BIT Numerical Mathematics
2005, to appear

LINEARLY IMPLICIT FINITE ELEMENT METHODS FOR THE TIME-DEPENDENT JOULE HEATING PROBLEM*

G. AKRIVIS[1†] and S. LARSSON[2]

June 27, 2005

[1] *Computer Science Department, University of Ioannina, 451 10 Ioannina, Greece. email: akrivis@cs.uoi.gr*

[2] *Department of Mathematics, Chalmers University of Technology and Göteborg University, SE-412 96 Göteborg, Sweden. email: stig@chalmers.se*

Abstract.

Completely discrete numerical methods for a nonlinear elliptic-parabolic system, the time-dependent Joule heating problem, are introduced and analyzed. The equations are discretized in space by a standard finite element method, and in time by combinations of rational implicit and explicit multistep schemes. The schemes are linearly implicit in the sense that they require, at each time level, the solution of linear systems of equations. Optimal order error estimates are proved under the assumption of sufficiently regular solutions.

AMS subject classification (2000): 65M30, 65M15, 35K60.

Key words: Error estimate, finite element, multistep, semi-implicit.

1 Introduction

In this paper we construct and analyze high-order numerical methods for the time-dependent Joule heating problem. In space we discretize by the finite element method; for the time stepping we use a combination of rational implicit and explicit multistep schemes.

We consider the following nonlinear elliptic-parabolic system: Given $T > 0$, Ω a bounded interval, convex polygonal or polyhedral domain in $\mathbb{R}^d$ for $d = 1, 2$ or 3, respectively, $\sigma : \mathbb{R} \to [\kappa, K]$, with two positive constants κ and K, a globally Lipschitz continuous function, $g : \Omega \times [0, T] \to \mathbb{R}$, and $u^0 : \Omega \to \mathbb{R}$, we seek two functions $u, \varphi : \Omega \times [0, T] \to \mathbb{R}$, satisfying

$$
\begin{cases}
u_t - \Delta u = \sigma(u)|\nabla\varphi|^2, & \text{in } \Omega \times [0,T],\\
-\nabla \cdot \big(\sigma(u)\nabla\varphi\big) = 0, & \text{in } \Omega \times [0,T],\\
u = 0, \quad \varphi = g, & \text{on } \partial\Omega \times [0,T],\\
u(\cdot, 0) = u^0, & \text{in } \Omega.
\end{cases}
\tag{1.1}
$$

*Received March 2005. Communicated by Axel Ruhe.

†G. A. was partially supported by a 'Pythagoras' grant funded by the Greek Ministry of National Education and the European Commission.

This system models the electric heating of a conductive body, with u being the temperature, φ the electric potential, and σ the temperature-dependent electric conductivity.

The existence and uniqueness of global weak solutions in two space dimensions, $d = 2$, was shown in [4]. The regularity of these solutions was further studied in [7], where regularity estimates involving, essentially, space and time derivatives of first and second order were proved. We are not aware of any existence and regularity result in the three-dimensional case. In the present work we need regularity of high order and we simply assume that problem (1.1) possesses a unique, sufficiently regular solution.

To motivate the definition of the finite element method for problem (1.1), we give a weak formulation of it: Find $u(\cdot,t) \in H_0^1$ and $\varphi(\cdot,t) \in H^1$, with $\varphi(\cdot,t) - g(\cdot,t) \in H_0^1$, such that

$$\begin{cases} (u_t, v) + (\nabla u, \nabla v) = (\sigma(u)|\nabla\varphi|^2, v) \quad \forall v \in H_0^1,\ t \in [0,T], \\ u(\cdot,0) = u^0, \end{cases}$$

and

$$(\sigma(u)\nabla\varphi, \nabla v) = 0 \quad \forall v \in H_0^1,\ t \in [0,T].$$

Here we use standard notation: $H^1 = H^1(\Omega)$ and $H_0^1 = H_0^1(\Omega)$ are Sobolev spaces, and $(\cdot,\cdot)$ denotes both the L^2 inner product and the duality pairing between H^{-1} and H_0^1.

Space discretization

Let $\{S_h\}_{0<h\le 1}$ be a family of finite dimensional subspaces of H^1, consisting of continuous piecewise polynomials of degree $r-1 \ge 1$ with respect to a quasi-uniform family of triangulations of Ω, and set $\mathring{S}_h := S_h \cap H_0^1$. Let $\pi_h : C(\bar\Omega) \to S_h$ be a linear interpolation operator such that $\pi_h v|_{\partial\Omega} = 0$ when $v|_{\partial\Omega} = 0$, and such that

$$\|v - \pi_h v\|_{W^{j,s}} \le C h^{r-j-d(\frac{1}{2}-\frac{1}{s})} \|v\|_{H^r}, \quad s \in [2,\infty],\ j = 0,1. \tag{1.2}$$

We consider the following semidiscrete problem: Find $u_h(\cdot,t) \in \mathring{S}_h$ and $\varphi_h(\cdot,t) \in S_h$ with $\varphi_h(\cdot,t) - \pi_h g(\cdot,t) \in \mathring{S}_h$, satisfying

$$\begin{cases} (u_{h,t}, \chi) + (\nabla u_h, \nabla\chi) = (\sigma(u_h)|\nabla\varphi_h|^2, \chi) \quad \forall \chi \in \mathring{S}_h,\ t \in [0,T], \\ u_h(\cdot,0) = u_h^0, \end{cases} \tag{1.3}$$

and

$$(\sigma(u_h)\nabla\varphi_h, \nabla\chi) = 0 \quad \forall \chi \in \mathring{S}_h,\ t \in [0,T], \tag{1.4}$$

where $u_h^0 \in \mathring{S}_h$ is an approximation to u^0 such that

$$\|u^0 - u_h^0\| \le C h^r, \tag{1.5}$$

with $\|\cdot\|$ denoting the norm of $L^2(\Omega)$.

