf{M}[F(\bar{U})], \quad t_n < t \leqslant t_{n+1},$$

$$[\bar{U}^n] \text{ known as initial condition at } t_n, \tag{9}$$

where $\mathbf{M}$ and $\mathbf{S}$ are sparse symmetric matrices the elements of which are given by

$$m_{ij} = \int_\Omega \phi_i \phi_j \,\mathrm{d}x, \quad i,j = 1,2,\ldots,M$$

and

$$s_{ij} = -\int_\Omega K\nabla\phi_i \cdot \nabla\phi_j \,\mathrm{d}x, \quad i,j = 1,2,\ldots,M,$$

respectively.

$$\left[\frac{\mathrm{d}\bar{U}}{\mathrm{d}t}\right] \equiv \left[\frac{\mathrm{d}\bar{U}_1}{\mathrm{d}t}, \frac{\mathrm{d}\bar{U}_2}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}\bar{U}_M}{\mathrm{d}t}\right]^{\mathrm{T}}, \quad [\bar{U}] \equiv [\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_M]^{\mathrm{T}}$$

and

$$[F(\bar{U})] \equiv [F_1(\bar{U}), F_2(\bar{U}), \ldots, F_M(\bar{U})]^{\mathrm{T}}$$

with

$$F_i(\bar{U}) = f(\bar{u}_h)_i, \quad i = 1,2,\ldots,M.$$

$\mathbf{M}$ is known as the mass matrix in finite element literature. It is a positive definite matrix with a low condition number, which is practically independent of h, so that it is very easy to invert it even by the diagonal preconditioned conjugate gradient. $\mathbf{S}$ is known as the stiffness matrix in the finite element literature. The condition number of $\mathbf{S}$ is $\mathrm{O}(C/h^2)$, where C is a bounded constant; this means that for small h, $\mathbf{S}$ may have a high condition number that makes difficult invert it. It is at this point where the explicit RKC methods play a useful role, because these methods do not need invert the matrix $\mathbf{S}$.

The third remark is concerned with the initial condition of (9). Note that the initial values $[\bar{U}^n]$ are the values of $u_h^n(x)$ at the feet $\{X_{hl}^n\}$ of the characteristics that go through the mesh points $\{x_l\}$ at time t_{n+1}. In the methods we propose in this paper, these values are obtained at the semi-Lagrangian step, whereas the solution of (9) is approximated via the RKC methods. So that, the application of the semi-Lagrangian RKC methods in each $[t_n, t_{n+1}]$ can be viewed as a splitting method of two steps. In the first step, termed *the transport step*, the semi-Lagrangian method is applied to obtain the trajectories of the mesh points and the initial condition for the second step, known as *reaction–diffusion step*, in which RKC methods are used.

3.1. The transport step: the quasi-monotone semi-Lagrangian method

Let $a_h(x,t) \in Z_h$ be the approximate flow velocity at time t. The application of the semi-Lagrangian method in each $[t_n, t_{n+1}]$ is performed as follows.

(1) For each mesh point compute the corresponding departure points and identify the elements where such points are located. That is, for $1 \leqslant j \leqslant M$, compute the points $\{X_{hj}^n\}$ and identify the elements $\{T_l\}$ where such points are located. Each X_{hj}^n is the foot at time t_n of the trajectory traced by the point that will arrive at the mesh point x_j at time t_{n+1}. So that, $X_{hj}^n = X_h(x_j, t_{n+1}; t)|_{t=t_n}$ where $X_h(x_j, t_{n+1}; t)$ satisfies

$$\frac{\mathrm{d}X_h(x_j, t_{n+1}; t)}{\mathrm{d}t} = a_h(X_h(x_j, t_{n+1}; t), t), \quad t_n < t \leqslant t_{n+1},$$
$$X_h(x_j, t_{n+1}; t_{n+1}) = x_j. \tag{10}$$

By virtue of A2 and given that $a_h(x,t) \in Z_h$, (10) has a unique solution. An efficient algorithm to compute the points $\{X_{hj}^n\}$ with order $\mathrm{O}(k^2)$ in unstructured meshes, is described in [1]. The same algorithm identifies the element T_l where each point X_{hj}^n is located.

(2) Assuming that for all j the pairs (X_{hj}^n, T_l) and the mesh point values $\{U_j^n\}$ are known, we compute the values $\{U_j^{*n}\}$ as

$$U_j^{*n} = u_h^n(X_{hj}^n) = \sum_{k=1}^{\mathrm{NH}} U_k^n \varphi_k^l(X_{hj}^n). \tag{11}$$

Note that U_j^{*n} is obtained by local interpolation in the elements T_l that contain the departure points X_{hj}^n. One could stop the transport step at this point and obtain a transport solution that we term *conventional semi-Lagrangian transport step solution*, which is expressed as

$$u_h^{*n} = \sum_{j=1}^{M} U_j^{*n} \phi_j. \tag{12}$$

But it is known that when polynomial interpolation is performed with polynomials of degree higher than one, the result may exhibit an oscillatory behavior. This kind of behavior of the solution obtained at the transport step is not admissible in reaction–diffusion problems where the exact solution admits a compact invariant region. So that, in order to get a nonoscillatory semi-Lagrange approximation we add a limiting procedure proposed in [4] and analyzed in [3], which is now described.

Limiting procedure

(3) Given the element T_l that contains X_{hj}^n, define

$$U^+ = \mathrm{Max}(U_1^n, \ldots, U_{\mathrm{NH}}^n),$$
$$U^- = \mathrm{Min}(U_1^n, \ldots, U_{\mathrm{NH}}^n),$$

where $\{U_1^n, \ldots, U_{\mathrm{NH}}^n\}$ is the set of values that the numerical solution takes at the vertices of the element T_t at time t_n.

(4) Compute

$$\bar{U}_j^n = \begin{cases} U^+ & \text{if } U_j^{*n} > U^+, \\ U^- & \text{if } U_j^{*n} < U^-, \\ U_j^{*n} & \text{otherwise.} \end{cases}$$

(5) Define the function $\bar{u}_h^n(x) \in S_h$ as

$$\bar{u}_h^n(x) = \sum_{j=1}^{M} \bar{U}_j^n \phi_j(x). \tag{13}$$

This completes the transport step. We shall denote $\bar{u}_h^n(x)$ as the *nonoscillatory semi-Lagrangian transport step solution*

For the numerical analysis of the transport step, we shall recall a result of [3] which establishes that the mesh point values $\{\bar{U}_j^n\}$ can be expressed as

$$\bar{U}_j^n = U_{Lj}^n + \beta_j^n (U_j^{*n} - U_{Lj}^n), \tag{14}$$

where

$$U_{Lj}^n = \sum_{k=1}^{NL} U_k^n \chi_k^l (X_{hj}^n) \tag{15}$$

is an approximation to $u_h^n(X_{hj}^n)$ in the low order subspace L_h. (Recall that the low order subspace L_h is the subspace S_h for $m = 1$). The limiting coefficients β_j^n, $0 \leqslant \beta_j^n \leqslant 1$, are defined as

$$\beta_j^n = \begin{cases} \text{Min}\left(1, \dfrac{Q^+}{P}\right) & \text{if } P > 0, \\ \text{Min}\left(1, \dfrac{Q^-}{P}\right) & \text{if } P < 0 \\ 1 & \text{otherwise,} \end{cases} \tag{16}$$

where

$$\begin{aligned} P &= U_j^{*n} - U_{Lj}^n, \\ Q^+ &= U^+ - U_{Lj}^n, \\ Q^- &= U^- - U_{Lj}^n. \end{aligned} \tag{17}$$

At this point it is convenient to write (14) in terms of the interpolant operators $I_1 : C^0(\bar{\Omega}) \to L_h$ and $I_m : C^0(\bar{\Omega}) \to S_h, m \geqslant 2$, as

$$\bar{U}_j^n = I_1 u_h^n(X_{hj}^n) + \beta_j^n (I_m u_h^n(X_{hj}^n) - I_1 u_h^n(X_{hj}^n)). \tag{14a}$$

where $I_1 u_h^n(X_{hj}^n)$ and $I_m u_h^n(X_{hj}^n)$ denote the values of $I_1 u_h^n$ and $I_m u_h^n$, respectively, at the point X_{hj}^n.

4. The reaction–diffusion step: the RKC schemes

In this step we compute the solution of (8) at each $[t_n, t_{n+1}]$ by means of a s-stage RKC scheme using as initial condition $\bar{u}_h^n$. Thus, following the description of the RKC schemes given in [13,14], we formulate the reaction–diffusion step as follows. Let s be the number of the stages of the RKC, then set

$$
\begin{aligned}
y_{h0} &= \bar{u}_h^n,\\
y_{h1} &= y_{h0} + \tilde{\mu}_1 k g_{h0},\\
y_{hj} &= \mu_j y_{hj-1} + \nu_j y_{hj-2} + (1-\mu_j-\nu_j) y_{h0}\\
&\quad + \tilde{\mu}_j k g_{hj-1} + \tilde{\gamma}_j k g_{h0}, \quad 2 \leqslant j \leqslant s, \qquad (18)\\
u_h^{n+1} &= y_{hs},
\end{aligned}
$$

where $g_{hj} = g_h(y_{hj}, t_n + c_j k) = -A_h y_{hj} + f_h(y_{hj})$. From a computational point of view, it is convenient to write (18) in terms of the mesh point values of $y_{hj} \in V_h$. Thus, we have

$$
\begin{aligned}
Y_0 &= \bar{U}^n,\\
Y_1 &= Y_0 + \tilde{\mu}_1 k G_0,\\
Y_j &= \mu_j Y_{j-1} + \nu_j Y_{j-2} + (1-\mu_j-\nu_j) Y_0 + \tilde{\mu}_j k G_{j-1} + \tilde{\gamma}_j k G_0, \quad 2 \leqslant j \leqslant s,\\
U^{n+1} &= Y_s, \qquad (18a)
\end{aligned}
$$

where the $\mathbb{R}^M$ vector valued functions Y_j and G_j have as entries their values at the mesh points. We now explain the meaning of the symbols that appear in (18). For $0 \leqslant m \leqslant s-1$,

$$G_m = G(Y_m, t_n + c_m k) = \mathbf{B} Y_m + F(Y_m), \qquad (18b)$$

with the matrix $\mathbf{B} = \mathbf{M}^{-1}\mathbf{S}$. Next, for $1 \leqslant j \leqslant s$, the coefficients c_j, $0 = c_0 < \cdots < c_s = 1, \tilde{\mu}_j, \mu_j, \nu_j$, and $\tilde{\gamma}_j$ are calculated in a recursive form using Chebyshev polynomials of the first kind of degree j. To calculate these coefficients the following criteria are taken into consideration. (i) Let $P_{hs}(z)$ be a function, which is defined below, called the genuine stability function of the corresponding RKC method, and let $\beta(s)$ be its real stability boundary defined as $\beta(s) = \max\{-z\colon z \leqslant 0, |P_{hs}(z)| \leqslant 1\}$, the coefficients $\tilde{\mu}_j, \mu_j, \nu_j$, and $\tilde{\gamma}_j$ are calculated in such a way as to make $\beta(s)$ be as large as possible to obtain good stability properties for parabolic problems. (ii) The application of the method with an arbitrary number of stages should not damage the convergence properties, that is, the accumulation of local errors does not grow without bound.

5. Analysis

We study in this section the stability and convergence properties of the finite element semi-Lagrangian RKC methods that we have just described.

5.1. Convergence of trajectories $X_h(x,t_{n+1};t)$

To establish the stability and convergence properties of the methods, we need calculate error estimates for the points $\{X_{hj}^n\}$, because the accuracy of the calculation of such points has a significant influence upon the overall accuracy of the method. In the analysis that follows we assume that the points $\{X_{hj}^n\}$ are calculated by one-step methods, because this is the kind of method we have implemented in our algorithm; although one can use multi-step methods if wished. So that, for any $x\in D_h$ we set, using Henrici's notation for the approximate solution of (10) obtained by one-step methods,

$$X_h^n(x)=x-k\Phi_{a_h}(t_{n+1},x;k),$$

where $\Phi_{a_h}(t_{n+1},x;k)$ is the incremental function. We assume that the following properties hold:

T1. There exist a real constant $1>k_0>0$ such that Φ_{a_h}: $[0,T]\times D_h\times(0,k_0)\to\mathbb{R}^d$ is a continuous function that only depends on a_h.

T2. For any t in $[0,T]$ and x in $D_h, \Phi_{a_h}(t,x;k)\to a_h(x,t)$ as $k\to 0$.

