Applications of Mathematics

Tomáš Vejchodský

Fully discrete error estimation by the method of lines for a nonlinear parabolic problem

Applications of Mathematics, Vol. 48 (2003), No. 2, 129–151

Persistent URL: `http://dml.cz/dmlcz/134523`

FULLY DISCRETE ERROR ESTIMATION BY THE METHOD OF LINES FOR A NONLINEAR PARABOLIC PROBLEM*

Tomáš Vejchodský, Praha

(Received March 2, 2001)

Abstract. A posteriori error estimates for a nonlinear parabolic problem are introduced. A fully discrete scheme is studied. The space discretization is based on a concept of hierarchical finite element basis functions. The time discretization is done using singly implicit Runge-Kutta method (SIRK). The convergence of the effectivity index is proven.

Keywords: a posteriori error estimates, finite elements, nonlinear parabolic problems, effectivity index, singly implicit Runge-Kutta methods (SIRK)

MSC 2000: 65M60, 65M20, 65M15

1. Introduction

This article deals with numerical solution of parabolic partial differential equations and in particular with error estimates. We will concentrate on the one-dimensional problem only. We will study the error estimates and especially their convergence to the true error.

The error estimate is a very important quantity because the numerical solution as a product of a numerical method is worthless without some information about the error. The error estimate gives us this information although it is only an approximation. And, moreover, most adaptive methods for solving parabolic equations are based on estimates of this kind.

The inspiration for this article is in the work of Moore [9], where two *fully discrete* schemes are described. These schemes differ in the time discretization. The

* The work was supported by the grant No. 201/01/1200 of the Grant Agency of the Czech Republic.
This work won the Babuška prize 2000 in the student category.

backward difference formula method (BDF) and the *singly implicit Runge-Kutta* method (SIRK) are used. The case of SIRK scheme is treated only for the *semilinear* (i.e., $a(u) \equiv 1$) problem in Moore [9]. We will examine a *nonlinear* equation.

In general, we use the notation of Moore [9]. In Section 2, the model problem and its weak formulation are stated. The space discretization and the definition of a semidiscrete solution is shown in Section 3. Definitions of semidiscrete error estimates and of the effectivity index are given in Section 4. More details about the semidiscrete problem can be found in Segeth [13]. In Section 5, the time discretization is done using the SIRK method. A fully discrete solution is defined as well and some auxiliary lemmas are proven. Finally, in Section 6, fully discrete error estimates are described and convergence of the effectivity index is proven.

2. Model problem

We use the same model problem and the same notation as Moore [9] and Segeth [13].

Consider the nonlinear equation

$$\partial_t u - \nabla(a(u)\nabla u) + f(u) = 0 \tag{2.1}$$

for an unknown scalar function $u(x,t)$ on a space interval $x \in [c,d]$ and on a time interval $t \in (0,T]$, where $T > 0$ is fixed. The symbols ∂_t and ∇ denote the partial derivatives $\partial/\partial t$ and $\partial/\partial x$, respectively. The coefficients a and f are smooth functions and, moreover, there exist constants μ, M and L which satisfy

$$0 < \mu \leqslant a(s) \leqslant M \quad \text{for all} \quad s \in \mathbb{R}, \tag{2.2}$$

$$|a(r) - a(s)| \leqslant L|r-s| \quad \text{for all} \quad r, s \in \mathbb{R}, \tag{2.3}$$

$$|f(r) - f(s)| \leqslant L|r-s| \quad \text{for all} \quad r,\ s \in \mathbb{R}. \tag{2.4}$$

Thus, a is positive and bounded and both the coefficients satisfy the global Lipschitz condition.

Let us introduce the homogeneous Dirichlet boundary condition

$$u(c,t) = u(d,t) = 0, \quad 0 \leqslant t \leqslant T, \tag{2.5}$$

and the initial condition

$$u(x,0) = u_0(x), \quad c < x < d, \tag{2.6}$$

where u_0 is a given smooth function. We assume that the boundary and initial conditions are consistent.

If f is constant, the existence and uniqueness of u follows immediately by applying the well-known Kirchhoff transformation (see [6], [8]). For non-constant f we can use the concept of pseudomonotone operators. The existence of u can be obtained as a weak limit of Galerkin approximations. If some coercivity on f is assumed, e.g.,

$$f(r)r \geqslant C_2 r^2 - C_3 \quad \text{for all } r \in \mathbb{R}, \tag{2.7}$$

where $C_2 > 0$ and $C_3 \in \mathbb{R}$, see e.g. Roubíček [12], then the assumptions on a and f are strong enough to ensure the existence of u. Sufficient conditions for uniqueness can be derived from the theory of monotone operators (cf. [7], p. 183).

Denote by

$$(v, w) = \int_c^d v(x)w(x)\,\mathrm{d}x$$

the L^2 inner product and by $\|w\|_0$ the corresponding norm. Let $H^k = H^k(c,d)$ stand for the Sobolev space of functions whose generalized derivatives up to order k are in $L^2(c,d)$, for an integer $k \geqslant 0$. The norm in this space is

$$\|w\|_k^2 = \sum_{i=0}^{k} \left\| \frac{\partial^i w}{\partial x^i} \right\|_0^2.$$

The case $k = 1$ is important for the weak formulation. We introduce the usual subspace $H_0^1 = H_0^1(c,d)$ of functions $w \in H^1$ satisfying the homogeneous Dirichlet boundary conditions. The constants C, C_1, C_2, etc. are generic, i.e., they may represent different constant quantities in different occurrences.

We will present a weak formulation of the above model problem. The finite element discretization is based on this weak formulation.

A function $v(x,t)$ is in space $H^1([0,T], X)$, where X is a Banach space, if $v(\cdot,t) \in X$ and $\partial_t v(\cdot,t) \in X$ for almost every $t \in [0,T]$ and if the term

$$\int_0^T (\|v\|_X^2 + \|\partial_t v\|_X^2)\,\mathrm{d}t$$

is finite.

We say that $u(x,t) \in H^1([0,T], H_0^1(c,d))$ is a *weak solution* of problem (2.1) with conditions (2.5) and (2.6) if the identity

$$(\partial_t u, v) + (a(u)\nabla u, \nabla v) + (f(u), v) = 0 \tag{2.8}$$

holds for almost every $t \in (0,T]$ and all functions $v \in H_0^1$, if $u_0 \in H_0^1$ and if the identity

$$(a(u_0)\nabla u, \nabla v) = (a(u_0)\nabla u_0, \nabla v) \tag{2.9}$$

holds for $t = 0$ and all functions $v \in H_0^1$. Throughout the paper we assume that the weak solution exists and is unique.

3. Discretization in space

Let us choose a positive integer p, which denotes the order of approximation. We solve problem (2.1) with conditions (2.5) and (2.6) or, in the weak formulation, (2.8) and (2.9), by the finite element method with a piecewise polynomial hierarchical basis functions of degree p. We introduce a partition

$$c = x_0 < x_1 < \ldots < x_{N-1} < x_N = d$$

of the interval $[c,d]$ into N subintervals (x_{j-1}, x_j), $j = 1, \ldots, N$. We further put $h_j = x_j - x_{j-1}$, $j = 1, \ldots, N$, and

$$h = \max_{j=1,\ldots,N} h_j.$$

Let this partition belong to the family of partitions which satisfies the so-called inverse assumption, i.e., there exists a constant $C_G > 0$ such that

$$C_G h \leqslant h_j \tag{3.1}$$

holds for $j = 1, 2, \ldots, N$.

We use the finite element concept described in Szabó, Babuška [14]. Let us construct a finite dimensional subspace $S_0^{N,p} \subset H_0^1$ in the following way. A function V belongs to $S_0^{N,p}$ if

$$V(x) = \sum_{j=1}^{N-1} V_{j1}\varphi_{j1}(x) + \sum_{j=1}^{N}\sum_{k=2}^{p} V_{jk}\varphi_{jk}(x),$$

where

$$\varphi_{j1}(x) = \begin{cases} (x - x_{j-1})/h_j, & x_{j-1} \leqslant x \leqslant x_j, \\ (x_{j+1} - x)/h_{j+1}, & x_j \leqslant x \leqslant x_{j+1}, \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

for $j = 1, \ldots, N-1$,

$$\varphi_{jk}(x) = \begin{cases} h_j^{-1}\sqrt{2(2k-1)}\int_{x_{j-1}}^{x} P_{k-1}(y)\,\mathrm{d}y, & x_{j-1} \leqslant x \leqslant x_j, \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

for $j = 1, \ldots, N$ and $k = 2, \ldots, p$, and where V_{jk} are coefficients. The function $P_k(y)$ is the kth degree Legendre polynomial linearly scaled to the subinterval $[x_{j-1}, x_j]$. Functions (3.2) and (3.3) form a *hierarchical basis* of the subspace $S_0^{N,p}$, see Szabó, Babuška [14]. To express a function $V(\cdot, t) \in S_0^{N,p}$ for a fixed $t \in [0, T]$ in the basis (3.2) and (3.3), we put $V_{jk}(t) = V_{jk}$, i.e.,

$$V(x,t) = \sum_{j=1}^{N-1} V_{j1}(t)\varphi_{j1}(x) + \sum_{j=1}^{N}\sum_{k=2}^{p} V_{jk}(t)\varphi_{jk}(x).$$

We will also use the local inner product

$$(v, w)_j = \int_{x_{j-1}}^{x_j} v(x)w(x)\,\mathrm{d}x$$

and the corresponding local norm $\|v\|_{0,j}$.