Fully discrete schemes

We discretize (1.3) in time by a class of linearly implicit schemes expressed in terms of bounded rational functions $\alpha_i, \beta_i : [0,\infty] \to \mathbb{R}$, $i = 0,\dots,q$, with $\alpha_q = 1, \beta_q = 0$, and the functions β_i vanishing at infinity, $\beta_i(\infty) = 0, i = 0,\dots,q-1$; these schemes were recently introduced and analyzed for nonlinear parabolic equations in [1], [2], [3].

To write the schemes in compact form, we introduce the discrete Laplacian $\Delta_h : \mathring{S}_h \to \mathring{S}_h$ and the L^2-projection $P_h : H^{-1} \to \mathring{S}_h$ by

$$-(\Delta_h \psi, \chi) = (\nabla\psi, \nabla\chi), \quad (P_h v - v, \chi) = 0 \quad \forall \psi, \chi \in \mathring{S}_h,\ v \in H^{-1}.$$

Letting $N \in \mathbb{N}$, $k = T/N$ be the time step, and $t^n = nk, n = 0,\dots,N$, we define a sequence of approximations U^n and Φ^n to $u^n := u(\cdot, t^n)$ and $\varphi^n := \varphi(\cdot, t^n)$, respectively, as follows: Find $U^n \in \mathring{S}_h$, $\Phi^n \in S_h$ with $\Phi^n - \pi_h g^n \in \mathring{S}_h$, such that, for $0 \le n \le N-q$,

$$\sum_{i=0}^{q} \alpha_i(-k\Delta_h) U^{n+i} = k \sum_{i=0}^{q-1} \beta_i(-k\Delta_h) P_h\big(\sigma(U^{n+i})|\nabla\Phi^{n+i}|^2\big), \tag{1.6}$$

$$(\sigma(U^{n+q})\nabla\Phi^{n+q}, \nabla\chi) = 0 \quad \forall \chi \in \mathring{S}_h. \tag{1.7}$$

Given $U^n,\dots,U^{n+q-1} \in \mathring{S}_h$ and $\Phi^n,\dots,\Phi^{n+q-1} \in S_h$, it is easily seen that $U^{n+q} \in \mathring{S}_h$ is well defined by (1.6); then, given $U^{n+q} \in \mathring{S}_h$, Φ^{n+q} is well defined by (1.7).

The scheme (1.6)–(1.7) is linearly implicit; the implementation of (1.6) requires, at every time level, the solution of several linear systems of dimension $\dim(\mathring{S}_h)$ with the same matrices for all time levels, while for the implementation of (1.7) a linear system with matrix varying with time has to be solved. The scheme (1.6)–(1.7) can be easily modified to allow partially parallel implementation, in the sense that U^{n+q} is not required for the computation of Φ^{n+q}, see Remark 3.3.

Stability assumptions

For $x \in [0,\infty]$ we introduce the polynomial $\alpha(x,\cdot)$ by

$$\alpha(x,\zeta) := \sum_{i=0}^{q} \alpha_i(x)\zeta^i.$$

We order the roots $\zeta_j(x),\ j = 0,\dots,q$, of $\alpha(x,\cdot)$ in such a way that the functions ζ_j are continuous in $[0,\infty]$ and the roots $\xi_j := \zeta_j(0),\ j = 1,\dots,s$, satisfy $|\xi_j| = 1$; these unimodular roots are the *principal roots* of $\alpha(0,\cdot)$ and the complex numbers $\lambda_j := \frac{\partial_1\alpha(0,\xi_j)}{\xi_j\partial_2\alpha(0,\xi_j)}$ (with ∂_1 denoting differentiation with respect to the first variable) are the *growth factors* of ξ_j. We assume throughout that the implicit method described by the rational functions $\alpha_0,\dots,\alpha_q$ is *strongly $A(0)$-stable* in the sense that for all $0 < x \le \infty$ and for all $j = 0,\dots,q$, there holds $|\zeta_j(x)| < 1$, and the principal roots of $\alpha(0,\cdot)$ are simple and their growth factors have positive real parts, $\operatorname{Re}\lambda_j > 0,\ j = 0,\dots,s$, see [1]. An additional hypothesis, namely that $\alpha_0,\dots,\alpha_{q-1}$ vanish at infinity,

$$\alpha_0(\infty) = \cdots = \alpha_{q-1}(\infty) = 0, \tag{H}$$

will be used in Theorem 3.4 and will allow us to derive error estimates under weaker approximation assumptions on the starting approximations.

Consistency assumptions

Let $p \geq 1$, and the functions $\varphi_\ell : [0,\infty) \to \mathbb{R},\ \ell = 0,\dots,p$, be defined by

$$\begin{aligned}
\varphi_\ell(x) &:= \sum_{i=0}^{q} \big[i^\ell \alpha_i(x) - (\ell i^{\ell-1} + x i^\ell)\beta_i(x) \big],\ \ell = 0,\dots,p-1, \\
\varphi_p(x) &:= \sum_{i=0}^{q} \big[i^p \alpha_i(x) - p i^{p-1}\beta_i(x) \big].
\end{aligned}$$

We assume that the order of the time-stepping scheme is p, i.e.,

$$\varphi_\ell(x) = O(x^{p+1-\ell}) \quad \text{as} \ \ x \to 0+,\ \ell = 0,\dots,p,$$

and its polynomial order is $p-1$, i.e.,

$$\varphi_\ell = 0,\ \ell = 0,\dots,p-2;$$

if, in addition, $\varphi_{p-1} = 0$, then the polynomial order is p, see [1].

