NUMERICAL MODELLING OF SOLIDIFICATION

A. D. Miller

§1. INTRODUCTION

In this paper we shall report on some preliminary investigations into numerical schemes for the two phase Stefan problem in two space dimensions. Interest in this problem has arisen in the course of trying to model the cooling and solidification of foundry castings. In the foundry industry a major design objective is the avoidance of shrinkage cavities and other forms of porosity in castings. These kinds of flaws can seriously affect the mechanical strength of the casting. They can however largely be avoided by ensuring that the solidification of the molten casting takes place in such a way that at any time the regions of still molten material remain connected to the feed points. If this is the case, molten material is able to flow freely from a feed point throughout the still molten region and make up for any local shrinkage caused by contraction of the material as it solidifies. Two of the most important design choices that are available to try to achieve this goal are the placement of the feed points (called "risers" in foundry terminology), and the positioning of "chills". Chills are metal inserts placed in the mould to remove heat very quickly from the casting, thus increasing the speed of local solidification.

A model casting arrangement is shown in Fig. 1. This could represent the cross section of a casting of an I-beam. The casting is being fed from the top. There is clearly the possibility that the narrow neck A could solidify before the lower part of the beam completely solidifies. If this were to occur, there would be an isolated still molten region in the lower part of the beam. No molten material could subsequently reach it from the feed point, and shrinkage

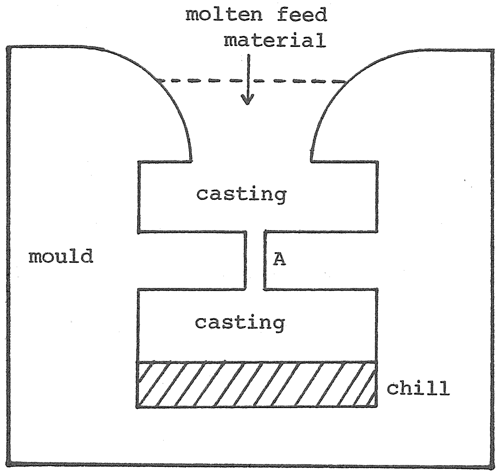

FIGURE 1

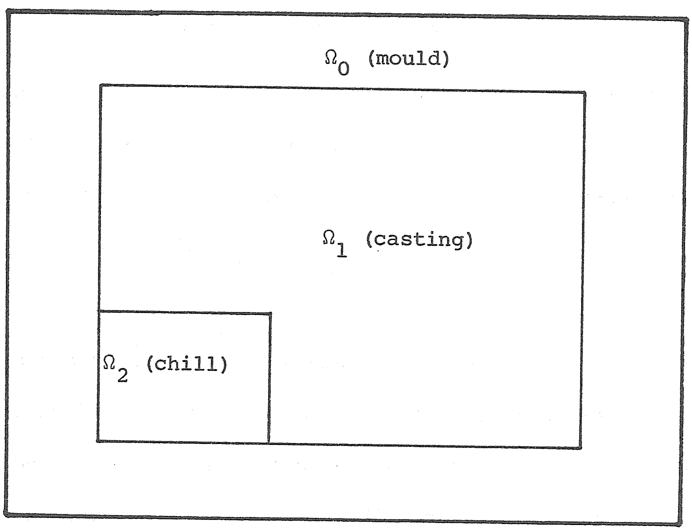

FIGURE 2

cavities would arise during later stages of the solidification. However, a chill placed against the lower part of the beam may hasten the local rate of solidification sufficiently to ensure that no such isolated molten region occurs.

In addition to this connectivity aspect, there is extensive experimental evidence that other features of the thermal history of the casting (e.g. rate of solidification) may also have a significant influence on the structure and mechanical properties of the resulting solid. For further details of the metallurgical aspects of the problem see [2].

From what has been said, the ability to predict the thermal history of a casting would seem a useful skill in foundry applications, particularly as a means of optimizing the design and placement of risers and chills. In setting up an initial mathematical model with this application in mind we have made a number of physical simplifications:

(a) The problem is spatially two-dimensional. It is posed on a bounded domain $\Omega \subset \mathbb{R}^2$. Ω is made up of subdomains $\Omega_i \subset \Omega$ $(i=0,\dots,N)$, $\bar{\Omega} = \bigcup_{i=0}^{N} \bar{\Omega}_i$. These subdomains represent the mould (Ω_0), the casting (Ω_1) and the chills $(\Omega_2,\dots,\Omega_N)$. Furthermore we suppose that each of the Ω_i are rectilinear in shape with sides parallel to the coordinate axes. (see Fig.2) Typically this set-up represents the section of a "long" right-angled prism with end effects neglected.

(b) Heat flow and phase changes are the only physical processes being considered. In particular, we do not consider any form of mass transfer. Although from what was said above some mass transfer effects such as the change of density after solidification (i.e. shrinkage) and the resulting flow of the molten material are very important as far as the mechanical properties of the casting are concerned, it does not seem unreasonable to suppose that they only have minor influences on the thermal history of the casting.

(c) Heat flow is governed by two processes. Firstly, in the interior of Ω_j $(j=0,\ldots,N)$, the rate of heat flow vector is given by

$$\underset{\sim}{q} = - k_j \nabla u$$

where u is the temperature and k_j is the conductivity. Secondly, on an interface between Ω_i and Ω_j $(i,j=0,\ldots,N;i\neq j)$ the rate of heat flow from Ω_i into Ω_j is given by

$$g_{ij}(u^{(i)}-u^{(j)})$$

where $u^{(i)} - u^{(j)}$ is the temperature jump across the interface and $g_{ij} = g_{ji}$ is the interface conductance.

(d) In the casting (Ω_1) solidification takes place at a fixed temperature u^* and is accompanied by the liberation of specific latent heat L . There are more sophisticated models of phase changes than this. They take account variously of nucleation, change in composition of alloys during solidification etc. (see [2]). The simple, classical description that we have chosen is thought to apply reasonably well, at least on a gross scale, in the case of some pure metals.

