SIAM J. NUMERICAL ANALYSIS
Vol. 16, No. 3, June 1979

0036-1429/79/1603-0010 $01.00/0

INCOMPLETE ITERATION FOR TIME-STEPPING A GALERKIN METHOD FOR A QUASILINEAR PARABOLIC PROBLEM*

JIM DOUGLAS, JR.†, TODD DUPONT† AND RICHARD E. EWING‡

Abstract. An iterative method is presented and analyzed which is based on using a preconditioned conjugate gradient iteration for approximately solving the linear equations produced at each time step by an extrapolated Crank–Nicolson–Galerkin procedure for time-stepping a quasilinear parabolic problem. Optimal order convergence rates are obtained for the iterative method which is shown to be (asymptotically) computationally more efficient than standard second-order-in-time correct methods.

1. Introduction. We shall consider a numerically efficient modification of an extrapolated Crank–Nicolson–Galerkin method [6], [14] for approximating the solution of the quasilinear parabolic problem given by

$$c(x,u)\frac{\partial u}{\partial t}-\nabla\cdot(a(x,u)\nabla u+b(x,u))=f(x,t,u),\qquad (x,t)\in Q, \tag{1.1a}$$

where $Q=\Omega\times J, J=(0,T]$, Ω is a bounded domain in $\mathbb{R}^d$, $d\leqq 3$, with boundary $\partial\Omega$, and $b=(b_1(x,u),\cdots,b_d(x,u))$; the solution $u\in C^1(\bar{Q})$ is subjected to the initial condition

$$u(x,0)=u_0(x),\qquad x\in\Omega, \tag{1.1b}$$

and the Neumann boundary condition

$$a(x,u)\frac{\partial u}{\partial \nu}+\nu\cdot b(x,u)=g(x,t),\qquad (x,t)\in\partial\Omega\times J, \tag{1.2}$$

where ν is outward unit normal to $\partial\Omega$.

Essentially, our modification of the extrapolated Crank–Nicolson–Galerkin method consists of using an iterative method based on a preconditioned conjugate gradient iteration employing a fixed preconditioning matrix to approximately solve the linear (extrapolation produces linear rather than nonlinear) algebraic equations at each time step. We preserve the accuracy inherent in the underlying Crank–Nicolson–Galerkin method, and we obtain very nearly optimal work estimates for the arithmetic required to produce the solution of a second-order-in-time correct method for (1.1)–(1.2).

In § 2 we introduce a finite element space, present the hypotheses on (1.1)–(1.2) and its solution u, discuss an elliptic projection of u, and recall the extrapolated Crank–Nicolson–Galerkin method. In § 3 we derive our modification of the above method and analyze the effect of the conjugate gradient iteration on a single time step. In § 4 we obtain global error estimates for any iterative method satisfying the local estimates already derived for the conjugate gradient procedure. In § 5 computational requirements are studied for our basic procedure and a number of variants that fall under the analysis of § 4.

2. Preliminaries. Let $(u,v)=\int_\Omega uv\,dx$, $\|u\|^2=(u,u)$, and $\langle u,v\rangle=\int_{\partial\Omega}uv\,d\sigma$. Also let $H^s=H^s(\Omega)$ be the Sobolev space of order s over Ω and denote by $\|u\|_s$ the corresponding norm. Let $\{\mathcal{M}_h\}$ be a family of finite-dimensional subsapces of $H^1(\Omega)$ with the following property:

* Received by the editors December 16, 1977.
† Department of Mathematics, University of Chicago, Chicago, Illinois 60637.
‡ Department of Mathematics, Ohio State University, Columbus, Ohio 43210.

There exist an integer $r \geqq 2$ and a constant K_0 such that, for $1 \leqq p \leqq r$ and $\varphi \in H^p(\Omega)$,

$$\inf_{\chi \in \mathcal{M}_h} \{\|\varphi - \chi\| + h\|\varphi - \chi\|_1\} \leqq K_0 \|\varphi\|_p h^p. \tag{2.1}$$

Assume that the family $\{\mathcal{M}_h\}$ also satisfies the following so-called "inverse hypotheses":

There exists a constant K_0, independent of h, such that for all $\varphi \in \mathcal{M}_h$,

$$\begin{aligned} &\text{a)} \quad \|\varphi\|_1 \leqq K_0 h^{-1} \|\varphi\|, \\ &\text{b)} \quad \|\varphi\|_{L^\infty} \leqq K_0 h^{-d/2} \|\varphi\|. \end{aligned} \tag{2.2}$$

We shall make one further restriction on $\mathcal{M}_h$ below in (2.10).

Restrict Ω as follows (with S denoting the collection of restrictions):

S: 1) The Neumann problem for $-\Delta + I$ on Ω is H^2-regular. (2.3)

2) The restricted cone property [1] holds on Ω; i.e., $\partial\Omega$ is Lipschitz.

The following regularity assumptions on a, b, c, f and the solution u of (1.1)–(1.2) are denoted collectively by R:

R: 1) There exist uniform constants c_*, c^*, a_*, a^*, and K_1 such that, for all $(x, t) \in \bar{Q}$ and $q \in \mathbb{R}$,

$$\begin{aligned} &\text{a)} \quad 0 < c_* \leqq c(x, q) \leqq c^*, \\ &\text{b)} \quad 0 < a_* \leqq a(x, q) \leqq a^*, \\ &\text{c)} \quad |b_i(x, q)| \leqq K_1, \qquad i = 1, \cdots, d, \\ &\text{d)} \quad |f(x, t, q)| \leqq K_1. \end{aligned} \tag{2.4}$$

2) The functions $a = a(x, u)$, $b_i = b_i(x, u)$, $c = c(x, u)$, and $f = f(x, t, u)$ are continuously differentiable with respect to u and have a uniform bound, K_1, for $(x, t) \in \bar{Q}$ and $q \in \mathbb{R}$:

$$|a|, |b_i|, |c|, |f|, \left|\frac{\partial a}{\partial u}\right|, \left|\frac{\partial b_i}{\partial u}\right|, \left|\frac{\partial c}{\partial u}\right|, \left|\frac{\partial f}{\partial u}\right|, \left|\frac{\partial f}{\partial t}\right|, \left|\frac{\partial^2 a}{\partial u^2}\right|, \left|\frac{\partial^2 b_i}{\partial u^2}\right|, \left|\frac{\partial^3 a}{\partial u^3}\right| \leqq K_1. \tag{2.5}$$

3) If

$$\|\varphi\|_{L^p((a,b);X)} \equiv \| \|\varphi(\cdot, t)\|_X \|_{L^p(a,b)}, \qquad 1 \leqq p \leqq \infty, \tag{2.6}$$

and u is the solution of (1.1)–(1.2), there exists a constant K_2 such that

$$\begin{aligned} \|u\|_{L^\infty(J;H^r)} + \left\|\frac{\partial u}{\partial t}\right\|_{L^2(J;H^{r-1})} + \left\|\frac{\partial u}{\partial t}\right\|_{L^\infty(J;H^2)} + \|u\|_{L^\infty(J;H^3)} \\ + \left\|\frac{\partial^2 u}{\partial t^2}\right\|_{L^\infty(J;H^1)} + \left\|\frac{\partial^3 u}{\partial t^3}\right\|_{L^2(J;L^2)} + \left\|\frac{\partial^3 u}{\partial t^3}\right\|_{L^1(J;H^1)} \leqq K_2. \end{aligned} \tag{2.7}$$

We note that under the hypotheses of the theorems and corollaries to follow, our approximations converge uniformly to u; thus (2.4) and (2.5) actually need hold only in a neighborhood of the solution.

Let $\Delta t > 0$, $N = T/\Delta t \in \mathbb{Z}$, and $t^\sigma = \sigma \Delta t$, $\sigma \in \mathbb{R}$. Also, let $\varphi^n \equiv \varphi^n(x) \equiv \varphi(x, t^n)$, $\varphi^{n+1/2} = (\varphi^{n+1} + \varphi^n)/2$, and $d_t \varphi^n = (\varphi^{n+1} - \varphi^n)/\Delta t$.

The analysis proceeds, following Wheeler [15] and Rachford [14], via an auxiliary elliptic problem. Define $W \in \mathcal{M}_h$ to be the unique function which, for $t \in \bar{J}$, satisfies

$$(a(\cdot, u(\cdot, t))\nabla[W(\cdot, t) - u(\cdot, t)], \nabla\chi) + ([W(\cdot, t) - u(\cdot, t)], \chi) = 0, \qquad \chi \in \mathcal{M}_h. \tag{2.8}$$

Thus W is a weighted H^1-projection of u, the solution of (1.1)–(1.2). As in [8], [9], (2.1) and the restrictions S imply the following result.

LEMMA 1. *For some p satisfying $2 \leqq p \leqq r$, let $u \in L^\infty(J; H^p)$ and $\partial u/\partial t \in L^2(J; H^{p-1})$. Under the assumption* R, *there exists a constant K_3, dependent on Ω, a_*, a^*, K_0, K_1, and K_2 such that if $\eta = u - W$ and $s = 0$ or 1,*

$$\text{(2.9)} \quad \begin{aligned} &\text{a)} \ \|\eta\|_{L^\infty(J;H^s)} \leqq K_3 h^{p-s} \|u\|_{L^\infty(J;H^p)}, \\ &\text{b)} \ \left\|\frac{\partial \eta}{\partial t}\right\|_{L^2(J;H^s)} \leqq K_3 h^{p-s} \left\{ \|u\|_{L^2(J;H^p)} + \left\|\frac{\partial u}{\partial t}\right\|_{L^2(J;H^p)} \right\}. \end{aligned}$$

We now make the assumption on $\{\mathcal{M}_h\}$ and u that there exists a constant K_4 such that

$$\text{(2.10)} \quad \|W\|_{L^\infty(J;L^\infty)} + \|\nabla W\|_{L^\infty(J;L^\infty)} + \left\|\frac{\partial W}{\partial t}\right\|_{L^\infty(J;L^\infty)} + \left\|\nabla \frac{\partial W}{\partial t}\right\|_{L^1(J;L^\infty)} \leqq K_4.$$

Given R and $\|\partial u/\partial t|_{L^1(J;H^3)} \leqq K_2$, a sufficient condition for (2.10) to hold is easily derivable from (2.9) and the following analogue of (2.1):