Remark 3.1. Let assumption (3.1) holds. Then there exists a positive constant C independent of h such that

$$\|\nabla\theta(x)\|_0 \leqslant C\frac{1}{h}\|\theta(x)\|_0 \tag{3.4}$$

holds for all $\theta \in S_0^{N,p}$.

This is the so called *inverse inequality*, which can be found for example in Ciarlet [4], p. 142.

To start the error analysis, we introduce an elliptic projection of the solution u.

Definition 3.1. A function $u^h(x,t)$ is called the *elliptic projection* of the solution $u(x,t)$ of problem (2.8) and (2.9) if $u^h \in H^1([0,T], S_0^{N,p})$, if the identity

$$(a(u)\nabla u^h, \nabla V) = (a(u)\nabla u, \nabla V) \tag{3.5}$$

holds for almost every $t \in (0, T]$ and all functions $V \in S_0^{N,p}$, and if the identity

$$(a(u_0)\nabla u^h, \nabla V) = (a(u_0)\nabla u_0, \nabla V)$$

holds for $t = 0$ and all functions $V \in S_0^{N,p}$. We further denote by

$$\varrho(x,t) = u(x,t) - u^h(x,t)$$

the error of the elliptic projection.

The following lemma shows the standard important properties of the elliptic projection u^h and its error.

Lemma 3.1. *Let $t \in [0,T]$ be fixed. Let $u(\cdot,t) \in H^{p+1} \cap H_0^1$ and $u^h(\cdot,t) \in S_0^{N,p}$ be the elliptic projection. Then there exists a constant $C(u)$, which does not depend on t, such that*

$$\|\varrho\|_0 + h\|\nabla\varrho\|_0 \leqslant C(u)h^{p+1}, \tag{3.6}$$

$$\|\partial_t\varrho\|_0 \leqslant C(u)h^{p+1},$$

$$\|\nabla u^h\|_\infty \leqslant C(u), \tag{3.7}$$

where $\|\cdot\|_\infty$ is the L^∞-norm.

Proof. See Thomée [15], p. 211 and Moore [9]. □

We say that a function $\overline{U}(x,t)$ is the *semidiscrete approximate solution* of problem (2.8) and (2.9) if $\overline{U} \in H^1([0,T], S_0^{N,p})$, if the identity

$$(\partial_t\overline{U}, V) + (a(\overline{U})\nabla\overline{U}, \nabla V) + (f(\overline{U}), V) = 0 \tag{3.8}$$

holds for almost every $t \in (0,T]$ and all functions $V \in S_0^{N,p}$, and if the identity

$$(a(u_0)\nabla\overline{U}, \nabla V) = (a(u_0)\nabla u_0, \nabla V)$$

holds for $t = 0$ and all functions $V \in S_0^{N,p}$.

Definition 3.2. Denote by

$$\overline{e}(x,t) = u(x,t) - \overline{U}(x,t) \tag{3.9}$$

the error of the semidiscrete solution.

4. Semidiscrete error estimation

From formula (3.9) we have $u = \overline{U} + \overline{e}$. Putting this into (2.8) and (2.9), we obtain the equation which motivates the following definitions, see e.g. Segeth [13].

Let us introduce the space $\hat{S}_0^{N,p+1}$ of functions $\hat{V}(x)$ such that

$$\hat{V}(x) = \sum_{j=1}^{N} \hat{V}_j\varphi_{j,p+1}(x).$$

We are looking for the error estimates $\overline{E}$ in the space $\hat{S}_0^{N,p+1}$, i.e.,

$$\overline{E}(x,t) = \sum_{j=1}^{N} \overline{E}_j(t)\varphi_{j,p+1}(x).$$

In the next definition we introduce four natural error estimates of the semidiscrete solution $\overline{U}$, which is supposed to be known.

Definition 4.1. The *parabolic nonlinear error estimate* ($\overline{E}_{\mathrm{PN}}$) is defined by the equation

$$\begin{aligned}(\partial_t\overline{E},\hat{V})_j + (a(\overline{U}+\overline{E})\nabla\overline{E},\nabla\hat{V})_j = &-(f(\overline{U}+\overline{E}),\hat{V})_j - (\partial_t\overline{U},\hat{V})_j \\ &-(a(\overline{U}+\overline{E})\nabla\overline{U},\nabla\hat{V})_j.\end{aligned} \tag{4.1}$$

The *elliptic nonlinear error estimate* ($\overline{E}_{\mathrm{EN}}$) is defined by the equation

$$(a(\overline{U}+\overline{E})\nabla\overline{E},\nabla\hat{V})_j = -(f(\overline{U}+\overline{E}),\hat{V})_j - (\partial_t\overline{U},\hat{V})_j - (a(\overline{U}+\overline{E})\nabla\overline{U},\nabla\hat{V})_j. \tag{4.2}$$

The *parabolic linear* ($\overline{E}_{\mathrm{PL}}$) and *elliptic linear error estimate* ($\overline{E}_{\mathrm{EL}}$) are defined by the equations

$$(\partial_t\overline{E},\hat{V})_j + (a(\overline{U})\nabla\overline{E},\nabla\hat{V})_j = -(f(\overline{U}),\hat{V})_j - (\partial_t\overline{U},\hat{V})_j - (a(\overline{U})\nabla\overline{U},\nabla\hat{V})_j \tag{4.3}$$

and

$$(a(\overline{U})\nabla\overline{E},\nabla\hat{V})_j = -(f(\overline{U}),\hat{V})_j - (\partial_t\overline{U},\hat{V})_j - (a(\overline{U})\nabla\overline{U},\nabla\hat{V})_j, \tag{4.4}$$

respectively. All these four equations hold for almost every $t \in (0,T]$, $j = 1,\dots,N$, and all functions $\hat{V} \in \hat{S}_0^{N,p+1}$. The initial condition for $\overline{E}_{\mathrm{PN}}$ and $\overline{E}_{\mathrm{PL}}$ is given by

$$(a(u_0)\nabla\overline{E},\nabla\hat{V})_j = (a(u_0)\nabla(u_0-\overline{U}),\nabla\hat{V})_j,$$

$t = 0$, $j = 1,\dots,N$ and all $\hat{V} \in \hat{S}_0^{N,p+1}$.

We introduce so called *effectivity index* of the respective error estimator. Is is ratio of error estimator to the exact error, i.e.,

$$\Theta_{\mathrm{PN}} = \frac{\|\overline{E}_{\mathrm{PN}}\|_1}{\|e\|_1}, \quad \Theta_{\mathrm{EN}} = \frac{\|\overline{E}_{\mathrm{EN}}\|_1}{\|e\|_1}, \quad \Theta_{\mathrm{PL}} = \frac{\|\overline{E}_{\mathrm{PL}}\|_1}{\|e\|_1}, \quad \text{and} \quad \Theta_{\mathrm{EL}} = \frac{\|\overline{E}_{\mathrm{EL}}\|_1}{\|e\|_1}.$$

5. Fully discrete solution

The time discretization of the semidiscrete problem leads to a fully discrete scheme.

We can write the semidiscrete solution $\overline{U}$ as a linear combination of basis functions (3.2) and (3.3) with coefficients depending on the time variable t. We can take advantage of this fact and rewrite equation (3.8) as a system of *ordinary differential equations* for these unknown coefficients. To obtain a fully discrete solution we have to solve this system using some suitable numerical method.

We will investigate the singly implicit Runge-Kutta method (SIRK) which is described e.g. in Butcher [2] and Burrage [3].

Let us denote by τ the length of the time step of the equidistant partition of the time interval $[0,T]$ and by t_i the nodes of this partition. Confine our attention to one time step $t_i \leqslant t \leqslant t_i + \tau$. Let us assume that there exists a positive constant C such that $\tau = Ch$. The stage s of the singly implicit Runge-Kutta method is chosen to be $s = p+1$. We denote by $U_\ell(x) = U(x, t_i + c_\ell \tau)$ the approximation of the solution in the respective SIRK stage, where c_ℓ, $\ell = 1, 2, \ldots, p+1$, are given by the SIRK method. Note that $U_0(x) = U(x, t_i)$.

