ALAN E. BERGER
HAIM BREZIS
JOËL C. W. ROGERS

A numerical method for solving the problem
\[ u_t - \Delta f(u) = 0 \]


<http://www.numdam.org/item?id=M2AN_1979__13_4_297_0>
A NUMERICAL METHOD
FOR SOLVING THE PROBLEM \( u_t - \Delta f(u) = 0 \) (*)

Alan E. Berger (1), Haim Brezis (2) and Joël C. W. Rogers (3)

Abstract. — A method is presented for solving a class of nonlinear evolution equations. This procedure involves solution of a corresponding linear problem together with simple algebraic operations. Stability and convergence of the algorithm are analyzed, and results of some numerical experiments are given.

Résumé. — On présente une méthode de résolution d'une famille d'équations d'évolution nonlinéaires. Ce procédé repose sur la solution du problème linéaire correspondant ainsi que sur des opérations algébriques simples. On analyse la stabilité et la convergence de l'algorithme, et on donne des résultats d'expériences numériques.

I. INTRODUCTION

We present an algorithm based on nonlinear semigroup theory which applies to a class of nonlinear evolution equations. This procedure is a generalization and simplification of the alternating phase truncation method for the Stefan problem studied by Rogers, Berger, and Ciment [1, 20], and is very closely related to the Laplace-modified forward Galerkin method considered by Douglas and Dupont [16].

II. THE NONLINEAR EVOLUTION EQUATION

An algorithm will be presented for the solution of the following problem. Let \( \Omega \subset \mathbb{R}^n \) be a bounded domain with smooth boundary \( \Gamma \). Let \( f : \mathbb{R} \to \mathbb{R} \) be a non-
decreasing function which is Lipschitz continuous on every bounded interval and such that \( f(0) = 0 \). Let \( L : D(L) \subset L^1(\Omega) \rightarrow L^1(\Omega) \) be an unbounded linear operator on \( L^1(\Omega) \) which satisfies the following conditions:

\( L \) is a closed operator with dense domain \( D(L) \) in \( L^1(\Omega) \); for every \( \lambda > 0 \), \( I + \lambda L \) maps \( D(L) \) one to one onto \( L^1(\Omega) \) and \( (I + \lambda L)^{-1} \) is a contraction in \( L^1(\Omega) \). In other words, \( -L \) generates a linear contraction semigroup in \( L^1(\Omega) \) denoted by \( S(t) \).

For any \( \lambda > 0 \) and \( \varphi \in L^1(\Omega) \),

\[
\sup_{\Omega} (I + \lambda L)^{-1} \varphi \leq \max \{ 0, \sup_{\Omega} \varphi \}
\]

[by \( \sup \) we mean the essential supremum; if \( \sup \varphi = \infty \), assumption (2) is empty]. There exists an \( \alpha > 0 \) such that

\[
\alpha \| u \|_1 \leq \| Lu \|_1 \quad \text{for all } u \in D(L)
\]

[throughout this paper, the \( L^p(\Omega) \) norm of any function \( \varphi \) will be denoted by \( \| \varphi \|_p \)].

Assumptions (1)-(3) are, for example, satisfied by \( Lu = -\Delta u \), \( D(L) = \{ u \in W^{1,1}_0(\Omega); Lu \in L^1(\Omega) \} \) (where \( Lu \) is understood in the distribution sense). More generally we may take

\[
Lu = -\sum_{i,j} \frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial u}{\partial x_j} \right) + \sum_i \frac{\partial}{\partial x_i} (a_i u) + au
\]

where \( a_{ij}, a_i \in C^1(\overline{\Omega}), a \in L^\infty(\Omega) \), both \( a \) and \( (a + \sum_i \partial a_i/\partial x_i) \) are nonnegative almost everywhere on \( \Omega \), and for some positive constant \( \alpha \),

\[
\sum_{i,j} a_{ij} \xi_i \xi_j \geq \alpha \| \xi \|^2 \quad \text{a.e. in } \Omega, \text{ for each } \xi \in \mathbb{R}^N
\]

(see, e.g., theorem 8 of Brezis and Strauss [11]).

We are concerned with solving the evolution equation

\[
\frac{du}{dt} + Lf(u) = 0, \quad u(0) = u_0.
\]

The nonlinear operator \( Au = Lf(u) \) defined as an operator in \( L^1(\Omega) \) with domain \( D(A) = \{ u \in L^1(\Omega); f(u) \in D(L) \} \) is \( m \)-accretive in \( L^1(\Omega) \), i.e., for every

R.A.I.R.O. Analyse numérique/Numerical Analysis
METHOD FOR SOLVING THE PROBLEM $u_t - \Delta f(u) = 0$