(e) In each of Ω_j $(j=0,2,\ldots,N)$ the material thermal parameters of density (ρ_j) , specific heat (c_j) and conductivity (k_j) are constant, that is, temperature independent. In the casting Ω_1 , ρ_1 and k_1 are constant whereas c_1 is assumed to be a constant in each phase separately. In addition, the interface conductances g_{ij} are all assumed to be constants $(i,j=0,\ldots,N;i\neq j)$.

(f) On the boundary of Ω, $\partial\Omega$, we shall suppose that the temperature is fixed at the ambient temperature which, without loss of generality, we take to be 0 .

Some of these simplifications may seem rather severe. However it should be borne in mind that our objective can really be no more than to obtain a qualitative impression of the effects of different casting designs. The precision with which the thermal properties of the materials involved are known does not seem to be great and may not justify the use of more complicated models.

In §2 we briefly describe two mathematical formulations of the phase change phenomenon. One of these formulations, the enthalpy formulation, is further considered in §3 where two discretizations based upon it are introduced. One of these is explicit, the other implicit. We discuss some features of these two discretizations in §4 and §5 and prove some L_1 stability results.

§2. MATHEMATICAL FORMULATION OF THE PROBLEM IN Ω_1 (MOLTEN/SOLID REGION)

In the mould region Ω_0 and the chill regions $\Omega_2,\ldots,\Omega_N$ our model is governed by the classical heat equation. However in the region Ω_1 we need to take account of the possible change of phase. A number of mathematical formulations have been suggested for handling this phenomenon. We shall briefly mention two such methods. The second of these is the one we shall be concerned with in this paper.

(a) Moving Boundary Formulation :

In this formulation it is supposed that at any time Ω_1 can be partitioned into two (a priori unknown) subregions : Ω_m , corresponding to the molten phase, and Ω_s , corresponding to the solid phase. Let Γ denote the (unknown) inter-phase boundary $\partial\Omega_m \cap \partial\Omega_s$ (see Fig.3). In Ω_m and Ω_s heat conduction is the only physical process occurring, so

$$\rho_1 c_{1,s} \frac{\partial u}{\partial t} = k_1 \nabla^2 u \qquad \text{in } \Omega_s$$

$$\rho_1 c_{1,m} \frac{\partial u}{\partial t} = k_1 \nabla^2 u \qquad \text{in } \Omega_m$$

are the appropriate heat balance equations where $c_{1,s}$ and $c_{1,m}$ are the specific heats in Ω_s , Ω_m respectively. Heat balance across Γ is expressed by the inter-phase condition
condition

$$\begin{cases} u = u^* \quad \text{on} \quad \Gamma \\ \underset{\sim}{q}^{(s)}\cdot\hat{n} - \underset{\sim}{q}^{(m)}\cdot\hat{n} = -\rho_1 L\, v \quad \text{on} \quad \Gamma , \end{cases} \tag{2.1}$$

where $\underset{\sim}{q}^{(\alpha)} = -k_1\nabla u \; (\alpha=s,m)$ is the rate of heat flow vector in Ω_α , L is the specific latent heat of solidification and v is the velocity of Γ (thought of as evolving along the unit normal $\hat{n}$) .

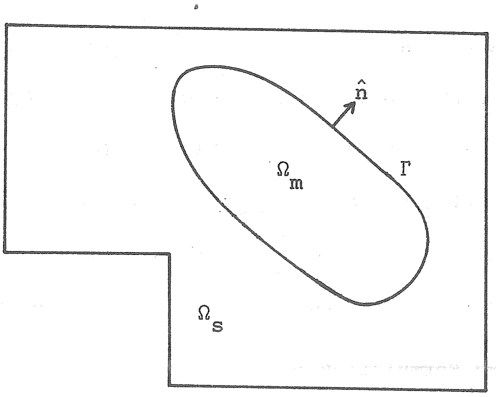

FIGURE 3

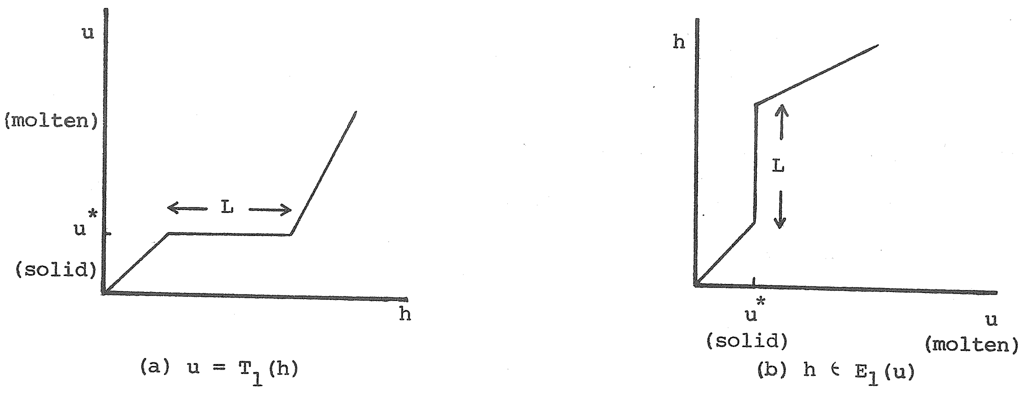

FIGURE 4

Physically this condition states that any imbalance in heat flow across Γ must be accounted for by the liberation of latent heat and consequent movement of the inter-phase boundary.

