

ALAN E. BERGER

HAIM BREZIS

JOËL C. W. ROGERS

**A numerical method for solving the problem**

$$u_t - \Delta f(u) = 0$$

*RAIRO. Analyse numérique*, tome 13, n° 4 (1979), p. 297-312

[http://www.numdam.org/item?id=M2AN\\_1979\\_\\_13\\_4\\_297\\_0](http://www.numdam.org/item?id=M2AN_1979__13_4_297_0)

© AFCET, 1979, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Analyse numérique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## A NUMERICAL METHOD FOR SOLVING THE PROBLEM $u_t - \Delta f(u) = 0$ (\*)

Alan E. BERGER <sup>(1)</sup>, Haim BREZIS <sup>(2)</sup> and Joël C. W. ROGERS <sup>(3)</sup>

Communiqué par P. G. CIARLET

*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  $f: \mathbb{R} \rightarrow \mathbb{R}$  be a non-

---

(\*) Reçu novembre 1978.

<sup>(1)</sup> Applied Mathematics Branch, Code R44, Naval Surface Weapons Center, Silver Spring, Md. U.S.A. This work was supported by the Naval Surface Weapons Center Independent Research Fund and the Office of Naval Research.

<sup>(2)</sup> Université de Paris-VI, Analyse numérique, Paris.

<sup>(3)</sup> Applied Physics Laboratory, Johns Hopkins University, Applied Mathematics Research Group, Laurel, Md. U.S.A. This work was supported by the Office of Naval Research.

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  $S(t)$ . (1)

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

$$\sup_{\Omega} (I + \lambda L)^{-1} \varphi \leq \max \{ 0, \sup_{\Omega} \varphi \} \quad (2)$$

[by sup we mean the essential supremum; if  $\sup \varphi = \infty$ , assumption (2) is empty].

There exists an  $\alpha > 0$  such that

$$\alpha \|u\|_1 \leq \|Lu\|_1 \quad \text{for all } u \in D(L) \quad (3)$$

[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  $Lu = -\Delta u$ ,  $D(L) = \{u \in W_{\delta}^{1,1}(\Omega); Lu \in L^1(\Omega)\}$  (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  $a_{ij}$ ,  $a_i \in C^1(\bar{\Omega})$ ,  $a \in L^\infty(\Omega)$ , both  $a$  and  $(a + \sum_i \partial a_i / \partial x_i)$  are nonnegative almost everywhere on  $\Omega$ , and for some positive constant  $\alpha$ ,

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

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

We are concerned with solving the evolution equation

$$\frac{du}{dt} + Lf(u) = 0, \quad u(0) = u_0. \quad (4)$$

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

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

$$u + \lambda L f(u) = \varphi,$$

and in addition, the mapping  $\varphi \rightarrow u$  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  $\varphi \in L^\infty(\Omega)$ , then  $u_\lambda = (I + \lambda A)^{-1} \varphi$  satisfies  $\|u_\lambda\|_2 \leq \|\varphi\|_2$  and so  $u_\lambda \rightarrow \varphi$  in  $L^2(\Omega)$  as  $\lambda \rightarrow 0$ ). It follows that

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

is a contraction semigroup on  $\overline{D(A)} = L^1(\Omega)$ ;  $S_g(t)u_0$  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 ( $c_i, c_w, k_i, k_w$ );

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

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

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

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

The proper energy balance conditions on the moving interface

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

is

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

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]) :

$$f(u) = \left\{ \begin{array}{l} \bar{z} + k_i(u - r_1)/c_i \text{ for } u \leq r_1, \\ \bar{z} \text{ for } r_1 \leq u \leq r_1 + \lambda, \\ \bar{z} + k_w(u - \lambda - r_1)/c_w \text{ for } u \geq r_1 + \lambda. \end{array} \right\} \quad (7)$$