The solution $U_\ell \in S_0^{N,p}$ at every SIRK stage can be obtained by solving the Galerkin problem

$$(\overline{\partial}_{t,\ell} U_\ell, V) + (a(U_\ell)\nabla U_\ell, \nabla V) + (f(U_\ell), V) = 0 \tag{5.1}$$

for all $V \in S_0^{N,p}$, $\ell = 1, \ldots, p+1$. The symbol $\overline{\partial}_{t,\ell} U_\ell$ means

$$\overline{\partial}_{t,\ell} U_\ell = \frac{1}{\tau} \sum_{m=0}^{p+1} \overline{a}_{\ell m} U_m, \quad \ell = 1, 2, \ldots, p+1, \tag{5.2}$$

where $\overline{a}_{\ell m}$ are elements of the matrix $\mathcal{A}^{-1}$ and the matrix $\mathcal{A}$ defines the SIRK method, see Moore and Flaherty [11]. We take the initial condition

$$U(x, t_i) = u^h(x, t_i), \tag{5.3}$$

where u^h is the elliptic projection of the exact solution.

Relation (5.1) is a system of $p+1$ finite-dimensional nonlinear Galerkin problems, which can be equivalently formulated in the form:

$$\text{find } \mathbf{U} \in [S_0^{N,p}]^{p+1} \text{ such that } \mathbf{F}(\mathbf{U}) = 0, \tag{5.4}$$

where $\mathbf{F}\colon [S_0^{N,p}]^{p+1} \to [S_0^{N,p}]^{p+1}$ is defined by the Riesz theorem so that

$$(\overline{\partial}_{t,\ell} U_\ell, V) + (a(U_\ell)\nabla U_\ell, \nabla V) + (f(U_\ell), V) = (F_\ell(\mathbf{U}), V)$$

holds for all $V \in S_0^{N,p}$ and $\ell = 1, 2, \ldots, p+1$. Note that $[S_0^{N,p}]^{p+1}$ is a Hilbert space equipped with the inner product

$$(\mathbf{v}, \mathbf{w})_{[S_0^{N,p}]^{p+1}} = \sum_{\ell=1}^{p+1} (v_\ell, w_\ell).$$

One of the stability conditions for the SIRK method is that the matrix $\mathcal{A}$ is positive definite. Thus $\mathcal{A}^{-1}$ is also positive definite. This fact together with conditions (2.2) and (2.7) ensures

$$\sum_{\ell=1}^{p+1} (F_\ell(\mathbf{U}), U_\ell) \geqslant \varepsilon > 0 \quad \text{for all } \mathbf{U} \in [S_0^{N,p}]^{p+1},\ \|\mathbf{U}\| = R,$$

where $R > 0$ is sufficiently large. The well known corollary of Brouwer's fixed-point theorem gives us existence of a solution of problem (5.4) and equivalently of (5.1). This corollary of Brouwer's fixed-point theorem can be found, e.g., in Fučík, Kufner [5], Theorem 30.6, which in fact demands the assumption

$$\frac{R^2}{\varepsilon} (\mathbf{F}(\mathbf{U}), \mathbf{U})_{[S_0^{N,p}]^{p+1}} \geqslant (\mathbf{U}, \mathbf{U})_{[S_0^{N,p}]^{p+1}} \quad \text{for all } \mathbf{U} \in [S_0^{N,p}]^{p+1},\ \|\mathbf{U}\| = R.$$

Let us define the local error

$$e_\ell(x) = u_\ell(x) - U_\ell(x), \quad \ell = 1, 2, \ldots, p+1, \tag{5.5}$$

and the error of elliptic projection

$$\theta_\ell(x) = u_\ell^h(x) - U_\ell(x), \quad \ell = 1, 2, \ldots, p+1,$$

where $e_\ell(x)$ stands for $e(x, t_i + c_\ell \tau)$, $\theta_\ell(x) = \theta(x, t_i + c_\ell \tau)$. Note that the notation with index ℓ is used also for other quantities, e.g., $u_\ell^h(x) = u^h(x, t_i + c_\ell \tau)$, ϱ_ℓ, $\hat{\varrho}_\ell$, η_ℓ, etc.

The key role in our analysis is played by the transformation T introduced by Butcher [2], and its inverse T^{-1}:

$$T_{m\ell} = L_{\ell-1}(\xi_m), \quad T_{m\ell}^{-1} = \frac{\xi_\ell L_{m-1}(\xi_\ell)}{[(p+1)L_p(\xi_\ell)]^2}, \quad m, \ell = 1, 2, \ldots, p+1, \tag{5.6}$$

where L_m denotes the Laguerre polynomial of degree m and $\xi_1, \xi_2, \ldots, \xi_{p+1}$ are the distinct zeros of L_{p+1}. The transformed quantities are denoted by a tilde:

$$\tilde{\chi}_m = \sum_{\ell=1}^{p+1} T_{m\ell}^{-1} \chi_\ell, \quad \chi_m = \sum_{\ell=1}^{p+1} T_{m\ell} \tilde{\chi}_\ell, \quad m = 1, 2, \ldots, p+1. \tag{5.7}$$

The matrix $\mathcal{A}^{-1}$ can be turned to lower triangular using the transformation T.

This fact, which can be found in Butcher [2] or in Moore and Flaherty [11], can be expressed, e.g., in the component notation:

$$\tag{5.8} \begin{aligned}\sum_{k=1}^{p+1}\sum_{m=1}^{p+1}\overline{a}_{\ell m}T_{mk}\tilde{\chi}_k &= \frac{1}{\lambda}\sum_{k=1}^{p+1}\sum_{m=k}^{p+1}T_{\ell m}\tilde{\chi}_k\\ &= \frac{1}{\lambda}\sum_{m=1}^{p+1}\sum_{k=1}^{m}T_{\ell m}\tilde{\chi}_k, \quad \ell = 1,2,\dots,p+1,\end{aligned}$$

where λ is the positive parameter connected with the SIRK method.

The following lemma solves the problem with the nonlinearity.

Lemma 5.1. *Let $V_m \in S_0^{N,p+1}$, $m = 1,2,\dots,p+1$. Let a function a satisfy the Lipschitz condition* (2.3) *with a constant L, and let $U_\ell(x)$, the solution of* (5.1) *with* (5.3), *satisfy an analogue of the Lipschitz condition:*

$$\tag{5.9} |U_\ell(x) - U_m(x)| \leqslant C_L|t_0 + c_\ell\tau - t_0 - c_m\tau| = C_L|c_\ell - c_m|\tau$$

for all $\ell, m = 1,2,\dots,p+1$ and for almost every $x \in [c,d]$. Then

$$\sum_{\ell=1}^{p+1}\sum_{m=1}^{p+1}T_{r\ell}^{-1}T_{\ell m}(a(U_\ell)\nabla V_m, \nabla V_r) \geqslant \mu\|\nabla V_r\|_0^2 - \Delta T L C_L C_c\tau\sum_{\substack{m\neq r\\ m=1}}^{p+1}\|\nabla V_m\|_0\|\nabla V_r\|_0,$$

where $r = 1,2,\dots,p+1$ and the constants ΔT and C_c depend just on p.

Proof. Fix $r = 1,2,\dots,p+1$. Let us separate the double sum and adjust the resulting sums using $T_{r\ell}^{-1}T_{\ell r} \geqslant 0$, (2.2) and the triangular inequality:

$$\tag{5.10} \begin{aligned}&\sum_{\ell=1}^{p+1}\sum_{m=1}^{p+1}T_{r\ell}^{-1}T_{\ell m}(a(U_\ell)\nabla V_m, \nabla V_r)\\ &\quad = \sum_{\ell=1}^{p+1}T_{r\ell}^{-1}T_{\ell r}(a(U_\ell)\nabla V_r, \nabla V_r) + \sum_{\ell=1}^{p+1}\sum_{\substack{m\neq r\\ m=1}}^{p+1}T_{r\ell}^{-1}T_{\ell m}(a(U_\ell)\nabla V_m, \nabla V_r)\\ &\quad \geqslant \mu\|\nabla V_r\|_0^2 - \sum_{\substack{m\neq r\\ m=1}}^{p+1}\left(\Bigl|\sum_{\ell=1}^{p+1}T_{r\ell}^{-1}T_{\ell m}a(U_\ell)\Bigr|\,|\nabla V_m|, |\nabla V_r|\right).\end{aligned}$$

