STABILITY OF ONE-STEP METHODS IN TRANSIENT NONLINEAR HEAT CONDUCTION

J. R. HUGHES

Mail Code 104-44, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, U.S.A.

SUMMARY

The purpose of the present work is to ascertain practical stability conditions for one-step methods commonly used in transient nonlinear heat conduction analyses.

The class of problems considered is governed by the temporally continuous, spatially discrete system

$$C(\theta, t)\,\dot{\theta} + K(\theta, t)\,\theta = R(t), \tag{1}$$

in which C is the capacity matrix, K is the conductivity matrix, R is the heat supply vector, θ is the temperature vector, t denotes time and a superposed dot indicates time differentiation. We assume throughout that C is symmetric and positive definite and K is symmetric and (at least) positive semi-definite. The initial-value problem consists of finding a function $\theta = \theta(t)$, $T \in [0, \tau]$, $\tau > 0$, satisfying (1) and the initial condition $\theta(0) = T$, where T is the given vector of initial temperatures.

In the linear case, in which K and C are constant, the stability behavior of one-step methods is well known.

The two major classes of one-step methods for (1) are the 'generalized trapezoidal family' and 'generalized midpoint family'. In the linear constant coefficient case these two families of methods become identical.

It has been known for some time that the stability behavior of the trapezoidal family in the linear regime does not carry over directly to nonlinear problems. In particular, according to the 'linearized' criterion of stability, almost all unconditionally stable members of the trapezoidal family become only conditionally stable in nonlinear applications.

On the other hand, the 'generalized midpoint family' seems to retain important stability characteristics in the nonlinear regime.

In this paper the concepts of stability, appropriate to the nonlinear problem, are thoroughly discussed. They of course reduce to the usual stability criterion for the linear, constant coefficient case. However, for nonlinear problems there are differences and these ideas are of key importance in obtaining practical stability conditions. Of particular importance is a recent result which indicates that, in a sense, the trapezoidal and midpoint families are equivalent. Thus, stability results for one family may be translated into a result for the other.

The main results obtained are summarized as follows:

- The stability behavior of the explicit Euler method in the nonlinear regime is analogous to that for linear problems. In particular, an *a priori* step size restriction may be determined for each time step.
- The precise time step restriction on implicit conditionally stable members of the trapezoidal and midpoint families is shown *not* to be determinable *a priori*.
- Of considerable practical significance, unconditionally stable members of the trapezoidal and midpoint families are identified.

All notions of stability employed are motivated and defined, and their interpretations in practical computing are indicated.

1. Introduction

We consider two families of one-step methods for solving the transient nonlinear heat conduction problem.

Generalized Trapezoidal Methods

Find $\underset{\sim}{T}_n$, $n \in \{0, 1, \ldots, N\}$ such that

$$\underset{\sim}{C}_n \underset{\sim}{U}_n + \underset{\sim}{K}_n \underset{\sim}{T}_n = \underset{\sim}{R}_n \,, \qquad n \in \{0, 1, \ldots, N\} \,, \tag{1}$$

$$\underset{\sim}{T}_{n+1} = \underset{\sim}{T}_n + \Delta t \underset{\sim}{U}_{n+\alpha} \,, \qquad n \in \{0, 1, \ldots, N-1\} \,, \tag{2}$$

$$\underset{\sim}{T}_0 = \underset{\sim}{T} \,, \tag{3}$$

in which

$$\underset{\sim}{C}_n = \underset{\sim}{C}(\underset{\sim}{T}_n, t_n) \,, \tag{4}$$

$$\underset{\sim}{K}_n = \underset{\sim}{K}(\underset{\sim}{T}_n, t_n) \,, \tag{5}$$

$$\underset{\sim}{R}_n = \underset{\sim}{R}(t_n) \,, \tag{6}$$

$$\underset{\sim}{U}_{n+\alpha} = (1-\alpha)\underset{\sim}{U}_n + \alpha \underset{\sim}{U}_{n+1} \,, \tag{7}$$

where $\underset{\sim}{T}_n$ is the approximation to $\underset{\sim}{\theta}(t_n)$, N is the total number of time steps, $\Delta t = \tau/N$ and $t_n = n\Delta t$. In this paper we limit our attention to parameter values of α in the interval $[0, 1]$.

With the exception of the case $\alpha = 0$, at each time step the problem to be solved is a nonlinear algebraic one and techniques such as the Newton-Raphson method, with suitable notions of convergence, must be resorted to. If $\alpha = 0$ (forward difference or Euler method) and $\underset{\sim}{C}$ is diagonal, the method is explicit and the solution may be constructed without solving systems of linear equations.