$q \in L^1(\Omega)$ and every $\lambda > 0$, there is a unique solution $u \in D(A)$ of the equation

$$u + \lambda Lf(u) = \phi,$$

and in addition, the mapping $\phi \to u$ is a contraction in $L^1(\Omega)$ (see for example theorem 1 of Brezis and Strauss [11]). On the other hand, $D(A)$ is dense in $L^1(\Omega)$ (note for example that if $q \in L^{\infty}(\Omega)$, then $u_\lambda = (I + \lambda A)^{-1} \phi$ satisfies $\|u_\lambda\|_2 \leq \|\phi\|_2$ and so $u_\lambda \to \phi$ in $L^2(\Omega)$ as $\lambda \to 0$). It follows that

$$S_g(t)u_0 = \lim_{m \to \infty} \left( I + \frac{t}{m} A \right)^{-m} u_0$$

is a contraction semigroup on $D(A) = L^1(\Omega)$; $S_g(t)u_0$ is the generalized solution of (4) in the sense of Crandall-Liggett and Benilan (see [4, 10, 12, 13, 14]).

The classical Stefan problem with homogeneous Dirichlet data can be cast into the form (4) as described below. Consider the Stefan problem posed in terms of temperature $(v)$, with the freezing point denoted by $z$, and for convenience, with constant material properties ($c_i, c_w, k_i, k_w$);

$$c_i v_t = k_i \Delta v \quad \text{for} \quad p \in I(t) \equiv \{ p \in \Omega; \, v(p, t)<z \}, \quad t > 0, \quad (5a)$$

$$c_w v_t = k_w \Delta v \quad \text{for} \quad p \in W(t) \equiv \{ p \in \Omega; \, v(p, t)>z \}, \quad t > 0, \quad (5b)$$

$$v(p, t) = \tilde{g}(p, t) \quad \text{for} \quad p \in \Gamma, \quad t > 0, \quad (5c)$$

$$v(p, t = 0) = \tilde{u}_0(p) \quad \text{for} \quad p \in \Omega. \quad (5d)$$

The proper energy balance conditions on the moving interface

$$M(t) \equiv \{ p \in \Omega; \, v(p, t) = z \}$$

is

$$-k_w v^+ + k_i v^- = \lambda V_v \quad \text{for} \quad p \in M(t), \quad t > 0. \quad (5e)$$

Here $v$ is the unit normal to $M(t)$ pointing into $W(t)$, $\lambda$ is the latent heat of the change of phase, $V_v$ is the velocity of the interface in the direction of $v$, and $v^+_v(p, t)$ denotes the limit of $v$ at $p$ approached from within $W(t)$, etc. Let

$$c(v) = \begin{cases} c_i & \text{for} \, v < z \\ c_w & \text{for} \, v \geq z \end{cases} \quad \text{and} \quad \eta(v) = \begin{cases} 0 & \text{for} \, v < z \\ 1 & \text{for} \, v \geq z \end{cases}.$$

Let the enthalpy $u(p, t) = u(v(p, t))$ be defined by

$$u(v) = \int_z^v c(\xi) d\xi + \lambda \eta(v) + r_1 \quad (r_1 \text{ is an arbitrary constant}). \quad (6)$$
Then when \( \bar{g} \equiv 0 \), (5) can be written in the form (4) with \( L = -\Delta \) and (see [9, 15]):

\[
\begin{cases}
\frac{\bar{z} + k_i(u - r_1)}{c_i} & \text{for } u \leq r_1, \\
\bar{z} & \text{for } r_1 \leq u \leq r_1 + \lambda, \\
\frac{\bar{z} + k_i(u - \lambda - r_1)}{c_i} & \text{for } u \geq r_1 + \lambda.
\end{cases}
\]

To obtain zero Dirichlet data for (4), the constants \( r_1 \) and \( z \) are chosen so that \( u(0) = 0 \) and \( f(0) = 0 \).

Problem (4) with \( L = -\Delta \) and \( f(u) = |u|^{\alpha - 1}u \) (\( \alpha > 1 \)) occurs in a model of gas diffusion through a porous medium (here \( u \) corresponds to gas density and \( \text{grad}|u|^{\alpha - 1} \) to velocity, see e.g. [1, 9, 17, 18].

III. THE ALGORITHM

We first present the "analytical" algorithm, i.e., the form of the algorithm which involves no spatial discretization. Let \( \sigma : (0, \infty) \rightarrow (0, \infty) \) be a function such that \( \lim \sigma = 0 \). Let \( t > 0 \) be fixed, and let the time step \( \tau = t/n \) where \( n \geq 1 \) is an integer. We consider the following algorithm;

\[
\frac{u^{k+1} - u^k}{\tau} + \left[ \frac{I - S(\sigma)}{\sigma} \right] f(u^k) = 0, \quad u^0 = u_0,
\]

that is \( u^{k+1} \) is determined from \( u^k \) by

\[
u^{k+1} = F(\tau) u^k,
\]

where

\[
F(\tau) \varphi = \varphi + \frac{\tau}{\sigma} [S(\sigma) f(\varphi) - f(\varphi)] \text{ for } \varphi \in L^1(\Omega).
\]

We define the approximation \( u_n(t) \) to \( u(t) \) [the generalized solution of (4)] by

\[
u_n(t) = u^n = \left[ F\left( \frac{t}{n} \right) \right]^n u_0.
\]

Concerning the convergence of \( u_n(t) \) to the solution \( u(t) \) of (4), one has.