T3. For any t in $[0,T]$, x and y in D_h and k in $(0,k_0)$, there exists a positive constant C such that

$$|\Phi_{a_h}(t,x;k)-\Phi_{a_h}(t,y;k)|\leqslant C|x-y|,$$

where $|\cdot|$ denotes a norm in $\mathbb{R}^d$.

T4. There exists k^*, with $0<k^*<k_0$, such that for k in $(0,k^*)$ and h in $(0,h_0)$ the method is absolutely stable.

T5. The method is of order p, p integer larger than 1.

The meaning of the latter property is the following. Let

$$X_h(x,t_{n+1};t_n)=x-\int_{t_n}^{t_{n+1}} a_h(X_h(x,t_{n+1};t),t)\,\mathrm{d}t \tag{19}$$

be the exact solution of (10) for any x in D_h. Assuming that $a_h(x,t)$ is sufficiently smooth in time, we have that for all k in $(0,k^*)$, h in $(0,h_0)$ and t_n in $(0,T]$

$$|X_h(x,t_{n+1}t_n)-X_h^n(x)|=\mathrm{O}(k^{p+1}). \tag{20}$$

Note that properties T1–T4 ensure the convergence of $X_h^n(x)$ to $X_h(x,t_{n+1};t_n)$.

Given k in $(0,k^*)$, we define the error committed in the calculation of $X_h^n(x)$ in each subinterval $[t_n,t_{n+1}]$ as

$$\varepsilon^n=X(x,t_{n+1};t_n)-X_h^n(x), \tag{21}$$

where $X(x,t_{n+1};t)$ satisfies the initial value problem

$$\frac{\mathrm{d}X(x,t_{n+1};t)}{\mathrm{d}t}=a(X(x,t_{n+1};t),t),\quad t_n\leqslant t<t_{n+1},\ x\in\bar{\Omega}$$
$$X(x,t_{n+1};t_{n+1})=x, \tag{22}$$

By noticing that $X_h(x, t_{n+1}; t)$ represents a space approximation to $X(x, t_{n+1}; t)$, we can further decompose the error ε^n as

$$\varepsilon^n = X(x, t_{n+1}; t_n) - X_h(x, t_{n+1}; t_n) \\ + X_h(x, t_{n+1}; t_n) - X_h^n(x) \equiv \varepsilon_1^n + \varepsilon_2^n. \quad (23)$$

In this decomposition of the error, the component ε_2^n denotes the local truncation error in time of the one-step method used to calculate $X_h^n(x)$. As pointed out above, such an error depends on the order of the method. As for ε_1^n, this component is originated by the approximation of the exact flow velocity $a(x, t)$ by $a_h(x, t)$. To estimate ε_1^n we note that by virtue of (14) and (17), and setting $s = t_{n+1}$ to simplify the notation, we have that

$$|\varepsilon_1^n| \leqslant \int_0^k |a_h(X_h(x, s; s-\tau), s-\tau) - a(X(x, s; s-\tau), s-\tau)| \, d\tau.$$

From here it follows that

$$|\varepsilon_1^n| \leqslant \int_0^k |a_h(X_h(x, s; s-\tau), s-\tau) - a(X_h(x, s; s-\tau), s-\tau)| \, d\tau \\ + \int_0^k |\nabla a(X_z)| \, |\varepsilon_1(x, \tau)| \, d\tau,$$

where $X_z = zX(x, s; s-\tau) + (1-z)X_h(x, s; s-\tau)$ with $0 < z < 1$. By virtue of A2 and Gronwall inequality it follows that

$$|\varepsilon_1^n| \leqslant , e^{kL} \int_0^k |a_h(X_h(x, s; s-\tau), s-\tau) - a(X_h(x, s; s-\tau), s-\tau)| \, d\tau,$$

where $L = |\nabla a|_{L^\infty(0,T;\Omega)}$. To estimate the L^2-norm we have by definition that

$$\|\varepsilon_1^n\|^2 \leqslant e^{2kL} \int_\Omega \left(\int_0^k |a_h(X_h(x, s; s-\tau), s-\tau) - a(X_h(x, s; s-\tau), s-\tau)| \, d\tau \right)^2 dx$$

and recalling the inequality

$$\int_a^b v \, dx \leqslant \sqrt{|b-a|} \left(\int_a^b |v|^2 \, dx \right)^{1/2},$$

it follows that

$$\|\varepsilon_1^n\|_{L^\infty(0,T;L^2)} \leqslant kC \|a - a_h\|_{L^\infty(0,T;L^2)}, \quad (24)$$

where $C = \exp k^* L$. We collect these results in the following lemma:

Lemma 2. *Assume that for each* $[t_n, t_{n+1}]$ *the points* $X_h^n(x)$ *are calculated by one-step methods such that the assumptions* T1–T5 *and* A2 *hold. Then*

$$\|X(x, t_{n+1}; t_n) - X_h^n(x)\|_{L^\infty(0,T;L^2)} \leqslant Ck \|a - a_h\|_{L^\infty(0,T;L^2)} + O(k^{p+1}). \quad (25)$$

Remark. The above estimate shows that if one takes $h = O(k)$, a is sufficiently regular and the degree m of the finite element space Z_h is at least equal to $p-1$, p being the order of the method used to solve (10), then $\|\varepsilon_1^n\|_{L^\infty(0,T;L^2)} = O(\|\varepsilon_2^n\|_{L^\infty(0,T;L^2)})$. In the rest of the paper we shall assume that this holds, unless otherwise stated.

Two auxiliary results concerning properties of the mappings $x \to X(x,s,t)$ and $x \to X_h(x,s,t)$ are stated next.

Lemma 3 (Bermejo [2]). *Assuming that* A2 *holds and* $|s-t|$ *is sufficiently small, the mapping* $x \to X(x,s;t)$ *is a quasi-isometric homeomorphism of* Ω *into itself with Jacobian determinant* $J(x,s;t)$ *in* $L^\infty(0,T;L^\infty(\Omega))$. *Moreover,*

$$L^{-1}\|x-y\| \leqslant \|X(x,s;t) - X(y,s;t)\| \leqslant L\|x-y\|,$$

where $L = |s-t|\|\nabla a|_{L^\infty(0,T;Q)}$ *and* $\|\cdot\|$ *denotes the Euclidean norm in* $\mathbb{R}^d$.

Lemma 4 (Süli [11]). *Assuming that* T1–T5 *hold, the mapping* $x \to X_h(x,t_{n+1};t)$ *is a quasi-isometric homeomorphism of* Ω *into itself with an a.e nonzero Jacobian determinant.*

Another interesting result related with the homeomorphisms of the previous lemmata is [8, Theorem 1.1.9] which, restricted to our problem, is stated in the following lemma:

Lemma 5. *As in Lemmata* 3 *and* 4, *let* $x \to X(x,s;t)$ *be a quasi-isometric homeomorphism of class* $C^{r-1,1}(\bar{\Omega}), r \geqslant 1$. *Let* $f \in W^{r,p}(\Omega^*)$ *and* $g = f(X(x,s;t))$. *Then,* $g(x) \in W^{r,p}(\Omega)$ *and there exist positive constants* K_1 *and* K_2 *such that*

$$K_1\|f\|_{r,p,\Omega^*} \leqslant \|g\|_{r,p,\Omega} \leqslant K_2\|f\|_{r,p,\Omega^*}.$$

5.2. Stability

Our next concern is to prove that semi-Lagrangian–RKC methods are stable in the L^2-norm. In [3, Theorem 3] the stability of the transport step for the pure advection equation in the L^∞-norm is proved; so that, we expect this property be also inherited in the L^2-norm. To see this is so, we consider the application of the quasi-monotone semi-Lagrangian scheme to the pure advection equation

$$\frac{Du}{Dt} = 0 \quad \text{in } \Omega \times (0,T]$$

$$u(x,0) = u_0(x) \quad \text{in } \Omega,$$

so that, following the notation for the transport step introduced above, the solution at time t_{n+1} is then expressed as $u_h^{n+1} = \bar{u}_h^n$. If we introduce the quasi-monotone transport step operator $\overline{\mathrm{TS}}(k,\beta^n) : V_h \to S_h$ defined as

$$u_h^{n+1} = \overline{\mathrm{TS}}(k,\beta^n)u_h^n,$$

where $\beta^n = (\beta_1^n, \ldots, \beta_M^n)$ is the vector of limiting coefficients for u_h^n.

It follows that

$$u_h^{n+1} = \overline{\mathrm{TS}}(k, \beta^n) u_h^n = \overline{\mathrm{TS}}(k, \beta^n) \circ \overline{\mathrm{TS}}(k, \beta^{n-1}) \circ \ldots \circ \overline{\mathrm{TS}}(k, \beta^0) u_{0h}.$$

Recalling the definition of stability of a difference scheme, we say that the transport step is stable (in $[0, T]$) in the norm $\|\cdot\|$ if for any h and k satisfying $0 < k \leqslant k^*$ and $0 < h \leqslant h_0$, there exists a positive constant C independent of k and h such that $\|\overline{\mathrm{TS}}(k, \beta^n) \circ \ldots \circ \overline{\mathrm{TS}}(k, \beta^0) u_{0h}\| \leqslant C$ for all n, assuming that the method used to compute the points X_{hj}^n is stable. We have the following result:

Lemma 6. *Assume that* T1–T5 *hold, then for all* t_n,

$$\|\bar{u}_h^n\| \leqslant \|u_h^n\|, \tag{26}$$

Proof. We prove the inequality (26) by contradiction. First, we recall that [3, Theorem 3] proves the stability of the quasi-monotone transport step in the L^∞-norm, i.e. for all t_{n+1}

$$\|u_h^{n+1}\|_{L^\infty(\Omega)} \leqslant C \|u_{0h}\|_{L^\infty(\Omega)}.$$

Next, we use the inverse estimate ([5])

$$\|v_h\| \leqslant C_1 \|v_h\|_{L^\infty(\Omega)}, \quad \forall v_h \in V_h,$$

where C_1 is a positive constant that does not depend on h, to show that the stability in the L^∞-norm implies

$$\|u_h^{n+1}\| \leqslant C_1 C \|u_{0h}\|_{L^\infty(\Omega)}.$$

Hence, u_h^{n+1} (and therefore, $\bar{u}_h^n$) is also uniformly bounded for all t_n in the L^2-norm and, therefore, the transport step is also stable in the L^2-norm. Knowing this, we proceed to prove (26). Let us assume that (26) is not true, or in other terms, for all t_n there exists a constant $K > 1$ independent of k and h such that $\|u_h^{n+1}\| = K\|u_h^n\| = K^2\|u_h^{n-1}\| = K^{n+1}\|u_{0h}\|$. Using the above inverse estimate together with the following one [5]:

$$\|v_h\|_{L^\infty(\Omega)} \leqslant C_1 h^{-d/2} \|v_h\|, \quad \forall v_h \in V_h,$$

where the positive constant C_1 does not depends on h, yields

$$\|u_h^{n+1}\|_{L^\infty(\Omega)} \geqslant C K^{n+1} \|u_{0h}\| \geqslant C K^{n+1} h^{d/2} \|u_{0h}\|_{L^\infty(\Omega)}.$$

The latter inequality implies that $\|u_h^{n+1}\|_{L^\infty(\Omega)}$ becomes unbounded. So that, $K \leqslant 1$ and this ends the proof of Lemma 6. □

Next, we proceed to study the L^2-norm stability for the reaction-diffusion step on a linearized version of (8) as is customary in numerical analysis of IVP. Thus, for all n and $t \in (t_n, t_{n+1}]$ we consider the equation

$$\frac{\mathrm{d}\bar{u}_h}{\mathrm{d}t} + A_h \bar{u}_h - \bar{f}_h = 0,$$

$$\bar{u}_h(t)|_{t=t_n} = \bar{u}_h^n. \tag{27}$$

Before going into the details of the stability analysis of (27), we recall some properties of the operator $A_h : V_h \to V_h$ which are needed below. First, A_h is a self-adjoint positive definite operator on V_h with spectrum $\{\lambda_j \in R\colon \lambda_{j+1} > \lambda_j,$ for $j=1,2,\dots,M\}$. The spectral radius of A_h, $\rho(A_h)=\max_j\{\lambda_j\}$, satisfies the inequality $\rho(A_h) \leqslant Ch^{-2}$, where C is a positive constant independent of h. Let $\{W_j\}_{j=1}^{M} \subset V_h$ be the corresponding orthonormal eigenfunctions of A_h, since they form a basis of V_h, then for any $v_h \in V_h$

$$v_h = \sum_{l=1}^{M} (v_h, W_l) W_l. \tag{28}$$

We need the r-power A_h^r of A_h, r being a positive integer, which is defined as

$$A_h^r v_h = \sum_{l=1}^{M} \lambda_l^r (v_h, W_l) W_l. \tag{28a}$$

Next, we calculate the stability function of scheme (18) when it is applied to approximate the solution of (27) in V_h with $\bar{f}_h = 0$. Following the arguments of [14], for integers $j = 0, 1, 2, \dots, s$, we define the operators $P_{hj}(-kA_h) : V_h \to V_h$ recursively as

$$P_{h0}(-kA_h)v_h = I_h v_h = v_h, \tag{29}$$

where $I_h : V_h \to V_h$ is the identity operator,

$$P_{h1}(-kA_h)v_h = v_h - \tilde{\mu}_1 k A_h v_h, \tag{29a}$$

for $2 \leqslant j \leqslant s$

$$\begin{aligned} P_{hj}(-kA_h)v_h = {} & (1 - \mu_j - v_j)v_h - \tilde{\gamma}_j k A_h v_h \\ & + (\mu_j I_h - \tilde{\mu}_j k A_h) \circ P_{hj-1}(-kA_h)v_h + v_j P_{hj-2}(-kA_h)v_h. \end{aligned} \tag{29b}$$