For examples of linearly implicit schemes, including implicit-explicit multistep schemes as well as the combination of Runge-Kutta schemes and extrapolation, satisfying our conditions, we refer to [1].

We introduce the elliptic projection $R_h : H_0^1 \to \mathring{S}_h$ by

$$(\nabla R_h v, \nabla \chi) = (\nabla v, \nabla \chi) \quad \forall \chi \in \mathring{S}_h\,,$$

and let $W(t) = R_h u(\cdot,t)$ for $t \in [0,T]$, $W^n = W(t^n)$ and $W^{(\ell)} = D_t^\ell W$. Assuming that the starting approximations $U^0,\dots,U^{p-1}$ are such that

$$\max_{0\leq j\leq q-1} \Big(\|W^j - U^j\| + k^{1/2}\|\nabla(W^j - U^j)\| \Big) \leq M_1(k^p + h^r),$$

which, in case assumption (H) is satisfied, can be relaxed to

$$\max_{0\leq j\leq q-1} \|W^j - U^j\| \leq M_1(k^p + h^r),$$

and that $k = O(h^{d/2p})$, we will establish the error estimate

$$\|u(t^n) - U^n\| + \|\varphi(t^n) - \Phi^n\| \leq C(k^p + h^r), \quad t^n \in [0,T].$$

Finite element methods for (1.1) have been analyzed in [7]. Both spatially semidiscrete schemes, as well as a first order time-stepping scheme, based on a combination of the backward and the forward Euler methods, have been considered. Optimal order error estimates have been derived under the weak mesh condition $k = O(h^{d/6})$. For a finite element analysis of the heating problem with more general boundary conditions see [8]. A similar system of equations arising in fluid mechanics is studied in [5], [6].

The outline of the paper is as follows: Section 2 is devoted to the consistency of the scheme (1.6)–(1.7) for the elliptic projection of the solution u of (1.1). In Section 3 optimal order error estimates are derived.

2 Consistency

Our main concern in this paper is to analyze the approximation properties of the sequences $\{U^n\}$, $\{\Phi^n\}$. As a preliminary step, we show consistency of the time-stepping scheme for the elliptic projection W of the solution u.

Let $E_h(t) \in \mathring{S}_h, t \in [0,T]$, denote the consistency error of the spatially discrete scheme for the elliptic projection $W = R_h u$, i.e.,

$$E_h(t) := W_t(t) - \Delta_h W(t) - P_h\left(\sigma(W(t))|\nabla\varphi(t)|^2\right), \quad t \in [0,T]. \tag{2.1}$$

Let $\|\cdot\|_{-1}$ denote the norm of H^{-1}, $\|v\|_{-1} := \sup_{w\in H_0^1} \frac{(v,w)}{\|\nabla w\|}$. Using the relation $\Delta_h R_h = P_h \Delta$ and (1.1), the well-known L^2 error estimate for W, a global Lipschitz condition for σ, and a $W^{1,\infty}$-bound for φ^n, we obtain

$$\max_{0\le t\le T} \|E_h(t)\|_{-1} \le C \max_{0\le t\le T} \|E_h(t)\| \le Ch^r. \tag{2.2}$$

We let $\tilde{W}^j = W^j,\ j = 0,\dots,q-1$, and apply the linearly implicit scheme (1.6) to (2.1) to define $\tilde{W}^m,\ m = q,\dots,N$, by the equations

$$\begin{aligned}&\sum_{i=0}^{q} \alpha_i(-k\Delta_h)\tilde{W}^{n+i}\\ &\quad = k\sum_{i=0}^{q-1} \beta_i(-k\Delta_h)\left[P_h\big(\sigma(\tilde{W}^{n+i})|\nabla\varphi^{n+i}|^2\big) + E_h(t^{n+i})\right],\end{aligned} \tag{2.3}$$

for $n = 0,\dots,N-q$. We define the consistency error E^n, $n = 0,\dots,N-q$, of (2.3) by

$$\begin{aligned}k(I - k\Delta_h)^{-1}E^n &= \sum_{i=0}^{q} \alpha_i(-k\Delta_h)W^{n+i}\\ &\quad - k\sum_{i=0}^{q-1} \beta_i(-k\Delta_h)\left[P_h\big(\sigma(W^{n+i})|\nabla\varphi^{n+i}|^2\big) + E_h(t^{n+i})\right],\end{aligned} \tag{2.4}$$

for $n = 0,\dots,N-q$. In this section we will derive an optimal order estimate for the consistency error E^n, see (2.10) below, assuming that the polynomial order of the scheme is p. We will also derive some preliminary consistency estimates for polynomial order $p-1$, which will be used in Section 3 to establish optimal order error estimates.

First, we use (2.1) to rewrite (2.4) in the form

$$
\begin{aligned}
k(I-k\Delta_h)^{-1}E^n &= \sum_{i=0}^{q}\alpha_i(-k\Delta_h)W^{n+i}\\
&\quad - k\sum_{i=0}^{q-1}\beta_i(-k\Delta_h)\big(W_t(t^{n+i})-\Delta_h W^{n+i}\big).
\end{aligned}
\tag{2.5}
$$