**Theorem 1:** Assume \( u_0 \in L^\infty(\Omega) \), set \( M = \| u_0 \|_\infty \), and let \( \mu \) denote the Lipschitz constant of \( f \) on the interval \([-M, M]\). Assume the following stability condition holds:

\[
\mu \tau / \sigma, \leq 1 \quad \text{for each } \tau > 0
\]

(for example, this is valid if \( \sigma, \equiv \mu \tau \). Then \( \lim_{n \rightarrow \infty} u_n(t) = u(t) \) [the solution of (4)] in

R.A.I.R.O. Analyse numérique/Numerical Analysis
METHOD FOR SOLVING THE PROBLEM \( u_t - \Delta f(u) = 0 \)

\( L^1(\Omega) \); in addition the convergence is uniform for \( t \) in any given bounded interval.

The proof will be obtained by demonstrating that \( F \) obeys a maximum principle (lemma 1), that \( F \) is contractive in \( L^1(\Omega) \) (lemma 2), and by then applying the nonlinear Chernoff formula.

**Lemma 1:** If (11) is valid, then \(-M \leq u^k \leq M\) for all \( k \).

**Proof:** We argue by induction; assume \(-M \leq u^k \leq M\). Since the function \( r \to r - \tau f(r)/\sigma \) is nondecreasing \([\text{by (11)}]\), it follows that

\[
-M - \tau f(-M)/\sigma \leq u^k - \tau f(u^k)/\sigma \leq M - \tau f(M)/\sigma.
\]

(12)

On the other hand, since \( f \) is nondecreasing one has \( f(-M) \leq f(u^k) \leq f(M) \). It follows from (2) that

\[
f(-M) \leq S(\sigma) f(u^k) \leq f(M).
\]

(13)

Combining (12) and (13), one obtains \(-M \leq u^{k+1} \leq M\). Therefore in performing the iteration (9), we can replace \( f \) by \( \tilde{f} \) where \( \tilde{f} = f \) for \(-M \leq r \leq M\), \( \tilde{f} = f(M) \) for \( r \geq M\), and \( \tilde{f} = f(-M) \) for \( r \leq -M\). In what follows we may thus assume that \( f \) is Lipschitz continuous with Lipschitz constant \( \mu \) on all of \( R^1 \).

**Lemma 2:** If (11) is valid, then \( F(\tau) \) is a contraction on \( L^1(\Omega) \), i.e.,

\[
\| F(\tau) \varphi - F(\tau) \psi \|_1 \leq \| \varphi - \psi \|_1 \quad \text{for} \quad \varphi, \psi \in L^1(\Omega).
\]

**Proof:** Indeed, we have

\[
\| F(\tau) \varphi - F(\tau) \psi \|_1 \leq \frac{\tau}{\sigma} \| f(\varphi) - f(\psi) \|_1 + \| (\varphi - \psi) - \frac{\tau}{\sigma} (f(\varphi) - f(\psi)) \|_1.
\]

(14)

Since the functions \( f(r) \) and \( r - \tau f(r)/\sigma \) are nondecreasing in \( r \), it follows that for \( r, s \) in \( R^1 \).

\[
\frac{\tau}{\sigma} \left| f(r) - f(s) \right| + \left| (r - s) - \frac{\tau}{\sigma} (f(r) - f(s)) \right| = |r - s|.
\]

(15)

Combining (14) and (15) gives the result.

We conclude the proof by applying theorem 3.2 of Brezis and Pazy [10] (this is the nonlinear Chernoff formula). It suffices to verify that for every \( \varphi \in L^1 \) and every \( \lambda > 0 \):

\[
\psi = (I + \frac{\lambda}{\tau} (I - F(\tau)))^{-1} \varphi \rightarrow \psi \equiv (I + \lambda A)^{-1} \varphi
\]

vol. 13, no 4, 1979
in $L^1(\Omega)$ as $\tau \to 0$. We have
\[ \psi_t + \frac{\lambda}{\tau} (I - F(\tau)) \psi = \phi. \] (17)

Set
\[ \psi + \frac{\lambda}{\tau} (I - F(\tau)) \psi = \phi_t. \] (18)

then $\phi_t \to \psi + \lambda L f(\psi) = \psi + \lambda A \psi = \phi$ in $L^1(\Omega)$ as $\tau \to 0$ since $\psi \in D(A)$ implies $f(\psi) \in D(L)$ and then
\[ \frac{I - S(\sigma_t)}{\sigma_t} f(\psi) \to L f(\psi) \quad \text{in } L^1(\Omega) \quad \text{as } \tau \to 0. \]

Finally, recalling that $F(\tau)$ is contractive, combining (17) and (18) gives
\[ \left(1 + \frac{\lambda}{\tau}\right) \| \psi_t - \psi \|_1 \leq \| \phi - \phi_t \|_1 + \frac{\lambda}{\tau} \| \psi_t - \psi \|_1. \]

Hence $\| \psi_t - \psi \|_1 \leq \| \phi - \phi_t \|_1 \to 0$ as $\tau \to 0$, and the proof of theorem 1 is complete.