Hence, it is easy to see that $P_{hj}(-kA_h)$ is a polynomial in $-kA_h$, so that by virtue of (28a) we have that

$$P_{hj}(-kA_h)v_h = \sum_{l=1}^{M} P_{hj}(z_l)(v_h, W_l) W_l, \tag{29c}$$

where the notation $z_l = -k\lambda_l$ has been used. The coefficients $\mu_j, v_j, \tilde{\gamma}_j$ and $\tilde{\mu}_j$ are calculated (for details, see [10,13,14] and references therein) in such a way that the stability boundary of $P_{hs}(z)$ is as large as possible, and therefore, for all $l, z_l \leqslant \beta(s) := \max\{-z : z \leqslant 0$ *and* $\max|P_{hs}(z)| \leqslant 1\}$. In [14], the value of $\beta(s)$ recommended for second order RKC is $\beta(s) = \frac{2}{3}(s^2 - 1)(1 - \frac{2}{15}\varepsilon)$, where $\varepsilon = 2/13$ is a small damping coefficient to enhance stability. Hence

$$\|P_{hs}(-kA_h)v_h\| \leqslant \max_l |P_{hs}(z_l)| \; \|v_h\| \leqslant \|v_h\|. \tag{30}$$

The operators $P_{hj}(-kA_h)$ *are termed stability functions of the RKC methods.*

We are now in a position to study the *internal stability* of the finite element semi-Lagrangian RKC methods. In this respect, we shall follow again the approach of [14]. For $j = 0, 1, 2, \dots, s$, let

$$y_{hj} = \sum_{i=1}^{M} Y_{ji} \phi_i \quad \text{and} \quad r_{hj} = \sum_{i=1}^{M} r_{ji} \phi_i$$

be the j-stage solution of the RKC schemes and a perturbation function introduced at stage j (e.g. round off errors) respectively. In what follows we write with an upper wavy symbol the perturbed variables, unless otherwise established. So that, the perturbed RKC scheme for (27), with $\bar{f}_h = 0$, reads

$$\tilde{y}_{h0} = \tilde{\bar{u}}_h^n,$$

$$\tilde{y}_{h1} = \tilde{y}_{h0} + \tilde{\mu}_1 k \tilde{\bar{g}}_{h0} + r_{h1}$$

$$\tilde{y}_{hj} = \mu_j \tilde{y}_{hj-1} + v_j \tilde{y}_{hj-2} + (1 - \mu_j - v_j)\tilde{y}_{h0} + \tilde{\mu}_j k \tilde{g}_{hj-1} + \tilde{\gamma}_j k \tilde{g}_{h0} + r_{hj}, \quad 2 \leqslant j \leqslant s,$$

$$\tilde{u}_h^{n+1} = \tilde{y}_{hs}, \quad n = 0, 1, \ldots,$$

where $\tilde{g}_{hj} = -A_h \tilde{y}_{hj}$.

For $0 \leqslant j \leqslant s$, let

$$\tilde{e}^n = \tilde{u}_h^n - u_h^n; \quad d_j = \tilde{y}_{hj} - y_{hj} \quad \text{with } d_0 = \tilde{\bar{e}}^n = \tilde{\bar{u}}_h^n - \bar{u}_h^n \text{ and } d_s = \tilde{e}^{n+1},$$

be the errors introduced by the perturbations. Since both the unperturbed and the perturbed schemes are linear, it follows that

$$d_0 = \tilde{\bar{e}}^n,$$

$$d_1 = d_0 - \tilde{\mu}_1 k A_h d_0 + r_{h1}$$

$$d_j = \mu_j d_{j-1} + v_j d_{j-2} + (1 - \mu_j - v_j) d_0 - \tilde{\mu}_j k A_h d_{j-1} - \tilde{\gamma}_j k A_h d_0 + r_{hj}, \quad 2 \leqslant j \leqslant s,$$

$$\tilde{e}^{n+1} = d_s, \quad n = 0, 1, \ldots .$$

Noticing that the expressions for all d_j are linear, then one obtains by substitution

$$d_j = P_{hj}(-kA_h)\tilde{\bar{e}}^n + \sum_{k=1}^{j} Q_{hjk}(-kA_h) r_{hk} \quad \text{for } 1 \leqslant j \leqslant s,$$

where the operators $Q_{hjk}(-kA_h) : V_h \to V_h$ are polynomials in $-kA_h$ of degree $j-k$. These operators determine the propagation of the perturbation over the stages j, because of this, they are called *internal stability functions*. We bound $\tilde{e}^{n+1}$ in the L^2-norm. Thus, for $j = s$

$$\|\tilde{e}^{n+1}\| \leqslant \max_{l=1,\ldots,M} |P_{hs}(z_l)| \; \|\tilde{\bar{e}}^n\| + \sum_{k=1}^{s} \|Q_{hsk}(-kA_h)\| \, \|r_{hk}\|.$$

In [14, Section 3] the following bound for the operator norm is proved

$$\sum_{k=1}^{s} \|Q_{hsk}(-kA_h)\| \leqslant (s - k + 1)(1 + C\varepsilon),$$

where C is a constant of moderate size and $\varepsilon > 0$ is a small damping parameter used in the RKC schemes. By substituting this bound and taking into account (30) and (26) it follows that

$$\|\tilde{e}^{n+1}\| \leqslant \|\tilde{e}^n\| + \tfrac{1}{2} s(s+1) C \max_k \|r_{hk}\|.$$

This proves the *internal stability* of the finite element semi-Lagrangian RKC methods as long as $\rho(-kA_h) \leqslant \beta(s)$, because the accumulation of the perturbation errors (e.g. round off errors) does not depend on the spectrum of the operator $-kA_h$. We state this property in the following lemma:

Lemma 7. *Assume that* T1–T5 *and* $\rho(-kA_h) \leqslant \beta(s)$ *hold. Then, the finite element semi-Lagrangian RKC methods possess internal stability in the* L^2*-norm.*

5.3. Convergence

We study in this section the *local* convergence of the methods presented in this paper. We define the error at time t_n as

$$e^n = u^n - u_h^n = (u^n - \mathbb{R}u^n) + (\mathbb{R}u^n - u_h^n) \equiv \eta^n + \theta^n, \tag{31}$$

where $\mathbb{R}u^n$ is the Ritz projection of the exact solution u into V_h, which is defined in Section 2, η^n is the approximation error, which is bounded in Lemma 1, and θ^n is the evolutionary error. Our analysis, therefore, is mainly concerned with the estimation of θ^n. To do so we shall use the specific properties of the transport step collected in Lemmata 2–5. Thus, we have that for all $t \in [t_n, t_{n+1}]$ the map $x \to X(x, t_{n+1}; t)$ is a quasi-isometric homeomorphism of Ω into $\Omega^*(t)$, so that we can set

$$u^*(x,t) := u(X(x,t_{n+1};t),t) \tag{32}$$

and introduce the *ephemeral* Ritz projection and L^2-projection operators, respectively, as follows. For $t \in [t_n, t_{n+1})$, let such operators be

$$Ru^*(x,t) \equiv w_h(x,t) = \sum_j^M W_j(t)\phi_j(x) \quad \text{and} \quad P_0u^*(x,t) \equiv c_h(x,t) = \sum_j^M C_j(t)\phi_j(x), \tag{33}$$

such that for all $\varphi \in V_h$

$$(K\nabla w_h, \nabla\varphi) = (K\nabla u^*, \nabla\varphi) = \left(f(u(X(x,t_{n+1},t),t),\varphi) - \left(\frac{Du(X(x,t_{n+1};t),t)}{Dt},\varphi\right)\right) \tag{33a}$$

and

$$(c_h, \varphi) = (u^*, \varphi) = (u(X(x,t_{n+1};t),t),\varphi). \tag{34}$$

Note that by virtue of the definitions of A_h, P_0 and Eqs. (33a) and (34) it follows that

$$A_h w_h = P_0\left(f\left(u(X(x,t_{n+1};t),t) - \frac{Du(X(x,t_{n+1};t),t)}{Dt}\right)\right). \tag{35}$$

Hence, for $t = t_{n+1}, w_h^{n+1} = Ru^{n+1}$ and $c_h^{n+1} = P_0u^{n+1}$ for all n. So that, by virtue of Lemma 1 and Lemma 5 we have the following results whose proofs are elementary and therefore omitted.

Lemma 8. *Assume that* $u \in L^\infty(0,T;H^{m+1}(\Omega))$, *then for any* $t \in [t_n, t_{n+1})$ *there exist constants* C_1 *and* C_2 *such that*

$$\|u^* - w_h\| + h\|u^* - w_h\|_1 \leqslant C_1 h^{m+1}\|u\|_{m+1},$$

$$\|u^* - c_h\| \leqslant C_2 h^{m+1}\|u\|_{m+1}. \tag{36}$$

To proceed further with the estimate of θ^{n+1} is convenient to obtain an expression for each $w_h(x,t_n+c_jk)$ that fits the RKC formula (18). To do so, we first introduce some notation that will help to simplify the writing of future formulas. Thus, for $j=0,1,\ldots,s$, we set $\omega_{hj}\equiv w_h(x,t_n+c_jk), u_j^*\equiv u^*(x,t_n+c_jk)=u(X(x,t_{n+1};t_n+c_jk),t_n+c_jk)$ and following the expressions of (18):

$$
\begin{aligned}
&\omega_{h0}=w_h^n,\\
&\omega_{h1}=\omega_{h0}+\tilde{\mu}_1 kb_{h0},\\
&\omega_{hj}=\mu_j\omega_{hj-1}+\nu_j\omega_{hj-2}+(1-\mu_j-\nu_j)\omega_{h0}\\
&\qquad +\tilde{\mu}_j kb_{hj-1}+\tilde{\gamma}_j kb_{h0},\quad 2\leqslant j\leqslant s, \qquad (37)\\
&w_h^{n+1}=\omega_{hs},
\end{aligned}
$$

where, by using (35), is a simple matter to see that the terms $\tilde{\mu}_1 kb_{hj}$ are as follows:

$$\tilde{\mu}_1 kb_{h0}=-\tilde{\mu}_1 kA_h\omega_{h0}+(\omega_{h1}-\omega_{h0})+\tilde{\mu}_1 kP_0\left(f(u^*)-\frac{Du^*}{Dt}\right)\Bigg|_{t=t_n}, \tag{38}$$

and for $2\leqslant j\leqslant s$

$$
\begin{aligned}
\tilde{\mu}_j kb_{hj-1}+\tilde{\gamma}_j kb_{h0}=&-\tilde{\mu}_j kA_h\omega_{hj-1}-\tilde{\gamma}_j kA_h\omega_{h0}+(\omega_{hj}-\mu_j\omega_{hj-1}-\nu_j\omega_{hj-2}-(1-\mu_j-\nu_j)\omega_{h0})\\
&+\tilde{\mu}_j kP_0\left(f(u^*)-\frac{Du^*}{Dt}\right)\Bigg|_{t=t_n+c_{j-1}k}+\tilde{\gamma}_j kP_0\left(f(u^*)-\frac{Du^*}{Dt}\right)\Bigg|_{t=t_n}. \qquad (39)
\end{aligned}
$$

Next, by subtracting (18) from (37), setting $\alpha_j=\omega_{hj}-y_{hj}$ and taking into account that for $v_h\in V_h$ $P_0v_h=v_h$, we obtain (see Appendix A for details)

$$\theta^{n+1}=(P_{hs}(z)+N_{hs}(k,z))\alpha_0+\sum_{l=1}^{s}Q_{hsl}(z)r_{hl}+\sum_{p=1}^{s-1}S_{hsp}(k,z)r_{hp}\equiv(B_1+B_2+B_3) \tag{40}$$

where the following items must be noticed:

(a) the notation $z=-kA_h$ has been used;
(b) for $0\leqslant j\leqslant s$, the operator $N_{hj}(k,z):V_h\to V_h$ can be recast as a polynomial in k of degree j whose expression is $N_{hj}(k,z)=\sum_{m=1}^{j}n_m(z)k^m$, and such that there exists a positive and bounded $C(u,u_h)$ satisfying

$$\|N_{hj}(k,z)\|\leqslant kC,\quad 1\leqslant j\leqslant s; \tag{41}$$

(c) for $2\leqslant j\leqslant s$ and $1\leqslant p\leqslant j-1$, the operator $S_{hjp}(k,z):V_h\to V_h$ can also be written as $S_{hjp}(k,z)=\sum_{m=1}^{j-p}s_m(z)k^m$, a polynomial in k of degree $j-p$, such that there exists another $C(u,u_h)$, positive and bounded, satisfying

$$\|S_{hjp}(k,z)\|\leqslant kC,\quad 1\leqslant j\leqslant s; \tag{42}$$

(d) for $2 \leqslant j \leqslant s$ and $1 \leqslant l \leqslant j$, the operator $Q_{hjl}: V_h \to V_h$ is also a polynomial in z which is a bounded as (see [14])

$$\|Q_{hjl}(z)\| \leqslant (j - l + 1)(1 + \varepsilon C), \tag{43}$$

here C is a positive constant of moderate size and $\varepsilon > 0$ is the damping coefficient of the expression of $\beta(s)$. We recall that a recommended value of ε for second order RKC methods is $\frac{2}{13}$;

(e) for $1 \leqslant j \leqslant s$, the terms r_{hj} are given as

$$r_{hj} = A_{j3} + A_{j4} + A_{j5} + A_{j6} \tag{44}$$

with

$$A_{13} = \tilde{\mu}_1 k P_0(f(u_0^*) - f_h(\omega_{h0})) \qquad A_{14} = P_0((\omega_{h1} - u_1^*) - (\omega_{h0} - u_0^*))$$

$$A_{15} = P_0(u_1^* - u_0^*) \qquad A_{16} = -P_0\left(\tilde{\mu}_1 k \left.\frac{Du^*}{Dt}\right|_{t=t_n}\right), \tag{44a}$$

and for $2 \leqslant j \leqslant s$

$$A_{j3} = P_0(\tilde{\mu}_j k(f(u_{j-1}^*) - f(\omega_{hj-1})) + \tilde{\gamma}_j k(f(u_0^*) - f(\omega_{h0}))),$$

$$A_{j4} = P_0\left(\sum_{l=0}^{2} \delta_l(\omega_{hj-l} - u_{j-l}^*) - (1 - \mu_j - \nu_j)(\omega_{h0} - u_0^*)\right),$$

$$A_{j5} = P_0(u_j^* - \mu_j u_{j-1}^* - \nu_j u_{j-2}^* - (1 - \mu_j - \nu_j)u_0^*),$$

$$A_{j6} = -P_0\left(\tilde{\mu}_j k \left.\frac{Du^*}{Dt}\right|_{t=t_n+c_{j-1}k} + \tilde{\gamma}_j k \left.\frac{Du^*}{Dt}\right|_{t=t_n}\right), \tag{44b}$$

the coefficients δ_l in the formula for A_{j4} take the following values: $\delta_0 = 1, \delta_1 = -\mu_j$ and $\delta_2 = -\nu_j$.

We are now in a position to state the local convergence result.

Theorem 9. *Assume that the grid is sufficiently fine to solve the local extrema and that the following hypotheses hold*:

(1) $k = O(h)$; *with* $k \in (0, k^*)$, $0 < k^* < k_0 < 1$ *and* $h \in (0, h_0)$, $h_0 < 1$,
(2) T1–T5,
(3) $\rho(-kA_h) \leqslant \beta(s)$,
(4) $u \in L^\infty(0, T; H^{m+1}(\Omega) \cap W^{m+1,\infty}(\Omega))$, $D^2u/Dt^2 \in L^\infty(0, T; L^2)$ *and* $D^3u/Dt^3 \in L^2(0, T; L^2)$.

Then, there exists positive constants C_1, C_2 *and* C_3 *such that*

$$\max_{0 \leqslant t_n \leqslant T} \|u(t_n) - u_h^n\| \leqslant C_2 h^{m+1} \|u\|_{L^\infty(0,T:H^{m+1})} + C_3 e^{C_1 t_n}(A_1 + A_2)$$

where

$$A_1 = \frac{1}{k}(h^{m+1} \max_n \max_j(|\tilde{\beta}_j^n - \beta_j^n|)h^2)\|u\|_{L^\infty(0,T\,;\,W^{m+1,\infty})}$$

$$+ \left\{\|a - a_h\|_{L^\infty(0,T\,;\,L^2)} + O(k^p)\right\} \|u\|_{L^\infty(0,T\,;\,H^{m+1})}$$

and

$$A_2 = \frac{k}{s^3}\left|\left|\frac{D^2u}{Dt^2}\right|\right|_{L^\infty(0,T;L^2)} + k^2\left|\left|\frac{D^3u}{Dt^3}\right|\right|_{L^{2\infty}(0,T;L^2)}.$$

Proof. We estimate the terms B_1, B_2 and B_3 of (40) in the L^2-norm.

5.3.1. Estimate of B_1

By virtue of the bounds $\|P_{hs}(z)\| \leqslant 1$ and (41) it follows, after using the triangle inequality, that

$$\|(P_{hs}(z) + N_{hs}(k,z))\alpha_0\| \leqslant (1 + C_1k)\|\alpha_0\|. \tag{45}$$

So that, it remains to bound $\|\alpha_0\|$. To do so, we recall that

$$\alpha_0 = \omega_{h0} - y_{h0} = w_h^n - \bar{u}_h^n = (w_h^n - \bar{w}_h^n) + (\bar{w}_h^n - \bar{u}_h^n) \equiv (w_h^n - \bar{w}_h^n) + \bar{\theta}^n,$$

where $\bar{w}_h^n$ and $\bar{\theta}^n$ are obtained from Ru^n and θ^n, respectively, by the procedure described in the transport step. On the other hand, for convergence we need the following hypothesis. For every j and n, $\beta_j^n \leqslant \bar{\beta}_j^n(1 + Ck)$, where C is a constant independent of k and h, and $\bar{\beta}_j^n$ denotes the limiting coefficient for θ^n. Then, by virtue of the stability of the transport step (Lemma 6) we have that

$$\|\bar{\theta}^n\| \leqslant (1 + Ck)\|\theta^n\|.$$

Hence,

$$\|\alpha_0\| \leqslant \|w_h^n - \bar{w}_h^n\| + (1 + Ck)\|\theta^n\|. \tag{46}$$

From Appendix B it follows that

$$\|w_h^n - \bar{w}_h^n\| \leqslant C[(h^{m+1} + \max_i|\tilde{\beta}_i^n - \beta_i^n|h^2)\|u^n\|_{m+1,\infty}] \\ + k\{\|a - a_h\|_{L^\infty(t_n,t_{n+1}:\ L^2)} + \mathrm{O}(k^p)\}\|u\|_{L^\infty(t_n,t_{n+1};H^1)}. \tag{47}$$

So that, from (45)–(47) it follows that

$$\|(P_{hs}(z) + N_{hs}(k,z))\alpha_0\| \leqslant (1 + C_1k)\|\theta^n\| + C_2(1 + C_1k)[(h^{m+1} + \max_i|\tilde{\beta}_i^n - \beta_i^n|h^2)\|u^n\|_{m+1,\infty} \\ + k\{\|a - a_h\|_{L^\infty(t_n,t_{n+1}:\ L^2)} + \mathrm{O}(k^p)\}\|u\|_{L^\infty(t_n,t_{n+1};H^1)}]. \tag{48}$$

Notice that $C_i = C_i(u, u_h)$, $i = 1, 2$, are bounded. (see Appendix B for details)

5.3.2. Estimates of B_2 and B_3

To estimate both B_2 and B_3 we evaluate first the terms r_{hl}. Thus, from (44a) and (44b) it follows that for $1 \leqslant j \leqslant s$

$$\|r_{hj}\| \leqslant \sum_{m=3}^{6}\|A_{jm}\|.$$

Hence, we need estimate each of the A_{jm} terms.

Estimate of A_{j3}:

From (44a) and (44b),

$$A_{j3} = P_0(\tilde{\mu}_jk(f(u_{j-1}^*) - f(\omega_{hj-1})) + \tilde{\gamma}_jk(f(u_0^*) - f(\omega_{h0}))).$$

Since P_0 is a bounded operator and the coefficients $\tilde{\mu}_j$ and $\tilde{\gamma}_j$ are also bounded, then there is a constant $C > 0$ such that for all j

$$\|A_{j3}\| \leqslant Ck(\|f(u^*_{j-1}) - f(\omega_{hj-1})\| + \|f(u^*_0) - f(\omega_{h0})\|)$$

and by virtue of (3) it follows that

$$\|A_{j3}\| \leqslant C(u)kh^{m+1}\|u(t)\|_{L^\infty(t_n,t_{n+1};H^{m+1})}. \tag{49}$$

Estimate of A_{j4}:
From (44a) and (44b),

$$A_{j4} = P_0\left(\sum_{l=0}^{2}\delta_l(\omega_{hj-l} - u^*_{j-l}) - (1-\mu_j-\nu_j)(\omega_{h0} - u^*_0)\right).$$

Using the same arguments as in the estimate of A_{j3} yields

$$\|A_{j4}\| \leqslant Ch^{m+1}\|u(t)\|_{L^\infty(t_n,t_{n+1};H^{m+1})}. \tag{50}$$

Estimate of $A_{j5} + A_{j6}$:
From (44a) and (44b)

$$A_{j5} + A_{j6} = P_0(u^*_j - \mu_j u^*_{j-1} - \nu_j u^*_{j-2} - (1-\mu_j-\nu_j)u^*_0)$$
$$- P_0\left(\tilde{\mu}_j k \left.\frac{Du^*}{Dt}\right|_{t=t_n+c_{j-1}k} + \tilde{\gamma}_j k \left.\frac{Du^*}{Dt}\right|_{t=t_n}\right).$$

By performing a Taylor series expansion of u^*_j, u^*_{j-1}, u^*_0 and $Du^*/Dt|_{t=t_n}$ about the point $t_n + c_{j-1}k$, and collecting terms of the same power in k, we get

$$A_{j5} + A_{j6} = p_{1j}k\left.\frac{Du^*}{Dt}\right|_{t=t_n+c_{j-1}k} + p_{2j}k^2\left.\frac{D^2u^*}{Dt^2}\right|_{t=t_n+c_{j-1}k}$$
$$+ p_{3j}\int_{t_n}^{t_{n+1}}(\tau - t_n)^2\left.\frac{D^3u^*}{D^3t}\right|_{t=\tau}\mathrm{d}\tau, \tag{51}$$

where p_{1j}, p_{2j} and p_{3j} correspond to the coefficients θ_{1j}, θ_{2j} and θ_{3j} respectively, in [14, Section 4]. The expressions of these coefficients are

$$p_{1j} = c_j - \mu_j c_{j-1} - \nu_j c_{-2} - \tilde{\mu}_j - \tilde{\gamma}_j = 0.$$

Notice that for any jth coefficients p_{1j} are identically zero due to the relationships among the coefficients of the RKC schemes,

$$p_{2j} = \tfrac{1}{2}(c_j^2 - \mu_j c_{j-1}^2 - \nu_j c_{j-2}^2) - \tilde{\mu}_j c_{j-1}, \quad 1 \leqslant j \leqslant s.$$

The coefficients p_{2j}, they are identically zero for $j \geqslant 4$, but

$$p_{21} = \frac{1}{2}c_1^2, \quad p_{22} = -\frac{1}{2}\mu_2 c_1^2, \quad p_{23} = -\frac{1}{2}\nu_3 c_1^2.$$

In general, the coefficients p_{3j} are bounded for the family of second and first order RKC schemes.

