A Posteriori Bounds in the Numerical Solution of Mildly Nonlinear Parabolic Equations*

By Alfred Carasso

Abstract. We derive a posteriori bounds for \( (V - \hat{V}) \) and its difference quotient \( (V - \hat{V})_x \), where \( V \) and \( \hat{V} \) are, respectively, the exact and computed solution of a difference approximation to a mildly nonlinear parabolic initial boundary problem, with a known steady-state solution. It is assumed that the computation is over a long interval of time. The estimates are valid for a class of difference approximations, which includes the Crank-Nicolson method, and are of the same magnitude for both \( (V - \hat{V}) \) and \( (V - \hat{V})_x \).

1. Introduction. Let \( \Omega \) be the strip \( \{(x, t) | 0 < x < 1, t > 0\} \) in the \((x, t)\) plane and consider the mixed problem

\[
\begin{align}
\frac{\partial u}{\partial t} &= \left[a(x, t)\frac{\partial u}{\partial x}\right]_x + b(x, t)\frac{\partial u}{\partial x} - f(x, t, u), & (x, t) \in \Omega, \\
\frac{\partial u(x, 0)}{\partial x} &= \chi(x), & 0 \leq x \leq 1, \\
u(0, t) &= \varphi_1(t), & u(1, t) = \varphi_2(t), & t > 0.
\end{align}
\]

We assume that \( a(x, t), b(x, t) \) are “smooth” bounded functions on \( \Omega \), with \( a(x, t) \geq a_0 > 0 \), and that \( f(x, t, w) \) is, at least once, continuously differentiable on \( \Omega \times [0, t] \) with \( \frac{\partial f}{\partial w} \geq 0 \). Moreover, \( \frac{\partial f}{\partial w} \) is to remain bounded if \( w \) stays bounded. The coefficients, data, and \( f \) are assumed such as to assure the existence and uniqueness of a solution \( u(x, t) \), four times boundedly differentiable in \( \Omega \), and converging to a steady state value \( u^*(x) \), as \( t \to \infty \). We assume \( u^*(x) \) is known and that, by means of asymptotic formulae and the like, one can estimate \( ||u(x, t) - u^*||_2 \) as a function of \( t \), for \( t \) sufficiently large. The analytical theory for such problems is discussed in Friedman [5].

Several finite-difference methods for the numerical computation of (1.1) have been shown to converge; see for example [4], [6], [8], [10], [3] and their references, and especially [9] for the linear case.

Because of round-off error, and the fact that one may need to use iterative methods at each time step to solve the nonlinear difference equations, only an approximation \( \hat{V}^* \) to the exact solution \( V^* \) of the difference equations can be computed in general. In [3], a “boundary-value” method for (1.1) was analyzed. This method yields an a posteriori estimate for \( V - \hat{V} \) by simply computing residuals. In the present note we make use of some of the results in [2] and [3] to derive such an estimate for a class of stable “marching” procedures for (1.1). Unlike the situation in [3], however, the estimate will involve bounds on the derivatives of \( u \). It is interesting that the estimate is of the same magnitude for both \( (V - \hat{V}) \) and its difference quotient \( (V - \hat{V})_x \).

Received September 18, 1969.
AMS 1969 subject classifications. Primary 6568, 6580.
Key words and phrases. Parabolic equations, Crank-Nicolson method, limiting steady state, computations over long times.
* Supported by NSF Grant GP-13024.

Copyright © 1971, American Mathematical Society
2. Notation. Let \( \mathcal{R}_T \) be the rectangle \( \{(x, t) \mid 0 < x < 1, 0 < t < T\} \) and let \( M \) and \( N \) be positive integers. Let \( \Delta x = 1/(M + 1), \Delta t = T/(N + 1) \), and introduce a mesh over \( \mathcal{R}_T \) by means of the lines \( x = k\Delta x, k = 0, 1, \ldots, M + 1, t = n\Delta t, n = 0, 1, \ldots, N + 1 \). Let \( v_k^n \) denote \( v(k\Delta x, n\Delta t) \). Define \( V^n \) to be the \( M \)-component vector
\[
V^n = \{v_1^n, v_2^n, \ldots, v_M^n\}^T
\]
and let \( V \) be the "block" vector of \( MN \) components
\[
V = \{V^1, V^2, \ldots, V^N\}^T.
\]
Although we will be dealing with real-valued mesh functions, it is convenient to define scalar products and norms for complex vectors. For any two \( M \) vectors \( V^n, W^n \) let
\[
\langle V^n, W^n \rangle = \Delta x \sum_{k=1}^{M} v_k^n w_k^n
\]
and let
\[
||V^n||_2^2 = \langle V^n, V^n \rangle.
\]
Let
\[
||V_k^n||_2^2 = \Delta x \sum_{k=0}^{M} \frac{|v_{k+1}^n - v_k^n|^2}{\Delta x^2}
\]
where \( v_0^n, v_{M+1}^n \) are defined to be zero.

For block vectors \( V, W \) define
\[
\langle V, W \rangle = \Delta t \sum_{n=1}^{N} \langle V^n, W^n \rangle
\]
and let
\[
||V||_2^2 = \langle V, V \rangle.
\]
\[
||V_k||_2^2 = \Delta t \sum_{n=1}^{N} ||V_k^n||_2^2.
\]
Finally, for a square matrix \( A \), define \( ||A|| \) in terms of vector norms, i.e., as
\[
||A|| = \sup_{||X|| = 1} ||AX||,
\]
the supremum being taken over all complex vectors.