Letting

$$
\begin{aligned}
E_1^n &:= \sum_{\ell=0}^{p}\frac{k^\ell}{\ell!}\varphi_\ell(-k\Delta_h)W^{(\ell)}(t^n),\\
E_2^n &:= \frac{1}{p!}\sum_{i=0}^{q}\alpha_i(-k\Delta_h)\int_{t^n}^{t^{n+i}}(t^{n+i}-s)^pW^{(p+1)}(s)\,ds,\\
E_3^n &:= -\frac{k}{(p-1)!}\sum_{i=0}^{q-1}\beta_i(-k\Delta_h)\int_{t^n}^{t^{n+i}}(t^{n+i}-s)^{p-1}\big(W^{(p+1)}-k\Delta_hW^{(p)}\big)(s)\,ds,
\end{aligned}
$$

and Taylor expanding the right-hand side of (2.5), we easily see that

$$
k(I-k\Delta_h)^{-1}E^n = E_1^n+E_2^n+E_3^n. \tag{2.6}
$$

Using the boundedness of α_i,β_i and $\hat\beta_i, \hat\beta_i(x):=x\beta_i(x)$, we easily obtain

$$
\|(I-k\Delta_h)E_2^n\|_{-1}\le Ck^p\int_{t^n}^{t^{n+q}}\big(\|W^{(p+1)}(s)\|_{-1}+k\|\nabla W^{(p+1)}(s)\|\big)\,ds, \tag{2.7}
$$

$$
\|(I-k\Delta_h)E_3^n\|_{-1}\le Ck^p\int_{t^n}^{t^{n+q-1}}\big(\|W^{(p+1)}(s)\|_{-1}+\|\nabla W^{(p)}(s)\|\big)\,ds. \tag{2.8}
$$

In the sequel we will distinguish two cases. First, assuming that the polynomial order is p, we have $E_1^n = k^p\varphi_p(-k\Delta_h)u^{(p)}(t^n)/p!$, i.e.,

$$
(I-k\Delta_h)E_1^n = -\frac{k^{p+1}}{p!}\big[(-k\Delta_h)^{-1}\varphi_p(-k\Delta_h)+\varphi_p(-k\Delta_h)\big]\Delta_hW^{(p)}(t^n).
$$

Using then the fact that φ_p and $\tilde\varphi_p$, $\tilde\varphi_p(x):=\varphi_p(x)/x$, are bounded, we easily conclude

$$
\|(I-k\Delta_h)E_1^n\|_{-1}\le Ck^{p+1}\|\nabla W^{(p)}(t^n)\|. \tag{2.9}
$$

Under obvious regularity hypotheses, from (2.6) and (2.7)–(2.9) we immediately obtain the desired consistency estimate

$$
\max_{0\le n\le N-q}\|E^n\|_{-1}\le Ck^p. \tag{2.10}
$$

Next we will consider the case of polynomial order $p-1$. Let $\alpha(x) := \alpha_0(x)+\cdots+\alpha_q(x)$. Using the fact that the function η, $\eta(x)=\varphi_{p-1}(x)/[x\,\rho(x)]$, is bounded in $[0,\infty]$, see [1], we will prove in Lemma 3.2 optimal order error estimates. As a preparation for Lemma 3.2, let us note that in this case there is an additional term in E_1^n, which can be written as

$$\begin{aligned}
&\frac{k^{p-1}}{(p-1)!}\varphi_{p-1}(-k\Delta_h)\,W^{(p-1)}(t^n)\\
&\quad=-\frac{k^p}{(p-1)!}\,\eta(-k\Delta_h)\,\rho(-k\Delta_h)\,\Delta_h W^{(p-1)}(t^n)\\
&\quad=k(I-k\Delta_h)^{-1}\tilde E^n\\
&\qquad-\frac{k^p}{(p-1)!}\,\eta(-k\Delta_h)\sum_{i=0}^{q}\rho_i(-k\Delta_h)\,\Delta_h\big[W^{(p-1)}(t^n)-W^{(p-1)}(t^{n+i})\big],
\end{aligned}$$

with $\tilde E^n$ such that

$$k(I-k\Delta_h)^{-1}\tilde E^n=-\frac{k^p}{(p-1)!}\,\eta(-k\Delta_h)\sum_{i=0}^{q}\alpha_i(-k\Delta_h)\,\Delta_h W^{(p-1)}(t^{n+i}).$$

Thus, in this case, since $\tilde\eta, \tilde\eta(x) := (1+x)\eta(x)$, is bounded, (2.10) is replaced by

$$\max_{0\le n\le N-q}\|E^n-\tilde E^n\|_{-1}\le C\,k^p. \tag{2.11}$$

3 Error estimates

In this section we shall derive optimal order error estimates. Let $\zeta^n := W^n - \tilde W^n$ and $\theta^n := \tilde W^n - U^n$, $n=0,\dots,N$. Then

$$u^n-U^n=(u^n-W^n)+\zeta^n+\theta^n. \tag{3.1}$$

Here $u^n-W^n=(I-R_h)u^n$ can be easily estimated,

$$\max_{0\le n\le N}\|u^n-W^n\|\le Ch^r. \tag{3.2}$$

Hence, it remains to estimate ζ^n and θ^n. We will estimate ζ^n in Lemmas 3.1 and 3.2 for polynomial order p and $p-1$, respectively. Estimates for θ^n will be derived in Theorem 3.3 for all methods under consideration, and in Theorem 3.4 for methods satisfying (H) under weaker approximation hypotheses on the starting approximations.

Lemma 3.1. *Assume that the polynomial order is p. Then*

$$\max_{0\le n\le N}\Big(\|W^n-\tilde W^n\|^2+k\sum_{\ell=q}^{n}\|\nabla(W^\ell-\tilde W^\ell)\|^2\Big)\le Ck^{2p}. \tag{3.3}$$

PROOF. Let $\tilde{b}^n := P_h[\big(\sigma(W^n) - \sigma(\tilde{W}^n)\big)|\nabla\varphi^n|^2], n = 0,\dots,N$. Subtracting (2.3) from (2.4), we obtain

$$\sum_{i=0}^{q} \alpha_i(-k\Delta_h)\zeta^{n+i} = k\sum_{j=0}^{q-1} \beta_j(-k\Delta_h)\tilde{b}^{n+j} + k(I-k\Delta_h)^{-1}E^n,$$