Next, recalling the expressions of B_2 and B_3, see (40), we can set

$$\|B_2\| \leqslant \sum_{l=1}^{s} \|Q_{hsl}(z)\| \|A_{l3}\| + \sum_{l=1}^{s} \|Q_{hsl}(z)\| \|A_{l4}\| + \sum_{l=1}^{s} \|Q_{hsl}(z)(A_{l5} + A_{l6})\|. \quad (52)$$

The first two terms on the right hand side of (52) are bounded by (49), (50) and the fact that for any l, $\|Q_{hsl}\|$ is bounded by (43). So that

$$\sum_{l=1}^{s} \|Q_{hsl}(z)\| \, \|A_{l3}\| + \sum_{l=1}^{s} \|Q_{hsl}(z)\| \, \|A_{l4}\| \leqslant C(1 + C_3(u)k)h^{m+1}\|u\|_{L^\infty(t_n,t_{n+1};H^{m+1})}. \quad (53)$$

To bound the remaining term in (52), we consider (51) and follow the argument of [14] to get

$$\sum_{l=1}^{s} \|Q_{hsl}(z)(A_{l5} + A_{l6})\| \leqslant C\frac{k^2}{s^3}\left\|\frac{D^2u}{Dt^2}\right\|_{L^\infty(t_n,t_{n+1};L^2)} + Ck^2\left\|\frac{D^3u}{Dt^3}\right\|_{L^2(t_n,t_{n+1};L^2)}. \quad (54)$$

Thus, collecting these estimates we have that

$$\|B_2\| \leqslant C(1 + C_3(u)k) \times \left[h^{m+1}\|u\|_{L^\infty(t_n,t_{n+1};H^{m+1})} + \frac{k^2}{s^3}\left\|\frac{D^2u}{Dt^2}\right\|_{L^\infty(t_n,t_{n+1};L^2)} + k^2\left\|\frac{D^3u}{Dt^3}\right\|_{L^2(t_n,t_{n+1};L^2)}\right]. \quad (55)$$

Similarly, applying the same arguments and considering the bound (42) for the operators S_{sp} we obtain that

$$\|B_3\| \leqslant kC_4(u)(1 + C_5(u)k) \times \left[h^{m+1}\|u\|_{L^\infty(t_n,t_{n+1};H^{m+1})} + \frac{k^2}{s^3}\left\|\frac{D^2u}{Dt^2}\right\|_{L^\infty(t_n,t_{n+1};L^2)} + k^2\left\|\frac{D^3u}{Dt^3}\right\|_{L^2(t_n,t_{n+1};L^2)}\right]. \quad (56)$$

From (40), (54), (55) and the fact that $\|u\|_{m+1} \leqslant C\|u\|_{m+1,\infty}$ holds for all t, it follows that

$$\|\theta^{n+1}\| \leqslant (1 + C_1k)\|\theta^n\| + C\Bigg[(h^{m+1} + \max_n \max_i |\tilde{\beta}_i^n - \beta_i^n|h^2)\|u\|_{L^\infty(0,T:W^{m+1,\infty})}$$
$$+k\{\|a - a_h\|_{L^\infty(0,T:L^2)} + \mathrm{O}(k^p)\}\|u\|_{L^\infty(0,T:H^1)}$$
$$+\frac{k^2}{s^3}\left\|\frac{D^2u}{Dt^2}\right\|_{L^\infty(t_n,t_{n+1};L^2)} + k^2\left\|\frac{D^3u}{Dt^3}\right\|_{L^2(t_n,t_{n+1};L^2)}\Bigg],$$

where $C = 2(1 + (k/2))\max\{C_2(u)(1 + C_1(u)k),\ C(1 + C_3(u)k),\ C_4(u)(1 + C_5(u)k)\}$. Hence, by virtue of Gronwall inequality we have that

$$\|\theta^{n+1}\| \leqslant e^{C_1t_{n+1}}\|\theta^0\| + Ce^{C_1t_{n+1}}\frac{1}{k}\Bigg[(h^{m+1} + \max_n \max_i |\tilde{\beta}_i^n - \beta_i^n|h^2)\|u\|_{L^\infty(0,T\,:\,W^{m+1,\infty})}$$
$$+\{\|a - a_h\|_{L^\infty(0,T\,:\,L^2)} + \mathrm{O}(k^p)\}\|u\|_{L^\infty(0,T\,:\,H^1)}$$
$$+\frac{k}{s^3}\left\|\frac{D^2u}{Dt^2}\right\|_{L^\infty(0,T;L^2)} + k^2\left\|\frac{D^3u}{Dt^3}\right\|_{L^2(0,T;L^2)}\Bigg]. \quad (57)$$

We end the proof of the theorem by applying the triangle inequality to (31) and using the estimates of Lemma 1 and (57). □

Few remarks are now in order. First, the estimate of Theorem 9 is local because $\exp(C_1 T)$ becomes unbounded as $T \to \infty$, making meaningless the estimate (57). However, if f were $f = f(x,t)$, then Theorem 9 would be valid for all T because the constant multiplying the error in (57) would not grow exponentially with T. Second, Theorem 9 does not show the classical second order in time estimate, for h and s fixed, because of the term $k/s^3 \|D^2u/Dt^2\|_{L^\infty(0,T;L^2)}$; however, it is pointed out in [14] that if $\partial^2 u/\partial t^2 = 0$ on the boundary one can recover for the second order RKC scheme the estimate $O(k^2)$, at least for linear parabolic problems. Although such considerations may not be valid for the semi-Lagrangian-RKC schemes, we note that one can take s sufficiently large in order to get $k/s^3 = O(k^2)$; in fact, this is what happens in practical computations.

6. Numerical examples

The first example is the problem

$$\frac{Du}{Dt} = K\Delta u \quad \text{in } \Omega \times (0,T],$$

$$u(x,y,0) = v_1(x,0)v_2(y,0) \quad \text{in } \Omega,$$

$$u(x,y,t)|_{\partial\Omega} = v_1(x,t)v_2(y,t), \quad (x,y) \in \partial\Omega,$$

where $\Omega = (0,1) \times (0,1)$,

$$v_i(\xi,t) = \frac{0.1e^{-A} + 0.5e^{-B} + e^{-C}}{e^{-A} + e^{-B} + e^{-C}},$$

with $\xi = x$ for $i = 1$ and $\xi = y$ for $i = 2$, $A = 0.05/K(\xi - 0.5 + 4.95t)$, $B = 0.25/K(\xi - 0.5 + 0.75t)$, $C = 0.50/K(\xi - 0.375)$. The velocity $a(x,y,t) = (v_1(x,t), v_2(y,t))$ and the analytical solution is

$$u(x,y,t) = v_1(x,t)v_2(y,t).$$

The partition D_h is composed of 25×25 squares which are divided into two quadratic triangles, this means that $h \simeq 0.05$. The time step $k = 0.01$ will be kept constant all the time, so that we determine the number of stages s to satisfy the linear stability criterium of the reaction–diffusion step

$$\rho(-kA_h) \leqslant \beta(s) = \frac{2}{3}(s^2 - 1)\left(1 - \frac{2}{15}\varepsilon\right), \quad \varepsilon = \frac{2}{13}$$

by the formula (see [14])

$$s = 1 + \text{integer part of } \left[\sqrt{1 + \frac{\rho(-kA_h)}{0.653}}\right]. \tag{58}$$

We must remark that this number of stages is necessary for stability, however Theorem 9 guarantees an error estimate of second order in time if the ratio $k/s^3 = k^2$. So that, if (58) gives a value for s that is insufficient for $k/s^3 = k^2$, then we increase the number of stages s to satisfy the order

Table 1
The discrete L_2-norm for the problem of fronts for the viscosity $v = 5 \times 10^{-4}$. Number between brackets refers to CPU time (in s) for the (CRKC) and (RKC) schemes

Time step k	Schemes	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$	$t = 0.6$
	(CRKC)	0.082(3.11)	0.100(4.36)	0.0948(5.65)	0.089(6.9)	0.103(8.17)
$k = 0.01$	(RKC)	0.173(3.12)	0.198(4.7)	0.187(5.8)	0.176(7.1)	0.221(8.61)

Table 2
The discrete L_2-norm for the problem of fronts for the viscosity $v = 10^{-2}$. Number between brackets refers to CPU time (in s) for the (CRKC) and (RKC) schemes

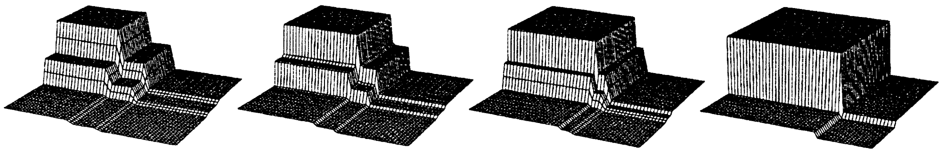
Time step k	Schemes	$t = 0.2$	$t = 0.3$	$t = 0.4$	$t = 0.5$	$t = 0.6$
	(CRKC)	0.032(3.13)	0.0356(4.41)	0.0385(5.7)	0.0414(6.98)	0.0446(8.27)
$k = 0.01$	(RKC)	0.041(3.62)	0.048(4.95)	0.0491(6.2)	0.053(7.8)	0.071(8.8)

condition. In this example, the matrix $-\mathbf{M}^{-1}\mathbf{S}$ of the discrete operator A_h is constant and an upper bound for $\rho(-kA_h)$ is estimated once and for all by using Gerschgorin theorem in the code of [10]. The departure points $\{X_{hj}^n\}$ are computed with the algorithm of [1]. In Table 1 we show the L^2-norm errors for two different diffusion coefficients, $K_1 = 10^{-2}$ and $K_2 = 5.0 \times 10^{-4}$, for the conventional Runge–Kutta–Chebyshev (RKC) and semi-Lagrangian Runge–Kutta–Chebyshev (CRKC) schemes at different time instants with the number of stages $s = 5$. For both diffusion coefficients the number of stages s for stability according to (58) is $s = 3$, but to obtain an error estimate of second order in time $k/s^3 = k^2$, so that, with $k = 0.01$ $s \geqslant \sqrt[3]{100}$; thus $s = 5$ will be a good number of stages to achieve both stability and second order in time error. Numbers inside the parentheses mean the CPU time in seconds. From the values of this table we conclude that semi-Lagrangian RKC schemes are able to produce more accurate solutions than conventional RKC schemes in convection dominated problems at the same computational cost. Fig. 1 illustrates the behavior of the numerical solutions as compared with the exact solution when $K = 5.0 \times 10^{-4}$. The semi-Lagrangian RKC schemes proposed in this paper resolve very well the regions of strong spatial variation that the solution exhibits as time progresses, while the conventional RKC produces large wiggles in a neighborhood of such regions, which contribute to increase the error. As a remark concerning the numerical examples, we must say that we use part the code of [10] for the calculations of conventional RKC methods and the reaction–diffusion step of the semi-Lagrangian RKC methods (Table 2).

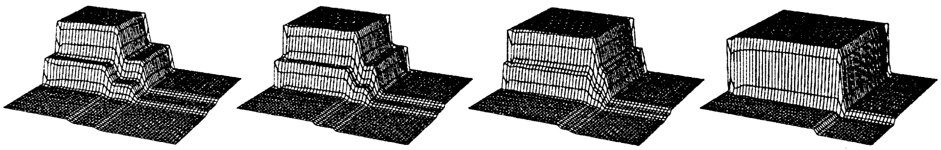
The second example is the Brusselator model with convection

$$\frac{Du}{Dt} = K\Delta u + f_1(u,v) \quad \text{in } \Omega \times (0,T],$$

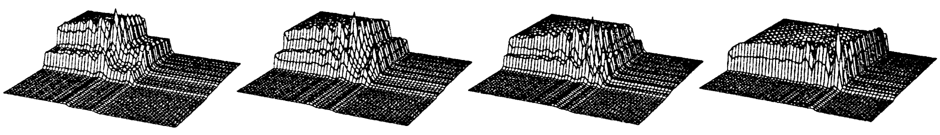
$$\frac{Dv}{Dt} = K\Delta v + f_2(u,v) \quad \text{in } \Omega \times (0,T],$$



The analytical solution at $t = 0.2$, $t = 0.3$, $t = 0.4$ and $t = 0.6$, respectively.



The numerical solution obtained by (CRKC) at $t = 0.2$, $t = 0.3$, $t = 0.4$ and $t = 0.6$, respectively.



The numerical solution obtained by (RKC) at $t = 0.2$, $t = 0.3$, $t = 0.4$ and $t = 0.6$, respectively.