Assume that there exist an integer $r \geqq 2$ and a constant $K_0 > 0$ such that for $2 \leqq p \leqq r$ when $d = 1$ or 2 and for $3 \leqq p \leqq r$ when $d = 3$,

$$\text{(2.11)} \quad \begin{aligned} &\inf_{\chi \in \mathcal{M}_h} \{\|\varphi - \chi\| + h\|\varphi - \chi\|_1 + h^{d/2}(\|\varphi - \chi\|_{L^\infty} + h\|\nabla(\varphi - \chi)\|_{L^\infty})\} \\ &\qquad \leqq K_0 \|\varphi\|_p h^p, \qquad \varphi \in H^p(\Omega). \end{aligned}$$

The hypothesis (2.7) together with (2.9) and (2.11) imply that

$$\text{(2.12)} \quad \begin{aligned} &\text{a)} \ \|\eta\|_{L^\infty(J;L^\infty)} \leqq Ch^{2-(d/2)} \|u\|_{L^\infty(J;H^2)}, \\ &\text{b)} \ \|\nabla \eta\|_{L^\infty(J;L^\infty)} \leqq Ch^{2-(d/2)} \|u\|_{L^\infty(J;H^3)}, \\ &\text{c)} \ \left\|\frac{\partial \eta}{\partial t}\right\|_{L^\infty(J;L^\infty)} \leqq Ch^{2-(d/2)} \left\{ \|u\|_{L^\infty(J;H^2)} + \left\|\frac{\partial u}{\partial t}\right\|_{L^\infty(J;H^2)} \right\}, \\ &\text{d)} \ \left\|\nabla \frac{\partial \eta}{\partial t}\right\|_{L^1(J;L^\infty)} \leqq Ch^{2-(d/2)} \left\{ \|u\|_{L^1(J;H^3)} + \left\|\frac{\partial u}{\partial t}\right\|_{L^1(J;H^3)} \right\}. \end{aligned}$$

The relations in (2.12), together with (2.7) and the assumption that $\partial u/\partial t \in L^1(J; H^3)$, imply that the terms in (2.10) are bounded. We can also adapt the proofs of corresponding results in [4], [9] to obtain the following lemma.

LEMMA 2. *There exists a constant K_5 depending on K_0, K_1, and K_2 such that*

$$\text{(2.13)} \quad \left\|\frac{\partial^2 W}{\partial t^2}\right\|_{L^\infty(J;H^1)} + \left\|\frac{\partial^3 W}{\partial t^3}\right\|_{L^2(J;L^2)} + \left\|\frac{\partial^3 W}{\partial t^3}\right\|_{L^1(J;H^1)} \leqq K_5.$$

Let $u_h : [0, T] \to \mathcal{M}_h$ be the approximate solution of (1.1)–(1.2) determined by (using the notation $c(u) = c(x, u(x, t))$, etc.)

$$\text{(2.14)} \quad \left(c(u_h)\frac{\partial u_h}{\partial t}, \chi\right) + (a(u_h)\nabla u_h + b(u_h), \nabla \chi) = (f(u_h), \chi) + \langle g, \chi \rangle, \qquad \chi \in \mathcal{M}_h, \qquad t \in J,$$

with $u_h(\cdot, 0) - u_0$ small in a sense to be specified later. It can be shown [15] (in fact it follows from the results in § 4) that

$$\text{(2.15)} \quad \|u - u_h\|_{L^\infty(J;L^2)} + h\|u - u_h\|_{L^\infty(J;H^1)} \leqq C(u)h^r,$$

provided u satisfies (2.7). A standard Crank–Nicolson–Galerkin approximation to the solution of (2.14) would result in a time discretization error of the order $(\Delta t)^2$, but it would require the solution of a nonlinear system of algebraic equations at each time step. Rachford [14] has analyzed the following variant of the Crank–Nicolson–Galerkin scheme. Let $U: \{0 = t_0, t_1, \cdots, t_N = T\} \to \mathcal{M}_h$ satisfy

$$\begin{aligned} &\text{a)}\ (c(EU^n)\, d_t U^n, \chi) + (a(EU^n)\nabla U^{n+1/2} + b(EU^n), \nabla\chi) \\ &\qquad = (f(t^{n+1/2}, EU^n), \chi) + \langle g(t^{n+1/2}), \chi\rangle, \qquad \chi \in \mathcal{M}_h, \qquad 0 < n < N, \\ &\text{b)}\ (a(u_0)\nabla[u_0 - U^0], \nabla\chi) = 0, \qquad \chi \in \mathcal{M}_h, \end{aligned} \tag{2.16}$$

where $EU^n \equiv \frac{3}{2}U^n - \frac{1}{2}U^{n-1}$. With this definition of EU^n, we see that information is required at two preceding time levels to advance in time. Thus a starting procedure is needed to define U^1 which will retain the overall accuracy of the method. Such a starting procedure will be discussed in § 3. We note that the method given by (2.16) requires the solution of exactly one linear system of algebraic equations at each time step; however the matrices generated by the linear equations of (2.16) are usually different for each time step. We shall consider a modification of (2.16) which will require the solution of equations associated with one common matrix at all time levels. (Our process actually provides a generalization of and new error estimates for (2.16).)

3. Approximate solution of the linear equations by iteration. In this section we shall present the linear equations to be treated and an iterative method for approximating their solution. We also define a predictor-corrector-corrector starting procedure.

The conjugate gradient procedure presented here provides only one example of the possible modifications of (2.16) that fall under the convergence analysis given in the next section. Any method that provides the norm reduction defined in this section will preserve the results of § 4. Several modifications of the iterative methods presented here are discussed in § 5.

Let $\{\varphi_i\}_{i=1}^M$ be a basis for $\mathcal{M}_h$ and denote the solution of (2.16) by

$$U^m = \sum_{i=1}^{M} \xi_i^m \varphi_i. \tag{3.1}$$

Let

$$\begin{aligned} &\text{a)}\ C^m(\theta) = (c_{ij}^m(\theta)) = \left(\left(c\left(E\sum_{l=1}^{M}\theta_l^m\varphi_l\right)\varphi_j, \varphi_i\right)\right), \\ &\text{b)}\ A^m(\theta) = (a_{ij}^m(\theta)) = \left(\left(a\left(E\sum_{l=1}^{M}\theta_l^m\varphi_l\right)\nabla\varphi_j, \nabla\varphi_i\right)\right), \\ &\text{c)}\ B^m(\theta) = (b_i^m(\theta)) = \left(\left(b\left(E\sum_{l=1}^{M}\theta_l^m\varphi_l\right), \nabla\varphi_i\right)\right), \\ &\text{d)}\ F^m(\theta) = (f_i^m(\theta)) = \left(\left(f\left(t^{m+1/2}, E\sum_{l=1}^{M}\theta_l^m\varphi_l\right), \varphi_i\right)\right), \\ &\text{e)}\ G^m(\theta) = (G_i^m(\theta)) = (\langle g(t^{m+1/2}), \varphi_i\rangle), \\ &\text{f)}\ C_0 = ((c_0\varphi_j, \varphi_i)) \quad \text{and} \quad A_0 = ((a_0\nabla\varphi_j, \nabla\varphi_i)), \end{aligned} \tag{3.2}$$

for $i = 1, \cdots, M$ and $j = 1, \cdots, M$. Here a_0 and c_0 can be chosen in a very arbitrary way. A good choice might be $a_0 = a(x, u_0(x))$ and $c_0 = c(x, u_0(x))$ or if an average value $\bar{u}$ is more or less known, evaluate a and c at $\bar{u}$.

We can write (2.16a) in the form

$$L^n(\xi)(\xi^{n+1}-\xi^n)\equiv\Big(C^n(\xi)+\frac{\Delta t}{2}A^n(\xi)\Big)(\xi^{n+1}-\xi^n) \tag{3.3}$$
$$=-\Delta t A^n(\xi)\xi^n+\Delta t[B^n(\xi)+F^n(\xi)+G^n(\xi)].$$

We shall not solve (3.3) exactly; instead, we shall (in this section; see also § 5) use a predetermined number of preconditioned conjugate gradient [2], [3], [7], [10] iterations to advance the solution one time step. The preconditioning matrix will be chosen to be independent of n, and consequently an appropriate "fast Poisson" solution technique, such as nested dissection [11] can be applied to obtain the solution of the linear equations arising below in (3.7c).

Denote by

$$V^m=\sum_{i=1}^{M}\gamma_i^m\varphi_i \tag{3.4}$$

the approximation to U^m, the solution of (2.16). We shall discuss a starting procedure for obtaining V^0 and V^1 later. We now find γ^{n+1} (and thus V^{n+1}) using a preconditioned conjugate gradient iteration. Let our preconditioner be defined by

$$L_0\equiv C_0+\frac{\Delta t}{2}A_0. \tag{3.5}$$

We shall use different initial guesses for $\xi^{n+1}-\xi^n$ for $n=1$ and for $n>1$. We shall use linear extrapolation for $n=1$ and quadratic extrapolation for $n>1$ (defined explicitly in (3.6a)). Specifically, we initialize our iteration as follows:

$$\begin{aligned}&\text{a)}\quad n=1\colon x_0\equiv x_0^2=\gamma^1-\gamma^0,\\ &\qquad n\geqq 2\colon x_0\equiv x_0^{n+1}=2\gamma^n-3\gamma^{n-1}+\gamma^{n-2},\\ &\text{b)}\quad n\geqq 1\colon q_0\equiv q_0^{n+1}=s_0\equiv s_0^{n+1}=L^n(\gamma)x_0+\Delta t A^n(\gamma)\gamma^n-\Delta t[B^n(\gamma)\\ &\qquad\qquad +F^n(\gamma)+G^n(\gamma)].\end{aligned} \tag{3.6}$$

Then, using the initialization x_0, q_0 and s_0 from (3.6), for $k=1,2,\cdots,\nu-1$, where the number of iterations ν will be chosen later independently of n, set

$$\begin{aligned}&\text{a)}\quad x_{k+1}=x_k+\alpha_k s_k,\quad \text{where } \alpha_k=\frac{-(L_0^{-1}q_k,q_k)_e}{(s_k,L^n(\gamma)s_k)_e},\\ &\text{b)}\quad q_{k+1}=q_k+\alpha_k L^n(\gamma)s_k,\\ &\text{c)}\quad s_{k+1}=L_0^{-1}q_{k+1}+\beta_k s_k,\quad \text{where } \beta_k=\frac{(L_0^{-1}q_{k+1},q_{k+1})_e}{(L_0^{-1}q_k,q_k)_e},\end{aligned} \tag{3.7}$$

where $(\cdot,\cdot)_e$ is the Euclidean inner product.