Now we have trouble only with the term $\Bigl|\sum_{\ell=1}^{p+1}T_{r\ell}^{-1}T_{\ell m}a(U_\ell)\Bigr|$. Let us introduce the notation

$$\beta_\ell = \beta_\ell^{m,r} = T_{r\ell}^{-1}T_{\ell m}.$$

We see that $\sum_{\ell=1}^{p+1}\beta_\ell = 0$ because $m \neq r$. If $\beta_\ell = 0$ held for $\ell = 1,2,\dots,p+1$, we would have no trouble. Unfortunately and naturally, the definition of T and T^{-1}, see (5.6), implies $\beta_\ell \neq 0$ for $\ell = 1, 2, \ldots, p+1$. Thus, we can define nonempty sets of indices

$$\mathcal{L}^+ = \{\ell\colon\ \beta_\ell > 0\} \quad \text{and} \quad \mathcal{L}^- = \{\ell\colon\ \beta_\ell < 0\}.$$

Let us define positive constant ΔT^{mr} by

$$\Delta T^{mr} = \sum_{\ell \in \mathcal{L}^+} \beta_\ell = \sum_{\ell \in \mathcal{L}^-} -\beta_\ell.$$

Let us find indices $\ell_{\text{Max}}, \ell_{\text{Min}} \in \{1, 2, \ldots, p+1\}$ such that

$$a(U_{\ell_{\text{Min}}}) \leqslant a(U_\ell) \leqslant a(U_{\ell_{\text{Max}}}) \quad \text{for all } \ell = 1, 2, \ldots, p+1.$$

The positivity of β_ℓ for $\ell \in \mathcal{L}^+$ and $-\beta_\ell$ for $\ell \in \mathcal{L}^-$ implies

$$\sum_{\ell \in \mathcal{L}^+} \beta_\ell a(U_\ell) - \sum_{\ell \in \mathcal{L}^-} (-\beta_\ell) a(U_\ell) \leqslant \Delta T^{mr}(a(U_{\ell_{\text{Max}}}) - a(U_{\ell_{\text{Min}}})), \tag{5.11}$$

$$-\sum_{\ell \in \mathcal{L}^+} \beta_\ell a(U_\ell) + \sum_{\ell \in \mathcal{L}^-} (-\beta_\ell) a(U_\ell) \leqslant \Delta T^{mr}(-a(U_{\ell_{\text{Min}}}) + a(U_{\ell_{\text{Max}}})). \tag{5.12}$$

Finally, using (5.11), (5.12), (2.3) and (5.9), we obtain

$$\begin{aligned} \left|\sum_{\ell=1}^{p+1} \beta_\ell a(U_\ell)\right| &\leqslant \Delta T^{mr} |a(U_{\ell_{\text{Max}}}) - a(U_{\ell_{\text{Min}}})| \leqslant \Delta T^{mr} L |U_{\ell_{\text{Max}}} - U_{\ell_{\text{Min}}}| \\ &\leqslant \Delta T^{mr} L C_L |t_0 + c_{\ell_{\text{Max}}}\tau - t_0 - c_{\ell_{\text{Min}}}\tau| \leqslant \Delta T L C_L C_c \tau, \end{aligned} \tag{5.13}$$

where

$$\Delta T = \max_{r \neq m} \Delta T^{mr} \quad \text{and} \quad C_c = \max_{\ell, m = 1, \ldots, p+1} |c_\ell - c_m|.$$

We complete the proof applying (5.13) and the Schwarz inequality in (5.10). □

To simplify the notation we denote the time derivative $\partial_t u$ by u_t.

Lemma 5.2. *Let $u_\ell(x) \in H^{p+1} \cap H_0^1$ be the solution of* (2.8) *with* (2.9) *at time instant $t_i + c_\ell\tau$ and let $U_\ell(x) \in S_0^{N,p}$ be the solution of* (5.1) *with* (5.3) *for $\ell = 1, 2, \ldots, p+1$. Let U_ℓ satisfy* (5.9).

Then there exist constants C and C_T independent of h such that

$$\|e_\ell\|_1 \leqslant Ch^p, \qquad \ell = 1, 2, \ldots, p+1, \tag{5.14}$$

$$\|\theta_\ell\|_1 \leqslant Ch^{p+1}, \quad \ell = 1, 2, \ldots, p+1, \tag{5.15}$$

$$\|\theta_\ell\|_0 \leqslant C_T \sum_{m=1}^{p+1} \|\tilde{\theta}_m\|_0, \quad \|\tilde{\theta}_\ell\|_0 \leqslant C_T \sum_{m=1}^{p+1} \|\theta_m\|_0, \quad \ell = 1, \ldots, p+1. \tag{5.16}$$

Proof. Inequalities (5.16) are evident from the definition of transformed quantities (5.7).

To prove inequalities (5.14) and (5.15) we employ equality (5.1), which can be adjusted with help of (2.8) and (3.5), to obtain

$$
\begin{aligned}
&(\overline{\partial}_{t,\ell}\theta_\ell, V) + (a(U_\ell)\nabla\theta_\ell, \nabla V)\\
&\qquad = -(\overline{\partial}_{t,\ell}\varrho_\ell, V) + (\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell, V) + (f(U_\ell) - f(u_\ell), V)\\
&\qquad\quad + ([a(U_\ell) - a(u_\ell)]\nabla u_\ell^h, \nabla V), \quad \ell = 1, 2, \ldots, p+1.
\end{aligned}
$$

We arrange the first two terms now. We use definition (5.2) of $\overline{\partial}_{t,\ell}$, the fact that $\theta_0(x) = \theta(x, t_i) = 0$ due to (5.3), then we replace θ_m by the transformed quantity $\tilde{\theta}_m$, see (5.7), and finally we apply formula (5.8) for the first term to obtain

$$
\frac{1}{\lambda\tau}\sum_{m=1}^{p+1} T_{\ell m}\sum_{k=1}^{m}(\tilde{\theta}_k, V) + \sum_{m=1}^{p+1} T_{\ell m}(a(U_\ell)\nabla\tilde{\theta}_m, \nabla V) = \ldots, \qquad \ell = 1, 2, \ldots, p+1,
$$

where the terms on the right-hand side remain unchanged. We can multiply this system of equations by the matrix T^{-1} from the left. This step can be exactly done taking $r = 1, 2, \ldots, p+1$, multiplying each equation by $T^{-1}_{r\ell}$ and summing over ℓ. Taking $V = \tilde{\theta}_r$ and rearranging slightly all terms, we obtain

$$
\begin{aligned}
(5.17)\qquad &\|\tilde{\theta}_r\|_0^2 + \lambda\tau\sum_{\ell=1}^{p+1}\sum_{m=1}^{p+1} T^{-1}_{r\ell}T_{\ell m}(a(U_\ell)\nabla\tilde{\theta}_m, \nabla\tilde{\theta}_r)\\
&\quad = -\sum_{k=1}^{r-1}(\tilde{\theta}_k, \tilde{\theta}_r) + \lambda\tau\sum_{\ell=1}^{p+1} T^{-1}_{r\ell}[-(\overline{\partial}_{t,\ell}\varrho_\ell, \tilde{\theta}_r) + (\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell, \tilde{\theta}_r)\\
&\qquad + (f(U_\ell) - f(u_\ell), \tilde{\theta}_r) + ([a(U_\ell) - a(u_\ell)]\nabla u_\ell^h, \nabla\tilde{\theta}_r)]
\end{aligned}
$$

for $r = 1, 2, \ldots, p+1$.

Note that (2.4), the Schwarz and triangle inequalities, (3.7), (2.3) and (5.16) imply

$$
|(f(U_\ell) - f(u_\ell), \tilde{\theta}_r)| \leqslant L(\|\theta_\ell\|_0 + \|\varrho_\ell\|_0)\|\tilde{\theta}_r\|_0
$$

and

$$
|([a(U_\ell) - a(u_\ell)]\nabla u_\ell^h, \nabla\tilde{\theta}_r)| \leqslant C(u)L\Big(C_T\sum_{m=1}^{p+1}\|\tilde{\theta}_m\|_0 + \|\varrho_\ell\|_0\Big)\|\nabla\tilde{\theta}_r\|_0.
$$

Applying these inequalities, Lemma 5.1 and the Young inequality to (5.17), we obtain

$$
\begin{aligned}
(5.18)\quad \|\tilde{\theta}_r\|_0^2 + \lambda\tau\mu\|\nabla\tilde{\theta}_r\|_0^2 \leqslant C_1 \sum_{k=1}^{r-1} \|\tilde{\theta}_k\|_0^2 + \tau^2 C_2 \sum_{\ell=1}^{p+1} \big[\|\overline{\partial}_{t,\ell}\varrho_\ell\|_0^2 + \|\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell\|_0^2 \\
+ \|\tilde{\theta}_\ell\|_0^2 + \|\varrho_\ell\|_0^2\big] + \tau C_3 \sum_{\ell=1}^{p+1}(\|\tilde{\theta}_\ell\|_0^2 + \|\varrho_\ell\|_0^2) \\
+ \tau^3 C_4 \sum_{\substack{m\neq r\\ m=1}}^{p+1} \|\nabla\tilde{\theta}_m\|_0^2
\end{aligned}
$$

for $r = 1, 2, \ldots, p+1$.