Fig. 1. Numerical results obtained by (CRKC) and (RKC) for the problem of fronts with $v = 5 \times 10^{-4}$.

$$u(x,y,0) = 0.5 + y, \quad v(x,y,0) = 1.0 + 5x \quad \text{in } \Omega,$$

$$\frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = 0 \quad \text{in } \partial\Omega \times (0,T),$$

where $\Omega = (0,1)\times(0,1)$, $f_1(u,v) = 1.0 + u^2v - 4.4u$, $f_2(u,v) = 3.4u - u^2v$, the velocity $a(x,y,t) = (-\partial\psi/\partial y, \partial\psi/\partial x)$ with $\psi(x,y,t) = (1-e^{-t})\sin^2\pi x \sin^2\pi y$. Notice that a is divergence free and $a\cdot n = 0$ on $\partial\Omega$. The parameter of the computations are $h = 0.01$, $k = 0.05$ (constant all the time) $K = 0.002$ and $s = 5$, so that we have a time error of second order. Fig. 2 is a three-dimensional representation of the semi-Lagrangian RKC solution of the component v at different time instants, whereas in Fig. 3 we show a cross section of the results of Fig. 2 along the main diagonal of Ω. In order to compare the solution of the semi-Lagrangian RKC method, we have also calculated the solution by the second order conventional RKC method. Fig. 4 shows a three-dimensional representation of component v computed by conventional RKC method at the same time instants as Fig. 2. Fig. 5 is a cross section of the results of Fig. 4. By simple visual inspection we observe that the conventional RKC solution has the same qualitative behavior as the solution of semi-Lagrangian RKC method, but for long time the latter is a little bit larger.

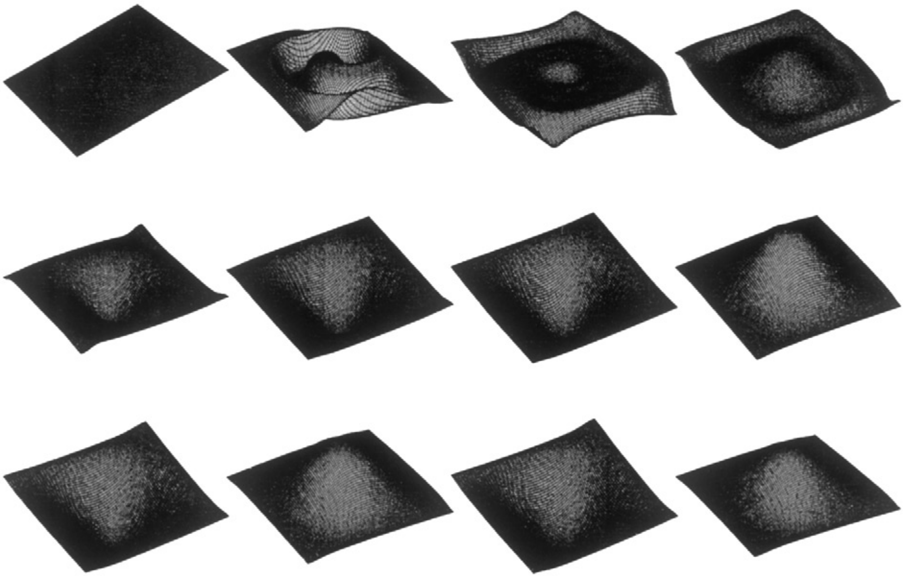


Fig. 2. (CRKC): The numerical result in 3D for the problem of Brusselator for v at $t = 0$, 1, 3, 7, 9, 11, 13, 15, 19, 23, 27, 31.

7. Conclusions

We have introduced and analyzed a numerical method to solve convection dominated convection–reaction–diffusion problems, in which the complex eigenvalues of the Jacobian matrix of the reaction–diffusion terms have small imaginary parts.

The features of such a method are: (i) a quasi-monotone, linearly unconditionally stable semi-Lagrangian treatment of the convection terms in the framework of C^0-finite elements of degree greater than 2, and (ii) the treatment of the reaction–diffusion terms by an s-stage Runge–Kutta–Chebyshev method that possesses an extended real stability interval. Both features play complementary roles that yield an efficient, accurate and stable method due to the following reasons. On one side, since the stability region of RKC methods is a narrow strip along the negative real axis, these methods may be a good choice to deal with those stiff problems, arising in the space discretization of parabolic problems, in which the eigenvalues of the Jacobian matrix of the discrete differential operator have small or zero imaginary parts; however, it is known that strong convection terms yield a Jacobian matrix whose eigenvalues may have large imaginary parts, which are a source of conflict for RKC methods. Our method circumvents such a difficulty by discretizing the material derivative via a quasi-monotone semi-Lagrangian scheme, which uses high degree piecewise polynomial interpolation to approximate the solution at the feet of the Characteristics of the convection (or transport) operator. The output of the semi-Lagrangian scheme is next used as input in the application of RKC methods to integrate the reaction–diffusion terms.

We have proved that the output of the semi-Lagrangian method is unconditionally stable in the L^2-norm and keeps the monotonicity properties of the analytical solution. We have made use of this result to show that the interval of the L^2-norm stability of the semi-Lagrangian RKC method is the

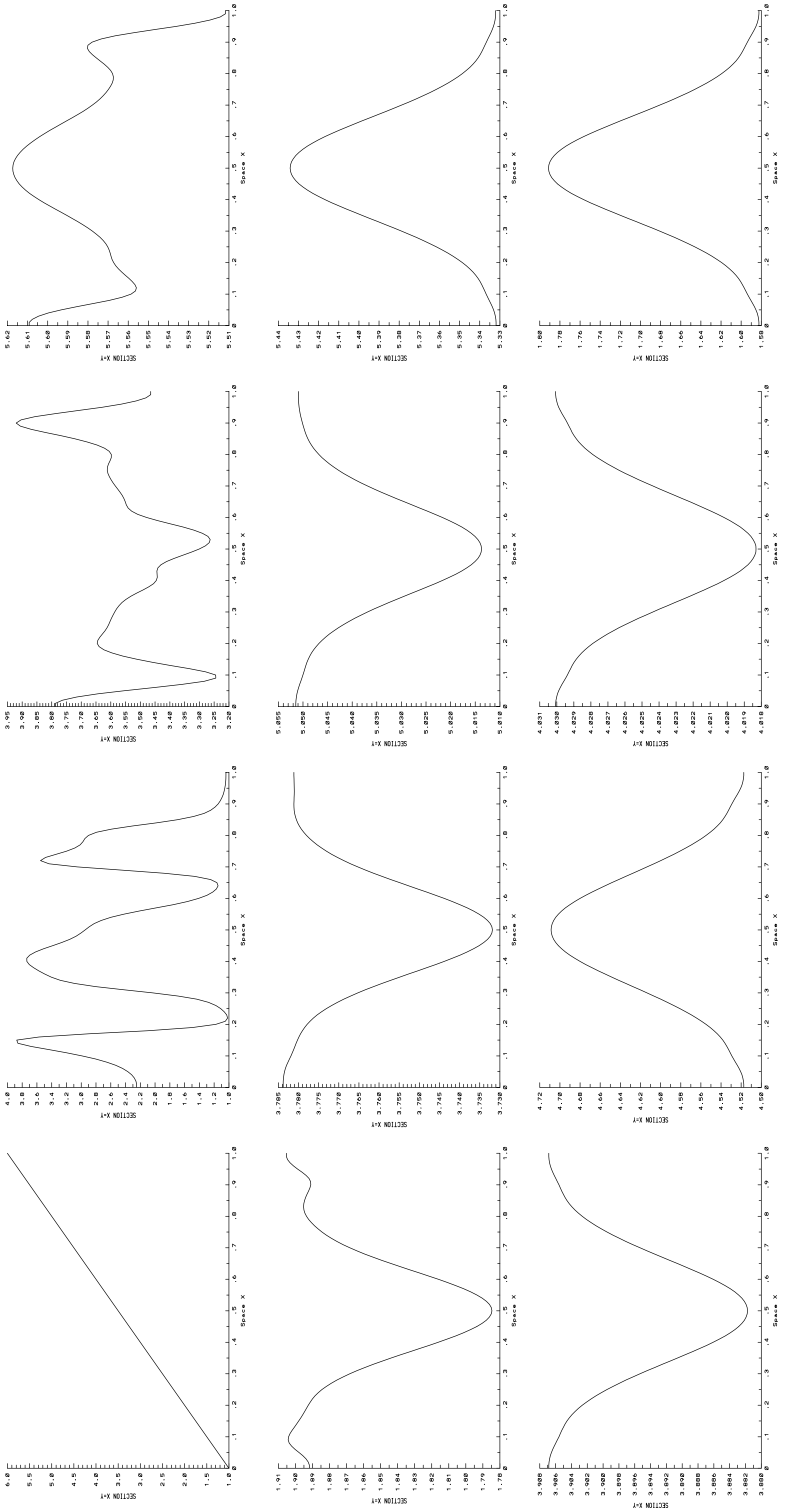


Fig. 3. The corresponding cross-sections $x \to v(x,x,t)$ of Fig. 2.

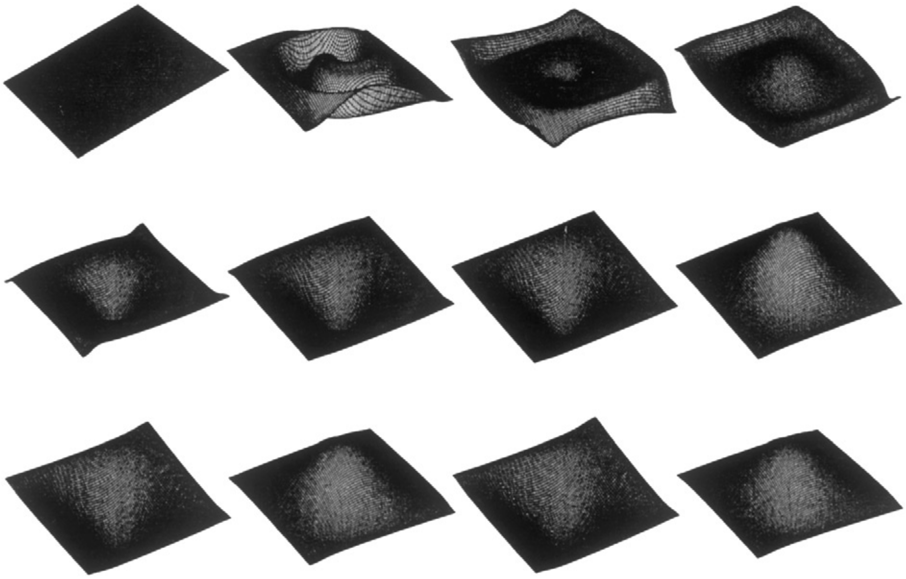


Fig. 4. (RKC): The numerical result in 3D for the problem of Brusselator for v at $t = 0$, 1, 3, 7, 9, 11, 13, 15, 19, 23, 27, 31.

same as the RKC method in absence of convection terms. After establishing the stability properties, we have obtained an error estimate $O(h^{m+1} + h^{m+1}/\Delta t + \Delta t^2)$, assuming that the analytical solution and the convection velocity are sufficiently smooth. We have run some difficult numerical examples to ascertain the behavior of the semi-Lagrangian RKC method in comparison with the conventional second order RKC method presented in [10,13,14].

Acknowledgements

This research has been partially funded by grant CLI98-1076 from Comisión Interministerial de Ciencia y Tecnología.