Remarks: The conclusion of theorem 1 holds (with the same proof) if in the scheme (9) one redefines $F(\tau)\phi$ to be
\[ \phi + \frac{\tau}{\sigma_t} (J(\sigma_t) f(\phi) - f(\phi)), \] (19)

where $J(\sigma) \equiv (I + \sigma L)^{-1}$. This is an "analytical" version of the Laplace modified forward Galerkin method considered by Douglas and Dupont in [16]. A similar approach may be used [3] to establish convergence of the truncation method ([5, 6, 8]) for obstacle variational inequalities.

IV. NUMERICAL IMPLEMENTATION

Note that while the proofs in the previous section only treat the case of zero boundary data, the extension of the algorithm given below for the situation with nonzero Dirichlet data seems intuitively reasonable. We will discuss implementation of the algorithm for the following problem (for simplicity taking $L = -\Delta$);
\begin{align*}
    u_t &= \Delta f(u) \quad \text{in } \Omega \quad \text{(20 a)} \\
    u(p, t=0) &= u_0(p) \quad \text{in } \Omega \quad \text{(20 b)} \\
    u(p, t) &= g(p, t) \quad \text{for } p \in \Gamma, \quad t > 0. \quad \text{(20 c)}
\end{align*}
Suppose one has approximate solution values \( \{ U_j^n \} \) at time \( t^n = n \Delta t \) on a set of \( J \) grid points \( \{ p_j \} = \bar{\Omega} \) (e.g., \( U_0^0 = u_0(p_j), j = 1, \ldots, J \)). To obtain the approximate solution values \( \{ U_j^{n+1} \} \) at the next time level \( t^{n+1} = t^n + \Delta t \), one performs the following "discretization of (9)" (here \( \alpha \) corresponds to \( \sigma_t / \tau \)):

\[
Q_j^n = \frac{f(U_j^n)}{\alpha}, \quad j = 1, \ldots, J, \tag{21a}
\]

set, solve, using any appropriate numerical method, the linear heat equation (21a)

\[
Q_t = \alpha \Delta Q \quad \text{in} \quad \Omega, \tag{22a}
\]

\[
Q(p, t) = \frac{f(g(p, t))}{\alpha} \quad \text{on} \quad \Gamma, \tag{22b}
\]

\[
Q(p_j, t^n) = Q_j^n \quad \text{for} \quad j = 1, \ldots, J, \tag{22c}
\]

obtaining values \( \{ Q_j^{n+1} \} \) at the underlying grid points \( \{ p_j \} \) at \( t = t^n + \Delta t \); then

\[
U_j^{n+1} = U_j^n + Q_j^{n+1} - \frac{f(U_j^n)}{\alpha} \quad \text{for} \quad j \text{ such that } p_j \in \Omega, \tag{21c}
\]

and

\[
U_j^{n+1} = g(p_j, t^{n+1}) \quad \text{when} \quad p_j \in \Gamma. \tag{21d}
\]

**Remarks:** Assume there is a constant \( M \) such that \( |g(p, t)| \leq M, |u_0(p)| \leq M \), and let \( \mu \) be a Lipschitz constant for \( f(r) \) on \( -M \leq r \leq M \). As before, by modifying \( f(r) \) on \( |r| > M \), we may suppose \( \mu \) is a Lipschitz constant for \( f \) on all of \( \mathbb{R}^1 \). Then analogous to lemmas 1 and 2; if \( \alpha \geq \mu \) and if the numerical method used to solve (22) is \( L^1 \) stable, then so is the entire algorithm (21). That is, if \( U^n = \{ U_j^n \} \) and \( \overline{U}^n \) are two different starting values, then

\[
|U^{n+1} - \overline{U}^{n+1}|_p \leq |U^n - \overline{U}^n|_p.
\]

If \( \alpha \geq \mu \) and the numerical scheme used to solve (22) satisfies a maximum principle, then so does the entire algorithm (21), i.e.,

\[
\min(u_0, g) \leq U^{n+1} \leq \max(u_0, g).
\]

For general \( L \) the appropriate maximum principle is \( |U^{n+1}|_{\infty} \leq M \). Sufficient conditions for \( L^1 \) stability and for a maximum principle for some standard finite difference schemes and for piecewise linear finite elements for (22) are given in, e.g., [7]. We note that \( L^1 \) stability or a maximum principle for the method used to solve (22) is not necessary for reasonable numerical behavior of the algorithm (21); similarly it is not always necessary to have \( \alpha \geq \mu \) (cf. the vol. 13, n° 4, 1979
numerical examples below). Indeed, in developing error estimates for the Laplace modified forward Galerkin equation when the solution \( u \) is "smooth". Douglas and Dupont only require \( \alpha > \mu / 2 \) \([b]\).