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  $\sigma_\tau : (0, \infty) \rightarrow (0, \infty)$  be a function such that  $\lim_{\tau \rightarrow 0} \sigma_\tau = 0$ . 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;

$$\frac{u^{k+1} - u^k}{\tau} + \left[ \frac{I - S(\sigma_\tau)}{\sigma_\tau} \right] f(u^k) = 0, \quad u^0 = u_0, \quad (8)$$

that is  $u^{k+1}$  is determined from  $u^k$  by

$$u^{k+1} = F(\tau) u^k, \quad (9 a)$$

where

$$F(\tau) \varphi \equiv \varphi + \frac{\tau}{\sigma_\tau} [S(\sigma_\tau) f(\varphi) - f(\varphi)] \quad \text{for } \varphi \in L^1(\Omega). \quad (9 b)$$

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

$$u_n(t) = u^n = \left[ F\left(\frac{t}{n}\right) \right]^n u_0. \quad (10)$$

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;

$$\mu\tau/\sigma_\tau \leq 1 \quad \text{for each } \tau > 0 \quad (11)$$

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

$L^1(\Omega)$ ; 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  $r \rightarrow r - \tau f(r)/\sigma_\tau$  is nondecreasing [by (11)], it follows that

$$-M - \tau f(-M)/\sigma_\tau \leq u^k - \tau f(u^k)/\sigma_\tau \leq M - \tau f(M)/\sigma_\tau. \quad (12)$$

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

$$f(-M) \leq S(\sigma_\tau) f(u^k) \leq f(M). \quad (13)$$

Combining (12) and (13), one obtains  $-M \leq u^{k+1} \leq M$ . Therefore in performing the iteration (9), we can replace  $f$  by  $\bar{f}$  where  $\bar{f} = f$  for  $-M \leq r \leq M$ ,  $\bar{f} = f(M)$  for  $r \geq M$ , and  $\bar{f} = f(-M)$  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.,

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

*Proof:* Indeed, we have

$$\|F(\tau)\varphi - F(\tau)\psi\|_1 \leq \frac{\tau}{\sigma_\tau} \|f(\varphi) - f(\psi)\|_1 + \|(\varphi - \psi) - \frac{\tau}{\sigma_\tau} (f(\varphi) - f(\psi))\|_1. \quad (14)$$

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

$$\frac{\tau}{\sigma_\tau} |f(r) - f(s)| + |(r - s) - \frac{\tau}{\sigma_\tau} (f(r) - f(s))| = |r - s|. \quad (15)$$

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  $\lambda > 0$ :

$$\psi_\tau \equiv \left( I + \frac{\lambda}{\tau} (I - F(\tau)) \right)^{-1} \varphi \rightarrow \psi \equiv (I + \lambda A)^{-1} \varphi \quad (16)$$

in  $L^1(\Omega)$  as  $\tau \rightarrow 0$ . We have

$$\psi_\tau + \frac{\lambda}{\tau}(I - F(\tau))\psi_\tau = \varphi. \quad (17)$$

Set

$$\psi + \frac{\lambda}{\tau}(I - F(\tau))\psi \equiv \varphi_\tau, \quad (18)$$

then  $\varphi_\tau \rightarrow \psi + \lambda L f(\psi) = \psi + \lambda A \psi = \varphi$  in  $L^1(\Omega)$  as  $\tau \rightarrow 0$  since  $\psi \in D(A)$  implies  $f(\psi) \in D(L)$  and then

$$\frac{I - S(\sigma_\tau)}{\sigma_\tau} f(\psi) \rightarrow L f(\psi) \quad \text{in } L^1(\Omega) \text{ as } \tau \rightarrow 0.$$

Finally, recalling that  $F(\tau)$  is contractive, combining (17) and (18) gives

$$\left(1 + \frac{\lambda}{\tau}\right) \|\psi_\tau - \psi\|_1 \leq \|\varphi - \varphi_\tau\|_1 + \frac{\lambda}{\tau} \|\psi_\tau - \psi\|_1.$$