Appendix A

We show in this appendix the technical details to obtain Eq. (40). Thus, by subtracting (18) from (37), setting $\alpha_j = \omega_{hj} - y_{hj}$ and taking into account that for $v_h \in V_h P_0 v_h = v_h$, we obtain

$$\alpha_0 = \omega_{h0} - y_{h0} \tag{A.1}$$

$$\begin{aligned} \alpha_1 = {} & \alpha_0 + (\tilde{\mu}_1(-kA_h)\alpha_0 + \tilde{\mu}_1 k B_{h0}\alpha_0) + \tilde{\mu}_1 k P_0(f(u_0^*) - f_h(\omega_{h0})) + P_0((\omega_{h1} - u_1^*) \\ & -(\omega_{h0} - u_0^*)) + P_0(u_1^* - u_0^*) - P_0\left(\tilde{\mu}_1 k \left.\frac{Du^*}{Dt}\right|_{t=t_n}\right) \equiv (A_{11} + \cdots + A_{16}), \end{aligned} \tag{A.2}$$

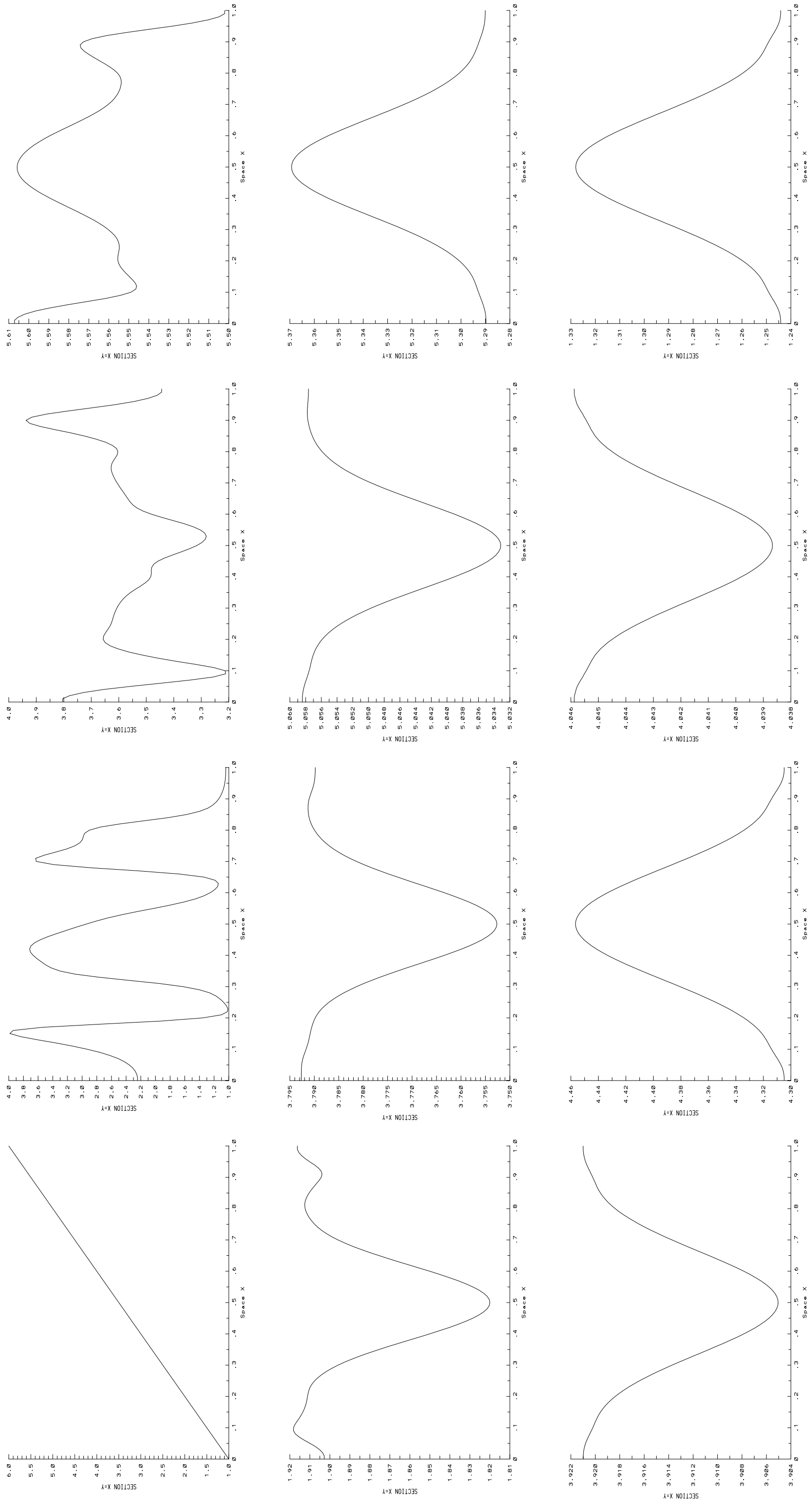


Fig. 5. The corresponding cross-sections $x \to v(x,x,t)$ of Fig. 4.

where $B_{h0}\alpha_0$ is a simplified notation for $B_h(\omega_{h0},\alpha_0)\alpha_0$, which is now defined. Thus, for $j=0,1,\dots,s$, given ω_{hj} and α_j in $V_h, \omega_{hj}\neq\alpha_j$ we define the operator $B_h(\omega_{hj},\alpha_j): V_h\to V_h$ as

$$B_h(\omega_{hj},\alpha_j)v_h=\left(\int_0^1 f_h'(\omega_{h_j}-s\alpha_j)\,\mathrm{d}s\right)v_h \tag{A.3}$$

for all $v_h\in V_h$. Hereafter, we shall denote by the simplified notation B_{hj} the operator $B_h(\omega_{hj},\alpha_j)$. Notice that the term $\tilde{\mu}_1 kB_{h0}\alpha_0+\tilde{\mu}_1 kP_0(f(u_0^*)-f_h(\omega_{h0}))$ in (A.2) is obtained from $\tilde{\mu}_1 k(P_0 f(u_0^*)-f_h(y_{h0}))$, which comes from substraction (37)–(18), as follows. Set

$$\tilde{\mu}_1 kP_0(f(u_0^*)-f_h(y_{h0}))=\tilde{\mu}_1 kP_0((f(u_0^*)-f_h(\omega_{h0}))+(f_h(\omega_{h0})-f_h(y_{h0}))),$$

and then for $j=0,1,\dots,s$, use the property

$$f_h(\omega_{hj})-f_h(y_{hj})=\left(\int_0^1 f_h'(\omega_{h_j}-s\alpha_j)\,\mathrm{d}s\right)(\omega_{hj}-y_{hj}).$$

The same idea is applied in the formulas of α_j when $j>1$. Therefore, for $2\leqslant j\leqslant s$

$$\begin{aligned}\alpha_j&=(\mu_j\alpha_{j-1}+v_j\alpha_{j-2}+(1-\mu_j-v_j)\alpha_0)\\&\quad+(\tilde{\mu}_j(-kA_h)\alpha_{j-1}+\tilde{\gamma}_j(-kA_h)\alpha_0+\tilde{\mu}_j kB_{hj-1}\alpha_{j-1}+\tilde{\gamma}_j kB_{h0}\alpha_0)\\&\quad+P_0(\tilde{\mu}_j k(f(u_{j-1}^*)-f(\omega_{hj-1}))+\tilde{\gamma}_j k(f(u_0^*)-f(\omega_{h0})))\\&\quad+P_0\left(\sum_{l=0}^{2}\delta_l(\omega_{hj-l}-u_{j-l}^*)-(1-\mu_j-v_j)(\omega_{h0}-u_0^*)\right)\\&\quad+P_0(u_j^*-\mu_j u_{j-1}^*-v_j u_{j-2}^*-(1-\mu_j-v_j)u_0^*)\\&\quad-P_0\left(\tilde{\mu}_j k\left.\frac{Du^*}{Dt}\right|_{t=t_n+c_{j-1}k}+\tilde{\gamma}_j k\left.\frac{Du^*}{Dt}\right|_{t=t_n}\right)\equiv(A_{j1}+\cdots+A_{j6}),\end{aligned} \tag{A.4}$$

where $\delta_0=1,\ \delta_1=-\mu_j$ and $\delta_2=-v_j$. To proceed further in our analysis we denote by r_{hj} the term

$$r_{hj}=A_{j3}+A_{j4}+A_{j5}+A_{j6}\quad\text{for }1\leqslant j\leqslant s, \tag{A.5}$$

and work on Eqs. (A.2)–(A.4) to derive the expressions

$$\begin{aligned}&\alpha_0=\omega_{h0}-y_{h0},\\&\alpha_1=P_1(z)\alpha_0+k\tilde{\mu}_1 B_{h0}\alpha_0+r_{h1}\\&\alpha_j=P_{hj}(z)\alpha_0+N_{hj}(k,z)\alpha_0+\sum_{l=1}^{hj}Q_{hjl}(z)r_{hl}+\sum_{p=1}^{j-1}S_{hjp}(k,z)r_{hp}\quad\text{for }2\leqslant j\leqslant s,\\&\theta^{n+1}=\alpha_s,\end{aligned} \tag{A.6}$$

where (a) the notation $z=-kA_h$ has been used,

(b) for $0 \leqslant j \leqslant s$:

(1) the operators $N_{hj}(k,z): V_h \to V_h$ are defined by

$$N_{h0}(k,z) = 0,$$

$$N_{h1}(k,z) = k\tilde{\mu}_1 B_{h0},$$

$$\begin{aligned} N_{hj}(k,z) &= k\tilde{\mu}_j B_{hj-1} \circ P_{hj-1}(z) + k\tilde{\gamma}_j B_{h0} \\ &\quad + (\mu_j + \tilde{\mu}_j z) \circ N_{hj-1}(k,z) + v_j N_{hj-2}(k,z) \\ &\quad + k\tilde{\mu}_j B_{hj-1} \circ N_{hj-1}(k,z) \quad \text{for } 2 \leqslant j \leqslant s; \end{aligned} \tag{A.7}$$

(2) the operators $Q_{hjl}(z): V_h \to V_h$ are given (see [13]) by the relations:

$$Q_{hll}(z) = 1, \quad Q_{hl+1l}(z) = \mu_{l+1} + \tilde{\mu}_{l+1} z \quad \text{for all } l,$$

and

$$Q_{hjl}(z) = (\mu_j + \tilde{\mu}_j z) \circ Q_{hj-1l}(z) + v_j Q_{hj-2l}(z) \quad \text{for } l+2 \leqslant j \leqslant s; \tag{A.8}$$

(3) finally, the operators $S_{hjl}(k,z): V_h \to V_h$ are defined by

$S_{h21}(k,z) = k\tilde{\mu}_2 B_{h1}$. For $l \geqslant 1$ $S_{hll} = 0$, and for $l > 1$ $S_{hll+1l}(k,z) = k\tilde{\mu}_{l+1} B_{hl}$,

$$\begin{aligned} S_{hjl}(k,z) &= (\mu_j + \tilde{\mu}_j z) \circ S_{hj-1l}(k,z) + v_j S_{hj-2l}(k,z) \\ &\quad + k\tilde{\mu}_j B_{hj-1} \circ (Q_{hj-1l}(z) + S_{hj-1l}(k,z)) \quad \text{for } l+2 \leqslant j \leqslant s. \end{aligned} \tag{A.9}$$

Notice that from (A.7) and (A.8) one can obtain, by substitution and rearrangement of terms, polynomial expressions for the operators $N_{hj}(k,z)$ and $S_{hjl}(k,z)$ such as $N_{hj}(k,z) = \sum_{m=1}^{j} n_m(z)k^m$ and $S_{hjl}(k,z) = \sum_{m=1}^{j-l} s_m(z)k^m$. From these expressions and taking into account (A.2) and the internal stability condition $\|P_{hj}(z)\| \leqslant 1$, it follows that $\|N_{hj}(k,z)\|$ and $\|S_{hjl}(k,z)\|$ are bounded as shown in (41) and (42), respectively.

Appendix B. Estimate of the term $w_h^n - \bar{w}_h^n$

To estimate $(w_h^n - \bar{w}_h^n)$ we shall express it as a combination of terms that we know how to estimate them. Thus, we set

$$\begin{aligned} (w_h^n - \bar{w}_h^n) &= u^{*n} - I_m u^{*n} - (u^{*n} - w_h^n) + (I_m u^{*n} - I_m \bar{u}^{*n}) \\ &\quad + (I_m \bar{u}^{*n} - \bar{u}_h^{*n}) + (\bar{u}_h^{*n} - \bar{w}_h^n), \end{aligned} \tag{B.1}$$

where u^{*n} is defined in (32), $\bar{u}^{*n} \equiv \bar{u}^*(x,t_n) = u(X_h(x,t_{n+1};t_n),t_n)$ and

$$\bar{u}_h^{*n} = \sum_{i=1}^{M} \bar{U}_i^{*n} \phi_i, \tag{B.2}$$

with (see (14a))

$$\bar{U}_i^{*n} = (1-\beta_i^n)I_1u^n(X_{hi}^n) + \beta_i^n I_m u^n(X_{hi}^n).$$

Here, $I_1u^n(X_{hi}^n)$ and $I_mu^n(X_{hi}^n)$ denote the values of $I_1u^n \in L_h$ and $I_mu^n \in S_h$, respectively, at the points $X_h(x_i, t_{n+1}; t_n)$. It is easy to estimate the first term on the right hand side of (B.1) by approximation theory and Lemma 5. Thus,

$$\|u^{*n} - I_mu^{*n}\| \leqslant Ch^{m+1}|u^n|_{m+1}. \tag{B.3}$$

Likewise, by Lemma 8 it follows that the second term is bounded as

$$\|u^{*n} - w_h^n\| \leqslant Ch^{m+1}|u^n|_{m+1}. \tag{B.4}$$

As for the third term, $(I_mu^{*n} - I_m\bar{u}^{*n})$, we have that

$$\|(I_mu^{*n} - I_m\bar{u}^{*n})\| \leqslant C\|u^{*n} - \bar{u}^{*n}\|$$

because I_m is a bounded operator. To estimate $\|u^{*n} - \bar{u}^{*n}\|$, we notice that

$$\begin{aligned} u^{*n}(x) - \bar{u}^{*n}(x) &= u(X(x, t_{n+1}; t_n), t_n) - u(X_h(x, t_{n+1}; t_n), t_n) \\ &= (X(x, t_{n+1}; t_n) - X_h(x, t_{n+1}; t_n)) \cdot \int_0^1 D_X u(X_\varsigma^n(x)) d\varsigma, \end{aligned}$$

where $X_\zeta^n(x) = X_h^n(x) + \zeta(X(x, t_{n+1}; t_n) - X_h^n(x)), 0 \leqslant \zeta \leqslant 1$. Hence, by applying Lemma 2 yields

$$\|(I_mu^{*n} - I_m\bar{u}^{*n})\| \leqslant Ck\{\|a - a_h\|_{L^\infty(t_n,t_{n+1}:\ L^2)} + \mathrm{O}(k^p)\}\|u\|_{L^\infty(t_n,t_{n+1};H^1)}. \tag{B.5}$$