Finally set

$$\gamma^{n+1}=\gamma^n+x_\nu. \tag{3.8}$$

We define $\bar\gamma^{n+1}$ to be the solution of (3.3) with ξ^n replaced by γ^n; i.e., let $\bar\gamma^{n+1}$ satisfy

$$L^n(\gamma)(\bar\gamma^{n+1}-\gamma^n)=-\Delta t A^n(\gamma)\gamma^n+\Delta t[B^n(\gamma)+F^n(\gamma)+G^n(\gamma)]. \tag{3.9}$$

It is well known [2], [3], [7], [10] that there exists a constant $\rho<1$ such that

$$\begin{aligned}&\text{a)}\quad \|L^1(\gamma)^{1/2}(\bar\gamma^2-\gamma^2)\|_e\leqq\rho\|L^1(\gamma)^{1/2}(\bar\gamma^2-2\gamma^1+\gamma^0)\|_e,\\ &\text{b)}\quad \|L^n(\gamma)^{1/2}(\bar\gamma^{n+1}-\gamma^{n+1})\|_e\leqq\rho\|L^n(\gamma)^{1/2}(\bar\gamma^{n+1}-3\gamma^n+3\gamma^{n-1}-\gamma^{n-2})\|_e,\end{aligned} \tag{3.10}$$

$n\geqq 2,$

where the subscript e indicates the Euclidean norm of the vector. Given c_0 and a_0 in (3.2f) there exist ψ_0 and ψ_1 such that, for $n \geqq 0$,

$$0 < \psi_0 \leqq \frac{x^T L^n(\gamma) x}{x^T L_0 x} \leqq \psi_1, \qquad 0 \neq x \in \mathbb{R}^M, \tag{3.11}$$

where the constants ψ_0 and ψ_1 are independent of h and depend only on the bounds for the coefficients in (2.4a) and (2.4b). Let

$$Q = \frac{1-(\psi_0/\psi_1)^{1/2}}{1+(\psi_0/\psi_1)^{1/2}}.$$

Then [2], [3], [7], [10] $\rho < 2Q^{\nu}$. If $\alpha > 0$ and

$$\nu \geqq \alpha \log \frac{1}{\Delta t} \Big/ \log \frac{1}{Q}, \tag{3.12}$$

then

$$\rho < 2(\Delta t)^{\alpha}. \tag{3.13}$$

Note that

$$\bar{V}^{n+1} = \sum_{i=1}^{M} \bar{\gamma}_i^{n+1} \varphi_i \tag{3.14}$$

satisfies

$$\begin{aligned} &\left(c(EV^n)\frac{\bar{V}^{n+1}-V^n}{\Delta t}, \chi\right) + (\tfrac{1}{2}a(EV^n)\nabla(\bar{V}^{n+1}+V^n) + b(EV^n), \nabla\chi) \\ &\quad = (f(t^{n+1/2}, EV^n), \chi) + \langle g(t^{n+1/2}), \chi\rangle, \qquad \chi \in \mathcal{M}_h. \end{aligned} \tag{3.15}$$

The following norms play an important role in our analysis:

$$\begin{aligned} &\text{a)} \ \|\varphi\|_{c^n}^2 \equiv (c(EV^n)\varphi, \varphi), \\ &\text{b)} \ \|\varphi\|_{a^n}^2 \equiv (\tfrac{1}{2}a(EV^n)\nabla\varphi, \nabla\varphi). \end{aligned} \tag{3.16}$$

By (2.4), $\|\cdot\|_{c^n}$ is equivalent to $\|\cdot\|$ and $\|\cdot\|_{a^n}$ is equivalent to $\|\nabla\cdot\|$ for each n. Thus $\|\cdot\|_{c^n}^2 + \|\cdot\|_{a^n}^2$ is equivalent to $\|\cdot\|_1^2$ for each n.

For future reference, we note that (3.10), the choice of ν given by (3.12) with $\alpha = 1$, and the triangle inequality yield the inequalities

$$\begin{aligned} &\text{a)} \ \|\bar{V}^2 - V^2\|_{c^1} + (\Delta t)^{1/2}\|\bar{V}^2 - V^2\|_{a^1} \leqq C_1 \Delta t\{\|\delta^2 V^1\|_{c^1} + (\Delta t)^{1/2}\|\delta^2 V^1\|_{a^1}\}, \\ &\text{b)} \ \|\bar{V}^{n+1} - V^{n+1}\|_{c^n} + (\Delta t)^{1/2}\|\bar{V}^{n+1} - V^{n+1}\|_{a^n} \\ &\qquad\qquad \leqq C_1 \Delta t\{\|\delta^3 V^n\|_{c^n} + (\Delta t)^{1/2}\|\delta^3 V^n\|_{a^n}\}, \qquad n \geqq 2, \end{aligned} \tag{3.17}$$

where

$$\begin{aligned} &\text{a)} \ \delta V^n = V^{n+1} - V^n, \\ &\text{b)} \ \delta^2 V^n = V^{n+1} - 2V^n + V^{n-1}, \\ &\text{c)} \ \delta^3 V^n = V^{n+1} - 3V^n + 3V^{n-1} - V^{n-2}. \end{aligned} \tag{3.18}$$

The convergence results of § 4 depend only on the norm reduction (3.10) which yields analogues of (3.17) for various choices of ρ and not on the particular iterative method used to achieve those norm reductions.

We shall now define a starting procedure which also uses the preconditioned conjugate gradient iteration. For V^0 we shall interpolate u_0 into $\mathcal{M}_h$ to obtain Iu_0. Then for some constant K_6 which depends upon $\|u_0\|_r$, we have

$$\|V^0 - W^0\| \leqq \|Iu_0 - u_0\| + \|u_0 - W^0\| \leqq K_6 h^r. \tag{3.19}$$

We note here that for Corollary 1 and Corollary 3 of § 4, in order to obtain the optimal order convergence through the predictor-corrector-corrector method to find V^1 to be described below, we must have the estimate

$$\|V^0 - W^0\|_1 \leqq K_6 h^r + (\Delta t)^2. \tag{3.20}$$

The method of computing V^0 to satisfy (3.20) would necessarily be more complicated than that described above. For example, if one is willing to factor one additional matrix, a V^0 satisfying (3.20) could be obtained by solving the equations generated by equation (2.8) at the initial time.

For V^1, we shall obtain an approximate solution of the following predictor-corrector-corrector Crank–Nicolson–Galerkin method by using the same preconditioned conjugate gradient iterative method as before. We first describe the exact equations for the Crank–Nicolson–Galerkin method. Let U^*, the prediction for U^1, be the unique solution of

$$\begin{aligned} &\left(c(U^0)\frac{U^* - U^0}{\Delta t}, \chi\right) + \left(\frac{1}{2}a(U^0)\nabla(U^* + U^0) + b(U^0), \nabla\chi\right) \\ &\quad = \langle g(t^{n+1/2}), \chi\rangle + (f(t^{n+1/2}, U^0), \chi), \qquad \chi \in \mathcal{M}_h. \end{aligned} \tag{3.21}$$

With $U^{*1/2} = (U^* + U^0)/2$, let U^{**} be the unique solution of

$$\begin{aligned} &\left(c(U^{*1/2})\frac{U^{**} - U^0}{\Delta t}, \chi\right) + \left(\frac{1}{2}a(U^{*1/2})\nabla(U^{**} + U^0) + b(U^{*1/2}), \nabla\chi\right) \\ &\quad = \langle g(t^{n+1/2}), \chi\rangle + (f(t^{n+1/2}, U^{*1/2}), \chi), \qquad \chi \in \mathcal{M}_h. \end{aligned} \tag{3.22}$$

With $U^{**1/2} = (U^{**} + U^0)/2$, let U^1 be the unique solution of

$$\begin{aligned} &\left(c(U^{**1/2})\frac{U^1 - U^0}{\Delta t}, \chi\right) + \left(\frac{1}{2}a(U^{**1/2})\nabla(U^1 + U^0) + b(U^{**1/2}), \nabla\chi\right) \\ &\quad = \langle g(t^{n+1/2}), \chi\rangle + (f(t^{n+1/2}, U^{**1/2}), \chi), \qquad \chi \in \mathcal{M}_h. \end{aligned} \tag{3.23}$$

Instead of solving (3.21)–(3.23) exactly, we shall approximate their solution by employing a preconditioned conjugate gradient method using the same preconditioning matrix L_0 defined in (3.5). If we iterate ν times, where ν is given by (3.12) with $\alpha = 1$ for the approximate solutions of (3.21)–(3.23), we obtain optimal order H^1 bounds. To obtain optimal order L^2 bounds, we choose ν as in (3.12) with $\alpha = 3/2$ for the approximations of the correctors (3.22) and (3.23). The following lemma can be proved using the arguments of § 4.

LEMMA 3. *Assume* S, R, (2.1), (2.2), *and* (2.10) *to hold. Let* ν *be given by* (3.12) *for* $\alpha = 1$. *Then, there exist positive constants* C_2 *and* τ *such that, if* $\Delta t \leqq \tau$,

$$\begin{aligned} &\text{a) } \|V^0 - W^0\| \leqq C_2 h^r, \\ &\text{b) } \|V^1 - W^1\|_1 + (\Delta t)^{1/2}\|d_t(V^0 - W^0)\| \leqq C_2\{(\Delta t)^2 + h^{r-1}\}, \end{aligned} \tag{3.24}$$

where C_2 *depends on* a_*, a^*, c_*, c^*, $K_0, \cdots, K_5$, *and* $\|u_0\|_r$.

4. A priori error estimates. In this section we develop *a priori* bounds for the error $V^n - u^n$ for the procedures defined in § 3. The first result, Theorem 1, states that optimal order H^1 estimates can be obtained for $r \geqq 3$ (piecewise quadratics or better) if the iterative process reduces the error in the solution of the algebraic problem by a factor proportional to Δt ($\rho = O(\Delta t)$) at each time step. The second result, Corollary 1, points out that with additional smoothness restrictions on $\partial u/\partial t$, the choice $\rho = O(\Delta t)$ is also sufficient to obtain optimal order L^2 estimates for $r \geqq 2$. If $\Delta t = O(h^{r/2})$, then the optimal order rate also applies in H^1 for $r \geqq 2$. The third result, Corollary 2, says that, if $r \geqq 3$ and $\Delta t \leqq Ch^2$, then, to obtain optimal order H^1 bounds, the error at each time step need only be reduced by a fixed (sufficiently small) factor that is independent of Δt and h. The next result, Corollary 3, extends this result to optimal order L^2 bounds for $r \geqq 2$ with additional smoothness on $\partial u/\partial t$. Theorem 2 shows that for $\rho = O((\Delta t)^{1/2})$, we can weaken the regularity assumptions used on u in Corollary 3 and still achieve optimal order L^2 bounds for $r \geqq 3$.