Hence  $\|\psi_\tau - \psi\|_1 \leq \|\varphi - \varphi_\tau\|_1 \rightarrow 0$  as  $\tau \rightarrow 0$ , 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  $F(\tau)\varphi$  to be

$$\varphi + \frac{\tau}{\sigma_\tau} (J(\sigma_\tau) f(\varphi) - f(\varphi)), \quad (19)$$

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  $L = -\Delta$ );

$$u_t = \Delta f(u) \quad \text{in } \Omega \quad (20a)$$

$$u(p, t=0) = u_0(p) \quad \text{in } \bar{\Omega} \quad (20b)$$

$$u(p, t) = g(p, t) \quad \text{for } p \in \Gamma, \quad t > 0. \quad (20c)$$

Suppose one has approximate solution values  $\{U_j^n\}$  at time  $t^n = n \Delta t$  on a set of  $J$  grid points  $\{p_j\} \subset \bar{\Omega}$  (e. g.  $U_j^0 = u_0(p_j)$ ,  $j = 1, \dots, J$ ). 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  $\sigma_\tau/\tau$ ):

$$\text{set} \quad Q_j^n = \frac{f(U_j^n)}{\alpha}, \quad j = 1, \dots, J, \quad (21 a)$$

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

$$Q_t = \alpha \Delta Q \quad \text{in } \Omega, \quad (22 a)$$

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

$$Q(p_j, t^n) = Q_j^n \quad \text{for } j = 1, \dots, J, \quad (22 c)$$

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

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

and

$$U_j^{n+1} \equiv g(p_j, t^{n+1}) \quad \text{when } p_j \in \Gamma. \quad (21 d)$$

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  $R^1$ . 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  $\bar{U}^n$  are two different starting values, then

$$|U^{n+1} - \bar{U}^{n+1}|_\mu \leq |U^n - \bar{U}^n|_\mu.$$

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



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

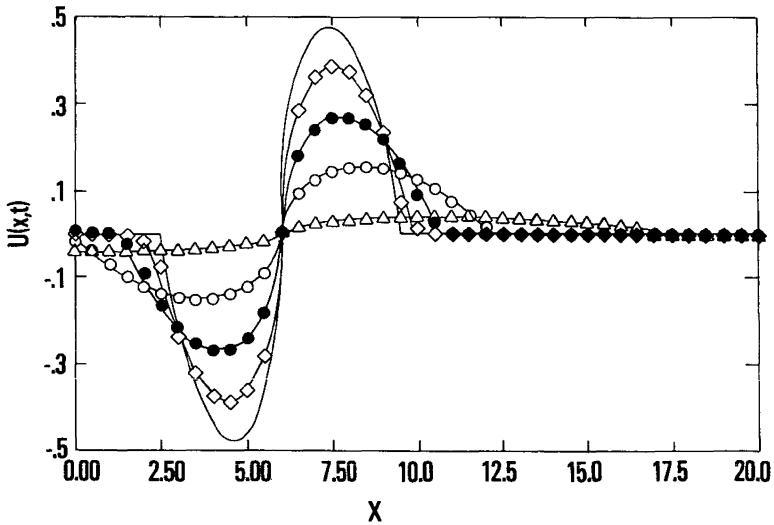


Figure 1. — Exact solution (solid lines) and numerical solution values (points) at  $t=0$ ,  $t=.559$  (diamonds),  $t=2.23$  (solid circles),  $t=8.94$  (open circles), and  $t=143$ . (triangles). The approximate solution values were obtained using the algorithm (21) with  $\alpha=1.$ , and using the standard Crank-Nicolson finite difference method with 41 grid points and  $\Delta t = 143./1024$  to solve (22).