The argument to bound the term $I_m\bar{u}^{*n} - \bar{u}_h^{*n}$ is more involved because although $I_m\bar{u}^{*n}$ and $\bar{u}_h^{*n}$ are both in S_h, their nodal values are not computed by the same procedure. In fact, the set of nodal values of $I_m\bar{u}^{*n}$ is the set $\{u(X_{hi}^n)\}$, whereas the nodal values of $\bar{u}_h^{*n}$ are calculated by interpolation plus the limiting procedure. So that, to account for the limiting procedure in our analysis we follow the argument of [3, Section 3.2]. Thus, for all n we define $\tilde{u}_h^n \in S_h$ as

$$\tilde{u}_h^n = I_m\bar{u}^{*n} - \bar{u}_h^{*n} = I_m(\bar{u}^{*n} - \bar{u}_h^{*n}), \tag{B.6}$$

such that the nodal values of $\tilde{u}_h^n$ can be expressed by virtue of (14a) as

$$\tilde{U}_i^n = (1-\beta_i^n)(u^n(X_{hi}^n) - I_1u(X_{hi}^n)) + \beta_i^n(u^n(X_{hi}^n) - I_mu(X_{hi}^n)). \tag{B.7}$$

Hence,

$$\|\tilde{u}_h^n\| \leqslant C\|\tilde{u}_h^n\|_{L^\infty(\Omega)} \leqslant C \max_i |\tilde{U}_i^n|$$
$$= C \max_i |(1-\beta_i^n)(u^n(X_{hi}^n) - I_1 u(X_{hi}^n)) + \beta_i^n(u^n(X_{hi}^n) - I_m u(X_{hi}^n))|$$
$$\leqslant C \max_i |(1-\beta_i^n)(u^n(X_{hi}^n) - I_1 u(X_{hi}^n))| + C \max_i |u^n(X_{hi}^n) - I_m u(X_{hi}^n)|. \quad \text{(B.8)}$$

We have to bound the two terms of the last inequality of (B.8). To do so, we shall apply (16), (17) and, depending upon the values that $P \equiv I_m u(X_{hi}^n) - I_1 u(X_{hi}^n)$ takes, distinguish the following cases: (a) $P = 0$, (b) $P > 0$ and (c) $P < 0$. Case (a) is the easiest one to deal with, for $\beta_i^n = 1$ and therefore

$$\|\tilde{u}_h^n\| \leqslant C \max_i |(u^n(X_{hi}^n) - I_m u^n(X_{hi}^n))| \leqslant C\|u^n - I_m u^n\|_{L^\infty(\Omega)}.$$

But, by approximation theory, $\|u^n - I_m u^n\|_{L^\infty(\Omega)} \leqslant Ch^{m+1}\|u(t_n)\|_{m+1,\infty}$, so that

$$\|\tilde{u}_h^n\| \leqslant Ch^{m+1}\|u^n\|_{m+1,\infty}. \quad \text{(B.9)}$$

Next, let us consider case (b). Now, β_i^n is equal either to 1 or to $Q^+/P < 1$, where $Q^+ = u^+ - I_1 u^n(X_{hi}^n) \geqslant 0$, with $u^+ = \mathrm{Max}(u_1^n, \ldots, u_{\mathrm{NH}}^n), (u_1^n, \ldots, u_{\mathrm{NH}}^n)$ being the values of $u^n(x)$ at the vertices of the element T_l that contains the point X_{hi}^n. If $\beta_i^n = 1$ for all i, then $\|\tilde{u}_h^n\|$ is bounded as in (B.9). However, if for some i $\beta_i^n = Q^+/P$, then assuming that $u^n(X_{hi}^n) - I_1 u(X_{hi}^n) \neq 0$, it follows that

$$(1-\beta_i^n)(u^n(X_{hi}^n) - I_1 u(X_{hi}^n)) = \frac{I_m u^n(X_{hi}^n) - u^+}{I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)}(u^n(X_{hi}^n) - I_1 u(X_{hi}^n))$$
$$= \frac{I_m u^n(X_{hi}^n) - u^+}{1 + (I_m u^n(X_{hi}^n) - u^n(X_{hi}^n))/(u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n))},$$

hence

$$\max_i |(1-\beta_i^n)(u^n(X_{hi}^n) - I_1 u(X_{hi}^n))| = \max_i \left| \frac{I_m u^n(X_{hi}^n) - u^+}{1 + \dfrac{(I_m u^n(X_{hi}^n) - u^n(X_{hi}^n))}{(u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n))}} \right|$$
$$\leqslant C(u^n) \max_i |I_m u^n(X_{hi}^n) - u^+|, \quad \text{(B.10)}$$

Notice that $P = I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)$ is by hypothesis greater than zero, so that $1 + (I_m u^n(X_{hi}^n) - u^n(X_{hi}^n))/(u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)) = P \neq 0$; in fact, for $u(x,t) \in W^{m+1,\infty}(\Omega), m > 1$, this term is equal to $1 - \mathrm{O}(h^{m-1})$, so that, defining

$$C(u^n) = \min_i \left| 1 + \frac{I_m u^n(X_{hi}^n) - u^n(X_{hi}^n)}{u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)} \right|^{-1} = 1 - \mathrm{O}(h^{m-1}),$$

we have that for u sufficiently smooth $C(u^n)$ is approximately equal to 1 for all time and, in general, $0 < C(u^n) < \infty$. The term $\max_i |I_m u^n(X_{hi}^n) - u^+|$ in (B.10) is bounded by considering the definition of u^+ from which it follows that if the grid is sufficiently fine to resolve local extrema, then

$$\max_i |I_m u^n(X_{hi}^n) - u^+| \leqslant \|I_m u^n - u^n\|_{L^\infty(\Omega)} \leqslant Ch^{m+1}|u^n|_{m+1,\infty}, \quad \text{(B.11)}$$

where the latter inequality is obtained by approximation theory. In (B.8), it remains to bound the term $\max_i|(u^n(X_{hi}^n) - I_m u(X_{hi}^n))|$. Again, by approximation theory

$$\max_i|(u^n(X_{hi}^n) - I_m u(X_{hi}^n))| \leqslant \|I_m u^n - u^n\|_{L^\infty(\Omega)} \leqslant Ch^{m+1}|u^n|_{m+1,\infty}. \tag{B.12}$$

Hence, collecting estimates (B.10)–(B.12) it follows that when $P > 0$

$$\|\tilde{u}_h^n\| \leqslant Ch^{m+1}|u^n|_{m+1,\infty}, \tag{B.13}$$

where C may depend on u^n as we have commented above.

The analysis for case (c) uses basically the same arguments as the analysis for case (b) yielding the same estimate for $\|\tilde{u}_h^n\|$, therefore, it will be omitted. Summarizing, we have the for all n there exists $C(u), 0 < C(u) < \infty$, such that

$$\|I_m\bar{u}^{*n} - \bar{u}_h^{*n}\| \leqslant Ch^{m+1}|u^n|_{m+1,\infty}. \tag{B.14}$$

Next, we turn our attention to estimate the term $\|\bar{u}_h^{*n} - \bar{w}_h^n\|$. To do so, we set $\bar{u}_h^{*n} - \bar{w}_h^n$ in terms of the transport step operator as

$$\bar{u}_h^{*n} - \bar{w}_h^n = \overline{\mathrm{TS}}(k, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)w_h^n$$
$$= \left(\overline{\mathrm{TS}}(k, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)u^n\right) + \left(\overline{\mathrm{TS}}(k, \beta^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)w_h^n\right).$$

The ith component of the first term on the right side is expressed as

$$\left(\overline{\mathrm{TS}}(k, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)u^n\right)_i = \left(\tilde{\beta}_i^n - \beta_i^n\right)(I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)).$$

If for all i, $\tilde{\beta}_i^n = \beta_i^n$, then this term is identically zero. So that, we assume that there are subscripts i for which $\tilde{\beta}_i^n \neq \beta_i^n$. Now, since $0 \leqslant \beta_i^n \leqslant 1$, then

$$\left|\left(\overline{\mathrm{TS}}(\Delta t, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(\Delta t, \beta^n)u^n\right)_i\right| \leqslant \left|\tilde{\beta}_i^n - \beta_i^n\right|\left|I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)\right|.$$

But

$$|I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)| = |I_m u^n(X_{hi}^n) - u^n(X_{hi}^n)| + |u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)|,$$

so by approximation theory it follows that

$$\max_i \mid I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n) \mid \leqslant C\|I_m u^n(X_{hi}^n) - I_1 u^n(X_{hi}^n)\|_{L^\infty(\Omega)} \leqslant Ch^2|u^n|_{m+1,\infty},$$

Hence, $\overline{\mathrm{TS}}(k, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)u^n$ is bounded as

$$\|\overline{\mathrm{TS}}(k, \tilde{\beta}^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)u^n\| \leqslant \max_i \mid \tilde{\beta}_i^n - \beta_i^n \mid Ch^2 \mid u^n \mid_{m+1,\infty}. \tag{B.15}$$

To bound $\overline{\mathrm{TS}}(k, \beta^n)u^n - \overline{\mathrm{TS}}(k, \beta^n)w_h^n$, we use the same hypothesis as for $\overline{\theta}^n$ together with the stability property of the transport step (Lemma 6) and then apply Lemma 1. The result that follows is

$$\|\bar{u}_h^{*n} - \bar{w}_h^n\| \leqslant C[\max_i \mid \tilde{\beta}_i^n - \beta_i^n \mid h^2|u^n|_{m+1,\infty} + h^{m+1} \parallel u \parallel_{m+1}]. \tag{B.16}$$

Hence, using the estimates (B.3)–(B.5) and (B.14)–(B.16) together with the triangle inequality in (B.1) it follows (47).

References

[1] A. Allievi, R. Bermejo, Finite element modified method of characteristics for Navier-Stokes equations, Int. J. Numer. Meth. Fluids 32 (2000) 439–464.

[2] R. Bermejo, An analysis of an algorithm for the Galerkin-characteristic method, Numer. Math. 60 (1991) 163–194.

[3] R. Bermejo, Analysis of a class of quasi-monotone and conservative semi-Lagrangian advection schemes, Numer. Math. 87 (2001) 597–623.

[4] R. Bermejo, A. Staniforth, The conversion of semi-Lagrangian advection schemes to quasi-monotone schemes, Mon. Wea. Rev. 120 (1992) 2622–2632.

[5] S. Brenner, R. Scott, The Mathematical Theory of Finite Element Methods, Springer, New York, 1994.

[6] P.J. van der Houwen, B.P. Sommeijer, On the internal stability of explicit m-stage Runge-Kutta methods for large m-values, Z. Angew Math. Mech. 60 (1980) 479–485.

[7] J.W. Jerome, Convection-dominated non linear systems: analysis of the Douglas-Russel algorithm based on approximate characteristics and invariant regions, SIAM J. Numer. Anal. 25 (1988) 815–836.

[8] V.G. Maz'ja, Sobolev Spaces, Springer, New York, 1979.

[9] O. Pironneau, On the transport-diffusion algorithm and its applications to the Navier-Stokes equations, Numer. Math. 38 (1982) 309–322.

[10] B.P. Sommeijer, L.F. Shampine, J.G. Verwer, RKC: an explicit solver for parabolic PDEs, Report MAS-R9715. CWI, Amsterdam, 1997.

[11] E. Süli, Convergence of the Lagrange-Galerkin method for the Navier-Stokes equations, Numer. Math. 53 (1988) 459–483.

[12] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems, Springer Series in Computational Mathematics, Springer, Berlin, 1997.

[13] J.G. Verwer, Explicit Runge–Kutta methods for parabolic partial differential equations, Appl. Numer. Math. 22 (1996) 359–379.

[14] J.G. Verwer, W.H. Hundsdofer, B.P. Sommeijer, Convergence properties of the Runge–Kutta–Chebyshev method, Numer. Math. 57 (1990) 157–178.