V. NUMERICAL EXPERIMENTS

The first numerical experiments discussed will be for a problem with \( f(u) = u |u| \) whose solution (on \( 0 \leq x \leq 20 \)) is depicted by the solid lines in figure 1. The exact solution \( u \) (and thus \( u_0 \) and \( g \)) was obtained from the top of page 363 of \([17]\) (note that the \( m \) appearing as an exponent of the term \((m - 1)/(2m(m + 1))\) in the definition of \( f(\eta) \) should be omitted). Here \( x = 6 \) corresponds to \( x = 0 \) in \([17]\), the solution reflects across \( x = 6 \) via \( u(6 - r) = -u(6 + r) \), and the values for the parameters of \([17]\) are \( m = 2, \tau = 1, \) and \( a = 3.5 \) (the Lipschitz constant \( \mu \) for this problem is a little less than 1.0).

The first numerical results for this problem (table I) that we will discuss were obtained by solving (22) using the standard Crank-Nicolson (C.N.) finite difference method using \( J = 433 \) (equally spaced) grid points on \([0, 20]\), and using

\[ \Delta t = \frac{143}{1024} \]
a variable number $N$ of time steps to reach time $T = 8.9375$ (thus $\Delta t = T/N$). The discrete $L^1$ error $E_1$ at time $T$ was calculated by

$$E_1 = \frac{20}{J-1} \sum_{j=1}^{J} |U_j^N - u(p_j, T)|.$$

Suppose $J$ is large enough so that the spatial discretization error is relatively negligible, and assume that $E_1$ is governed by

$$E_1 = C_1 (\Delta t)^\rho,$$

where $C_1$ is a constant independent of $\Delta t$. Then the numerical rate of convergence $\rho$ computed from the results of using two successive values of $N$ (i.e. $N$ and $N = 2N$ with corresponding errors $E_1$ and $E_1$) is

$$\rho = \frac{\ln(E_1) - \ln(E_1)}{\ln(2)}.$$

Results of taking $\alpha = 1$, $\alpha = .75$, and $\alpha = .5$ are given in table I (setting $\alpha = .4$ resulted in a fatal instability (exponent overflow) at $N = 256$). Note the decrease in $E_1$ (i.e. decrease in $C_1$ of (23) while $\rho$ remains roughly constant) as $\alpha$ is decreased (until instability develops). The last result in table I was obtained by

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\alpha = 1$ C.N.</th>
<th>$\alpha = .75$ C.N.</th>
<th>$\alpha = .5$ C.N.</th>
<th>$\alpha = .5$ Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.2794</td>
<td>.1666</td>
<td>.0996</td>
<td>.2747</td>
</tr>
<tr>
<td>8</td>
<td>.1580</td>
<td>.1026</td>
<td>.0541</td>
<td>.1578</td>
</tr>
<tr>
<td>16</td>
<td>.0875</td>
<td>.0583</td>
<td>.0306</td>
<td>.0873</td>
</tr>
<tr>
<td>32</td>
<td>.0475</td>
<td>.0321</td>
<td>.0171</td>
<td>.0474</td>
</tr>
<tr>
<td>64</td>
<td>.0256</td>
<td>.0174</td>
<td>.0093</td>
<td>.0255</td>
</tr>
<tr>
<td>128</td>
<td>.0136</td>
<td>.0093</td>
<td>.0050</td>
<td>.0136</td>
</tr>
<tr>
<td>256</td>
<td>.0071</td>
<td>.0049</td>
<td>.0026</td>
<td>.0071</td>
</tr>
</tbody>
</table>

using the standard fully implicit finite difference scheme (instead of C.N.) with $J = 433$ to solve (22), and by taking $\alpha = .5$ (this corresponds to an
implementation of a form of the Laplace modified forward Galerkin method — c.f. p. 154 of [16]). Notice the almost exact correspondence between the results for \( (\alpha = .5, \text{implicit}) \) and those for \( (\alpha = 1, \text{C.N.}) \). This correspondence also occurs between \( (\alpha = .375, \text{implicit}) \) and \( (\alpha = .75, \text{C.N.}) \), and between \( (\alpha = .25, \text{implicit}) \) and \( (\alpha = .5, \text{C.N.}) \) \((J = 433 \text{ throughout})\). Taking \( \alpha = .2 \) with the implicit method \((J = 433)\) yields instability at \( N = 256 \). The plausibility of this correspondence may be seen by writing out the algorithm (21) using the appropriate matrix \( K \) for the scheme being used for (22), and here taking \( U^n \) to be the vector of values \( \{ U^n_j; p_j \in T \} \), etc. and assuming zero Dirichlet data

\[
U^{n+1} = U^n + (I + \theta \Delta t \alpha K)^{-1} (I - (1 - \theta) \Delta t \alpha K) \frac{f(U^n)}{\alpha} - \frac{f(U^n)}{\alpha}.
\]

For C.N., \( \theta = 1/2 \), while for implicit, \( \theta = 1 \) in this expression.