For linear problems, in which $\underset{\sim}{C}$ and $\underset{\sim}{K}$ are constant matrices, the stability properties of this family of algorithms is well-known (see for example Wood and Lewis [1] and Taylor [2]). For instance, if $\alpha < \frac{1}{2}$ the algorithm in question is conditionally stable, i.e., stability considerations limit the maximum size of the time step employed. To be precise the time step must satisfy the condition

$$\lambda \Delta t \leq 2/(1 - 2\alpha) \,, \tag{8}$$

in which λ is the maximum eigenvalue of the matrix $\underset{\sim}{C}^{-1}\underset{\sim}{K}$. On the other hand, if $\alpha \geq \frac{1}{2}$ the algorithm in question is unconditionally stable, i.e., there is no restriction on the maximum size of time step. Condition (8) is a stringent one in practice and for this reason unconditionally stable algorithms are generally preferred.

Generalized Midpoint Family

Find $\underset{\sim}{T}_n$, $n \in \{0, 1, \ldots, N\}$ such that

$$\underset{\sim}{C}_{n+\alpha} \underset{\sim}{U}_{n+\alpha} + \underset{\sim}{K}_{n+\alpha} \underset{\sim}{T}_{n+\alpha} = \underset{\sim}{R}_{n+\alpha} \,, \qquad n \in \{0, 1, \ldots, N-1\} \,, \tag{9}$$

$$\underset{\sim}{T}_{n+1} = \underset{\sim}{T}_n + \Delta t \underset{\sim}{U}_{n+\alpha} , \qquad n \in \{0, 1, \ldots, N-1\} , \tag{10}$$

$$\underset{\sim}{T}_0 = \underset{\sim}{T} , \tag{11}$$

where

$$\underset{\sim}{C}_{n+\alpha} = \underset{\sim}{C}(\underset{\sim}{T}_{n+\alpha}, t_{n+\alpha}) , \tag{12}$$

$$\underset{\sim}{K}_{n+\alpha} = \underset{\sim}{K}(\underset{\sim}{T}_{n+\alpha}, t_{n+\alpha}) , \tag{13}$$

$$\underset{\sim}{T}_{n+\alpha} = (1-\alpha)\underset{\sim}{T}_n + \alpha \underset{\sim}{T}_{n+1} , \tag{14}$$

$$\underset{\sim}{R}_{n+\alpha} = (1-\alpha)\underset{\sim}{R}_n + \alpha \underset{\sim}{R}_{n+1} , \tag{15}$$

$$\underset{\sim}{U}_{n+\alpha} = (1-\alpha)\underset{\sim}{U}_n + \alpha \underset{\sim}{U}_{n+1} , \tag{16}$$

$$t_{n+\alpha} = (n+\alpha)\Delta t . \tag{17}$$

Notice that for the case in which $\underset{\sim}{C}$ and $\underset{\sim}{K}$ are constant matrices this family of algorithms is identical to the preceding one.

2. Stability Analysis

Concern over the stability of the generalized trapezoidal family has emanated from the following analysis (see Gourlay [3]):

Consider the single-degree-of-freedom nonlinear model equation

$$\dot{\theta} + \lambda(\theta, t)\theta = 0 , \tag{18}$$

in which it is assumed $\lambda > 0$. Applying the generalized trapezoidal algorithm to (18) and employing the obvious notation for the single-degree-of-freedom case, we obtain the recursion relation:

$$T_{n+1} = A\, T_n , \tag{19}$$

where

$$A = \frac{1 - \Delta t(1-\alpha)\lambda_n}{1 + \Delta t\, \alpha\, \lambda_{n+1}} , \tag{20}$$

in which $\lambda_n = \lambda(T_n, t_n)$ and $\lambda_{n+1} = \lambda(T_{n+1}, t_{n+1})$. In keeping with the common definition of stability for equations of the type considered here, we require that

$$|A| \leq 1 . \tag{21}$$

(In the linear case (21) leads to the stability conditions cited above.) In addition, we stipulate that (21) must hold for all possible combinations of λ_n and λ_{n+1}. For example, if $\alpha = \frac{1}{2}$ (trapezoidal rule or Crank-Nicolson method) and $\lambda_n > \lambda_{n+1}$, then (21) imposes the time step restriction

$$\Delta t \leq 4/(\lambda_n - \lambda_{n+1}) . \tag{22}$$

Thus the unconditional stability of the trapezoidal rule in linear problems does not carry over to the nonlinear regime. In fact, only the case $\alpha = 1$ (backward difference method) is unconditionally stable for nonlinear problems. Precise stability conditions for the various cases are summarized as follows:

$$\alpha = 0: \qquad \Delta t \le 2/\lambda_n \tag{23}$$

$$\alpha \in (0, 1): \quad \begin{cases} \text{stable}, & \lambda_{n+1} \ge \dfrac{(1-\alpha)}{\alpha}\lambda_n \quad (24) \\ \Delta t \le \dfrac{2}{(1-\alpha)\lambda_n - \alpha\lambda_{n+1}}, & \lambda_{n+1} < \dfrac{(1-\alpha)}{\alpha}\lambda_n \quad (25) \end{cases}$$

$$\alpha = 1: \qquad \text{stable} \tag{26}$$

As can be seen from (24) and (25), increasing α tends to stabilize the algorithm. Since conditions (24) and (25) involve λ_{n+1} they are not suitable for establishing a priori time step estimates. All algorithms for which $\alpha < 1$ become conditionally stable in the nonlinear regime.