THEOREM 1. *Let* S *and* R *and the restrictions on* $\{\mathcal{M}_h\}$ *of* § 2 *hold. Let* V^n *satisfy* (3.17) *and* (3.24). *Then there exist constants* τ *and* C_3, *where* C_3 *is dependent on the constants in* R, K_0, K_3, K_4, K_5 *and* C_2, *such that, if* $r \geqq 3$, $\Delta t \leqq \tau$, *and* $\Delta t \leqq h^{d/3}$,

$$\sup_{t^n} \|u - V\|_1 \leqq C_3\{(\Delta t)^2 + h^{r-1}\}. \tag{4.1}$$

Proof. Letting $\zeta^n = V^n - W^n$, we see that

$$\begin{aligned}
&(c(EV^n)\, d_t\zeta^n, \chi) + (a(EV^n)\nabla\zeta^{n+1/2}, \nabla\chi) \\
&\quad = -([c(EV^n) - c(u(t^{n+1/2}))]\, d_tW^n, \chi) - \left(c(u(t^{n+1/2}))\left[d_tW^n - \frac{\partial u}{\partial t}\right], \chi\right) \\
&\qquad + ([a(u(t^{n+1/2}))\nabla W(t^{n+1/2}) - a(EV^n)\nabla W^{n+1/2}], \nabla\chi) \\
&\qquad - (\eta(t^{n+1/2}), \chi) + ([b(u(t^{n+1/2})) - b(EV^n)], \nabla\chi) \\
&\qquad + ([f(t^{n+1/2}, EV^n) - f(t^{n+1/2}, u(t^{n+1/2}))], \chi) \\
&\qquad + \left(c(EV^n)\frac{V^{n+1} - \bar{V}^{n+1}}{\Delta t}, \chi\right) + \frac{1}{2}(a(EV^n)\nabla(V^{n+1} - \bar{V}^{n+1}), \nabla\chi), \quad \chi \in \mathcal{M}_h.
\end{aligned} \tag{4.2}$$

We shall obtain estimates in the H^1-norm by using $\chi = \zeta^{n+1} - \zeta^n = \Delta t\, d_t\zeta^n$ as a test function. Clearly

$$\begin{aligned}
&(c(EV^n)\, d_t\zeta^n, \zeta^{n+1} - \zeta^n) + (a(EV^n)\nabla\zeta^{n+1/2}, \nabla(\zeta^{n+1} - \zeta^n)) \\
&\quad = \Delta t\|d_t\zeta^n\|^2_{c^n} + \|\zeta^{n+1}\|^2_{a^n} - \|\zeta^n\|^2_{a^n}.
\end{aligned} \tag{4.3}$$

We use the assumptions in R of § 2 to obtain estimates for the first two terms on the right hand side of (4.2). Thus,

$$\begin{aligned}
&\Delta t|([c(EV^n) - c(Eu^n) + c(Eu^n) - c(u(t^{n+1/2}))]\; d_tW^n, d_t\zeta^n)| \\
&\qquad + \Delta t\left|\left(c(u(t^{n+1/2}))\left[d_t\eta^n + d_tu^n - \frac{\partial u}{\partial t}(t^{n+1/2})\right], d_t\zeta^n\right)\right| \\
&\quad \leqq C_4\,\Delta t\{\|\zeta^n\|^2 + \|\zeta^{n-1}\|^2 + \|\eta^n\|^2 + \|\eta^{n-1}\|^2 + \|d_t\eta^n\|^2\} \\
&\qquad + \varepsilon_5\,\Delta t\|d_t\zeta^n\|^2_{c^n} + C(\Delta t)^4\sigma_n,
\end{aligned} \tag{4.4}$$

where

$$\sigma_n = \int_{t^{n-1}}^{t^{n+1}} \left(\left\| \frac{\partial^3 u}{\partial t^3}(\cdot, s) \right\|^2 + \left\| \frac{\partial^2 u}{\partial t^2}(\cdot, s) \right\|^2 \right) ds = \sigma_{1,n} + \sigma_{2,n}. \tag{4.5}$$

From (2.7), $\sum_{n=1}^{N-1} \sigma_n < C$. We note that C_4 depends on c^*, K_1 and the bound K_4 for $\|\partial W/\partial t\|_{L^\infty(J;L^\infty)}$. The fourth and sixth terms on the right hand side of (4.2) can be treated similarly. We shall now estimate the last two terms on the right of (4.2) using (3.17). We first note that from (2.13), we have

$$\begin{aligned} &\text{a) } \|\delta^2 W^1\|_1 \leqq C(\Delta t)^2, \\ &\text{b) } \|\delta^3 W^n\|_1 \leqq C(\Delta t)^2 \int_{t^{n-2}}^{t^{n+1}} \left\| \frac{\partial^3 W}{\partial t^3}(\cdot, s) \right\|_1 ds. \end{aligned} \tag{4.6}$$

The constants appearing here will then sum in the proper fashion to achieve the desired results of the theorem. Since different starting procedures were used in the conjugate gradient iteration to obtain V^2 and V^m for $m \geqq 3$, we shall estimate each case separately. From (3.17) we see that for $n = 1$,

$$\begin{aligned} &\left| \left(c(EV^1) \frac{V^2 - \bar{V}^2}{\Delta t}, \zeta^2 - \zeta^1 \right) + \frac{1}{2}(a(EV^1)\nabla(V^2 - \bar{V}^2), \nabla(\zeta^2 - \zeta^1)) \right| \\ &\quad \leqq \|V^2 - \bar{V}^2\|_{c^1} \|d_t \zeta^1\|_{c^1} + \|V^2 - \bar{V}^2\|_{a^1} \|\zeta^2 - \zeta^1\|_{a^1} \\ &\quad \leqq C\,\Delta t \{\|\delta^2 V^1\|_{c^1} + (\Delta t)^{1/2} \|\delta^2 V^1\|_{a^1}\} \|d_t \zeta^1\|_{c^1} \\ &\qquad + C(\Delta t)^{1/2} \{\|\delta^2 V^1\|_{c^1} + (\Delta t)^{1/2} \|\delta^2 V^1\|_{a^1}\} \{\|\zeta^2\|_{a^1} + \|\zeta^1\|_{a^1}\} \\ &\quad \leqq C\{\Delta t[\|d_t \zeta^1\|_{c^1} + \|d_t \zeta^0\|_{c^0}] + (\Delta t)^2 + (\Delta t)^{1/2}[\|\zeta^2\|_1 + \|\zeta^1\|_1 + \|\zeta^0\|_1]\} \\ &\qquad \cdot \{\Delta t \|d_t \zeta^1\|_{c^1} + (\Delta t)^{1/2}[\|\zeta^2\|_1 + \|\zeta^1\|_1]\} \\ &\quad \leqq \varepsilon_5\, \Delta t \{\|d_t \zeta^1\|_{c^1}^2 + \|d_t \zeta^0\|_{c^0}^2\} + C\{(\Delta t)^4 + \Delta t[\|\zeta^2\|_1^2 + \|\zeta^1\|_1^2 + \|\zeta^0\|_1^2]\}. \end{aligned} \tag{4.7}$$

From (3.17) we obtain the estimate for $n \geqq 2$ in a similar manner. For $n \geqq 2$,

$$\begin{aligned} &\left| \left(c(EV^n) \frac{V^{n+1} - \bar{V}^{n+1}}{\Delta t}, \zeta^{n+1} - \zeta^n \right) + \frac{1}{2}(a(EV^n)\nabla(V^{n+1} - \bar{V}^{n+1}), \nabla(\zeta^{n+1} - \zeta^n)) \right| \\ &\quad \leqq \varepsilon_5\, \Delta t \{\|d_t \zeta^n\|_{c^n}^2 + \|d_t \zeta^{n-1}\|_{c^{n-1}}^2 + \|d_t \zeta^{n-2}\|_{c^{n-2}}^2\} + C(\Delta t)^4[\sigma_{3,n}^2 + \sigma_{4,n}] \\ &\qquad + C_5\, \Delta t \{\|\zeta^{n+1}\|_1^2 + \|\zeta^n\|_1^2 + \|\zeta^{n-1}\|_1^2 + \|\zeta^{n-2}\|_1^2\}, \end{aligned} \tag{4.8}$$

where

$$\begin{aligned} &\text{a) } \sigma_{3,n} = \int_{t^{n-2}}^{t^{n+1}} \left\| \frac{\partial^3 W}{\partial t^3}(\cdot, s) \right\|_1 ds, \\ &\text{b) } \sigma_{4,n} = \int_{t^{n-2}}^{t^{n+1}} \left\| \frac{\partial^3 W}{\partial t^3}(\cdot, s) \right\|^2 ds. \end{aligned} \tag{4.9}$$

The third term on the right side of (4.2) can be split in the following manner:

$$\begin{aligned} &([a(u(t^{n+1/2}))\nabla W(t^{n+1/2}) - a(EV^n)\nabla W^{n+1/2}], \nabla \chi) \\ &\quad = (a(u(t^{n+1/2}))[\nabla W(t^{n+1/2}) - \nabla W^{n+1/2}], \nabla \chi) \\ &\quad + (\nabla W^{n+1/2}[a(u(t^{n+1/2})) - a(EV^n)]\nabla \chi) \\ &\quad = A_{1,n} + A_{2,n}. \end{aligned} \tag{4.10}$$

We shall use summation by parts in time to treat $A_{1,n}$. We see that

$$
\begin{aligned}
\left|\sum_{n=1}^{l-1} A_{1,n}\right| &\leqq \Bigg|\sum_{n=2}^{l-1} (\{a(t^{n+1/2})\nabla[W(t^{n+1/2}) - W^{n+1/2}] \\
&\qquad - a(t^{n-1/2})\nabla[W(t^{n-1/2}) - W^{n-1/2}]\}, \nabla\zeta^n)\Bigg| \\
&\quad + |(a(u(t^{3/2}))\nabla[W(t^{3/2}) - W^{3/2}], \nabla\zeta^1)| \\
&\quad + |(a(u(t^{l-1/2}))\nabla[W(t^{l-1/2}) - W^{l-1/2}], \nabla\zeta^l)| \\
&\leqq \left|\sum_{n=2}^{l-1} \Delta t\left(\left[\frac{a(u(t^{n+1/2})) - a(u(t^{n-1/2}))}{\Delta t}\right]\nabla[W(t^{n+1/2}) - W^{n+1/2}], \nabla\zeta^n\right)\right| \\
&\quad + \left|\sum_{n=2}^{l-1} \Delta t\left(a(u(t^{n-1/2}))\nabla\left[\frac{W(t^{n+1/2}) - W^{n+1/2} - \{W(t^{n-1/2}) - W^{n-1/2}\}}{\Delta t}\right], \nabla\zeta^n\right)\right| \\
&\quad + C\{\|\zeta^1\|_1^2 + (\Delta t)^4\} + \frac{1}{20}\|\zeta^l\|_{a^{l-1}}^2 \\
&\leqq C_6\left\{\|\zeta^1\|_1^2 + (\Delta t)^4 + \Delta t \sum_{n=1}^{l-1} \|\zeta^n\|_1^2\right\} + C\sum_{n=2}^{l-1} \sigma_{3,n}\|\zeta^n\|_1^2 + \frac{1}{20}\|\zeta^l\|_{a^{l-1}}^2.
\end{aligned} \tag{4.11}
$$