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

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

$$E_1 = C_1 (\Delta t)^{\bar{\rho}}, \tag{23}$$

where  $C_1$  is a constant independent of  $\Delta t$ . Then the numerical rate of convergence  $\bar{\rho}$  computed from the results of using two successive values of  $N$  (i.e.  $N$  and  $\bar{N} = 2N$  with corresponding errors  $E_1$  and  $\bar{E}_1$ ) is

$$\bar{\rho} \equiv (\ln(E_1) - \ln(\bar{E}_1)) / \ln(2).$$

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  $\bar{\rho}$  remains roughly constant) as  $\alpha$  is decreased (until instability develops). The last result in table I was obtained by

TABLE I  
Discrete  $L^1$  errors and numerical convergence rates  
for several implementations of the algorithm (21)

$N$	$\alpha = 1$ C.N.		$\alpha = .75$ C.N.		$\alpha = .5$ C.N.		$\alpha = .5$ Implicit	
	Error $E_1$	Rate $\bar{\rho}$	Error $E_1$	Rate $\bar{\rho}$	Error $E_1$	Rate $\bar{\rho}$	Error $E_1$	Rate $\bar{\rho}$
4	.2749	.80	.1666	.70	.0996	.88	.2747	.80
8	.1580	.85	.1026	.82	.0541	.82	.1578	.85
16	.0875	.88	.0583	.86	.0306	.84	.0873	.88
32	.0475	.90	.0321	.89	.0171	.87	.0474	.90
64	.0256	.91	.0174	.91	.0093	.91	.0255	.91
128	.0136	.93	.0093	.93	.0050	.92	.0136	.93
256	.0071		.0049		.0026		.0071	

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  $(\alpha = .5, \text{C.N.})$  ( $J = 433$  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  $\{U_j^n; p_j \in T\}$ , 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 (C.N.,  $\alpha = \bar{\alpha}$ ) is identical to taking (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 (C.N.,  $\alpha = \bar{\alpha}$ ) and (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;

$J = 55$	109	217	433
$E_1 = .0177$	.0044	.0042	.0026

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  $D_x^2 U_j \equiv (U_{j-1} - 2U_j + U_{j+1})/\Delta x^2$  and  $\Delta x$  is the (uniform) mesh length];

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

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

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

$J = 55$	109	217	433
$E_1 = .0163$	.0028	.0023	.0005

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)

$N$	$\beta = .1$		$\beta = .5$		$\beta = .25$		$\beta = .2$	
	Error $E_1$	Rate $\rho$	Error $E_1$	Rate $\rho$	Error $E_1$	Rate $\rho$	Error $E_1$	Rate $\rho$
4 . . . . .	1.572		.7560		.3175		.2537	
8 . . . . .	.4853	1.7	.2160	1.8	.1179	1.4	.1043	1.3
16 . . . . .	.1637	1.6	.0854	1.3	.0485	1.3	.0475	1.1
32 . . . . .	.0609	1.4	.0349	1.3	.0212	1.2	.0193	1.3
64 . . . . .	.0224	1.4	.0139	1.3	.0084	1.3	.0072	1.4
128 . . . . .	.0079	1.5	.0048	1.5	.0024	1.8	.0018	2.0
256 . . . . .	.0021	1.9	.0009	2.4	.0005	2.3	.0005	2.0

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  $z = \bar{z} = r_1 = 0$ ). 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