One point to note about this formulation is that the existence of Ω_m, Ω_s and a "smooth" inter-phase boundary Γ (smooth, so that $\hat{n}$, v etc. can be defined in a reasonable way) is in fact a physical assumption. Indeed, it need not be true. Instances where a phase suddenly appears or disappears, or where an originally connected molten region breaks up into a number of disconnected pieces are common occurrences in the physical process we are seeking to model. Admittedly, away from such "singularities" it is often physically plausible that (2.1) should apply, while extra conditions can be added to (2.1) to deal with the singularities. However, all this is certainly complicating the formulation of the problem. More critical is the fact that even with these modifications the model may still not be physically realistic in some settings. If, for instance, there is a body heat source, (due, for example, to electric current passing through the material or absorption of radiation) there may develop a proper two-dimensional (in our setting) region which is at the phase change temeprature u^*. In such a region the heat balance relation will involve only the body heat source and the absorption of latent heat. To persevere with a moving boundary formulation in this setting would demand adding to Ω_m and Ω_s this so called "mushy" region together with conditions governing heat balance across all the possible boundary combinations for these regions.

Apart from the above physical objections to a moving boundary formulation, from a computational viewpoint, working directly with (2.1) in any more than one space dimension does not appear easy. Even more so if it has to be supplemented with special conditions to handle the possible singularities etc. mentioned earlier.

(b) Enthalpy Formulation :

This formulation is based upon treating the specific enthalpy as the primary quantity of interest. Heat flow is still driven by temperature gradients, but with temperature now being related to enthalpy by a non-linear relation. In the enthalpy formulation no explicit mention is made of the "inter-phase region" (in particular, no a priori assumptions concerning its character are made). Whatever conditions apply at the phase interface are "natural interface conditions" in the enthalpy formulation. Since the enthalpy can be expected to be discontinuous at the phase interface one cannot expect a differential formulation to hold in a classical sense. Some kind of weak formulation needs to be considered. To motivate one such formulation consider the integral relation

$$\int_A \rho \frac{\partial h}{\partial t} = - \int_{\partial A} \underset{\sim}{q} \cdot \hat{n} + \int_A Q \ , \tag{2.2}$$

expressing the heat balance in an arbitrary region $A \subseteq \Omega_1$. Here h is the (specific) enthalpy, $\hat{n}$ is the outward unit normal to ∂A , as usual $\underset{\sim}{q}$ is the rate of heat flow vector, and Q is a body heat source term. Proceeding in the usual (informal) way, we approximate any $\phi \in C_0^\infty(\Omega_1)$ by step functions. After making use of (2.2) we are led in the limit to

$$\int_{\Omega_1} \rho_1 \frac{\partial h}{\partial t} \phi = \int_{\Omega_1} \underset{\sim}{q} \cdot \nabla\phi + \int_{\Omega_1} Q\phi \ . \tag{2.3}$$

This is the basis of our weak enthalpy formulation together with the classical heatflow law

$$\underset{\sim}{q} = - \underline{k}_1 \nabla u$$

and the temperature-enthalpy relation

$$u = T_1(h) \ .$$

In accord with our earlier assumptions, T_1 is piecewise constant as shown in Fig. 4a. The part of the curve where $u < u^*$ corresponds to the solid phase while $u > u^*$ corresponds to the molten phase. (The slopes of T_1 in these parts are simply $\frac{1}{c_{1,s}}$ and $\frac{1}{c_{1,m}}$ respectively.) The flat part of the curve at $u = u^*$ corresponds to the phase transition, with the (specific) latent heat L. Later on we shall also need the "inverse" of T_1. Clearly this is not single-valued. We shall consider it as a set valued function E_1 as illustrated in Fig. 4(b).

In that (2.3) is posed on the fixed domain Ω_1, it is referred to as a fixed domain formulation to contrast it with formulations like (a). There are other fixed domain formulations. One is based on defining a new quantity called the freezing index. For details see [4].

Let us mention that there is an even weaker formulation than (2.3), in which the test function $\varphi = \varphi(x,t) \in C_0^\infty(\Omega_1 \times (0,T))$. (2.3) is integrated over the time interval $(0,T)$, and a formal integrations by parts with respect to time is carried out on the left hand side. This gives

$$-\int_0^T \int_{\Omega_1} \rho h \frac{\partial \phi}{\partial t}\, dxdt = \int_0^T \int_{\Omega_1} \underset{\sim}{q} \cdot \nabla_x \phi \, dxdt + \int_0^T \int_{\Omega_1} Q\phi \, dxdt \qquad (2.4)$$

This is the weak formulation that is often considered in theoretical discussions of the Stefan problem ([6],[7],[8]).

§3. THE DISCRETIZATION OF THE ENTHALPY FORMULATION

To further simplify the exposition we shall suppose from now on that $N = 1$, that is there are only two subdomains, Ω_0 (the mould) and Ω_1 (the casting). This arrangement displays all the features of the more general problem. In the light of this simplification there is only one kind of interface, and we shall write g in place of $g_{01} = g_{10}$ for the interface conductance.

An enthalpy formulation is trivially possible in the mould region Ω_0 where no phase change occurs. In this case the enthalpy-temperature relations are simply linear functions, $h = E_0(u) = c_0 u$ and $u = T_0(h) = \frac{h}{c_0}$.

Introducing the interface conditions into the enthalpy formulation leads to the following tentative weak formulation of our problem : Find functions u and h defined on Ω for all time $t \geq 0$ which satisfy

(i) $u = 0$, $h = 0$ on $\partial\Omega$

(ii) $h(\cdot,0) = h^0$ on Ω

(iii) $m(\frac{\partial h}{\partial t},\phi) = - a(u,\phi) + \ell(Q,\phi)$

for all test functions ϕ which are defined on Ω, vanish on $\partial\Omega$ and are "smooth" on Ω_0 and Ω_1. Here we have used the notation

$$m(\frac{\partial h}{\partial t},\phi) = \sum_j \int_{\Omega_j} \rho_j \frac{\partial h}{\partial t} \phi$$

$$a(u,\phi) = \sum_j \left[\int_{\Omega_j} k_j \nabla u \nabla \phi - \sum_{i \neq j} \int_{\partial\Omega_0 \cap \partial\Omega_1} g(u^{(i)} - u^{(j)})\phi^{(j)}\right]$$

$$\ell(Q,\phi) = \sum_j \int_{\Omega_j} Q\phi$$

(iv) $u(x,t) = T_j(h(x,t))$ for $x \in \Omega_j$, $t \geq 0$.