We note that C_6 depends on a^*, K_1, K_4, and K_5. We next sum by parts to estimate $A_{2,n}$ from (4.10).

$$
\begin{aligned}
\left|\sum_{n=1}^{l-1} A_{2,n}\right| &\leqq \Bigg|\sum_{n=2}^{l-1} (\{[a(u(t^{n+1/2})) - a(EV^n)]\nabla W^{n+1/2} \\
&\qquad - [a(u(t^{n-1/2})) - a(EV^{n-1})]\nabla W^{n-1/2}\}, \nabla\zeta^n)\Bigg| \\
&\quad + |([a(u(t^{3/2})) - a(EV^1)]\nabla W^{3/2}, \nabla\zeta^1)| \\
&\quad + |([a(u(t^{l-1/2})) - a(EV^{l-1})]\nabla W^{l-1/2}, \nabla\zeta^l)| \\
&\leqq C_7\{\|\zeta^1\|_1^2 + \|\zeta^0\|^2 + \|\zeta^{l-2}\|^2 + \|\zeta^{l-1}\|^2 + \|\eta^0\|^2 + \|\eta^1\|^2 + \|\eta^{l-2}\|^2 + \|\eta^{l-1}\|^2 + (\Delta t)^4\} \\
&\quad + \frac{1}{20}\|\zeta^l\|_{a^{l-1}}^2 + \left|\sum_{n=2}^{l-1} ([a(u(t^{n+1/2})) - a(EV^n)][\nabla W^{n+1/2} - \nabla W^{n-1/2}], \nabla\zeta^n)\right| \\
&\quad + \Bigg|\sum_{n=2}^{l-1} (\nabla W^{n-1/2}[a(u(t^{n+1/2})) - a(EV^n) \\
&\qquad - \{a(u(t^{n-1/2})) - a(EV^{n-1})\}], \nabla\zeta^n)\Bigg|.
\end{aligned} \tag{4.12}
$$

We then bound the next to the last term on the right of (4.12) as follows:

$$
\begin{aligned}
&\left|\sum_{n=2}^{l-1} ([a(u(t^{n+1/2})) - a(EV^n)][\nabla W^{n+1/2} - \nabla W^{n-1/2}], \nabla\zeta^n)\right| \\
&\qquad \leqq C\sum_{n=1}^{l-1} [\|\zeta^n\|_1^2 + \|\eta^n\|^2 + (\Delta t)^4]\sigma_{5,n},
\end{aligned} \tag{4.13}
$$

where

$$
\sigma_{5,n} = \int_{t^{n-1}}^{t^{n+1}} \left\|\nabla \frac{\partial W}{\partial t}(\cdot, s)\right\|_{L^\infty} ds.
$$

Note that $\sum_{n=1}^{l-1} \sigma_{5,n} \leqq 2K_4$ by (2.10). Next define

$$\text{(4.14)}\quad \begin{aligned} &\text{a)}\ a'_{1,n}(x) = \int_0^1 \frac{\partial a}{\partial u}(x, \theta u(t^{n+1/2}) + (1-\theta)u(t^{n-1/2}))\, d\theta, \\ &\text{b)}\ a'_{2,n}(x) = \int_0^1 \frac{\partial a}{\partial u}(x, \theta EV^n + (1-\theta)EV^{n-1})\, d\theta. \end{aligned}$$

We can now treat the last term on the right of (4.12) in the following manner:

$$\text{(4.15)}\quad \begin{aligned} &\left| \sum_{n=2}^{l-1} (\nabla W^{n-1/2}[a(u(t^{n+1/2})) - a(u(t^{n-1/2})) - \{a(EV^n) - a(EV^{n-1})\}], \nabla\zeta^n) \right| \\ &= \left| \sum_{n=2}^{l-1} (\nabla W^{n-1/2}[a'_{1,n}\{u(t^{n+1/2}) - u(t^{n-1/2})\} - a'_{2,n}\{E\,\delta V^{n-1}\}], \nabla\zeta^n) \right| \\ &\leqq \left| \sum_{n=2}^{l-1} \left(\left\{ \nabla W^{n-1/2}[a'_{1,n} - a'_{2,n}] \left[\frac{u(t^{n+1/2}) - u(t^{n-1/2})}{\Delta t} \right] \Delta t + \nabla W^{n-1/2} \right. \right. \right. \\ &\quad \left. \left. \left. \cdot a'_{2,n}[u(t^{n+1/2}) - u(t^{n-1/2}) - Eu^n + Eu^{n-1} + E\,\delta\eta^{n-1} - E\,\delta\zeta^{n-1}] \right\}, \nabla\zeta^n \right) \right| \\ &\leqq C\,\Delta t \sum_{n=0}^{l-1} \{\|\zeta^n\|_1^2 + \|\eta^n\|^2 + \|d_t\eta^n\|^2\} + \varepsilon_5\,\Delta t \sum_{n=1}^{l-1} \|d_t\zeta^n\|_{c^n}^2 \\ &\quad + C\,\Delta t\|d_t\zeta^0\|^2 + C_8(\Delta t)^4. \end{aligned}$$

Clearly bounds for the fifth term on the right of (4.2) can be obtained as above. By choosing $\varepsilon_5 < 1/32$ and $\Delta t < (20C_5)^{-1}$, and combining the above bounds, we see that

$$\text{(4.16)}\quad \begin{aligned} &\frac{3\,\Delta t}{4} \sum_{n=1}^{l-1} \|d_t\zeta^n\|_{c^n}^2 + \sum_{n=1}^{l-1} \{\|\zeta^{n+1}\|_{a^n}^2 - \|\zeta^n\|_{a^n}^2\} \\ &\quad \leqq \tfrac{1}{2}\|\zeta^l\|_{a^{l-1}}^2 + C\{\|\zeta^1\|_1^2 + \|\zeta^0\|^2 + \Delta t\|d_t\zeta^0\|^2\} + C_7\{\|\zeta^{l-1}\|^2 + \|\zeta^{l-2}\|^2\} \\ &\qquad + C \sum_{n=1}^{l-1} (\Delta t + \sigma_{3,n} + \sigma_{5,n})\|\zeta^n\|_1^2 \\ &\qquad + C\{\|\eta\|_{L^\infty(J;L^2)}^2 + \|d_t\eta\|_{L^2(J;L^2)}^2 + (\Delta t)^4\}. \end{aligned}$$

Note that

$$\text{(4.17)}\quad \|\zeta^{n+1}\|^2 - \|\zeta^n\|^2 = 2\,\Delta t(d_t\zeta^n, \zeta^n) + (\Delta t)^2\|d_t\zeta^n\|^2 \leqq \varepsilon\,\Delta t\|d_t\zeta^n\|^2 + C\,\Delta t\|\zeta^n\|^2.$$

Sum this inequality from $n = 1$ to the upper limits $l-3$, $l-2$, and $l-1$; then multiply the resulting inequalities by $C_7 + \frac{1}{2}$ and add them to (4.16), after choosing ε so that

$$\text{(4.18)}\quad 3\varepsilon\left(C_7 + \frac{1}{2}\right) \sum_{n=1}^{l-1} \|d_t\zeta^n\|^2 \leqq \frac{1}{4} \sum_{n=1}^{l-1} \|d_t\zeta^n\|_{c^n}^2.$$

Thus,

$$\text{(4.19)}\quad \begin{aligned} &\frac{\Delta t}{2} \sum_{n=1}^{l-1} \|d_t\zeta^n\|_{c^n}^2 + \left(C_7 + \frac{1}{2}\right)\{\|\zeta^l\|^2 + \|\zeta^{l-1}\|^2 + \|\zeta^{l-2}\|^2 - 3\|\zeta^1\|^2\} \\ &\quad + \sum_{n=1}^{l-1} \{\|\zeta^{n+1}\|_{a^n}^2 - \|\zeta^n\|_{a^n}^2\} \leqq \frac{1}{2}\|\zeta^l\|_{a^{l-1}}^2 + C_7\{\|\zeta^{l-1}\|^2 + \|\zeta^{l-2}\|^2\} \\ &\quad + C\{\|\zeta^1\|_1^2 + \|\zeta^0\|^2 + \Delta t\|d_t\zeta^0\|^2\} + C \sum_{n=1}^{l-1} (\Delta t + \sigma_{3,n} + \sigma_{5,n})\|\zeta^n\|_1^2 \\ &\quad + C\{\|\eta\|_{L^\infty(J;L^2)}^2 + \|d_t\eta\|_{L^2(J;L^2)}^2 + (\Delta t)^4\}. \end{aligned}$$

We shall now use ideas of Douglas [5] and Rachford [14] to establish comparability between $\|\zeta^n\|^2_{a^n}$ and $\|\zeta^n\|^2_{a^{n-1}}$ to obtain telescoping sums on the left side of (4.19). Note that

$$\begin{aligned}\|\zeta^n\|^2_{a^n} &= \|\zeta^n\|^2_{a^{n-1}} + ([a(EV^n) - a(EV^{n-1})]\nabla\zeta^n, \nabla\zeta^n) \\ &= \|\zeta^n\|^2_{a^{n-1}} + \left(\left[\frac{\partial a}{\partial u}E(\zeta^n - \zeta^{n-1}) + \frac{\partial a}{\partial u}E(W^n - W^{n-1})\right]\nabla\zeta^n, \nabla\zeta^n\right) \\ &\leqq \|\zeta^n\|^2_{a^{n-1}} + C\{\|\delta\zeta^{n-1}\|_{L^\infty} + \|\delta\zeta^{n-2}\|_{L^\infty} + \Delta t\}\|\zeta^n\|^2_1.\end{aligned} \tag{4.20}$$

Thus, as in [13],

$$\begin{aligned}&\sum_{n=2}^{l-1}(\|\zeta^{n+1}\|^2_{a^n} - \|\zeta^n\|^2_{a^{n-1}}) \\ &\leqq \sum_{n=2}^{l-1}(\|\zeta^{n+1}\|^2_{a^n} - \|\zeta^n\|^2_{a^n}) + C\sum_{n=2}^{l-1}(\|\delta\zeta^{n-1}\|_{L^\infty} + \|\delta\zeta^{n-2}\|_{L^\infty} + \Delta t)\|\zeta^n\|^2_1.\end{aligned} \tag{4.21}$$