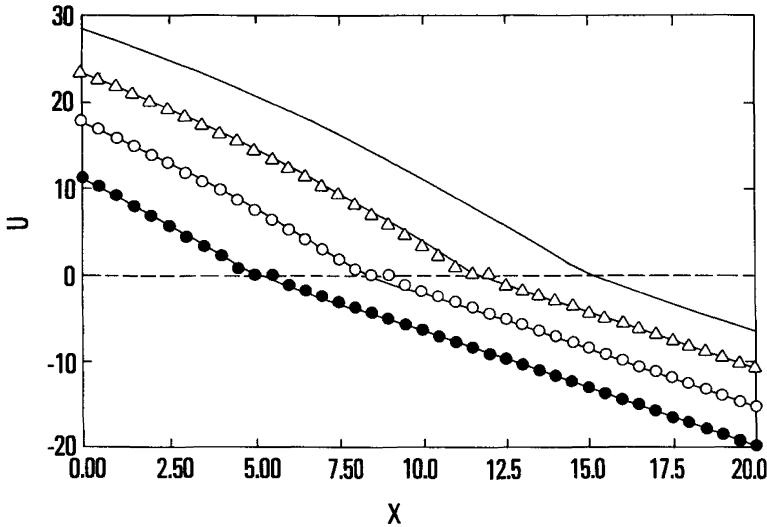


Figure 2. — Exact solution (temperature) of a two phase Stefan problem (solid lines) and numerical solution values (points) at  $t = 0$ ,  $t = T/3$  (triangles),  $t = 2T/3$  (open circles), and  $t = T$  (solid circles) where  $T = 200,000$ . The approximate solution values were obtained using the algorithm (21) with  $\alpha = k_i/c_i$ , and using the standard Crank-Nicolson finite difference method with 41 grid points and  $\Delta t = T/2,595$  to solve (22).

error in the temperature value and in the interface location could be bounded by

$$C(\Delta t \cdot \ln(1 + T/\Delta t))^{1/2}, \tag{25}$$

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  $t^N \equiv N \Delta t = T$ . 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

$$e_1 = \frac{20}{J-1} \sum_{j=1}^J |V_j^N - v(p_j, t^N)|, \tag{26}$$

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

$$C \equiv e_1 / (\Delta t \cdot \ln(1 + T/\Delta t))^{1/2}. \tag{27}$$

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  $T = (2/3)200,000$ . For table III, the standard Crank-Nicolson (C N) method

TABLE III

Discrete  $L^1$  errors  $e_1$  and values of the constant  $C$  of (27) for several implementations of the algorithm (21) and an implementation of the APT method for the two phase Stefan problem

N	$\alpha = 1/3 \mu$ C N		$\alpha = \mu$ C N		$\alpha = 96 \mu$ C N		APT C N	
	$e_1$	C	$e_1$	C	$e_1$	C	$e_1$	C
216	9 10	16	10 16	18	84 6	1 5	12 26	21
432	6 81	16	6 63	15	89 8	2 1	8 69	20
864	4 75	15	4 40	14	88 9	2 8	5 87	18
1 728	3 46	14	3 06	13	88 2	3 7	4 20	18
3 456	2 43	14	2 22	13	82 1	4 6	2 99	17

TABLE IV

Discrete  $L^1$  errors  $e_1$  and values of the constant  $C$  of (27) for several implementations of the algorithm (21) for the two phase Stefan problem. The standard implicit method with  $J = 161$  was used to solve (22)

N	$\alpha = \mu$		$\alpha = 65 \mu$		$\alpha = 5 \mu$		$\alpha = 48 \mu$	
	$e_1$	C	$e_1$	C	$e_1$	C	$e_1$	C
216	11 25	20	9 12	16	10 25	18	92 0	1 6
432	8 05	19	6 83	16	6 76	16	92 3	2 1
864	5 82	18	4 75	15	4 39	14	89 7	2 8
1 728	4 19	17	3 46	14	3 02	13	87 4	3 6
3 456	3 00	17	2 43	14	2 22	13	82 1	4 6

with  $J = 161$  was used to solve (22), results are given for  $\alpha = 1/3, 1$ , 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  $J = 161$ ,  $\beta = .3 \mu$ ,  $N = 3,456$ .