We replace the last term using (3.4). Now we use the nonnegativity of the term $\lambda\tau\mu\|\nabla\tilde{\theta}_r\|_0^2$ and move the terms with $\tilde{\theta}$ to the left-hand side to obtain

$$
\begin{aligned}
(5.19)\quad -(C_1 + \tau^2 C_2 + \tau C_{34}) \sum_{k=1}^{r-1} \|\tilde{\theta}_k\|_0^2 + (1 - \tau^2 C_2 - \tau C_3)\|\tilde{\theta}_r\|_0^2 \\
- (\tau^2 C_2 + \tau C_{34}) \sum_{k=r+1}^{p+1} \|\tilde{\theta}_k\|_0^2 \\
\leqslant \sum_{\ell=1}^{p+1} (\tau^2 C_2 \big[\|\overline{\partial}_{t,\ell}\varrho_\ell\|_0^2 + \|\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell\|_0^2 + \|\varrho_\ell\|_0^2\big] + \tau C_3 \|\varrho_\ell\|_0^2) \\
\leqslant C\tau^{2p+3}, \quad r = 1, 2, \ldots, p+1.
\end{aligned}
$$

In order to bound the right-hand part terms, we have used the results of Moore and Flaherty [10]:

$$(5.20)\qquad \|\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell\|_0 \leqslant C(u)h^{p+1},$$

$$(5.21)\qquad \|\overline{\partial}_{t,\ell}\varrho_\ell\|_0 \leqslant C(u)h^{p+1},$$

$$(5.22)\qquad \|\overline{\partial}_{t,\ell}\hat{\varrho}_\ell\|_0 \leqslant C(u)h^{p+1},$$

the result (3.6) and the assumption $\tau = Ch$. Inequality (5.22) will be used later.

Relation (5.19) is a system of $p+1$ inequalities. Our aim is to obtain a bound for $\|\tilde{\theta}_\ell\|_0$ for all $\ell = 1, 2, \ldots, p+1$. The matrix of this system is not lower triangular but its elements in the upper triangle are small, i.e. of order $O(\tau)$. Thus, we can use the Gaussian elimination to obtain a lower triangular matrix. The important fact is that the diagonal elements are positive for a sufficiently small τ and the off-diagonal elements are negative or zero. Thus, every step in the Gaussian elimination is correct and all inequalities are preserved. The resulting lower triangular system is

$$(1-\tau^2C_5-\tau C_6)\|\tilde{\theta}_r\|_0^2 \leqslant (C_7+\tau^2C_8+\tau C_9)\sum_{k=1}^{r-1}\|\tilde{\theta}_k\|_0^2 + C\tau^{2p+3} \tag{5.23}$$

for $r=1,2,\dots,p+1$. Using $\tau = Ch$ and solving (5.23) by forward substitution yields

$$\|\tilde{\theta}_r\|_0 \leqslant Ch^{p+3/2}, \quad r=1,2,\dots,p+1. \tag{5.24}$$

Returning with (5.24) to (5.18), we obtain

$$\|\nabla\tilde{\theta}_r\|_0 \leqslant Ch^{p+1}, \quad r=1,2,\dots,p+1.$$

Using (5.16), we have

$$\|\theta_r\|_0 \leqslant Ch^{p+3/2}, \quad r=1,2,\dots,p+1, \tag{5.25}$$

$$\|\nabla\theta_r\|_0 \leqslant Ch^{p+1}, \qquad r=1,2,\dots,p+1. \tag{5.26}$$

The inequality (5.14) is now easy:

$$\|e_\ell\|_1 \leqslant \|\theta_\ell\|_1 + \|\varrho_\ell\|_1 \leqslant Ch^p, \quad \ell=1,2,\dots,p+1.$$

□

6. Fully discrete error estimation

The equations which define the semidiscrete error estimates (see Definition 4.1) can be equivalently written as some systems of ordinary differential equations. Solving this systems by the SIRK method we obtain the *fully discrete* estimates. The SIRK method can be implemented using the discrete derivative $\overline{\partial}_{t,\ell}$.

Definition 6.1. The respective error estimates at $t_i + c_\ell\tau$, $\ell = 1,2,\dots,p+1$, are obtained by solving the following systems of uncoupled problems. The system for the *parabolic nonlinear error estimate* $E_{\mathrm{PN},\ell} = E_\ell \in \hat{S}_0^{N,p+1}$, cf. (4.1), is

$$\begin{aligned}(\overline{\partial}_{t,\ell}E_\ell,\hat{V})_j + (a(U_\ell+E_\ell)\nabla E_\ell,\nabla\hat{V})_j = &-(f(U_\ell+E_\ell),\hat{V})_j - (\overline{\partial}_{t,\ell}U_\ell,\hat{V})_j\\ &-(a(U_\ell+E_\ell)\nabla U_\ell,\nabla\hat{V})_j.\end{aligned} \tag{6.1}$$

The *parabolic linear error estimate* $E_{\mathrm{PL},\ell} = E_\ell \in \hat{S}_0^{N,p+1}$ is obtained by solving the system, cf. (4.3):

$$(6.2)\qquad \begin{aligned}(\overline{\partial}_{t,\ell}E_\ell, \hat{V})_j + (a(U_\ell)\nabla E_\ell, \nabla\hat{V})_j = &-(f(U_\ell), \hat{V})_j - (\overline{\partial}_{t,\ell}U_\ell, \hat{V})_j\\ &-(a(U_\ell)\nabla U_\ell, \nabla\hat{V})_j.\end{aligned}$$

And finally, the *elliptic linear error estimate* $E_{\mathrm{EL},\ell} = E_\ell \in \hat{S}_0^{N,p+1}$, cf. (4.4), is obtained by solving

$$(6.3)\qquad (a(U_\ell)\nabla E_\ell, \nabla\hat{V})_j = -(f(U_\ell), \hat{V})_j - (\overline{\partial}_{t,\ell}U_\ell, \hat{V})_j - (a(U_\ell)\nabla U_\ell, \nabla\hat{V})_j.$$

These equalities hold for all $\hat{V} \in \hat{S}_0^{N,p+1}$, $j = 1, 2, \ldots, N$ and $\ell = 1, 2, \ldots, p+1$, with

$$(6.4)\qquad E_0(x) = e_0^h(x),$$

where $e_0^h(x)$ stands for $e^h(x, t_i)$ and the definition of $e^h(x,t)$ follows.

Note that the existence and uniqueness of solutions of problems (6.1)–(6.3) is ensured again by Brouwer's fixed-point theorem.

Definition 6.2. The function $e^h(x,t) \in H^1([0,T], \hat{S}_0^{N,p+1})$ is a projection of the error $e(x,t)$ if the equality

$$(6.5)\qquad (a(u)\nabla(u^h + e^h), \nabla\hat{V}) = (a(u)\nabla u, \nabla\hat{V})$$

holds for all $\hat{V} \in \hat{S}_0^{N,p+1}$. We set

$$(6.6)\qquad \hat{\varrho}(x,t) = u(x,t) - u^h(x,t) - e^h(x,t).$$

Note that $e^h(x,t) = \sum\limits_{j=1}^{N} C_j(t)\varphi_{j,p+1}(x)$.

Lemma 6.1. *Let $u \in H^1([0,T], H^{p+2} \cap H_0^1)$ be a solution of* (2.8) *with* (2.9). *Let u^h, e^h and $\hat{\varrho}$ be defined by* (3.5), (6.5) *and* (6.6), *respectively. Then*

$$(6.7)\qquad \|\hat{\varrho}\|_0 + h\|\nabla\hat{\varrho}\|_0 \leqslant C(u)h^{p+2}$$

and

$$(6.8)\qquad \|\nabla(u^h + e^h)\|_\infty \leqslant C(u),$$

independently of t and h.