Next, if $u \in L^\infty(J; H^r)$ and $(\partial u/\partial t) \in L^2(J; H^{r-1})$, then (2.9), (3.24), (4.19), and (4.21) imply that

$$\begin{aligned}&\|\zeta^l\|^2_1 + \|\zeta^{l-1}\|^2 + \|\zeta^{l-2}\|^2 + \Delta t\sum_{n=1}^{l-1}\|d_t\zeta^n\|^2 \\ &\leqq C_{10}\sum_{n=1}^{l-1}(\Delta t + \sigma_{3,n} + \sigma_{5,n})\|\zeta^n\|^2_1 + C_{11}\sum_{n=2}^{l-1}(\|\delta\zeta^{n-1}\|_{L^\infty} + \|\delta\zeta^{n-2}\|_{L^\infty})\|\zeta^n\|^2_1 \\ &\quad + C_{12}\{h^{2r-2} + (\Delta t)^4\}.\end{aligned} \tag{4.22}$$

We note, for example, that C_{12} depends upon a^*, c^*, K_1, K_3, and bounds on

$$\|W_t\|_{L^\infty(J;L^\infty)}, \quad \|\nabla W|_{L^\infty(j;L^\infty)}, \quad \left\|\frac{\partial^2 u}{\partial t^2}\right\|_{L^\infty(J;H^1)}, \quad \left\|\frac{\partial^2 u}{\partial t^3}\right\|_{L^2(J;L^2)}, \quad \left\|\frac{\partial^3 u}{\partial t^3}\right\|_{L^1(J;H^1)},$$

$$\|u\|_{L^\infty(J;H^r)}, \quad \text{and} \quad \left\|\frac{\partial u}{\partial t}\right\|_{L^2(J;H^{r-1})}.$$

In order to apply the discrete Gronwall lemma to (4.22), we wish to show that there exists $C_{13} > 0$ such that

$$\sum_{n=0}^{l-2}\|\delta\zeta^n\|_{L^\infty} < C_{13}. \tag{4.23}$$

The predictor-corrector-corrector starting method yields

$$\|\delta\zeta^0\|_{L^\infty} < C_{14}. \tag{4.24}$$

We shall use an induction argument as in [14] to yield (4.23) with the summation starting at $n = 1$. For $l = 2$, the inequality (4.22) and the estimate (3.24) imply that

$$\begin{aligned}\Delta t\|d_t\zeta^1\|^2 &\leqq C_9\,\Delta t\|\zeta^1\|^2_1 + C_{10}(K_4 + K_5)\|\zeta^1\|^2_1 + C_{12}\{h^{2r-2} + (\Delta t)^4\} \\ &\leqq C_{15}\{(\Delta t)^4 + h^{2r-2}\}.\end{aligned} \tag{4.25}$$

If

$$\text{(4.26)}\qquad \text{a) } h \leqq (2C_{15})^{-3/d},\quad \text{b) } \Delta t \leqq h^{d/3},\quad \text{c) } r \geqq 3 \geqq \tfrac{2}{3}d+1,$$

then

$$\text{(4.27)}\qquad \Delta t \|d_t \zeta^1\|^2 \leqq h^d.$$

Assume the following induction hypothesis:

$$\text{(4.28)}\qquad \Delta t \sum_{n=1}^{k} \|d_t \zeta^n\|^2 \leqq h^d \quad \text{for } 1 \leqq k \leqq l-2.$$

If we use the inverse hypothesis (2.2c) assumed satisfied by $\mathcal{M}_h$, (4.28) and the fact that $N\,\Delta t = T$, we see that

$$\text{(4.29)}\qquad \begin{aligned} \sum_{n=1}^{l-2} \|\delta\zeta^n\|_{L^\infty} &\leqq (l-2)^{1/2}\Big(\sum_{n=1}^{l-2} \|\delta\zeta^n\|_{L^\infty}^2\Big)^{1/2} \\ &\leqq N^{1/2} K_0 h^{-d/2}\Big(\sum_{n=1}^{l-2} \|\delta\zeta^n\|^2\Big)^{1/2} \\ &\leqq (\Delta t)^{-1/2} T^{1/2} K_0 h^{-d/2} (\Delta t)^{1/2}\Big(\Delta t \sum_{n=1}^{l-2} \|d_t\zeta^n\|^2\Big)^{1/2} \\ &\leqq T^{1/2} K_0. \end{aligned}$$

Then, with $C_{13} = T^{1/2}K_0$, we apply the discrete Gronwall lemma in (4.22) to obtain

$$\text{(4.30)}\qquad \|\zeta^l\|_1^2 + \Delta t \sum_{n=1}^{l-1} \|d_t\zeta^n\|^2 \leqq C_{16}\{(\Delta t)^4 + h^{2r-2}\}$$

where

$$\text{(4.31)}\qquad C_{16} \leqq C_{12} \exp\{C_9 T + C_{10}[K_4 + K_5] + 2C_{11} T^{1/2} K_0\}.$$

Then with (4.26a) replaced by $h \leqq (2C_{16})^{-3/d}$ we see that our induction argument is completed. Since (4.30) holds for each l from $l = 1$ to $l = N$, we have

$$\text{(4.32)}\qquad \sup_{t^n} \|V - W\|_1 \leqq C_{17}\{(\Delta t)^2 + h^{r-1}\}.$$

Then for $u \in L^\infty(J; H^r)$, (2.9) and the triangle inequality yield the desired result (4.1). □

Note that if $\partial u/\partial t \in L^2(J; H^r)$ and if V^0 and V^1 are determined such that

$$\text{(4.33)}\qquad \|V^0 - W^0\|_1 + \|V^1 - W^1\|_1 + (\Delta t)^{1/2}\|d_t(V^0 - W^0)\| \leqq C_2\{(\Delta t)^2 + h^r\},$$

then by using (2.9), h^{2r-2} can be replaced by h^{2r} in the above proof. In this case it suffices to assume that $r \geqq 2 \geqq \frac{2}{3}d$.

COROLLARY 1. *Let all the hypotheses of Theorem* 1 *except the assumption on r be satisfied. Assume that* $r \geqq 2$, (4.33) *is satisfied, and* $(\partial u/\partial t) \in L^2(J; H^r)$. *Then there exist constants* τ *and* C_{18} *such that if* $\Delta t \leqq \tau$ *and* $\Delta t \leqq h^{d/3}$, *then*

$$\sup_{t^n} \{\|u - V\| + h\|u - V\|_1\} \leqq C_{18}\{(\Delta t)^2 + h^r\}.$$

The constant C_{18} has the same dependencies as C_3 with the addition of a bound on $\|\partial u/\partial t\|_{L^\infty(J;H^r)}$.

Next, we show that a further restriction on Δt can reduce the number of iterations for $n \geqq 2$ from (3.12) to a number independent of Δt. Assume that

$$\Delta t \leqq C_{19} h^2. \tag{4.34}$$

We note that (4.34) is really no restriction if $\mathcal{M}_h$ is a space of piecewise cubic (or higher order) polynomials ($r \geqq 4$) since the optimal choice of Δt is $O(h^{r/2})$ for such choices of $\mathcal{M}_h$. If the quadratic extrapolation given by (3.6b) is used to initialize the iteration for $n \geqq 2$, then (2.2a), (3.10), (3.17), and (3.18) imply that

$$\begin{aligned}
&\left|\left(c(EV^n)\frac{V^{n+1}-\bar V^{n+1}}{\Delta t}, \zeta^{n+1}-\zeta^n\right)\right.\\
&\quad\left.+\frac{1}{2}(a(EV^n)\nabla(V^{n+1}-\bar V^{n+1}), \nabla(\zeta^{n+1}-\zeta^n))\right|\\
&\quad\leqq \|V^{n+1}-\bar V^{n+1}\|_{c^n}\|d_t\zeta^n\|_{c^n}+\|V^{n+1}-\bar V^{n+1}\|_{a^n}\|d_t\zeta^n\|_{a^n}\,\Delta t\\
&\quad\leqq C_1\rho\{\|\delta^3 V^n\|_{c^n}+(\Delta t)^{1/2}\|\delta^3 V^n\|_{a^n}\}\{\|d_t\zeta^n\|_{c^n}+(\Delta t)^{1/2}\|d_t\zeta^n\|_{a^n}\}\\
&\quad\leqq C_1\rho\left(1+\frac{a^*K_0(\Delta t)^{1/2}}{c_* h}\right)^2\{\Delta t\|\delta^2 d_t\zeta^n\|_{c^n}+\|\delta^3 W^n\|_{c^n}\}\|d_t\zeta^n\|_{c^n} \qquad (4.35)\\
&\quad\leqq C_1\rho\left(1+\frac{a^*K_0C_{19}}{c_*}\right)^2\{\|\delta^2 d_t\zeta^n\|_{c^n}+C\,\Delta t\sigma_{6,n}\}\,\Delta t\|d_t\zeta^n\|_{c^n}\\
&\quad\leqq \Delta t C_1\rho\left(1+\frac{a^*K_0C_{19}}{c_*}\right)^2\frac{c^*}{c_*}\{\|d_t\zeta^n\|_{c^n}+2\|d_t\zeta^{n-1}\|_{c^{n-1}}+\|d_t\zeta^{n-2}\|_{c^{n-2}}\}\|d_t\zeta^n\|_{c^n}\\
&\qquad+C(\Delta t)^2\sigma_{6,n}\|d_t\zeta^n\|_{c^n},
\end{aligned}$$

where $\sigma_{6,n}=\int_{t^{n-2}}^{t^{n+1}}\|\partial^3 W/\partial t^3(\cdot, s)\|\,ds$. Iterate sufficiently many times such that

$$\rho<\left\{8\frac{C_1c^*}{c_*}\left(1+\frac{a^*K_0C_{19}}{c_*}\right)\right\}^{-1}\equiv K_8, \tag{4.36}$$

where

$$\rho\leqq\rho_1=2Q^\nu. \tag{4.37}$$

Then for some $\varepsilon_7>0$,

$$\begin{aligned}
&\left|\left(c(EV^n)\frac{V^{n+1}-\bar V^{n+1}}{\Delta t}, \zeta^{n+1}-\zeta^n\right)+\frac{1}{2}(a(EV^n)\nabla(V^{n+1}-\bar V^{n+1}), \nabla(\zeta^{n+1}-\zeta^n))\right|\\
&\quad\leqq C(\Delta t)^4\sigma_{4,n}+\left(\varepsilon_7+\frac{9}{32}\right)\Delta t\|d_t\zeta^n\|_{c^n}^2+\frac{\Delta t}{8}\{\|d_t\zeta^{n-1}\|_{c^{n-1}}^2+\|d_t\zeta^{n-2}\|_{c^{n-2}}^2\}.
\end{aligned} \tag{4.38}$$

Inequality (4.38) will replace inequality (4.8). Note that Lemma 2 yields $\sum_{n=1}^{N-1}\sigma_{4,n}\leqq 3K_5$. We recall from (3.10) that the general form of (3.17) is

$$\begin{aligned}
&\|\bar V^{n+1}-V^{n+1}\|_{c^n}+(\Delta t)^{1/2}\|\bar V^{n+1}-V^{n+1}\|_{a^n}\\
&\qquad\leqq C_1\rho\{\|\delta^3 V^n\|_{c^n}+(\Delta t)^{1/2}\|\delta^3 V^n\|_{a^n}\},\quad n\geqq 2.
\end{aligned} \tag{4.39}$$

We have thus shown the following corollary.