This is the weak formulation of the problem upon which our discretization will be based.

As a first step towards describing this discretization we place a uniform square finite element mesh on Ω . Let δ be the mesh size. (We suppose that Ω_0 and Ω_1 consist of whole elements) Let F_δ be the space of all functions v which satisfy

(i) v is bilinear on each element of the mesh

(ii) the restriction of v to Ω_j lies in $C^0(\Omega_j)$ $(j=0,1)$.

Note that functions in F_δ may be discontinuous across $\partial\Omega_0 \cap \partial\Omega_1$. Let

$$F_\delta^0 = \{v \in F_\delta : v = 0 \text{ on } \partial\Omega\}$$

For any mesh point $\bar{x}$ with $\bar{x} \in \bar{\Omega}_j$ let $\psi(\bar{x},j)$ be the (unique) function in F_δ which satisfies

(i) $\psi(\bar{x},j) = 0$ in $\Omega - \Omega_j$

(ii) if $\bar{y}$ is a mesh point, $\bar{y} \in \bar{\Omega}_j$, then

$$\psi(\bar{x},j)(\bar{y}) = \begin{cases} 1 & \text{if } \bar{y} = \bar{x} \\ \\ 0 & \text{otherwise .} \end{cases}$$

(The index j of the mesh point pair $(\bar{x},j)$ only need be given if $\bar{x}$ is on the interface $\partial\Omega_0 \cap \partial\Omega_1$. Otherwise, there is no ambiguity if it is dropped, and we shall often do so.) The set of all such $\psi(\bar{x},j)$ constitutes a local basis for F_δ , while the set of all $\psi(\bar{x},j)$ where $\bar{x} \notin \partial\Omega$ forms a basis for F_δ^0 . If v lies in F_δ we shall denote its coordinates relative to these basis functions by $v(\bar{x},j)$, that is

$$v = \sum_{(\bar{x},j)} v(\bar{x},j)\psi(\bar{x},j) .$$

We also need to discretize time. We shall break up $[0,\infty)$ into equal time intervals of duration Δ .

Finally we define a quadrature operator I_δ in both one and two dimensions as follows : For any sufficiently smooth function f

(i) If γ is an edge of an element of the mesh

$$\underset{\gamma}{I_\delta} f = \frac{\delta}{2} \sum_{z} f(z) .$$
$$(z \text{ an endpoint of } \gamma)$$

If $\Gamma = \bigcup_{j=1}^{N} \gamma_j$, where the γ_j are (distinct) edges of elements of the mesh

$$\underset{\Gamma}{I_\delta} f = \sum_{j=1}^{N} \underset{\gamma_j}{I_\delta} f .$$

(ii) If S is an element of the mesh

$$\underset{S}{I_\delta} f = \frac{\delta^2}{4} \sum_{z} f(z)$$
$$(z \text{ a vertex of } S) .$$

where $f(z)$ is the limit from within S .

If $A = \bigcup_{i=1}^{N} S_i$, where the S_i are (distinct) elements of the mesh

$$\underset{A}{I_\delta} f = \sum_{i=1}^{N} \underset{S_i}{I_\delta} f .$$

Associated with these quadrature operators are the discrete norms defined for $v \in F_\delta$ by

$$\|v\|_{p,\delta} = \left(\underset{\Omega}{I_\delta}(|v|^p) \right)^{1/p} \qquad 1 \le p < \infty$$

We are now in a position to describe the two discretizations that we wish to consider : Find $U^n, H^n \in F^0_\delta$ $n = 0,1,2,\ldots$ such that

(i) $U^n(\bar{x},j) = T_j(H^n(\bar{x},j))$

(ii-E) Explicit method :

$$m_\delta\left(\frac{H^{n+1}-H^n}{\Delta}, \phi\right) = - a_\delta(U^n,\phi) + \ell_\delta(Q^n,\phi) \qquad \forall\phi \in F^0_\delta\ , \quad n = 1,2,\ldots \qquad (3.1.E)$$

(ii-I) Implicit method :

$$m_\delta\left(\frac{H^{n+1}-H^n}{\Delta}, \phi\right) = - a_\delta(U^{n+1},\phi) + \ell_\delta(Q^n,\phi) \qquad \forall\phi \in F^0_\delta\ , \quad n = 1,2,\ldots \qquad (3.1.I)$$

where $$m_\delta(H,\phi) = \sum_j \underset{\Omega_j}{I_\delta}(\rho_j\ H\phi)$$

$$a_\delta(U,\phi) = \sum_j \left[\underset{\Omega_j}{I_\delta}(k_j \nabla U \nabla \phi) - \sum_{i \ne j} \underset{\partial\Omega_0 \cap \partial\Omega_1}{I_\delta} (g(U^{(i)} - U^{(j)})\phi^{(j)}) \right]$$

$$\ell_\delta(Q^n,\phi) = \sum_j \underset{\Omega_j}{I_\delta}(Q(\cdot,n\Delta)\phi)$$

(iii) U^0 , H^0 are specified initial data (satisfying (i)).

§4. THE EXPLICIT METHOD

Existence and Uniqueness :

Letting ϕ in (3.1.E) run through the basis functions $\psi(\bar{x},j)$ of F_δ^0 we obtain

$$H^{n+1}(\bar{x},j)m_\delta(\psi(\bar{x},j),\psi(\bar{x},j))$$

$$= m_\delta(H^n,\psi(\bar{x},j) - \Delta a_\delta(U^n,\psi(\bar{x},j)) + \Delta\ell_\delta(Q^n,\psi(\bar{x},j))$$

which may be explicitly solved for $H^{n+1}(\bar{x},j)$ in terms of H^n, U^n and Q^n. This together with $U^{n+1}(\bar{x},j) = T_j(H^{n+1}(\bar{x},j))$ provides the (unique) solution of the explicit discretization.