A little algebra shows that taking \( (\text{C.N.}, \alpha = \bar{\alpha}) \) is identical to taking \( (\text{implicit, } \alpha = \bar{\alpha}/2) \). Some more algebra shows that including the effect of nonzero Dirichlet data yields a formally \( O(\Delta t) \) difference between \( (\text{C.N.}, \alpha = \bar{\alpha}) \) and \( (\text{implicit, } \alpha = \bar{\alpha}/2) \). Since the boundary data for this example is zero until \( t \) is near \( T = 8.9375 \), the numerical agreement is very pronounced, despite the fact that the observed numerical error of the algorithm itself is behaving almost like \( O(\Delta t) \). As an indication of the effect of the size of \( J \) on the error; with \( N = 256 \), \( \alpha = .5 \), and using C.N. for (22), the values of \( E_1 \) corresponding to several values of \( J \) were;

<table>
<thead>
<tr>
<th>( J )</th>
<th>( J = 55 )</th>
<th>( J = 109 )</th>
<th>( J = 217 )</th>
<th>( J = 433 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_1 )</td>
<td>( .0177 )</td>
<td>( .0044 )</td>
<td>( .0042 )</td>
<td>( .0026 )</td>
</tr>
</tbody>
</table>

For comparison, we also implemented the 3 level (centered) Laplace modified Galerkin equation (see p. 155 of [16]) for (20), in the following form [here \( D_x^2 U_j = (U_{j-1} - 2U_j + U_{j+1})/\Delta x^2 \) and \( \Delta x \) is the (uniform) mesh length];

\[
(I - 2 \Delta t \beta D_x^2) U_j^{n+1} = (I + 2 \Delta t \beta D_x^2) U_j^n - 4 \Delta t \beta D_x^2 U_j^n + 2 \Delta t D_x^2 (f(U_j^n)) \quad \text{for } j \text{ such that } 0 < p_j < 20. \tag{24}
\]

The exact solution was used to provide the values \( U_j^1 \) at \( t = \Delta t \). Results are given in table II (again \( J = 433 \) and \( T = 8.9375 \)). The condition on the parameter \( \beta \) given in [16] (for “smooth” problems) is \( \beta > \mu/4 \) (here \( \beta = .25 \) satisfies this while...
METHOD FOR SOLVING THE PROBLEM $u_t - \Delta f(u) = 0$

$\beta = .2$ does not). Taking $\beta = .15$ yields instability (exponent overflow) at $N = 256$. For $N = 256$ and $\beta = .25$ one has:

<table>
<thead>
<tr>
<th>$J = 55$</th>
<th>109</th>
<th>217</th>
<th>433</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1 = .0163$</td>
<td>.0028</td>
<td>.0023</td>
<td>.0005</td>
</tr>
</tbody>
</table>

Even for this (nonsmooth) problem, the 3 level scheme produces a higher numerical rate of convergence.

**Table II**

*Discrete $L^1$ errors and numerical convergence rates for several implementations of the Laplace modified centered equation (24)*

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\beta = .1$</th>
<th>$\beta = .5$</th>
<th>$\beta = .25$</th>
<th>$\beta = .2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error $E_1$</td>
<td>Rate $\rho$</td>
<td>Error $E_1$</td>
<td>Rate $\rho$</td>
</tr>
<tr>
<td>4</td>
<td>1.572</td>
<td>1.7</td>
<td>0.756</td>
<td>1.3</td>
</tr>
<tr>
<td>8</td>
<td>0.4853</td>
<td>1.6</td>
<td>0.2160</td>
<td>1.3</td>
</tr>
<tr>
<td>16</td>
<td>0.1637</td>
<td>1.4</td>
<td>0.0854</td>
<td>1.3</td>
</tr>
<tr>
<td>32</td>
<td>0.0609</td>
<td>1.4</td>
<td>0.0349</td>
<td>1.3</td>
</tr>
<tr>
<td>64</td>
<td>0.0224</td>
<td>1.5</td>
<td>0.0139</td>
<td>1.3</td>
</tr>
<tr>
<td>128</td>
<td>0.0079</td>
<td>1.5</td>
<td>0.0048</td>
<td>1.5</td>
</tr>
<tr>
<td>256</td>
<td>0.0021</td>
<td>1.9</td>
<td>0.0009</td>
<td>2.4</td>
</tr>
</tbody>
</table>

The second problem to be considered is a two phase Stefan problem whose solution (in terms of temperature) is depicted in figure 2 (the exact solution is given in both [7] and [20] — note that in this particular example $\mu = k_i/c_i$ and $z = z_i = r_1 = 0$). In order to suppress the effect of the jump discontinuity from 0 to $\lambda$ in enthalpy at the interface, errors discussed here for the Stefan problem will be errors in temperature values. The temperature $v$ corresponding to an enthalpy value $u$ is easily obtained by inverting (6).