$n = 0,\dots,N-q$. Then, we have, see [1],

$$\begin{aligned} &\|\zeta^n\|^2 + k\sum_{\ell=q}^{n}\|\nabla\zeta^\ell\|^2 \\ &\qquad \le C\Big\{\sum_{j=0}^{q-1}\big(\|\zeta^j\|^2 + k\|\nabla\zeta^j\|^2\big) + k\sum_{\ell=0}^{n-q}\Big(\|\tilde{b}^\ell\|_{-1}^2 + \|E^\ell\|_{-1}^2\Big)\Big\}, \end{aligned} \tag{3.4}$$

$n = q,\dots,N$. Using here (2.10) and the fact that $\zeta^0 = \cdots = \zeta^{q-1} = 0$ and $\|\tilde{b}^\ell\|_{-1} \le C\|\zeta^\ell\|$, a discrete Gronwall argument leads to (3.3) and the proof is complete. □

Next we relax the polynomial order hypothesis.

LEMMA 3.2. *Assume that the polynomial order is $p-1$. Then* (3.3) *is valid.*

PROOF. Let $\hat{\zeta}_2$ and $\tilde{\zeta}_2^j, j = 0,\dots,N$, be such that $\hat{\zeta}_2^j = \tilde{\zeta}_2^j = 0, j = 0,\dots,q-1$, and

$$\begin{aligned} \sum_{i=0}^{q} \alpha_i(-k\Delta_h)\hat{\zeta}_2^{n+i} &= k(I-k\Delta_h)^{-1}(E^n - \tilde{E}^n), \\ \sum_{i=0}^{q} \alpha_i(-k\Delta_h)\tilde{\zeta}_2^{n+i} &= k(I-k\Delta_h)^{-1}\tilde{E}^n, \end{aligned}$$

$n = 0,\dots,N-q$. Let further $\zeta_1^m := \zeta^m - \hat{\zeta}_2^m - \tilde{\zeta}_2^m$, $m = 0,\dots,N$. Then

$$\sum_{i=0}^{q} \alpha_i(-k\Delta_h)\zeta_1^{n+i} = k\sum_{j=0}^{q-1} \beta_j(-k\Delta_h)\tilde{b}^{n+j},$$

$n = 0,\dots,N-q$. Now ζ_1^n and, in view of (2.11), also $\hat{\zeta}_2^n$, can be easily estimated in the desired form, see (3.4). Further, $\tilde{\zeta}_2^n$ can be estimated as in the proof of Theorem 4.2 in [1]. □

Next we estimate $\theta^n = \tilde{W}^n - U^n$ and obtain a bound for the total error.

THEOREM 3.3. *Assume that $r \ge 3$, $d \le 3$. Let U^n, Φ^n and $\tilde{W}^n$ be solutions of* (1.6)–(1.7) *and* (2.3)*, respectively, with starting approximations $U^0,\dots,U^{q-1} \in \mathring{S}_h$ such that*

$$\max_{0\le j\le q-1}\Big(\|W^j - U^j\| + k^{1/2}\|\nabla(W^j - U^j)\|\Big) \le M_1(k^p + h^r). \tag{3.5}$$

Let $k = O(h^{d/2p})$. Then, for k and h small enough,

$$\|u(t^n) - U^n\| + \|\varphi(t^n) - \Phi^n\| \le C(k^p + h^r), \quad t^n \in [0,T], \tag{3.6}$$

with a constant C independent of k and h.

PROOF. We first recall an argument which is based on an inverse inequality and the interpolation error bounds in (1.2). Let $s \in [2, \infty]$. Then, for $v_h \in S_h$,

$$\begin{aligned}\|v - v_h\|_{W^{1,s}} &\le \|v - \pi_h v\|_{W^{1,s}} + \|\pi_h(v - v_h)\|_{W^{1,s}}\\ &\le \|v - \pi_h v\|_{W^{1,s}} + Ch^{-(\frac{d}{2}-\frac{d}{s})}\|\pi_h(v - v_h)\|_{W^{1,2}}\\ &\le \|v - \pi_h v\|_{W^{1,s}}\\ &\quad + Ch^{-(\frac{d}{2}-\frac{d}{s})}\big(\|v - \pi_h v\|_{W^{1,2}} + \|v - v_h\|_{W^{1,2}}\big)\\ &\le Ch^{r-1-(\frac{d}{2}-\frac{d}{s})}\|v\|_{H^r} + Ch^{-(\frac{d}{2}-\frac{d}{s})}\|v - v_h\|_{W^{1,2}}.\end{aligned} \tag{3.7}$$

We have $\|\nabla\varphi\|_{L^\infty([0,T],L^\infty)} \le M_2$ for some constant M_2. In order to prove a preliminary error estimate, we replace the nonlinearity $(u,\varphi) \mapsto \sigma(u)|\nabla\varphi|^2$ by $(u,\varphi) \mapsto \sigma(u)f(\nabla\varphi)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a globally Lipschitz continuous function such that

$$f(x) = |x|^2 \quad \text{for } x \in \mathbb{R}^d \text{ with } |x| \le M_2 + 2.$$

This does not affect the solution of (1.1), since $|\nabla\varphi| \le M_2$ implies $f(\nabla\varphi) = |\nabla\varphi|^2$. We begin by proving a preliminary error bound, see (3.13) and (3.11) below, for the solution of the modified equation (1.6)–(1.7), still denoted U^n, Φ^n. By the inverse inequality argument (3.7), applied with $v = \varphi^n$, $v_h = \Phi^n$, $s = \infty$, using $k = O(h^{d/2p})$, $r \ge 3$, $d \le 3$, we may then show a posteriori that

$$\sup_{t^n \in [0,T]} \|\nabla\Phi^n\|_{L^\infty} \le M_2 + 1, \tag{3.8}$$

if k and h are small enough, so that $\sigma(U^n)f(\nabla\Phi^n) = \sigma(U^n)|\nabla\Phi^n|^2$ and hence U^n, Φ^n is identical to the solution of the original equation (1.6)–(1.7).