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

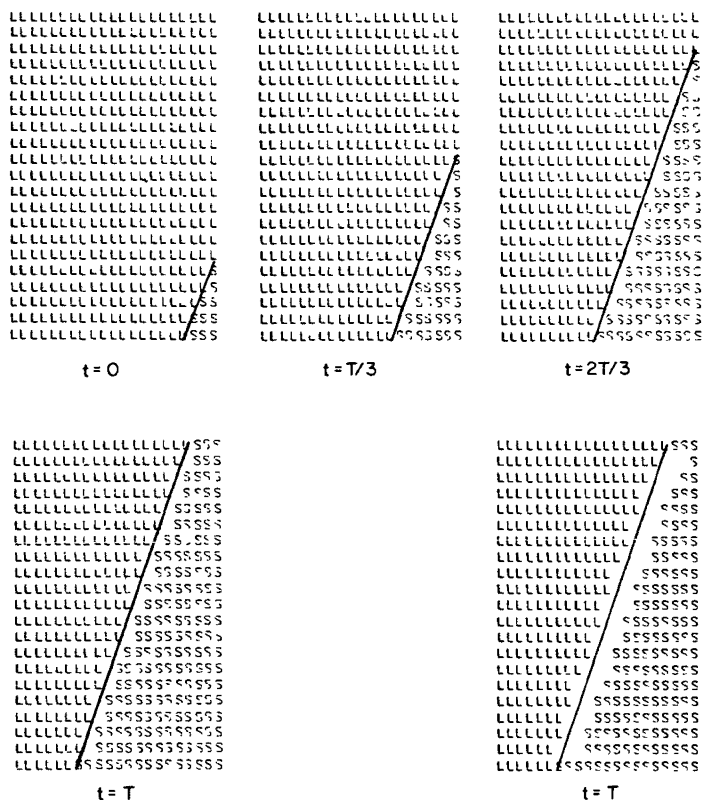


Figure 3. — Numerical solution of a two phase Stefan problem on the region  $0 \leq x, y \leq 20$  using the algorithm (21) with  $\alpha = k_i/c_i$ . In each frame, an  $L(S)$  was printed at points where the approximate solution value was  $\geq \lambda$  ( $\leq 0$ ). A blank was printed for values 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  $u(p \equiv x \cos(30^\circ) - y \cdot \sin(30^\circ), t)$  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  $z = \dot{z} = r_1 = 0$ ).

TABLE V

Discrete  $L^1$  errors  $e_1$ , 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  $J$ .

N	J=41		J=81		J=161	
	$e_1$	C	$e_1$	C	$e_1$	C
216. ....	8.64	.15	9.33	.16	10.16	.18
432. ....	6.44	.15	6.15	.14	6.63	.15
864. ....	4.12	.13	4.28	.13	4.40	.14
1,728. ....	3.67	.15	3.12	.13	3.06	.13
3,456. ....	2.51	.14	2.34	.13	2.22	.13

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).

N	$\beta = \mu$ J=161	$\beta = .7\mu$ J=161	$\beta = .5\mu$ J=161	$\beta = .4\mu$ J=161	$\beta = .7\mu$ J=81	$\beta = .5\mu$ J=81	$\beta = .5\mu$ J=321
	$e_1$	$e_1$	$e_1$	$e_1$	$e_1$	$e_1$	$e_1$
216. ....	48.88	57.30	39.22	68.56	34.31	38.41	52.63
432. ....	27.34	27.81	23.05	28.69	9.78	6.16	41.25
864. ....	10.35	5.47	3.69	4.16	2.05	1.38	18.39
1,728. ....	1.48	1.29	1.59	.87	1.40	.92	3.63
3,456. ....	.88	.51	.44	.27	.65	.60	1.02

REFERENCES

1. D. G. ARONSON, *Regularity Properties of Flows through Porous Media*, S.I.A.M., J. Appl. Math., Vol. 17, 1969, pp. 461-467.
2. D. R. ATTHEY, *A finite Difference Scheme for Melting Problems*, J. Inst. Math. Appl., Vol. 13, 1974, pp. 353-366.
3. J. R. BAILLON and B. MERCIER, *Convergence of Approximations to Nonlinear Semigroups*, Ecole Polytechnique Internal Report Number 24, December 1977.
4. Ph. BENILAN, *Equations d'évolution dans un espace de Banach quelconque et applications*, Thèse, Publications Mathématiques d'Orsay, n° 25, 1972.
5. A. E. BERGER, *The Truncation Method for the Solution of a Class of Variational Inequalities*, R.A.I.R.O., Analyse Numérique, Vol. 10, 1976, pp. 29-42.
6. A. E. BERGER, M. CIMENT and J. C. W. ROGERS, *Numerical Solution of a Diffusion Consumption Problem with a Free Boundary*, S.I.A.M. J. Numer. Anal., Vol. 12, 1975, pp. 646-672.