COROLLARY 2. *Let* S *and* R *and the restrictions on $\{\mathcal{M}_h\}$ of* § 2 *hold. Let V^0 and V^1 satisfy* (3.24) *and V^2 satisfy* (3.17a). *For $n\geqq 2$, assume that* (4.39) *is satisfied with ρ*

satisfying (4.36). *Let* $r \geqq 3$ *and* $\Delta t \leqq C_{19}h^2$. *Then, there exist constants* τ *and* C_{20} *such that if* $\Delta t \leqq \tau$ *and* $\Delta t \leqq h^{d/3}$,

$$\sup_{t^n} \|u - V\|_1 \leqq C_{20}\{(\Delta t)^2 + h^{r-1}\}. \tag{4.40}$$

C_{20} *has the same dependencies as* C_3.

COROLLARY 3. *Let all the hypotheses of Corollary* 2 *be satisfied except the assumption on r. Assume that* $r \geqq 2$ *and that* (4.33) *is satisfied. If* $\partial u/\partial t \in L^2(J; H^r)$, *then there exist constants* τ *and* C_{21} *such that, if* $\Delta t \leqq C_{19}h^2$, $\Delta t \leqq \tau$ *and* $\Delta t \leqq h^{d/3}$,

$$\sup_{t^n} \{\|u - V\| + h\|u - V\|_1\} \leqq C_{21}\{(\Delta t)^2 + h^r\}. \tag{4.41}$$

We shall now use Corollary 2 to obtain an optimal order L^2-estimate with the smoothness assumptions on $\partial u/\partial t$ of Theorem 1 if $\Delta t \leqq C_{19}h^2$. Let

$$\|\varphi\|_{-1} \equiv \sup\left\{\int_\Omega \varphi\psi \, dx : \|\psi\|_1 = 1\right\}. \tag{4.42}$$

THEOREM 2. *Let* S *and* R *and the restrictions on* $\mathcal{M}_h$ *of* § 2 *hold. Assume that the Neumann problem for* $-\Delta + I$ *on* Ω *is* H^3*-regular. Let* $r \geqq 3$ *and* $\Delta t \leqq C_{19}h^2$. *Let* V^0 *and* V^1 *satisfy* (3.24), V^2 *satisfy* (3.17a), *and* V^{n+1} *for* $n \geqq 2$ *satisfy* (3.17b) *with* Δt *replaced by* $(\Delta t)^{1/2}$. *Then there exist constants* τ *and* C_{22} *such that, if* $\Delta t \leqq \tau$ *and* $\Delta t \leqq h^{d/3}$,

$$\sup_{t^n} \{\|u - V\| + h\|u - V\|_1\} \leqq C_{22}\{(\Delta t)^2 + h^r\}. \tag{4.43}$$

C_{22} *has the same dependencies as* C_{20} *or* C_3.

Proof. To obtain an L^2-estimate, we use the test function $\chi = \zeta^{n+1} + \zeta^n = 2\zeta^{n+1/2}$ instead of $\chi = \zeta^{n+1} - \zeta^n$ in (4.2). The bounds for most of the terms follow more easily than in the proof of Theorem 1 without using summation by parts in time. The definition (4.42) is used in the bound for the third term on the left hand side of (4.4). After multiplying by Δt, we have

$$|\Delta t(c(EV^n)d_t\eta^n, 2\zeta^{n+1/2})| \leqq C\|d_t\eta^n\|_{-1}\|\zeta^{n+1/2}\|_1 \, \Delta t \leqq \frac{\Delta t}{64}\|\zeta^{n+1/2}\|_1^2 + C\,\Delta t\|d_t\eta^n\|_{-1}^2. \tag{4.44}$$

Then, as noted in [8], since the Neumann problem is H^3-regular, for each $t \in [0, T]$,

$$\|d_t\eta^n\|_{-1} \leqq h^2\|d_t\eta^n\|_1 \leqq C_{23}h^r\left\{\|u\|_{r-1} + \left\|\frac{\partial u}{\partial t}\right\|_{r-1}\right\}. \tag{4.45}$$

C_{23} has the same dependencies as K_3. We now note that from the proof of Corollary 2 we obtain the estimate

$$\Delta t \sum_{n=0}^{l-1} \|d_t\zeta^n\|^2 \leqq C\{(\Delta t)^4 + h^{2r-2}\}, \tag{4.46}$$

which, with the assumption that $\Delta t \leqq C_{19}h^2$, yields

$$\begin{aligned} \sum_{n=0}^{l-1} \|\delta\zeta^n\|^2 &\leqq C\{(\Delta t)^5 + \Delta t h^{2r-2}\} \\ &\leqq C\{(\Delta t)^5 + h^{2r}\}. \end{aligned} \tag{4.47}$$

We use (2.2a), (4.47) and ν from (3.12) with $\alpha = 1/2$ (a norm reduction of $O((\Delta t)^{1/2})$), to see that for $n \geqq 2$,

$$\begin{aligned}
&\Big|\sum_{n=2}^{l-1} \Delta t\Big[\Big(c(EV^n)\frac{V^{n+1}-\bar{V}^{n+1}}{\Delta t}, \zeta^{n+1}+\zeta^n\Big)\\
&\qquad +\frac{1}{2}(a(EV^n)\nabla(V^{n+1}-\bar{V}^{n+1}), \nabla(\zeta^{n+1}+\zeta^n))\Big]\Big|\\
&\leqq C\sum_{n=2}^{l-1}(\Delta t)^{1/2}\{\|\delta^3 V^n\|+(\Delta t)^{1/2}\|\delta^3 V^n\|_1\}\{\|\zeta^{n+1}+\zeta^n\|+(\Delta t)^{1/2}\|\zeta^{n+1}+\zeta^n\|_1\}\\
&\leqq C\sum_{n=2}^{l-1}(\Delta t)^{1/2}\Big(1+\frac{K_0(\Delta t)^{1/2}}{h}\Big)\{\|\delta^3\zeta^n\|\\
&\qquad +(\Delta t)^2\sigma_{6,n}\}\Big\{\Big(1+\frac{K_0(\Delta t)^{1/2}}{h}\Big)[\|\zeta^{n+1}\|+\|\zeta^n\|]\Big\} \qquad (4.48)\\
&\leqq C\sum_{n=2}^{l-1}(\Delta t)^{1/2}\{\|\delta\zeta^n\|+\|\delta\zeta^{n-1}\|+\|\delta\zeta^{n-2}\|+(\Delta t)^2\sigma_{6,n}\}\{\|\zeta^{n+1}\|+\|\zeta^n\|\}\\
&\leqq C\sum_{n=2}^{l-1}\Delta t\{\|\zeta^{n+1}\|^2+\|\zeta^n\|^2\}+C(\Delta t)^4+C\sum_{n=0}^{l-1}\|\delta\zeta^n\|^2\\
&\leqq C\sum_{n=2}^{l-1}\Delta t\{\|\zeta^{n+1}\|^2+\|\zeta^n\|^2\}+C\{(\Delta t)^4+h^{2r}\}.
\end{aligned}$$

The rest of the proof follows in a manner similar to that of Theorem 1. We note in particular that (4.23) holds for $l = 1, \cdots, N$ without a further induction argument since the hypotheses of Corollary 2 are assumed for this theorem. Thus, a norm-comparability argument for $\|\cdot\|_{c^n}$ similar to that in (4.20) and (4.21) will follow.

5. Computational considerations. In this section we reconsider the preconditioned conjugate gradient (PCG) method of § 3. It is computationally wasteful to iterate exactly ν times each time step if ν is determined using the pessimistic bound (3.12) and the results of § 4. In this section we present some additional criteria which are very easy to apply and which can terminate the iteration in fewer than ν steps. Next, to illustrate the effect of using incomplete iteration in terms of computational work, we give some rough operation counts for the linearized Crank–Nicolson procedure (3.3) and the PCG method. In one case, one can obtain work estimates for the PCG method which are of optimal order in the sense that the number of operations is proportional to the number of unknowns that define the solution. In another case we modify the basic process by changing the preconditioning matrix each $(\Delta t)^{-1/2}$ time steps; this gives a norm reduction of $O(\sqrt{\Delta t})$ with one iteration and $O(\Delta t)$ with two.

For $\varphi \in H^1$ let

$$|||\varphi|||^2 = |||\varphi|||_n^2 = \|\varphi\|_{c^n}^2 + \Delta t\|\varphi\|_{a^n}^2. \qquad (5.1)$$

If we let V_k^{n+1} correspond to the vector x_k of (3.7), then, from (3.11),

$$\psi_0 \leqq |||V_k^{n+1} - \bar{V}^{n+1}|||^2/(L_0^{-1}q_k, q_k)_e \leqq \psi_1. \qquad (5.2)$$

Since the denominator in (5.2) is computed during the course of the PCG procedure, we can easily estimate the size of the error $|||V_k^{n+1} - \bar{V}^{n+1}|||$, and, by comparing $(L_0^{-1}q_0, q_0)_e$ and $(L_0^{-1}q_k, q_k)_e$, we can observe the actual factor by which the norm is reduced. We can use these two quantities to stop the iteration if either is sufficiently small.