We now turn to $\theta^n = \tilde{W}^n - U^n$, which by (1.6) and (2.3), satisfies the equation:

$$\sum_{i=0}^{q} \alpha_i(-k\Delta_h)\theta^{n+i} = k\sum_{i=0}^{q-1} \beta_i(-k\Delta_h)\left(P_h b^{n+i} + E_h(t^{n+i})\right), \quad n = 0, \dots, N-q,$$

where, since $f(\nabla\varphi^\ell) = |\nabla\varphi^\ell|^2$,

$$b^\ell := \sigma(\tilde{W}^\ell)f(\nabla\varphi^\ell) - \sigma(U^\ell)f(\nabla\Phi^\ell).$$

Using the fact that $\|P_h v\|_{-1} \le C\|v\|_{-1}$ for $v \in H^{-1}$, it easily follows from Lemma 2.1 of [1] that

$$\begin{aligned}\|\theta^n\|^2 + k\sum_{\ell=q}^{n} \|\nabla\theta^\ell\|^2 \le C\Big(&\sum_{j=0}^{q-1}\big(\|\theta^j\|^2 + k\|\nabla\theta^j\|^2\big)\\ &+ k\sum_{\ell=0}^{n-q}\|b^\ell\|_{-1}^2 + k\sum_{\ell=0}^{n-q}\|E_h(t^\ell)\|_{-1}^2\Big),\end{aligned} \tag{3.9}$$

for $n = q, \dots, N$, and thus, in view of (2.2) and (3.5),

$$\|\theta^n\|^2 + k\sum_{\ell=q}^{n} \|\nabla\theta^\ell\|^2 \le C\Big((k^p + h^r)^2 + k\sum_{\ell=0}^{n-q} \|b^\ell\|_{-1}^2\Big). \tag{3.10}$$

Now $\varphi^\ell - \Phi^\ell$ can, in view also of (3.3) and the fact that $k = O(h^{d/6p})$, be estimated in terms of θ^ℓ as follows:

$$\|\nabla(\varphi^\ell - \Phi^\ell)\| \le C\big(h^{r-1} + k^p + \|\theta^\ell\|\big), \tag{3.11}$$

$$\|\varphi^\ell - \Phi^\ell\| \le C\big(h^r + k^p + \|\theta^\ell\| + h^{-\frac{d}{6}}\|\theta^\ell\|^2\big), \tag{3.12}$$

see Lemma 3.2 in [7].

We use this to show a preliminary low order estimate,

$$\|\theta^n\| \le C\big(k^p + h^{r-1}\big). \tag{3.13}$$

We have, since $f(\nabla\varphi^\ell) = |\nabla\varphi^\ell|^2$,

$$b^\ell = \big(\sigma(\tilde W^\ell) - \sigma(U^\ell)\big)|\nabla\varphi^\ell|^2 + \sigma(U^\ell)\big(f(\nabla\varphi^\ell) - f(\nabla\Phi^\ell)\big).$$

Hence, by (3.11) and the boundedness of σ and $\nabla\varphi^\ell$,

$$\begin{aligned}\|b^\ell\|_{-1} &\le C\|b^\ell\|\\ &\le C\|\theta^\ell\|\|\nabla\varphi^\ell\|_{L^\infty}^2 + C\|\sigma(U^\ell)\|_{L^\infty}\|\nabla(\varphi^\ell - \Phi^\ell)\|\\ &\le C\big(h^{r-1} + k^p + \|\theta^\ell\|\big).\end{aligned}$$

Together with (3.10) and Gronwall's lemma this proves (3.13).

From this point on we will not use f, since we now know that Φ^n satisfies (3.8).

In order to prove the optimal order estimate (3.6), using the relation

$$|\nabla\varphi^\ell|^2 - |\nabla\Phi^\ell|^2 = 2\nabla\varphi^\ell\cdot\nabla(\varphi^\ell - \Phi^\ell) - |\nabla(\varphi^\ell - \Phi^\ell)|^2,$$

we can split b^ℓ as $b^\ell = R_1^\ell + R_2^\ell + R_3^\ell$ with

$$\begin{aligned}R_1^\ell &:= \big[\sigma(\tilde W^\ell) - \sigma(U^\ell)\big]|\nabla\varphi^\ell|^2 + 2\big[\sigma(U^\ell) - \sigma(u^\ell)\big]\nabla\varphi^\ell\cdot\nabla(\varphi^\ell - \Phi^\ell),\\ R_2^\ell &:= 2\sigma(u^\ell)\nabla\varphi^\ell\cdot\nabla(\varphi^\ell - \Phi^\ell),\\ R_3^\ell &:= -\sigma(U^\ell)|\nabla(\varphi^\ell - \Phi^\ell)|^2.\end{aligned}$$

First, obviously, in view of (3.2), (3.3), and the Lipschitz continuity of σ,

$$\|R_1^\ell\|_{-1} \le C\big(\|\theta^\ell\| + k^p + h^r\big).$$

Further, for $v \in H_0^1$,

$$\begin{aligned}(R_2^\ell, v) &= 2(\nabla(\varphi^\ell - \Phi^\ell), \sigma(u^\ell)\nabla\varphi^\ell v)\\ &= -2(\varphi^\ell - \Phi^\ell, \nabla\cdot(\sigma(u^\ell)\nabla\varphi^\ell v))\\ &= -2(\varphi^\ell - \Phi^\ell, \sigma(u^\ell)\nabla\varphi^\ell\cdot\nabla v),\end{aligned}$$

since $\nabla\cdot(\sigma(u^\ell)\nabla\varphi^\ell)=0$ by (1.1). Hence

$$\|R_2^\ell\|_{-1}\le C\|\varphi^\ell-\Phi^\ell\|,$$

and thus, in view of (3.12), (3.13), and the assumption that $k=O(h^{d/6p})$,

$$\begin{aligned}\|R_2^\ell\|_{-1}&\le C\big(h^r+k^p+\|\theta^\ell\|+h^{-\frac{d}{6}}\|\theta^\ell\|^2\big)\\&\le C\big(h^r+k^p+(1+h^{r-1-\frac{d}{6}}+k^ph^{-\frac{d}{6}})\|\theta^\ell\|\big)\le C\big(h^r+k^p+\|\theta^\ell\|\big).\end{aligned}$$