The previous results are somewhat disconcerting despite the fact that at least one of the algorithms considered (backward differences) is unconditionally stable. The reasons for this are as follows: The only second-order accurate algorithm among those considered is the Crank-Nicolson method. The remaining algorithms all possess only first-order accuracy. The ones with the largest error constants are $\alpha = 0$ and $\alpha = 1$. Thus to attain unconditional stability, in the sense of (21), we must be content with a significant loss of accuracy.

An analogous stability analysis of the generalized midpoint family yields more pleasant results. In this case

$$A = \frac{1 - \Delta t(1-\alpha)\lambda_{n+\alpha}}{1 + \Delta t\,\alpha\,\lambda_{n+\alpha}}, \tag{27}$$

where $\lambda_{n+\alpha} = \lambda(T_{n+\alpha}, t_{n+\alpha})$ and condition (21) requires that

$$\lambda_{n+\alpha}\Delta t \le 2/(1-2\alpha), \tag{28}$$

for $\alpha < \frac{1}{2}$, whereas for $\alpha \ge \frac{1}{2}$ the algorithm in question is unconditionally stable. Thus the midpoint family has the advantage that unconditionally stable methods for linear problems maintain this property in the nonlinear regime. In addition, for $\alpha = \frac{1}{2}$ (midpoint rule) the method is second-order accurate. Again, (28) does not suffice to establish a priori time step estimates.

We now shall show that, for $\alpha \ge \frac{1}{2}$, the midpoint methods are unconditionally stable for multi-degree-of-freedom systems. Our hypothesis on $\underset{\sim}{C}$ and $\underset{\sim}{K}$ insure that the matrix $\underset{\sim}{C}^{-1}_{n+\alpha}\underset{\sim}{K}_{n+\alpha}$ has a complete set of orthonormal eigenvectors $\underset{\sim}{\phi}^{i}_{n+\alpha}$ and corresponding eigenvalues $\lambda^{i}_{n+\alpha} \ge 0$. Thus we can write

$$\underset{\sim}{T}_n = \sum_i T_n^i \underset{\sim}{\phi}_{n+\alpha}^i \ , \tag{29}$$

$$\underset{\sim}{T}_{n+1} = \sum_i T_{n+1}^i \underset{\sim}{\phi}_{n+\alpha}^i \ , \tag{30}$$

where

$$T_n^i = \underset{\sim}{T}_n^T \underset{\sim}{C}_{n+\alpha} \underset{\sim}{\phi}_{n+\alpha}^i \ , \tag{31}$$

$$T_{n+1}^i = \underset{\sim}{T}_{n+1}^T \underset{\sim}{C}_{n+\alpha} \underset{\sim}{\phi}_{n+\alpha}^i \ . \tag{32}$$

Assuming $\underset{\sim}{R} = \underset{\sim}{0}$ and employing (29-32) in (9) and (10) yields

$$\underset{\sim}{T}_{n+1} = \sum_i \left\{ \frac{1 - \Delta t(1-\alpha)\lambda_{n+\alpha}^i}{1 + \Delta t \alpha \lambda_{n+\alpha}^i} \ T_n^i \underset{\sim}{\phi}_{n+\alpha}^i \right\} \ , \tag{33}$$

from which it follows that

$$\| \underset{\sim}{T}_{n+1} \| \leq \| \underset{\sim}{T}_n \| \ , \tag{34}$$

where $\| \ \|$ denotes the norm defined by $\| \underset{\sim}{X} \|^2 = \underset{\sim}{X}^T \underset{\sim}{C}_{n+\alpha} \underset{\sim}{X}$. Thus if $\lambda_{n+\alpha}$ is interpreted as the maximum $\lambda_{n+\alpha}^i$, the results for the model problem provide both necessary and sufficient conditions for the stability of the midpoint methods. Note, however, that a similar conclusion cannot be drawn for the trapezoidal methods since the matrices $\underset{\sim}{C}$ and $\underset{\sim}{K}$ must be evaluated at two different steps. Because of this the modal decomposition argument does not work for these equations and we can only conclude that the stability analysis of the model equation for $\alpha \in (0, 1)$ provides necessary stability conditions - they may not be sufficient.