Uniform Boundedness :

Theorem 4.1 $\exists c > 0$ such that if U^n, H^n are solutions of (3.1.E) then

$$\left(\sum_{k=0}^{n} \Delta\left\|\frac{U^{k+1}-U^k}{\Delta}\right\|_{2,\delta}^2\right)^{\frac{1}{2}} + \|U^{n+1}\|_{2,\delta} + \|\nabla U^{n+1}\|_{2,\delta}$$
$$\le c\left[\|U^0\|_{2,\delta} + \|\nabla U^0\|_{2,\delta} + \left(\sum_{k=0}^{n} \Delta\|Q^k\|_{2,\delta}^2\right)^{\frac{1}{2}}\right] \tag{4.1}$$

provided $$\lambda\frac{\Delta}{\delta^2} + \mu\frac{\Delta}{\delta} \le 1 \tag{4.2}$$

where $$\lambda = \frac{4\max(k_0,k_1)}{\min(\rho_0,\rho_1)\min(c_0,c_{1,m},c_{1,s})}$$

and $$\mu = \frac{4g}{\min(\rho_0,\rho_1)\min(c_0,c_{1,m},c_{1,s})}$$

Proof The proof of (4.1) follows by arguments similar to those found in [5] or [3], for instance.

The proof essentially tries to obtain a discrete analog in the Stefan problem setting of the relation

$$\int_0^T \int_\Omega \left(\frac{\partial u}{\partial t}\right)^2 + \frac{1}{2}\int_\Omega |\nabla u(\cdot,T)|^2 = \frac{1}{2}\int_\Omega |\nabla u(\cdot,0)|^2 + \int_0^T \int_\Omega q\,\frac{\partial u}{\partial t}$$

which holds for the classical heat conduction equation $\frac{\partial u}{\partial t} = \nabla^2 u + q$. A crucial role is played in the proof by the monotonicity of E_j . ///

Remarks (i) The stability condition (4.1) bears obvious similarities to the stability condition for explicit methods for the heat conduction equation.

(ii) Once uniform boundedness is established, compactness arguments can be used to show that the solutions of (3.1.E) converge in some sense as Δ and δ approach zero. The limit may be identified as the (unique) solution of a weak formulation of the problem arising from (2.4). The convergence is in terms of norms such as $L_2((0,T)\times\Omega)$ and $L_2((0,T),W_0^{1,2}(\Omega))$. For details of the method of proof see, for instance, [1], [3], [5] and the references therein.

One would, of course, hope for a stronger convergence result than this. A reasonable first step towards this should be some form of continuous dependence with respect to data. Since the problem is non-linear, continuous dependence does not follow directly from boundedness.

<u>Uniform Continuous Dependence</u> :

<u>Theorem 4.2</u> $\exists c > 0$ such that if U^n , H^n and $\tilde{U}^n$, $\tilde{H}^n$ are solutions of (3.1.E) corresponding to data Q^n , H^0 and $\tilde{Q}^n$, $\tilde{H}^0$ respectively then

$$\|H^n-\tilde{H}^n\|_{1,\delta} \le c\left(\|H^0-\tilde{H}^0\|_{1,\delta} + \sum_{k=0}^{n-1} \Delta\|Q^k-\tilde{Q}^k\|_{1,\delta}\right) \qquad n = 1,2,\ldots$$

provided (4.2) is satisfied.

<u>Proof</u> Let $h^n = H^n - \tilde{H}^n$, $u^n = U^n - \tilde{U}^n$ and $q^n = Q^n - \tilde{Q}^n$. Clearly the relation

$$m_\delta\left(\frac{h^{n+1}-h^n}{\Delta}, \phi\right) = - a_\delta(u^n,\phi) + \ell_\delta(q^n,\phi) \quad \forall\phi \in F_\delta^0 , \quad n = 1,2,\ldots \qquad (4.3)$$

is satisfied. It follows from the form of E_j , that for any mesh point pair $(\bar{x},j)$

$$u^n(\bar{x},j) = \eta(\bar{x},j,n)h^n(\bar{x},j) \qquad (4.4)$$

where $$0 \le \eta(\bar{x},j,n) \le \max\left(\frac{1}{c_0}, \frac{1}{c_{1,s}}, \frac{1}{c_{1,m}}\right) .$$

Rearranging (4.3) gives

$$m_\delta(h^{n+1},\phi) = m_\delta(h^n,\phi) - \Delta a_\delta(u^n,\phi) + \Delta\ell_\delta(q^n,\phi) \quad \forall\phi \in F_\delta^0 \quad n = 1,2,\ldots \qquad (4.5)$$

Now choose $\phi \in F_\delta^0$ defined by

$$\phi(\bar{x},j) = \begin{cases} 1 & \text{if } h^{n+1}(\bar{x},j) > 0 \\ 0 & \text{if } h^{n+1}(\bar{x},j) = 0 \\ -1 & \text{if } h^{n+1}(\bar{x},j) < 0 . \end{cases} \qquad (4.6)$$