Finally, concerning R_3^ℓ, we first note that

$$\|R_3^\ell\|_{-1}\le C\|R_3^\ell\|_{L^{6/5}},$$

which follows from Hölder's and Sobolev's inequalities (for $d\le 3$):

$$(R_3^\ell,v)\le\|R_3^\ell\|_{L^{6/5}}\|v\|_{L^6}\le C\|R_3^\ell\|_{L^{6/5}}\|\nabla v\|,\quad\forall v\in H_0^1.$$

Thus, using (1.2), (3.7) with $s=12/5$, and (3.11),

$$\begin{aligned}\|R_3^\ell\|_{-1}&\le C\|R_3^\ell\|_{L^{6/5}}\\&\le C\|\sigma(U^\ell)\|_{L^\infty}\|\nabla(\varphi^\ell-\Phi^\ell)\|^2_{L^{12/5}}\le C\|\nabla(\varphi^\ell-\Phi^\ell)\|^2_{L^{12/5}}\\&\le Ch^{2r-2-\frac{d}{6}}\|\varphi^\ell\|_{H^r}+Ch^{-\frac{d}{6}}\|\nabla(\varphi^\ell-\Phi^\ell)\|^2\\&\le Ch^{2r-2-\frac{d}{6}}+Ch^{-\frac{d}{6}}k^{2p}+Ch^{-\frac{d}{6}}\|\theta^\ell\|^2,\end{aligned}$$

i.e., in view of (3.13) and the assumptions that $k=O(h^{d/6p})$ and $r\ge 3$, $d\le 3$,

$$\|R_3^\ell\|_{-1}\le C(h^r+k^p).$$

Hence

$$\|b^\ell\|_{-1}\le C\big(h^r+k^p+\|\theta^\ell\|\big).\tag{3.14}$$

Together with (3.10) and Gronwall's lemma this proves an $O(h^r+k^p)$ bound for $\|\theta^n\|$, and by (3.12) also for $\|\varphi^n-\Phi^n\|$. Combined with (3.1), (3.2), and Lemmas 3.1–3.2 this proves (3.6). ∎

For schemes satisfying (H), (3.9) can be replaced by

$$\|\theta^n\|^2+k\sum_{\ell=q}^n\|\nabla\theta^\ell\|^2\le C\Big(\sum_{j=0}^{q-1}\|\theta^j\|^2+k\sum_{\ell=0}^{n-q}\|b^\ell\|^2_{-1}+k\sum_{\ell=0}^{n-q}\|E_h(t^\ell)\|^2_{-1}\Big),$$

see Remark 7.2 and Lemma 2.1 in [1]. Therefore, we have the following result:

Theorem 3.4. *Assume that* (H) *is satisfied, and that the starting approximations* $U^0,\dots,U^{q-1}\in\mathring{S}_h$ *are such that*

$$\max_{0\le j\le q-1}\|W^j-U^j\|\le M_1(k^p+h^r),$$

and all hypotheses of Theorem 3.3 except (3.5) *are satisfied. Then* (3.6) *is valid.*

Remark 3.1. The mesh condition $k = O(h^{d/2p})$ in Theorems 3.3 and 3.4 becomes stringent for $d = 3$ and $p = 1$. For a first-order scheme, a combination of the backward and the forward Euler methods, the estimate (3.6) is established in [7] under the weaker condition $k = O(h^{d/6})$; the major difference between the approach of [7] and our present approach is what smoothing property is used; in [7] a stronger smoothing property for the backward Euler method is used while here we use (3.9), which is valid for any strongly $A(0)$-stable scheme.

Remark 3.2. The optimal order error estimate (3.6) in Theorems 3.3 and 3.4 is derived for $r \geq 3$. Here we briefly discuss the case $r = 2$. The assumption $r \geq 3$ is used in the proof only in the proof of (3.8) and in the estimation of $\|R_3^\ell\|_{-1}$. For $r = 2$ and $d = 3$ this estimate yields

$$\|R_3^\ell\|_{-1} \leq C\big(k^p + h^{3/2}\big).$$

However, our proof of (3.8) does not work in this case.

For $d = 1$ or 2, $\|R_3^\ell\|_{-1}$ is dominated (modulo a constant factor) by $\|R_3^\ell\|_{L^s}$ with $s = 1$ for $d = 1$, and any $s > 1$ for $d = 2$. Estimating $\|R_3^\ell\|_{-1}$ by $\|R_3^\ell\|_{L^s}$ instead of $\|R_3^\ell\|_{L^{6/5}}$ and proceeding along the lines of our proof we obtain (for $r = 2$)

$$\begin{aligned}
&\max_{0\leq n\leq N} \|\theta^n\| \leq C(k^p + h^2) \quad \text{for} \quad d = 1,\\
&\max_{0\leq n\leq N} \|\theta^n\| \leq C(k^p + h^{2-\varepsilon}) \quad \text{for} \quad d = 2,
\end{aligned}$$

for any positive ε. Thus, for $d = 1$ the estimate (3.6) holds for $r = 2$ as well, while for $r = 2$ and $d = 2$ we obtain an estimate of the form (3.6) with h^2 replaced by $h^{2-\varepsilon}$, for any positive ε. The proof of (3.8) works for $r = 2$ and $d \leq 2$.