Proof. The proof of (6.7) follows from the work of Adjerid, Flaherty, Wang [1]. Let $\hat{V}$ be an arbitrary function from $\hat{S}_0^{N,p+1}$. Using (2.2), (6.5) and the Schwarz inequality we obtain

$$\begin{aligned}\mu\|\nabla(u^h+e^h-u)\|_0^2 &\leqslant (a(u)\nabla(u^h+e^h-u),\nabla(u^h+e^h-u))\\ &= (a(u)\nabla(u^h+e^h-u),\nabla(u^h+\hat{V}-u))\\ &\leqslant M\|\nabla(u^h+e^h-u)\|_0\,\|\nabla(u^h+\hat{V}-u)\|_0,\end{aligned}$$

i.e.,

$$\|\nabla(u^h+e^h-u)\|_0 \leqslant \frac{M}{\mu}\|\nabla(u^h+\hat{V}-u)\|_0 \tag{6.9}$$

holds. According to Adjerid, Flaherty, Wang [1], Lemma 3.3 and Lemma 3.5, there exists a suitable function $\varphi \in \hat{S}_0^{N,p+1}$, which can be substituted for $\hat{V}$ in (6.9), such that

$$\|\nabla(u^h+e^h-u)\|_0 \leqslant C(u)h^{p+1}.$$

The estimate of (u^h+e^h-u) in the L^2 norm can be obtained by the duality argument.

The proof of (6.8) for $p=1$ can be found in Thomée [15], p. 212. We rewrite this proof for $p=1,2,\ldots$. Let

$$\Pi\colon\ H_0^1 \mapsto S_0^{N,p+1}$$

denote the standard interpolation operator. The inverse inequality, see for example Ciarlet [4], p. 142, converts the L^∞ norm into the L^2 norm:

$$\begin{aligned}\|\nabla(u^h+e^h-\Pi u)\|_\infty &\leqslant \frac{C}{h^{1/2}}\|\nabla(u^h+e^h-\Pi u)\|_0\\ &\leqslant \frac{C}{h^{1/2}}(\|\nabla(u^h+e^h-u)\|_0+\|\nabla(u-\Pi u)\|_0).\end{aligned}$$

Inequality (6.7) and the well known estimate $\|\nabla(u-\Pi u)\|_0 \leqslant C(u)h^{p+1}$ give us

$$\|\nabla(u^h+e^h-\Pi u)\|_\infty \leqslant C(u)h^{p+1/2}. \tag{6.10}$$

Using the estimate of the interpolation error, see Ciarlet [4], p. 122, we obtain

$$\|\nabla\Pi u\|_\infty \leqslant C\|\nabla u\|_\infty. \tag{6.11}$$

The combination of results (6.10) and (6.11) gives us estimate (6.8). □

Definition 6.3. We put

$$\eta_{\mathrm{PN},\ell}(x) = e_\ell^h(x) - E_{\mathrm{PN},\ell}(x), \tag{6.12}$$

$$\eta_{\mathrm{PL},\ell}(x) = e_\ell^h(x) - E_{\mathrm{PL},\ell}(x), \tag{6.13}$$

$$\eta_{\mathrm{EL},\ell}(x) = e_\ell^h(x) - E_{\mathrm{EL},\ell}(x), \quad \ell = 1, 2, \ldots, p+1, \tag{6.14}$$

where, apparently, $\eta \in \hat{S}_0^{N,p+1}$. We omit the indices of η if no ambiguity can occur. Error estimates $E_{\mathrm{PN},\ell}$, $E_{\mathrm{PL},\ell}$ and $E_{\mathrm{EL},\ell}$ are defined using (6.1), (6.2) and (6.3), respectively, and $e_\ell^h \in \hat{S}_0^{N,p+1}$ is given by (6.5).

Lemma 6.2. *Let $u_\ell(x) \in H^{p+2} \cap H_0^1$ be the solution of* (2.8) *with* (2.9) *at time instants $t_i + c_\ell\tau$, $\ell = 1, 2, \ldots, p+1$. Let $E_{\mathrm{PN},\ell} \in \hat{S}_0^{N,p+1}$ be the error estimate given by* (6.1) *with* (6.4)*, and let $e_\ell^h \in \hat{S}_0^{N,p+1}$ and $\eta_{\mathrm{PN},\ell} \in \hat{S}_0^{N,p+1}$, $\ell = 1, 2, \ldots, p+1$, be given by* (6.5) *and* (6.12)*, respectively. Let U_ℓ, given by* (5.1) *with* (5.3)*, satisfy* (5.9) *and let $E_{\mathrm{PN},\ell}$ satisfy*

$$|E_{\mathrm{PN},\ell}(x) - E_{\mathrm{PN,m}}(x)| \leqslant C_L|t_0 + c_\ell\tau - t_0 - c_m\tau| = C_L|c_\ell - c_m|\tau \tag{6.15}$$

for all $\ell, m = 1, 2, \ldots, p+1$ and for almost every $x \in [c, d]$.

Then

$$\|\eta_{\mathrm{PN},\ell}\|_1 \leqslant Ch^{p+1}, \quad \ell = 1, 2, \ldots, p+1.$$

Proof. We start with the equation (6.1). We subtract the terms $(\overline{\partial}_{t,\ell}e_\ell^h, \hat{V})_j$ and $(a(U_\ell+E_\ell)\nabla e_\ell^h, \nabla\hat{V})_j$ on both sides of (6.1), add and subtract the terms $(\overline{\partial}_{t,\ell}u_\ell^h, \hat{V})_j$, $(\overline{\partial}_{t,\ell}u_\ell, \hat{V})_j$ and $(a(U_\ell + E_\ell)\nabla u_\ell^h, \nabla\hat{V})_j$ on the right-hand side of (6.1), adding the weak formulation (2.8) tested by $\hat{V}$, and using (6.5) we obtain

$$\begin{aligned}
(\overline{\partial}_{t,\ell}\eta_\ell, \hat{V})_j &+ (a(U_\ell + E_\ell)\nabla\eta_\ell, \nabla\hat{V})_j \\
= &- (\overline{\partial}_{t,\ell}\theta_\ell, \hat{V})_j - (\overline{\partial}_{t,\ell}\hat{\varrho}_\ell, \hat{V})_j + (\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell, \hat{V})_j \\
&+ (f(U_\ell + E_\ell) - f(u_\ell), \hat{V})_j + (a(U_\ell + E_\ell)[\nabla U_\ell - \nabla u_\ell^h], \nabla\hat{V})_j \\
&+ ([a(U_\ell + E_\ell) - a(u_\ell)][\nabla u_\ell^h + \nabla e_\ell^h], \nabla\hat{V})_j
\end{aligned}$$

for all $\hat{V} \in \hat{S}_0^{N,p+1}$, $j = 1, 2, \ldots, N$ and $\ell = 1, 2, \ldots, p+1$.

Using the definition of $\overline{\partial}_{t,\ell}$, (5.2), the fact that $\eta_0 = 0$ and $\theta_0 = 0$ due to (6.4) and (5.3), substituting η_ℓ and θ_ℓ by the transformed quantities according to (5.7), we arrive at

$$\begin{aligned}
\frac{1}{\tau}\sum_{m=1}^{p+1}\sum_{k=1}^{p+1}(\overline{a}_{\ell m}T_{mk}\tilde{\eta}_k, \hat{V})_j &+ \sum_{k=1}^{p+1}(a(U_\ell + E_\ell)T_{\ell k}\nabla\tilde{\eta}_k, \nabla\hat{V})_j \\
&= -\sum_{m=1}^{p+1}\sum_{k=1}^{p+1}(\overline{a}_{\ell m}T_{mk}\tilde{\theta}_k, \hat{V})_j + \ldots
\end{aligned}$$

for all $\hat{V} \in \hat{S}_0^{N,p+1}$, $j = 1, 2, \ldots, N$ and $\ell = 1, 2, \ldots, p+1$. The symbol $\ldots$ at the end of the above formula means that the other terms remain unchanged.

Employing formula (5.8), multiplying this system of equalities by the matrix T^{-1} from the left, i.e. multiplying each equality by $T_{r\ell}^{-1}$ and summing over ℓ, putting $\hat{V} = \tilde{\eta}_r$ and rearranging, we have

$$\begin{aligned}(6.16)\qquad &\|\tilde{\eta}_r\|_{0,j} + \lambda\tau \sum_{\ell=1}^{p+1} \sum_{m=1}^{p+1} T_{r\ell}^{-1} T_{\ell m} (a(U_\ell + E_\ell)\nabla\tilde{\eta}_m, \nabla\tilde{\eta}_r)_j \\ &= -\sum_{k=1}^{r-1} (\tilde{\eta}_k, \tilde{\eta}_r)_j - \sum_{k=1}^{r} (\tilde{\theta}_k, \tilde{\eta}_r)_j \\ &+ \lambda\tau \sum_{\ell=1}^{p+1} T_{r\ell}^{-1} \bigl[-(\overline{\partial}_{t,\ell}\hat{\varrho}_\ell, \tilde{\eta}_r)_j + (\overline{\partial}_{t,\ell} u_\ell - (u_t)_\ell, \tilde{\eta}_r)_j \\ &+ (f(U_\ell + E_\ell) - f(u_\ell), \tilde{\eta}_r)_j + (a(U_\ell + E_\ell)[\nabla U_\ell - \nabla u_\ell^h], \nabla\tilde{\eta}_r)_j \\ &+ ([a(U_\ell + E_\ell) - a(u_\ell)][\nabla u_\ell^h + \nabla e_\ell^h], \nabla\tilde{\eta}_r)_j\bigr]\end{aligned}$$

for $j = 1, 2, \ldots, N$ and $r = 1, 2, \ldots, p+1$.