We then have

$$m_\delta(h^{n+1},\phi) = \|\rho h^{n+1}\|_{1,\delta} \qquad (4.7)$$

Next consider $a_\delta(u^n,\phi)$. By a discrete integration by parts this may be written as $\sum_{\bar{x},j} u^n(\bar{x},j)\phi^*(\bar{x},j)$ where $\phi^*(\bar{x},j)$ takes the form $\phi^*(\bar{x},j) = \sum_{(\bar{y},i)} \alpha(\bar{x},j,\bar{y},i)\,\phi(\bar{y},i)$. (Space does not permit us to give details of the $\alpha(\bar{x},j,\bar{y},i)$ here). Letting $\theta(\bar{x},j)$ denote the area of the support of $\psi(\bar{x},j)$, we may write

$$m_\delta(h^n,\phi) - \Delta a_\delta(u^n,\phi)$$

$$= \sum_{\bar{x},j} \left[\rho_j h^n(\bar{x},j)\phi(\bar{x},j)\frac{\delta^2}{4}\theta(\bar{x},j) - \Delta u^n(\bar{x},j)\phi^*(\bar{x},j)\right]$$

$$= \sum_{\bar{x},j} \rho_j h^n(\bar{x},j)\frac{\delta^2}{4}\theta(\bar{x},j)\left[\phi(\bar{x},j) - \frac{\Delta}{\delta^2}\frac{4\eta(\bar{x},j,n)}{\rho_j\theta(\bar{x},j)}\phi^*(\bar{x},j)\right] .$$

It may be shown that (4.2) ensures that $|[\ \dots\ ]| \le 1$, and so we have

$$|m_\delta(h^n,\phi) - \Delta a_\delta(u^n,\phi)| \le \sum_{\bar{x},j} \rho_j |h^n(\bar{x},j)|\frac{\delta^2}{4}\theta(\bar{x},j) = \|\rho h^n\|_{1,\delta}$$

From (4.5) we then have

$$\|\rho h^{n+1}\|_{1,\delta} \le \|\rho h^n\|_{1,\delta} + \Delta\|q^n\|_{1,\delta}$$

Iterating this estimate gives the result. ///

Remark : (i) Clearly we also have a related stability estimate for the temperature since $\|U^n-\tilde{U}^n\|_{1,\delta} \le c\|H^n-\tilde{H}^n\|_{1,\delta}$.

(ii) The stability condition for this theorem is the same as the condition for boundedness in Theorem 4.1. The norms employed are however weaker than those dealt with in Theorem 4.1. Nonetheless, L_1-type norms of the enthalpy are in some sense physically natural as they represent the "energy" in this model.

§5. THE IMPLICIT METHOD

Existance and Uniqueness :

Let some numbering of the mesh points $(\bar{x},j)$ be established, with the convention that the indices $1 \le k \le M$ correspond to non boundary mesh points. Write H_k^{n+1} , U_k^{n+1} , ψ_k etc. for $H^{n+1}(\bar{x},j)$, $U^{n+1}(\bar{x},j)$, $\psi(\bar{x},j)$ etc.

On letting ϕ in (3.1.I) run through the basis functions ψ_k (k=1,...,M) of $\mathcal{F}_\delta^0$ we obtain a system of equations for H_k^{n+1} , U_k^{n+1} . We may write these in the form

$$\mathcal{M}Z = \mathcal{A}W + D \qquad (5.1)$$

where (i) $\mathcal{M}$ is the $M \times M$ diagonal "mass" matrix

$$\mathcal{M}_{kk} = m_\delta(\psi_k,\psi_k) > 0$$

(ii) Z is the M vector

$$Z_k = H_k^{n+1}$$

(iii) $\mathcal{A}$ is the $M \times M$ "stiffness matrix"

$$\mathcal{A}_{k\ell} = -\Delta \;\; a_\delta(\psi_k,\psi_\ell)$$

It follows almost immediately that $\mathcal{A}$ is symmetric and negative definite

(iv) W is the M vector

$$W_k = U_k^{n+1}$$

The components of W and Z are related by

$$Z_k \in E_{(k)}(W_k) \tag{5.2}$$

where $E_{(k)}$ is either E_0 or E_1 .

(v) D is a M vector involving only data and quantities known in terms of H^n . The particular form of D need not concern us here.

<u>Theorem 5.1</u> There exists a unique solution for W and Z to (5.1), (5.2). This solution pair can be obtained as the limits of sequences $W^0, W^1, \ldots$ and $Z^1, Z^2, \ldots$ of Gauss-Seidel iterates which are constructed as follows :

1) for $j = 1,\ldots,M$

   W_j^0 arbitrary

2) for $s = 0,1,2,\ldots$

   for $j = 1,\ldots,M$

   Find the solution W_j^{s+1} of the scalar equation

$$\sum_{\substack{i \\ i<j}} A_{ij} W_i^{s+1} + A_{jj} W_j^{s+1} + \sum_{\substack{i \\ i>j}} A_{ij} W_i^s + D_j \in M_{jj} E_{(j)}(W_j^{s+1}) \tag{5.3}$$

   Define

$$Z_j^{s+1} = \frac{1}{M_{jj}} \left(\sum_{\substack{i \\ i\leq j}} A_{ij} W_i^{s+1} + \sum_{\substack{i \\ i>j}} A_{ij} W_i^s + D_j \right)$$

$$W^{s+1} = (W_1^{s+1},\ldots,W_M^{s+1})$$

$$Z^{s+1} = (Z_1^{s+1},\ldots,Z_M^{s+1})$$

Proof   Firstly note that the scalar equation (5.3) has a unique solution (the left hand side is a strictly decreasing function of W_j^{s+1}, with limits of $\mp\infty$ as $W_j^{s+1} \to \pm\infty$, while right hand side is a strictly increasing set valued function with limits $\pm\infty$ as $W_j^{s+1} \to \pm\infty$)

Consider the function $\Lambda : \mathbb{R}^M \to \mathbb{R}$,

$$\Lambda(W) = \frac{1}{2}\sum_{i,j=1}^{M} A_{ij}W_iW_j - \sum_{j=1}^{M} M_{jj}\int_0^{W_j} E_{(j)}(s)ds + \sum_{j=1}^{M} D_jW_j$$

which is easily seen to have the following properties :

(i) $\Lambda \in C^0(\mathbb{R}^M)$

(ii) $\Lambda(W) \to -\infty$ as $|W| \to \infty$

(iii) Λ is strictly concave.