For the one phase Stefan problem, the alternating phase truncation (APT) method ([7, 20]) reduces to (21) with $\alpha = \mu$. In [20], for the "analytical" APT method for a one dimensional one phase Stefan problem, it was shown that the

vol. 13, n° 4, 1979
error in the temperature value and in the interface location could be bounded by

$$C(\Delta t \ln (1 + T/\Delta t))^{1/2},$$

(25)

where $T$ is the time at which one is bounding the error, and $C$ is a constant independent of $\Delta t$ (but depending on $T$ and the data of the problem). Let $N$ denote the integer $T/\Delta t$, so then $t^N = N \Delta t = T$. We have tested to see if the temporal error behavior for (21) applied to the two phase Stefan problem is consistent with (25) by fixing $T$, taking $J$ "large", and calculating numerical values for $C$ for several values of $\Delta t$. Let $v(x, t^N)$ denote the exact temperature solution at time $t^N = T$, and let $\{ V_j^N \}$ denote the approximate temperature solution values at the grid points $\{ p_j \}$ at time $t^N$ obtained via (21) [i.e. $V_j^N$ is defined to be the temperature corresponding (by (6)) to $U_j^N$]. The discrete $L^1$ error $e_1$ is defined to be

$$e_1 = \frac{20}{J-1} \sum_{j=1}^J | V_j^N - v(p_j, t^N) |,$$

(26)

and the numerical value of $C$ corresponding to $N$ (i.e. to $\Delta t = T/N$) is

$$C = e_1 / (\Delta t \ln (1 + T/\Delta t))^{1/2}.$$  

(27)

R.A.I.R.O. Analyse numérique/Numerical Analysis
METHOD FOR SOLVING THE PROBLEM $u_t - \Delta f(u) = 0$ 309

Results of applying the algorithm (21) to the two phase Stefan problem are given in tables III and IV. In all the tables for the Stefan problem $T = (2/3) \times 200,000$. For table III, the standard Crank-Nicolson (C N ) method

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\alpha = 1 \mu$ C N</th>
<th>$\alpha = \mu$ C N</th>
<th>$\alpha = 96 \mu$ C N</th>
<th>APT C N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e_t$</td>
<td>$C$</td>
<td>$e_t$</td>
<td>$C$</td>
</tr>
<tr>
<td>216</td>
<td>9.10</td>
<td>16</td>
<td>10.16</td>
<td>18</td>
</tr>
<tr>
<td>432</td>
<td>6.81</td>
<td>16</td>
<td>6.63</td>
<td>15</td>
</tr>
<tr>
<td>864</td>
<td>4.75</td>
<td>14</td>
<td>4.40</td>
<td>14</td>
</tr>
<tr>
<td>1728</td>
<td>3.46</td>
<td>14</td>
<td>3.06</td>
<td>13</td>
</tr>
<tr>
<td>3456</td>
<td>2.43</td>
<td>14</td>
<td>2.22</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\alpha = \mu$</th>
<th>$\alpha = 65 \mu$</th>
<th>$\alpha = 5 \mu$</th>
<th>$\alpha = 48 \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e_t$</td>
<td>$C$</td>
<td>$e_t$</td>
<td>$C$</td>
</tr>
<tr>
<td>216</td>
<td>11.25</td>
<td>20</td>
<td>9.12</td>
<td>16</td>
</tr>
<tr>
<td>432</td>
<td>8.05</td>
<td>19</td>
<td>6.83</td>
<td>16</td>
</tr>
<tr>
<td>864</td>
<td>5.82</td>
<td>18</td>
<td>4.75</td>
<td>15</td>
</tr>
<tr>
<td>1728</td>
<td>4.19</td>
<td>17</td>
<td>3.46</td>
<td>14</td>
</tr>
<tr>
<td>3456</td>
<td>3.00</td>
<td>17</td>
<td>2.43</td>
<td>14</td>
</tr>
</tbody>
</table>

with $J = 161$ was used to solve (22), results are given for $\alpha = 1 \mu$, 3, 1, and 95 $\mu$, and for comparison, results for the APT method (C N , $J = 161$) are given in the last column. The error behavior appears consistent with (25). The error decreases as $\alpha$ decreases until the abrupt deterioration below $\alpha = \mu$. Results when using the standard implicit method for (22) with $J = 161$ are given in table IV. The correspondence between (C N , $\alpha$) and (implicit, $\alpha/2$) is again observed [this is not surprising since $O(\Delta t)$ is "small" relative to (25)]. The data in table V indicates that for the values of $N$ being considered, the spatial discretization error is relatively negligible even for $J = 41$. We note that the method for the Stefan problem discussed in [2] and [21] corresponds to using a purely explicit method to solve (20). The 3 level Laplace modified Galerkin method (24) when
applied to the Stefan problem in the form (20) produced very erratic behavior with respect to variations in both $\beta$ and $J$ (typical behavior is presented in table VI). Fatal instability occurs with $J = 161$, $\beta = .3$, $N = 3,456$.