We bound the second term employing an analogue of Lemma 5.1 together with assumption (6.15). This step is correct because

$$|U_\ell + E_\ell - U_m - E_m| \leqslant |U_\ell - U_m| + |E_\ell - E_m| \leqslant 2C_L|c_\ell - c_m|\tau.$$

Let us bound some terms in (6.16). Employing (2.2), (2.3), (2.4), (6.8), the triangle and the Schwarz inequalities and the definitions of θ, η and $\hat{\varrho}$, we obtain

$$(6.17)\qquad |(f(U_\ell + E_\ell) - f(u_\ell), \tilde{\eta}_r)_j| \leqslant L(\|\theta_\ell\|_{0,j} + \|\eta_\ell\|_{0,j} + \|\hat{\varrho}_\ell\|_{0,j})\|\tilde{\eta}_r\|_{0,j},$$

$$(6.18)\qquad |(a(U_\ell + E_\ell)[\nabla U_\ell - \nabla u_\ell^h], \nabla\tilde{\eta}_r)_j| \leqslant M\|\nabla\theta_\ell\|_{0,j}\|\nabla\tilde{\eta}_r\|_{0,j}$$

and finally

$$\begin{aligned}(6.19)\qquad &|([a(U_\ell + E_\ell) - a(u_\ell)][\nabla u_\ell^h + \nabla e_\ell^h], \nabla\tilde{\eta}_r)_j| \\ &\leqslant C(u, L)(\|\theta_\ell\|_{0,j} + \|\eta_\ell\|_{0,j} + \|\hat{\varrho}_\ell\|_{0,j})\,\|\nabla\tilde{\eta}_r\|_{0,j}.\end{aligned}$$

Employing inequalities (6.17), (6.18), (6.19) and the Young inequality in (6.16), we arrive at

(6.20)

$$
\begin{aligned}
&\|\tilde{\eta}_r\|_{0,j}^2 + \lambda\tau\mu\|\nabla\tilde{\eta}_r\|_{0,j}^2 \\
&\quad\leqslant C_1\Big(\sum_{k=1}^{r-1}\|\tilde{\eta}_k\|_{0,j}^2 + \sum_{k=1}^{r}\|\tilde{\theta}_k\|_{0,j}^2\Big) \\
&\qquad + \tau^2 C_2\sum_{\ell=1}^{p+1}\big[\|\overline{\partial}_{t,\ell}\hat{\varrho}_\ell\|_{0,j}^2 + \|\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell\|_{0,j}^2 + \|\theta_\ell\|_{0,j}^2 + \|\tilde{\eta}_\ell\|_{0,j}^2 + \|\hat{\varrho}_\ell\|_{0,j}^2\big] \\
&\qquad + \tau C_3\sum_{\ell=1}^{p+1}(\|\nabla\theta_\ell\|_{0,j}^2 + \|\theta_\ell\|_{0,j}^2 + \|\tilde{\eta}_\ell\|_{0,j}^2 + \|\hat{\varrho}_\ell\|_{0,j}^2) \\
&\qquad + \tau^3 C_4\sum_{\substack{m\neq r\\ m=1}}^{p+1}\|\nabla\tilde{\eta}_m\|_{0,j}^2
\end{aligned}
$$

for $j = 1, 2, \ldots, N$ and $r = 1, 2, \ldots, p+1$.

Summing over j, applying (3.4) to the last term, using (5.24), (5.22), (5.20), (5.25), (6.7), (5.26) and $\tau = Ch$, we have

$$
\begin{aligned}
(1 - \tau^2 C_2 - \tau C_3)\|\tilde{\eta}_r\|_0^2 - (C_1 + \tau^2 C_2 + \tau C_{34})\sum_{k=1}^{r-1}\|\tilde{\eta}_k\|_0^2 \\
- (\tau^2 C_2 + \tau C_{34})\sum_{k=r+1}^{p+1}\|\tilde{\eta}_k\|_0^2 \leqslant C\tau^{2p+3}
\end{aligned}
$$

for $r = 1, 2, \ldots, p+1$.

This is a system of inequalities in the same form as (5.19) and we can employ the Gaussian elimination to obtain the lower triangular system

$$
(1 - \tau^2 C_5 - \tau C_6)\|\tilde{\eta}_r\|_0^2 \leqslant (C_7 + \tau^2 C_8 + \tau C_9)\sum_{k=1}^{r-1}\|\tilde{\eta}_k\|_0^2 + C\tau^{2p+3}
$$

for $r = 1, 2, \ldots, p+1$.

Solving this system by forward substitution, we finally find that

$$
\|\tilde{\eta}_r\|_0 \leqslant Ch^{p+3/2}, \quad r = 1, 2, \ldots, s. \tag{6.21}
$$

Using (5.16), we have

$$
\|\eta_r\|_0 \leqslant Ch^{p+3/2}, \quad r = 1, 2, \ldots, s. \tag{6.22}
$$

Returning with (6.21) and (6.22) to (6.20) summed over j, we obtain

$$\|\nabla\tilde{\eta}_r\|_0 \leqslant Ch^{p+1}, \quad r = 1, 2, \ldots, s.$$

Finally, relation (5.16) gives us

$$\|\nabla\eta_r\|_0 \leqslant Ch^{p+1}, \quad r = 1, 2, \ldots, s.$$

□

Lemma 6.3. *Let* $u_\ell(x) \in H^{p+2} \cap H_0^1$ *be the solution of* (2.8) *with* (2.9) *at time instants* $t_i + c_\ell\tau$, $\ell = 1, 2, \ldots, p+1$. *Let* $E_{\mathrm{PL},\ell} \in \hat{S}_0^{N,p+1}$ *be the error estimate given by* (6.2) *with* (6.4), *and let* $e_\ell^h \in \hat{S}_0^{N,p+1}$ *and* $\eta_{\mathrm{PL},\ell} \in \hat{S}_0^{N,p+1}$, $\ell = 1, 2, \ldots, p+1$, *be given by* (6.5) *and* (6.13), *respectively. Let* U_ℓ, *given by* (5.1) *with* (5.3), *satisfy* (5.9).

Then

$$\|\eta_{\mathrm{PL},\ell}\|_1 \leqslant Ch^{p+1}, \quad \ell = 1, 2, \ldots, p+1.$$

Proof. The proof of this lemma is very similar to that of the previous lemma. The starting equality which follows from (6.2), (2.8) and (6.5) is

$$\begin{aligned}(\overline{\partial}_{t,\ell}\eta_\ell, \hat{V})_j + (a(U_\ell)\nabla\eta_\ell, \nabla\hat{V})_j = {} & -(\overline{\partial}_{t,\ell}\theta_\ell, \hat{V})_j - (\overline{\partial}_{t,\ell}\hat{\varrho}_\ell, \hat{V})_j \\ & + (\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell, \hat{V})_j + (f(U_\ell) - f(u_\ell), \hat{V})_j \\ & + (a(U_\ell)[\nabla U_\ell - \nabla u_\ell^h], \nabla\hat{V})_j \\ & + ([a(U_\ell) - a(u_\ell)][\nabla u_\ell^h + \nabla e_\ell^h], \nabla\hat{V})_j\end{aligned}$$

for all $\hat{V} \in \hat{S}_0^{N,p+1}$, $j = 1, 2, \ldots, N$ and $\ell = 1, 2, \ldots, p+1$.

Proceeding in the same way as in the previous proof, i.e., applying the following steps:

- definition of $\overline{\partial}_{t,\ell}$,
- transformation (5.7) and formula (5.8),
- multiplying the resulting system by the matrix T^{-1},
- putting $\hat{V} = \tilde{\eta}_r$,
- bounding the terms with a and f using (2.2), (2.3), (2.4), etc.,
- applying Lemma 5.1,
- summing over j, using (3.4), (5.24), (5.22), (5.20), (5.25), (3.6), (5.26),
- performing Gaussian elimination,

we obtain the result. □

Lemma 6.4. *Let $u_\ell(x) \in H^{p+2} \cap H_0^1$ be the solution of* (2.8) *with* (2.9) *at time instants $t_i + c_\ell\tau$, $\ell = 1, 2, \ldots, p+1$. Let $E_{\mathrm{EL},\ell} \in \hat{S}_0^{N,p+1}$ be the error estimate given by* (6.3), *and let $e_\ell^h \in \hat{S}_0^{N,p+1}$ and $\eta_{\mathrm{EL},\ell} \in \hat{S}_0^{N,p+1}$, $\ell = 1, 2, \ldots, p+1$, be given by* (6.5) *and* (6.14), *respectively. Then*

$$\|\eta_{\mathrm{EL},\ell}\|_1 \leqslant Ch^{p+1/2}, \quad \ell = 1, 2, \ldots, p+1.$$

Proof. The proof of this lemma is very similar to those of the previous two lemmas but we need not use the Gaussian elimination nor the transformation T.