Remark 3.3. To advance in time the scheme (1.6)–(1.7) a number of linear systems have to be solved for U^{n+q} and one linear system for Φ^{n+q}. These systems cannot be solved in parallel since Φ^{n+q} depends on U^{n+q}. (The scheme (1.6)–(1.7) can be implemented in parallel for Φ^{n+q} and U^{n+q+1} if $\beta_{q-1} = 0$.) However, the scheme (1.6)–(1.7) can be easily modified to allow partially parallel implementation. To this end we replace (1.7) by

$$(\sigma(\widehat{U}^{n+q})\nabla\Phi^{n+q}, \nabla\chi) = 0 \quad \forall \chi \in \mathring{S}_h. \tag{3.15}$$

with

$$\widehat{U}^{n+q} := \sum_{j=0}^{q-1} L_j^{n,q}(t^{n+q})\, U^{n-j}, \tag{3.16}$$

where $L_j^{n,q} \in \mathbb{P}_{q-1}$ are the Lagrange polynomials associated with the points $t^n, \dots, t^{n+q-1}$, i.e., $L_j^{n,q}(t^{n+i}) = \delta_{ij}, i, j = 0, \dots, q-1$. It is easily seen that all error estimates for the scheme (1.6)–(1.7) established in this paper are also valid for the scheme (1.6)–(3.15)–(3.16).

REMARK 3.4. The error estimate (3.6) is established under the condition that σ is globally Lipschitz continuous. This may be relaxed to a local Lipschitz condition if we assume that $\max_{0\le j\le q-1}\|W^j-U^j\|_{L^\infty}\le 1/2$ and

$$\begin{aligned} h &= o(k^{\frac{1}{2r}}) && \text{for } d=1,\\ h|\log h|^{\frac{1}{2r}} &= o(k^{\frac{1}{2r}}) && \text{for } d=2,\\ h &= o(k^{\frac{1}{2r-1}}) && \text{for } d=3. \end{aligned} \tag{3.17}$$

To see this, we replace σ everywhere by a globally Lipschitz continuous function $\tilde\sigma$ coinciding with σ in the interval

$$I := [\min_{x,t} u(x,t)-1, \max_{x,t} u(x,t)+1]. \tag{3.18}$$

This does not affect the solution of (1.1). Further, using the same notation for the numerical approximations given by (1.6)–(1.7) (with $\tilde\sigma$ instead of σ), all our estimates are valid and we have, in view of (3.10) and (3.14),

$$\max_{q\le n\le N}\sqrt{k}\,\|\nabla\theta^n\|\le C(k^p+h^r).$$

By an inverse inequality, from (3.2) and (3.3) we obtain, for sufficiently small h,

$$\max_{0\le n\le N}\|u^n-\tilde W^n\|_{L^\infty}\le\frac12. \tag{3.19}$$

Similarly,

$$\begin{aligned} &\max_{q\le n\le N}\|\theta^n\|_{L^\infty}\le C\big(k^{p-\frac12}+h^rk^{-\frac12}\big) && \text{for } d=1,\\ &\max_{q\le n\le N}\|\theta^n\|_{L^\infty}\le C\big(k^{p-\frac12}+h^rk^{-\frac12}\big)|\log h|^{\frac12} && \text{for } d=2,\\ &\max_{q\le n\le N}\|\theta^n\|_{L^\infty}\le C\big(k^{p-\frac12}h^{-\frac12}+h^{r-\frac12}k^{-\frac12}\big) && \text{for } d=3. \end{aligned}$$

Thus, under the mesh conditions (3.17), for k and h sufficiently small,

$$\max_{q\le n\le N}\|\theta^n\|_{L^\infty}\le\frac12. \tag{3.20}$$

From (3.19) and (3.20) we easily see that $U^n(x)\in I, n=0,\dots,N, x\in\Omega$, and thus $\sigma(U^n)=\tilde\sigma(U^n)$, and conclude easily that (3.6) is valid for the approximations given by the scheme (1.6)–(1.7).

REFERENCES

1. G. Akrivis and M. Crouzeix, *Linearly implicit methods for nonlinear parabolic equations*, Math. Comp. **73** (2004), 613–635.
2. G. Akrivis, M. Crouzeix, and Ch. Makridakis, *Implicit-explicit multistep finite element methods for nonlinear parabolic problems*, Math. Comp. **68** (1998), 457–477.

3. G. Akrivis, M. Crouzeix, and Ch. Makridakis, *Implicit-explicit multistep methods for quasilinear parabolic equations*, Numer. Math. **82** (1999), 521–541.

4. G. Cimatti, *Existence of weak solutions for the nonstationary problem of the Joule heating of a conductor*, Ann. Mat. Pura Appl. (4) **162** (1992), 33–42.

5. J. Douglas, Jr., R. E. Ewing, and M. F. Wheeler, *The approximation of the pressure by a mixed method in the simulation of miscible displacement*, RAIRO Anal. Numér. **17** (1983), 17–33.

6. J. Douglas, Jr., R. E. Ewing, and M. F. Wheeler, *A time-discretization procedure for a mixed finite element approximation of miscible displacement in porous media*, RAIRO Anal. Numér. **17** (1983), 249–265.

7. C. M. Elliott and S. Larsson, *A finite element model for the time-dependent Joule heating problem*, Math. Comp. **64** (1995), 1433–1453.

8. A. Sunmonu, *Galerkin method for a nonlinear parabolic-elliptic system with nonlinear mixed boundary conditions*, Numer. Methods PDE **9** (1993), 235–259.