As an illustration of the simplicity of using (21) in several space dimensions, numerical results for a two dimensional Stefan problem are given in figure 3. The

Figure 3. -- Numerical solution of a two phase Stefan problem on the region $0 \leq x, y \leq 20$ using the algorithm (21) with $\alpha = k_i / c_i$. In each frame, an $L(S)$ was printed at points where the approximate solution value was $\geq \lambda$ ($\leq 0$). A blank was printed for values between 0 and $\lambda$. The solid line is the exact interface location. The standard ADI method [19] with $\Delta x = \Delta y = 1$, was used to solve (22). For the first four frames the value of $\Delta t$ used was $T/5,184$ ($T = 200,000$). For the frame on the lower right the value of $\Delta t$ was $T/648$.

classical ADI method [19] was used to solve (22). The exact solution at $(x, y, t)$ is given by $u(p \equiv x \cos (30^\circ) - y \sin (30^\circ))$, $t$ where $u$ is the solution in [7], [20] ($p$ is the projection of $(x, y)$ on the line $30^\circ$ below the positive $x$-axis, note also that $\mu = k_i / c_i$ and $z = \hat{z} = r_1 = 0$).
METHOD FOR SOLVING THE PROBLEM $u_t - \Delta f(u) = 0$

Table V

Discrete $L^1$ errors $e_x$ and values of the constant $C$ of (27) for several implementations of the algorithm (21) with $\alpha = \mu$ for the two phase Stefan problem. The Crank-Nicolson method was used to solve (22), with several different values for $J$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$J = 41$</th>
<th>$J = 81$</th>
<th>$J = 161$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e_1$</td>
<td>$C$</td>
<td>$e_1$</td>
</tr>
<tr>
<td>216.</td>
<td>8.64</td>
<td>.15</td>
<td>9.33</td>
</tr>
<tr>
<td>432.</td>
<td>6.44</td>
<td>.15</td>
<td>6.15</td>
</tr>
<tr>
<td>864.</td>
<td>4.12</td>
<td>.13</td>
<td>4.28</td>
</tr>
<tr>
<td>1,728.</td>
<td>3.67</td>
<td>.15</td>
<td>3.12</td>
</tr>
<tr>
<td>3,456.</td>
<td>2.51</td>
<td>.14</td>
<td>2.34</td>
</tr>
</tbody>
</table>

Table VI

Discrete $L^1$ errors for several implementations of the Laplace modified centered equation (24) for the two phase Stefan problem in the form (20).

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\beta = \mu$</th>
<th>$\beta = 0.7\mu$</th>
<th>$\beta = 0.5\mu$</th>
<th>$\beta = 0.4\mu$</th>
<th>$\beta = 0.7\mu$</th>
<th>$\beta = 0.5\mu$</th>
<th>$\beta = 0.5\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$J = 161$</td>
<td>$J = 161$</td>
<td>$J = 161$</td>
<td>$J = 161$</td>
<td>$J = 81$</td>
<td>$J = 81$</td>
<td>$J = 321$</td>
</tr>
<tr>
<td></td>
<td>$e_1$</td>
<td>$e_1$</td>
<td>$e_1$</td>
<td>$e_1$</td>
<td>$e_1$</td>
<td>$e_1$</td>
<td>$e_1$</td>
</tr>
<tr>
<td>216.</td>
<td>48.88</td>
<td>39.22</td>
<td>68.56</td>
<td>34.31</td>
<td>38.41</td>
<td>52.63</td>
<td></td>
</tr>
<tr>
<td>432.</td>
<td>27.34</td>
<td>23.05</td>
<td>28.69</td>
<td>9.78</td>
<td>6.16</td>
<td>41.25</td>
<td></td>
</tr>
<tr>
<td>864.</td>
<td>10.35</td>
<td>3.69</td>
<td>4.16</td>
<td>2.05</td>
<td>1.38</td>
<td>18.39</td>
<td></td>
</tr>
<tr>
<td>1,728.</td>
<td>1.48</td>
<td>1.59</td>
<td>.87</td>
<td>1.40</td>
<td>.92</td>
<td>3.63</td>
<td></td>
</tr>
<tr>
<td>3,456.</td>
<td>.88</td>
<td>.51</td>
<td>.44</td>
<td>.27</td>
<td>.65</td>
<td>.60</td>
<td>1.02</td>
</tr>
</tbody>
</table>

REFERENCES