7. A. E. BERGER, M. CIMENT and J. C. W. ROGERS, *Numerical Solution of a Stefan Problem by a Technique of Alternating Phase Truncation*, Séminaires IRIA, Analyse et contrôle des systèmes, Rocquencourt, Institut de Recherche d'Informatique et d'automatique, 1976, pp. 21-34.
8. A. E. BERGER and R. S. FALK, *An Error Estimate for the Truncation Method for the Solution of Parabolic Obstacle Variational Inequalities*, Math. Comp., Vol. 31, 1977, pp. 619-628.
9. H. BRÉZIS, *On Some Degenerate Nonlinear Parabolic Equations*, *Nonlinear Functional Analysis*, F. BROWDER, Ed., Proc. Symp. in pure math., Vol. 18, A.M.S., 1970, pp. 28-38.
10. H. BRÉZIS and A. PAZY, *Convergence and Approximation of Semigroups of Nonlinear Operators in Banach Spaces*, J. Func. Anal., Vol. 9, 1972, pp. 63-74.
11. H. BRÉZIS and W. A. STRAUSS, *Semi-Linear Second Order Elliptic Equations in  $L^1$* , J. Math. Soc. Japan, Vol. 25, 1973, pp. 565-590.
12. M. G. CRANDALL, *An Introduction to Evolution Governed by Accretive Operators*, Dynamical systems vol. 1, Proc. of the Int. Symp. on Dyn. Sys. at Brown U. August 12-16, 1974, L. CESARI, J. K. HALE and J. P. LASALLE, Eds, New York, Academic Press, 1976, pp. 131-165.
13. M. G. CRANDALL, *Semigroups of Nonlinear Transformations in Banach Spaces*, *Contributions to nonlinear Functional Analysis*, E. ZARANTONELLO, Ed., New York, Academic Press, 1971, pp. 157-179.
14. M. G. CRANDALL and T. M. LIGGETT, *Generation of Semi-groups of Nonlinear Transformations on General Banach Spaces*, Amer. J. Math., Vol. 93, 1971, pp. 265-298.
15. A. DAMLAMIAN, *Some Results on the Multi-phase Stefan Problem*, Comm. on P.D.E., Vol. 2, 1977, pp. 1017-1044.
16. J. DOUGLAS Jr. and T. DUPONT, *Alternating-Direction Galerkin Methods on Rectangles*, *Numerical Solution of Partial Differential Equations-II*, SYNSPADE 1970, B. HUBBARD, Ed., New York, Academic press, 1971, pp. 133-214.
17. B. H. GILDING and L. A. PELETIER, *On a Class of Similarity Solutions of the Porous Media Equation*, J. Math. Anal. Appl., Vol. 55, 1976, pp. 351-364.
18. B. H. GILDING and L. A. PELETIER, *On a Class of Similarity Solutions of the Porous Media Equation II*, J. Math. Anal. Appl., Vol. 57, 1977, pp. 522-538.
19. D. W. PEACEMAN and H. H. RACHFORD Jr., *The Numerical Solution of Parabolic and Elliptic Differential Equations*, J. Soc. Indust. Appl. Math., Vol. 3, 1955, pp. 28-41.
20. J. C. W. ROGERS, A. E. BERGER and M. CIMENT, *The Alternating Phase Truncation Method for Numerical Solution of a Stefan Problem*, to appear in S.I.A.M. J. Num. Anal.
21. M. ROSE, *A Method for Calculating Solutions of Parabolic Equations with a Free Boundary*, Math. Comp., Vol. 14, 1960, pp. 249-256.