It follows that Λ attains its maximum value at a unique point, W^* say.

We shall show that (a) W satisfies (5.1), (5.2) iff $W = W^*$

(b) $W^s = (W_1^s,\ldots,W_M^s) \to W^*$ as $s \to \infty$.

Let us note at this point the basic identity

$$\Lambda(W+V) - \Lambda(W)$$

$$= \sum_{i,j} A_{ij}W_iV_j + \frac{1}{2}\sum_{i,j} A_{ij}V_iV_j - \sum_j M_{jj}\int_{W_j}^{W_j+V_j} E_{(j)}(s)ds + \sum_j D_jV_j \qquad (5.4)$$

If (5.1), (5.2) hold then

$$\Lambda(W+V) - \Lambda(W) = \frac{1}{2}\sum A_{ij}V_iV_j - \sum_j M_{jj}\left(\int_{W_j}^{W_j+V_j} E_{(j)}(s)ds - \xi_jV_j\right)$$

for some $\xi_j \in E_{(j)}(W_j)$. The monotonicity of E_0 and E_1 guarantees that the term in brackets is non-negative. Since A is negative definite, we conclude that $\Lambda(W+V) \le \Lambda(W)$ for all $V \in \mathbb{R}^M$. Therefore $W = W^*$.

Conversely, if $W = W^*$ and $V = te_k$ (e_k is a unit vector in the kth coordinate direction) then (5.4) gives

$$\Lambda(W^*+te_k) - \Lambda(W^*)$$

$$= \left(\sum_i A_{ik}W_i^*\right)t + \frac{1}{2}A_{kk}t^2 - M_{kk}\int_{W_k^*}^{W_k^*+t} E_{(k)}(s)ds + D_k t \le 0 \qquad (5.5)$$

for any $t \in \mathbb{R}$. By the form of $E_{(k)}$ we know that as $t \to 0\pm$

$$\int_{W_k^*}^{W_k^*+t} E_k(s)ds = \xi_\pm t + O(t^2)$$

where $\xi_\pm \in E_{(k)}(W_k^*)$. Thus, dividing (5.5) by $t(\neq 0)$ and letting $t \to 0\pm$ gives

$$\sum_i A_{ik}W_i^* - M_{kk}\xi_+ + D_k \le 0$$

and

$$\sum_i A_{ik}W_i^* - M_{kk}\xi_- + D_k \ge 0 .$$

Since $E_k(W_k^*)$ is either an interval or a point, it follows that

$$\sum_i A_{ik}W_i^* - M_{kk}\xi + D_k = 0$$

for some $\xi \in E_{(k)}(W_k^*)$. That is W^* satisfies (5.1), (5.2).

Turning now to (b). The Gauss-Seidel iteration is based upon successive maximizations in the coordinate directions. Consider a typical instance of (5.3). Define W' and W" by

$$W_i' = \begin{cases} W_i^{s+1} & \text{if} \quad i < j \\ W_i^s & \text{if} \quad i \ge j \end{cases}$$

$$W_i'' = \begin{cases} W_i^{s+1} & \text{if} \quad i \le j \\ W_i^{s} & \text{if} \quad i > j \end{cases}$$

Employing the basic identity (5.4) gives

$$\Lambda(W'') - \Lambda(W') = \sum_i A_{ij} W_i' (W_j'' - W_j') + \frac{1}{2} A_{jj} (W_j'' - W_j')^2$$

$$- M_{jj} \int_{W_j'}^{W_j''} E_{(j)}(s)\,ds + D_j (W_j'' - W_j') \qquad (5.6)$$

since $W_i'' = W_i'$ for $i \neq j$. However (5.3) may be rewritten as

$$\sum_i A_{ij} W_i' + D_j = - A_{jj} (W_j'' - W_j') + M_{jj} \xi''$$

where $\xi'' \in E_{(j)}(W_j'')$. Multiplying this by $(W_j'' - W_j')$ and substituting into (5.6) gives

$$\Lambda(W'') - \Lambda(W') = - \frac{1}{2} A_{jj} (W_j'' - W_j')^2 - M_{jj} \left[\int_{W_j'}^{W_j''} E_{(j)}(s)\,ds - \xi'' (W_j'' - W_j') \right] .$$

The monotonicity of $E_{(j)}$ this time ensures that the term in square brackets is non-positive. Thus by the negative definiteness of A

$$\Lambda(W'') - \Lambda(W') \ge \alpha |W_j'' - W_j'|^2$$

for some $\alpha > 0$, independent of j and s. Since $W_j^{s+1} - W_j^{s} = W_j'' - W_j'$, after running through the indices $j = 1, \ldots, M$

$$\Lambda(W^{s+1}) - \Lambda(W^{s}) \ge \alpha \sum_j |W_j^{s+1} - W_j^{s}|^2 \qquad (5.7)$$

for some $\alpha > 0$, independent of s.

Clearly, the sequence $\Lambda(W^s)$ is non-decreasing, and since $\Lambda(W^s) \leq \Lambda(W^*)$, it follows that $\Lambda(W^{s+1}) - \Lambda(W^s) \to 0$. Thus by (5.7)

$$W^{s+1} - W^s \to 0 \qquad (s \to \infty) . \tag{5.8}$$

Now (5.3) may be rewritten as

$$\sum_i A_{ij} W_i^{s+1} + D_j = M_{jj}\xi_j + \varepsilon_j$$

where $\xi_j \in E_{(j)}(W_j^{s+1})$ and $\varepsilon_j = \sum_{\substack{i \\ i<j}} A_{ij}(W_i^{s+1} - W_i^s)$ $\quad j = 1,\ldots,M$.