We wish to make some comments on (34). Note that, in general, (34) will be a strict inequality; equality only holding in trivial cases such as equilibrium. If $\underset{\sim}{C}$ is a constant matrix, the norm $\| \ \|$ does not change from step to step and (34) implies

$$\| \underset{\sim}{T}_n \| \leq \| \underset{\sim}{T}_0 \| \ , \qquad n \in \{0, 1, \ldots, N\} \ . \tag{35}$$

This decay is in keeping with the behavior of the governing equations and precludes instability. In the general case, in which $\underset{\sim}{C}$ is not constant, decay in each step is with respect to a changing norm (i.e., one defined by $\underset{\sim}{C}_{n+\alpha}$, as above). For this case $\underset{\sim}{T}_{n+1}$ may actually be greater than $\underset{\sim}{T}_n$ with respect to some other norm (e.g., the euclidean norm). However, we still view (34) as the appropriate stability condition since it again mimics the behavior of the governing equations whose solution may grow or decay with respect to some arbitrary norm depending on the nature of $\underset{\sim}{C}(\underset{\sim}{\theta}, t)$. Specifically, it can be shown that

$$\underset{\sim}{\theta}_{n+1}^{T}\underset{\sim}{C}(\theta_{n+\alpha}, t_{n+\alpha})\underset{\sim}{\theta}_{n+1} \leq \underset{\sim}{\theta}_{n}^{T}\underset{\sim}{C}(\underset{\sim}{\theta}_{n+\alpha}, t_{n+\alpha})\underset{\sim}{\theta}_{n} \tag{36}$$

for small enough Δt. Thus the decay behavior of the midpoint methods for arbitrary Δt is analogous to the small time step behavior of the governing equations.

In a recent work it has been shown that the trapezoidal and midpoint rules are, in a sense, equivalent; see [4]. For example, in the case in which $\underset{\sim}{C}$ is constant, [4] and (34) imply that, for trapezoidal rule, the following condition holds:*

$$\underset{\sim}{T}_{n+1}^{T}(\underset{\sim}{C} + \Delta t\underset{\sim}{K}_{n+1} + \frac{\Delta t^2}{4}\underset{\sim}{K}_{n+1}\underset{\sim}{C}^{-1}\underset{\sim}{K}_{n+1})\underset{\sim}{T}_{n+1} \leq \underset{\sim}{T}_{n}^{T}(\underset{\sim}{C} + \Delta t\underset{\sim}{K}_{n} + \frac{\Delta t^2}{4}\underset{\sim}{K}_{n}\underset{\sim}{C}^{-1}\underset{\sim}{K}_{n})\underset{\sim}{T}_{n} \ , \tag{37}$$

and thus

$$\| \underset{\sim}{T}_{n} \|^2 \leq \underset{\sim}{T}_{0}^{T}(\underset{\sim}{C} + \Delta t\underset{\sim}{K}_{0} + \frac{\Delta t^2}{4}\underset{\sim}{K}_{0}\underset{\sim}{C}^{-1}\underset{\sim}{K}_{0})\underset{\sim}{T}_{0} \ . \tag{38}$$

Eq. (38) guarantees stability as $\| \underset{\sim}{T}_{n} \|$ is uniformly bounded by a constant for all n. However, the bound is greater than for the corresponding situation involving midpoint rule [cf. (35)] and, indeed, is greater than for the exact solution. Nevertheless, this result does indicate that the single-degree-of-freedom model equation stability condition, eq. (21), may yield overly pessimistic results for some algorithms.

Whether or not the above example can be generalized to other members of the trapezoidal family and to the case of variable $\underset{\sim}{C}$ requires further research. Even if this proves to be the case, it appears that the midpoint methods are still to be preferred to the trapezoidal methods in nonlinear heat conduction since the bounds obtained are more in keeping with those for the exact governing equations.

3. Acknowledgment

We would like to acknowledge the support for this work provided by the United States Energy Research and Development Administration through Lawrence Berkeley Laboratory.

4. References

[1] WOOD, W.L., LEWIS, R.W., "A Comparison of Time Marching Schemes for the Transient Heat Conduction Equation," International Journal for Numerical Methods in Engineering, 9, 679-689 (1975).

[2] TAYLOR, R.L., "*HEAT*, A Finite Element Computer Program for Heat Conduction Analysis," unpublished.

[3] GOURLAY, A.R., "A Note on Trapezoidal Methods for the Solution of Initial Value Problems," Mathematics of Computation, 24, 629-633 (1970).

[4] DAHLQUIST, G., LINDBERG, B., "On Some Implicit One-Step Methods for Stiff Differential Equations," Report No. TRITA-NA-7302, Department of Information Processing, The Royal Institute of Technology, Stockholm.

---

*This may be verified by direct computation.