Suppose that

$$|||V^{n+1} - \bar{V}^{n+1}|||^2 \leqq \mu. \tag{5.3}$$

(If the iteration is stopped at the kth step, then (5.3) follows if $\psi_1(L_0^{-1}q_k, q_k)_e \leqq \mu$). Then the left hand side of (4.8) can be bounded by

$$\varepsilon \|d_t\zeta^n\|^2 \Delta t + C[\|\zeta^n\|_1^2 + \|\zeta^{n+1}\|_1^2] \Delta t + C\mu(\Delta t)^{-2}. \tag{5.4}$$

Thus, if at each time step either a norm reduction factor of $O(\Delta t)$ is achieved or if

$$\mu(\Delta t)^{-2} \leqq C(h^{2r-2} + (\Delta t)^4) \Delta t, \tag{5.5}$$

then the hypotheses of Theorem 1 are satisfied. Hence in the program one could set a parameter $\kappa = O((h^{2r-2} + (\Delta t)^4)(\Delta t)^3)$ and stop iterating if

$$(L_0^{-1}q_k, q_k)_e \leqq \kappa. \tag{5.6}$$

Similarly, a parameter $\rho_1 = O(\Delta t)$ could be defined and the iteration could be halted if

$$(L_0^{-1}q_k, q_k)_e / (L_0^{-1}q_0, q_0)_e \leqq \rho_1^2. \tag{5.7}$$

With this additional test, a modification of Theorem 1 holds with the same error bounds.

For each result in § 4, there correspond appropriate choices of κ and ρ_1 such that if the iteration is terminated whenever (5.6) *or* (5.7) holds, then the error bounds still apply. The following table summarizes these choices:

TABLE 1

Result	$\kappa \leqq$	$\rho_1 \leqq$
Theorem 1	$C(h^{2r-2} + (\Delta t)^4)(\Delta t)^3$	$C\,\Delta t$
Corollary 1	$C(h^{2r} + (\Delta t)^4)(\Delta t)^3$	$C\,\Delta t$
Corollary 2	$C(h^{2r-2} + (\Delta t)^4)(\Delta t)^2$	$K_8(\psi_1/\psi_0)^{1/2}$
Corollary 3	$C(h^{2r} + (\Delta t)^4)(\Delta t)^2$	$K_8(\psi_1/\psi_0)^{1/2}$
Theorem 2	$C(h^{2r} + (\Delta t)^4)(\Delta t)^2$	$C(\Delta t)^{1/2}$

As an aside we remark that if one chooses parameters κ and ρ_1 given by the bounds indicated for Theorem 2 and if the solution is actually so smooth that Corollary 3 applies, then, as h and Δt go to zero, the κ test will almost always stop the iteration before the ρ_1 test will. (By "almost always" we mean that the fraction of the timesteps stopped using the ρ_1 test goes to zero.) This follows from the bounds used in (4.8) and the conclusions of Corollary 3.

We shall now restrict our attention to spaces of piecewise cubic polynomials over quasi-regular meshes and give estimates of number of arithmetic operations needed to compute the extrapolated-Crank–Nicolson and the PCG approximate solutions. The heuristic arguments presented below can currently be made precise only in cases in which the meshes have very special structure, such as a uniform mesh on a square. However, numerical experiments indicate that the assumptions we use appear to be valid more generally.

Since we are using cubic polynomials as our example, $r = 4$. Restrict Ω to be a domain in the plane ($d = 2$) for the moment. The quasi-regularity of the meshes (all elements in a given mesh are assumed to be about the same size and shape) implies that

$$M \approx h^{-2}. \tag{5.8}$$

Balancing the h^4 and $(\Delta t)^2$ terms in the L^2 error bounds leads naturally to

$$\Delta t \approx h^2; \tag{5.9}$$

hence the number of time steps is

$$N = T/\Delta t \approx M. \tag{5.10}$$

We shall assume that the work to factor a matrix with the structure of L_0 is

$$FW \approx M^{3/2}. \tag{5.11}$$

In the case of a rectangular mesh on a rectangle this can be achieved by using the nested dissection process of George [11]; the results of Hoffman, Martin, and Rose [12] indicate that it cannot be improved. Experimentally, it has been found that minimal degree orderings (they are not unique) give the same form of work estimate. We shall also assume that the factors of L_0 have

$$SW \approx M \ln M \tag{5.12}$$

nonzero elements. Note that the work to perform one preconditioned conjugate gradient iteration is $O(SW+M) \approx SW$.

Combining these results shows that the work to compute the solution of the linearized-Crank–Nicolson scheme is

$$N(FW+SW) = O(M^{5/2}). \tag{5.13}$$

In the context of Theorem 1, Corollary 1 or Theorem 2 we would need $NI = O(|\ln \Delta t|) = O(\ln M)$ iterations at each step. Thus the expected work to compute the PCG approximation in these cases is

$$FW + N * NI * SW = O(M^2(\ln M)^2). \tag{5.14}$$

The processes analyzed by Corollaries 2 and 3 only require a fixed number of iterations per time step; hence in these cases the work is

$$FW + N * NI * SW = O(M^2 \ln M). \tag{5.15}$$

Note that $N * M = O(M^2)$ parameters are used to define the solution. We see that the work estimates (5.14) and (5.15) are very close to optimal order since any process that deals with each parameter at least once must do at least $N * M$ operations of some type.

We now indicate how the PCG scheme of § 3 can be modified to give an optimal order work estimate of $O(M^2)$ while still achieving the norm reduction necessary for Corollaries 2 and 3.

Suppose, as is frequently the case, that C_0 from (3.2f) is comparable with its diagonal; i.e., assume that there exists $\beta > 0$, independent of h, such that, if $D_0 = \operatorname{diag}(C_0)$ and $0 \neq x \in \mathbb{R}^M$, then

$$\beta^{-1} \leqq (C_0 x, x)_e/(D_0 x, x)_e \leqq \beta. \tag{5.16}$$

In the case of Lagrange type cubic elements over a quasi-regular family of meshes this relation follows from a simple homogeneity argument, but it holds for some other element types as well. It follows from (2.2a), the fact that $\Delta t \leqq Ch^2$, and (2.4) that there exists $\hat{\beta} > 0$, independent of h, such that for $0 \neq x \in \mathbb{R}^M$

$$\hat{\beta}^{-1} \leqq ((C^n + \Delta t A^n)x, x)_e/(D_0 x, x)_e \leqq \hat{\beta}. \tag{5.17}$$

If the iteration (3.7) uses D_0 as the preconditioning matrix instead of L_0, then (5.17) implies that a fixed number NI (independent of h) of iterations can be used to reduce the error by the factor K_8 needed for Corollaries 2 and 3. Thus the work required to compute the corresponding approximate solution is $O(M^2)$, since each iteration necessitates only $O(M)$ operations.

Another interesting modification of the basic process in § 3 is obtained by changing L_0 during the course of the computation. If we take NS to be approximately $N^{1/2}$ and, if after each NS steps we set L_0 equal to the current L^n, then it follows that $\psi_0 = 1 - c\sqrt{\Delta t}$ and $\psi_1 = 1 + c\sqrt{\Delta t}$. For this process we see that the $O(\Delta t)$ norm reduction needed for Theorem 1 and Corollary 1 requires only two iterations per time step, while the $O(\sqrt{\Delta t})$ needed in Theorem 2 can be obtained with one. With the assumptions made above we see that the work for this process is

$$\frac{N}{NS} * FW + N * SW = O(M^2 \ln M). \tag{5.18}$$

We briefly consider the case of $d = 3$. The best conjectures we know of say that

$$FW = M^2, \qquad SW = M^{4/3}. \tag{5.19}$$

Since we have $r = 4$, we still use

$$\Delta t \approx h^2$$

and

$$M \approx h^{-3},$$
$$N = T/\Delta t = M^{2/3}.$$

Thus, an optimal order process would require $O(M^{5/3})$ operations and this can be achieved using the procedure described above that utilizes a diagonal preconditioning matrix. Even with the weaker smoothness constraints of Theorem 1 or 2 the expected work is the nearly optimal $O(M^{5/3} \ln M)$. If we use the basic PCG procedure of § 3 we see that the work is $O(M^2 \ln M)$; this is far superior to the $O(M^{8/3})$ operations needed to carry out the extrapolated-Crank–Nicolson process.

REFERENCES

[1] S. AGMON, *Lectures on Elliptic Boundary Value Problems*, Van Nostrand, Princeton, NJ, 1965.

[2] O. AXELSSON, *On preconditioning and convergence acceleration in sparse matrix problems*, CERN European Organization for Nuclear Research, Geneva, 1974.

[3] ———, *On the computational complexity of some matrix iterative algorithms*, Report 74.06, Dept. of Computer Science, Chalmers University of Technology, Göteberg, Sweden, 1974.

[4] J. E. DENDY, JR., *An analysis of some Galerkin schemes for the solution of nonlinear time-dependent problems*, this Journal, 12 (1975), pp. 541–565.

[5] J. DOUGLAS, JR., *A survey of numerical methods for parabolic differential equations*, Advances in Computers, vol. II, Academic Press, New York, 1961.

[6] J. DOUGLAS, JR. AND T. DUPONT, *Galerkin methods for parabolic equations*, this Journal, 7 (1970), pp. 575–626.

[7] ———, *Preconditioned conjugate gradient iteration applied to Galerkin methods for a mildly nonlinear Dirichlet problem*, Sparse Matrix Computations, Academic Press, Inc., New York, 1976, pp. 333–348.

[8] T. DUPONT, *L_2 error estimates for projection methods for parabolic equations in approximating domains*, Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press, New York, 1974, pp. 313–352.

[9] T. DUPONT, G. FAIRWEATHER, AND J. P. JOHNSON, *Three-level Galerkin methods for parabolic equations*, this Journal, 11 (1974), pp. 392–410.
[10] M. ENGELI, TH. GINSBURG, H. RUTISHAUSER, AND E. STIEFEL, *Refined iterative methods for the computation of the solution and the eigenvalues of self-adjoint boundary value problems*, Mitteilungen aus dem Institut für Angewandte Mathematik, nr. 8, ETH, Zurich, 1950.
[11] A. GEORGE, *Nested dissection on a regular finite element mesh*, this Journal, 10 (1973), pp. 345–363.
[12] A. J. HOFFMAN, M. S. MARTIN, AND D. J. ROSE, *Complexity bounds for regular finite difference and finite element grids*, this Journal, 10 (1973), pp. 364–369.
[13] M. LUSKIN, *A Galerkin method for nonlinear parabolic equations with nonlinear boundary conditions*, this Journal, to appear, April 1979.
[14] H. H. RACHFORD, JR., *Two-level discrete-time Galerkin approximations for second order nonlinear parabolic partial differential equations*, this Journal, 10 (1973), pp. 1010–1026.
[15] M. F. WHEELER, *A priori L_2 error estimates for Galerkin approximations to parabolic partial differential equations*, this Journal, 10 (1973), pp. 723–759.