Subtracting from each of these equations the corresponding component equation of (5.1) leads to

$$\sum_i A_{ij}(W_i^{s+1} - W_i^*) = M_{jj}(\xi_j - \xi_j^*) + \varepsilon_j \qquad (j = 1,\ldots,M)$$

where $\xi_j^* \in E_{(j)}(W_j^*)$. However the monotonicity of $E_{(j)}$ implies that $(\xi_j - \xi_j^*)(W_j^{s+1} - W_j^*) \geq 0$. Therefore multiplying through by $(W_j^{s+1} - W_j^*)$, summing over $j = 1,\ldots,M$ and noting that A is negative definite gives

$$\sum_j (W_j^{s+1} - W_j^*)^2 \leq \alpha \sum_j (-\varepsilon_j)(W_j^{s+1} - W_j^*)$$

$$\leq c \left(\sum_j (W_j^{s+1} - W_j^s)^2 \sum_j (W_j^{s+1} - W_j^*)^2 \right)^{\frac{1}{2}}$$

for α , c depending only on A . The result now follows from (5.8). ///

Uniform Boundedness

The estimate (4.1) holds for the implicit method without any condition on δ and Δ such as (4.2). Again the proof follows the lines of that in [5] or [3].

Uniform Continuous Dependence :

Theorem 5.3 $\exists c > 0$ such that if U^n, H^n and $\tilde{U}^n$, $\tilde{H}^n$ are solutions of (5.1), (5.2) corresponding to data Q^n, H^0 and $\tilde{Q}^n$, $\tilde{H}^0$ respectively, then

$$\|H^n-\tilde{H}^n\|_{1,\delta} \le c(\|H^0-\tilde{H}^0\|_{1,\delta} + \sum_{k=0}^{n-1} \Delta\|Q^n-\tilde{Q}_n\|_{1,\delta}) \qquad n = 1,2,\ldots$$

Proof Let $h^n = H^n - \tilde{H}^n$, $u^n = U^n - \tilde{U}^n$ and $q^n = Q^n - \tilde{Q}^n$.

Clearly the relation

$$m_\delta\left(\frac{h^{n+1}-h^n}{\Delta}, \phi\right) = - a_\delta(u^{n+1},\phi) + \ell_\delta(q^n,\phi) \quad \forall\phi \in F_\delta^0 \; ; \quad n = 0,1,2,\ldots \tag{5.9}$$

is satisfied.

Rearranging (5.9) gives

$$m_\delta(h^{n+1},\phi) + \Delta a_\delta(u^{n+1},\phi) = m_\delta(h^n,\phi) + \Delta\ell_\delta(q^n,\phi) \; , \tag{5.10}$$

and proceeding much as in the proof of Theorem 4.2 choose $\phi \in F_h^0$ given by (4.6). It follows just as in (4.7) that

$$m_\delta(h^{n+1},\phi) = \|\rho h^{n+1}\|_{1,\delta}$$

We claim that

$$a_\delta(u^{n+1},\phi) \ge 0 \tag{5.11}$$

Accepting this for the moment, the right hand side of (5.10) can be estimated to give

$$\|\rho h^{n+1}\|_{1,\delta} \le |m_\delta(h^n,\phi)| + |\Delta \ell_\delta(q^n,\phi)|$$

$$\le \|\rho h^n\|_{1,\delta} + \Delta\|q^n\|_{1,\delta}$$

Iterating this gives the desired result.

Returning now to (5.11). Notice that the monotonicity properties of T_j imply that

i) if $u^{n+1}(\bar{x},j) > 0$ then $h^{n+1}(\bar{x},j) > 0$ and so $\phi(\bar{x},j) = 1$

ii) if $u^{n+1}(\bar{x},j) < 0$ then $h^{n+1}(\bar{x},j) < 0$ and so $\phi(\bar{x},j) = -1$.

By considering the possible signs of $u^{n+1}(\bar{x},j)$ and $u^{n+1}(\bar{y},i)$, it follows that

i) if $u^{n+1}(\bar{x},j) - u^{n+1}(\bar{y},i) > 0$ then $\phi(\bar{x},j) - \phi(\bar{y},i) \ge 0$

ii) if $u^{n+1}(\bar{x},j) - u^{n+1}(\bar{y},i) < 0$ then $\phi(\bar{x},j) - \phi(\bar{y},i) \le 0$,

and so in particular

$$(u^{n+1}(\bar{x},j) - u^{n+1}(\bar{y},i))(\phi(\bar{x},j) - \phi(\bar{y},i)) \ge 0 \ . \tag{5.12}$$

Upon considering the form of $a_\delta(u^{n+1},\phi)$ it follows readily from (5.12) that (5.11) must hold. ///

References

[1] Atthey, D.R., "A finite difference scheme for melting problems", J. Inst. Maths Applics (1974) 13, pp 353-366.

[2] Chalmers, B, "Physical Metallurgy", John Wiley and Sons, New York, London, 1959.

[3] Ciavaldini, J.F, "Analyse numérique d'un problème de Stefan à deux phases par une méthode d'éléments finis", SIAM J. Num. Anal. (1975) 12, pp 464-487.

[4] Duvant G, "The solution of a two-phase Stefan problem by a variational inequality" in "Moving Boundary Problems in Heat Flow and Diffusion" (Eds J.R. Ockendon and W.R. Hodgkins), Clarendon Press, Oxford, 1975, pp 173-181.

[5] Elliot C.M. and Ockendon J.R., "Weak and Variational Methods for Moving Boundary Problems", Pitman Research Notes in Mathematics No.59.

[6] Friedman, A, "The Stefan problem in several space variables", Trans. Amer. Math. Soc. (1968), 132, pp 51-87.

[7] Kamenomostskaja, S.L., "On Stefan's problem", Mat. Sb., (1961), 53 (94) pp 489-514.

[8] Oleinik, O.A., "A method of solution of the general Stefan problem", Soviet Math. Dokl, (1960), 1, pp 1350-1354.

Centre for Mathematical Analysis,
Australian National University,
Canberra, A.C.T.
AUSTRALIA