The starting equality which follows from (6.3), (2.8) and (6.5) is

$$\begin{aligned}(a(U_\ell)\nabla\eta_\ell, \nabla\hat{V})_j = &-(\overline{\partial}_{t,\ell}\theta_\ell, \hat{V})_j - (\overline{\partial}_{t,\ell}\varrho_\ell, \hat{V})_j + (\overline{\partial}_{t,\ell}u_\ell - (u_t)_\ell, \hat{V})_j \\ &+ (f(U_\ell) - f(u_\ell), \hat{V})_j + (a(U_\ell)[\nabla U_\ell - \nabla u_\ell^h], \nabla\hat{V})_j \\ &+ ([a(U_\ell) - a(u_\ell)][\nabla u_\ell^h + \nabla e_\ell^h], \nabla\hat{V})_j\end{aligned}$$

for all $\hat{V} \in \hat{S}_0^{N,p+1}$, $j = 1, 2, \ldots, N$ and $\ell = 1, 2, \ldots, p+1$.

Proceeding in the same way as in the previous proofs, i.e., applying the following steps:

- putting $\hat{V} = \eta_\ell$,
- using $0 \leqslant \mu\|\nabla\eta_\ell\|_{0,j}^2 \leqslant (a(U_\ell)\nabla\eta_\ell, \nabla\eta_\ell)_j, \quad \ell = 1, 2, \ldots, p+1$,
- bounding the terms with a and f using (2.2), (2.3), (2.4), etc.,
- applying the Friedrichs inequality utilizing $\eta_\ell \in \hat{S}_0^{N,p+1} \subset H_0^1(c,d)$,
- using the fact that (5.25) implies

$$\|\overline{\partial}_{t,\ell}\theta_\ell\|_0 = \tau^{-1}\Bigl\|\sum_{m=1}^{p+1}\overline{a}_{\ell m}\theta_m\Bigr\| \leqslant \tau^{-1}C_a\sum_{m=1}^{p+1}\|\theta_m\| \leqslant Ch^{p+1/2}$$
$$\text{for } \ell = 1, 2, \ldots, p+1,$$

- summing over j, using the previous step, (5.21), (5.20), (5.25), (3.6) and (5.26) we obtain the result. □

The following theorem about the convergence of the effectivity indices is now easy to prove.

Theorem 6.1. *Let $u_\ell \in H^{p+2} \cap H_0^1$ and $U_\ell \in S_0^{N,p}$ be solutions of* (2.8) *with* (2.9), *and* (5.1) *with* (5.3) *at time instants $t_i + c_\ell\tau$, $\ell = 1, 2, \ldots, p+1$. Let $E_\ell \in \hat{S}_0^{N,p+1}$ be the solution of* (6.1) *with* (6.4) *(for E_{PN}), or* (6.2) *with* (6.4) *(for E_{PL}), or* (6.3) *(for E_{EL}). Let U_ℓ satisfy* (5.9) *and let $E_{\mathrm{PN},\ell}$ satisfy* (6.15) *for $\ell = 1, 2, \ldots, p+1$. Further, let*

$$\|e_\ell\|_1 \geqslant Ch^p, \quad \ell = 1, 2, \ldots, p+1. \tag{6.23}$$

Then

$$\lim_{h\to 0}\Theta_\ell = \lim_{h\to 0}\frac{\|E_\ell\|_1}{\|e_\ell\|_1} = 1, \quad \ell = 1,2,\dots,p+1, \tag{6.24}$$

where Θ *is* Θ_{PN}, Θ_{PL} *or* Θ_{EL}.

Proof. Rewrite e given by (5.5) as

$$e_\ell = u_\ell - (u_\ell^h + e_\ell^h) + (u_\ell^h - U_\ell) + (e_\ell^h - E_\ell) + E_\ell, \quad \ell = 1,2,\dots p+1.$$

Then

$$e_\ell = E_\ell + \hat{\varrho}_\ell + \theta_\ell + \eta_\ell \quad \text{and} \quad E_\ell = e_\ell - \hat{\varrho}_\ell - \theta_\ell - \eta_\ell$$

for $\ell = 1,2,\dots p+1$, and

$$\|E_\ell\|_1 \geqslant \|e_\ell\|_1 - \|\hat{\varrho}_\ell\|_1 - \|\theta_\ell\|_1 - \|\eta_\ell\|_1 \geqslant \|e_\ell\|_1 - C_1 h^{p+\alpha}, \tag{6.25}$$

$$\|E_\ell\|_1 \leqslant \|e_\ell\|_1 + \|\hat{\varrho}_\ell\|_1 + \|\theta_\ell\|_1 + \|\eta_\ell\|_1 \leqslant \|e_\ell\|_1 + C_2 h^{p+\alpha} \tag{6.26}$$

for $\ell = 1,2,\dots p+1$, where $\alpha = 1$ for E_{PN} and E_{PL}, $\alpha = 1/2$ for E_{EL}. Dividing (6.25) and (6.26) by $\|e_\ell\|_1$ and taking (6.23) into account, we see that

$$1 - C_1 h^\alpha \leqslant \Theta_\ell \leqslant 1 + C_2 h^\alpha, \quad \ell = 1,2,\dots p+1.$$

Then (6.24) holds for $h \to 0$. □

Note that assumption (6.23) implies $C_1 h^p \leqslant \|e_\ell\|_1 \leqslant C_2 h^p$, see (5.14).

References

[1] *A. Adjerid, J. E. Flaherty and Y. J. Wang*: A posteriori error estimation with finite element methods of lines for one-dimensional parabolic systems. Numer. Math. *65* (1993), 1–21.

[2] *J. C. Butcher*: A transformed implicit Runge-Kutta method. J. Assoc. Comput. Mach. *26* (1979), 731–738.

[3] *K. Burrage*: A special family of Runge-Kutta methods for solving stiff differential equations. BIT *18* (1978), 22–41.

[4] *P. G. Ciarlet*: The Finite Element Method for Elliptic Problems. North-Holland Publishing Company, Amsterdam, New York, Oxford, 1978.

[5] *S. Fučík, A. Kufner*: Nonlinear Differential Equations. Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York, 1980.

[6] *H. Gajevski, K. Gröger and K. Zacharias*: Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen. Akademie-Verlag, Berlin, 1974.

[7] *I. Hlaváček, M. Křížek and J. Malý*: On Galerkin approximations of a quasilinear nonpotential elliptic problem of a nonmonotone type. J. Math. Anal. Appl. *184* (1994), 168–189.

[8] *S. Larsson, V. Thomée and N. Y. Zhang*: Interpolation of coefficients and transformation of the dependent variable in the finite element methods for the nonlinear heat equation. Math. Methods Appl. Sci. *11* (1989), 105–124.

[9] *P. K. Moore*: A posteriori error estimation with finite element semi- and fully discrete methods for nonlinear parabolic equations in one space dimension. SIAM J. Numer. Anal. *31* (1994), 149–169.

[10] *P. K. Moore, J. E. Flaherty*: High-order adaptive solution of parabolic equations I. Singly implicit Runge-Kutta methods and error estimation. Rensselaer Polytechnic Institute Report 91-12. Troy, NY, Department of Computer Science, Rensselaer Polytechnic Institute, 1991.

[11] *P. K. Moore, J. E. Flaherty*: High-order adaptive finite element-singly implicit Runge-Kutta methods for parabolic differential equations. BIT *33* (1993), 309–331.

[12] *T. Roubíček*: Nonlinear differential equations and inequalities. Mathematical Institute of Charles University, Prague, in preparation.

[13] *K. Segeth*: A posteriori error estimation with the finite element method of lines for a nonlinear parabolic equation in one space dimension. Numer. Math. *33* (1999), 455–475.

[14] *B. Szabó, I. Babuška*: Finite Element Analysis. John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1991.

[15] *V. Thomée*: Galerkin Finite Element Methods for Parabolic Problems. Springer, Berlin, 1997.

Author's address: *Tomáš Vejchodský*, Mathematical Institute of the Academy of Sciences of the Czech Republic, Žitná 25, CZ-115 67 Praha 1, Czech Republic, e-mail: `vejchod@math.cas.cz`.