3. Difference Approximations to (1.1). Let \( U^n \) be the \( M \)-vector consisting of the solution to (1.1) evaluated at the interior mesh points of the line \( t = n\Delta t \) and let \( V^n \) be the corresponding exact solution of the difference equations used to approximate (1.1). Define \( E^n = V^n - U^n \). We will consider the class of marching schemes which lead to a priori estimates of the form
\[
||E^n||_2^2 + ||E_k^n||_2^2 \leq K(T)(\Delta t^{r+1} + \Delta x^{s+1}), \quad n\Delta t \leq T,
\]
where \( r \) and \( s \) are positive integers and \( K(T) \) is known. An example of a difference
scheme for (1.1) satisfying (3.1) with \( r = s = 1 \), is the Crank-Nicolson version analyzed in [8]. In general \( K(T) \) will involve bounds on \( a, b, f, u \) and their derivatives, as well as a growth factor. The reason for the latter is that, even if the exact solution to (1.1) decays asymptotically to a steady state, the exact solution of a stable, consistent, difference approximation may grow exponentially as \( n\Delta t \to \infty \), \( \Delta t \) fixed. Hence, we cannot expect \( K(T) \) to remain bounded as \( T \to \infty \), in general. We remark, however, that in [7], Kreiss and Widlund have shown how to construct schemes (for linear time-dependent problems with periodic boundary conditions) which preserve the asymptotic behavior of \( u(x, t) \) provided \( |b| \Delta t / \Delta x < 1 \). In the following we will derive bounds for \( \| \tilde{V} - V \|_2 \) and \( \| \tilde{V}_x - V_x \|_2 \) for computations of (1.1) up to some "large" but fixed time \( T \). These bounds will depend on \( K(T) \).

We begin by deriving new finite-difference equations for the exact solution \( \{ V^n \} \) of a difference scheme used to approximate (1.1). Since \( V^n \) = \( U^n + E^n \), we have

\[
\begin{align*}
\frac{v_{k}^{n+1} - v_{k}^{n-1}}{2\Delta t} &= \frac{u_{k}^{n+1} - u_{k}^{n-1}}{2\Delta t} + \frac{e_{k}^{n+1} - e_{k}^{n-1}}{2\Delta t} \\
&= \left( \frac{\partial u}{\partial t} \right)_{k}^{n} + \left( \frac{e_{k}^{n+1} - e_{k}^{n-1}}{2\Delta t} \right) + \frac{\Delta t^2}{6} \left( u_{t_{tt}} \right)_{k},
\end{align*}
\]

where "\( \tilde{V} \)" represents a mean value of \( \psi \) called for by Taylor's theorem. From (1.1) we have

\[
\begin{align*}
\left( \frac{\partial u}{\partial t} \right)_{k}^{n} + \frac{\Delta t^2}{6} \left( u_{t_{tt}} \right)_{k} &= \frac{a_{k+1/2}^{n} (u_{k+1}^{n} - u_{k}^{n}) - a_{k-1/2}^{n} (u_{k}^{n} - u_{k-1}^{n})}{\Delta x^2} \\
&\quad + b_{k}^{n} \frac{(u_{k+1}^{n} - u_{k-1}^{n})}{2\Delta x} - f(k\Delta x, n\Delta t, u_{k}^{n}) + \tau_{k}^{n},
\end{align*}
\]

where

\[
\tau_{k}^{n} = \frac{\Delta t^2}{6} \left( u_{t_{tt}} \right)_{k}
\]

\[
\begin{align*}
&\Delta x^2 \left\{ \frac{(u_{x}^{n} (a_{xx})^{n})}{3} + \frac{(u_{xx}^{n} (a_{x})^{n})}{2} + \frac{(u_{xxx}^{n} (a_{x})^{n})}{6} + \frac{(a_{u} u_{x}^{n})}{12} + b_{k}^{n} \left( u_{x}^{n} \right) \right\}.
\end{align*}
\]

From (3.2) and (3.3) we have

\[
\begin{align*}
\frac{v_{k}^{n+1} - v_{k}^{n-1}}{2\Delta t} &= \frac{a_{k+1/2}^{n} (v_{k+1}^{n} - v_{k}^{n}) - a_{k-1/2}^{n} (v_{k}^{n} - v_{k-1}^{n})}{\Delta x^2} \\
&\quad + b_{k}^{n} \frac{(v_{k+1}^{n} - v_{k-1}^{n})}{2\Delta x} - f(k\Delta x, n\Delta t, u_{k}^{n}) \\
&\quad + \frac{e_{k}^{n+1} - e_{k}^{n-1}}{2\Delta t} - b_{k}^{n} \frac{(e_{k+1}^{n} - e_{k-1}^{n})}{2\Delta x} \\
&\quad - \frac{(a_{k+1/2}^{n} (e_{k+1}^{n} - e_{k}^{n}) - a_{k-1/2}^{n} (e_{k}^{n} - e_{k-1}^{n}))}{\Delta x^2} + \tau_{k}^{n},
\end{align*}
\]

\[k = 1, \ldots, M, \quad n = 1, 2, \ldots,\]
with the initial boundary data

\[ \begin{align*}
\nu_0 &= \chi(k\Delta x), \quad k = 1, \ldots, M, \\
\nu_0^\circ &= \psi_1(n\Delta t), \quad \nu_{M+1}^\circ &= \psi_2(n\Delta t), \quad n = 1, 2, \ldots.
\end{align*} \]

With \( T = (N + 1)\Delta t \) we now consider the system formed by equations \((3.5)\) for \( n = 1, 2, \ldots, N \). It is convenient to write this system in matrix-vector notation.

Let \( L^n \) and \( B^n \) be the tridiagonal \( M \times M \) matrices defined by

\[ L^n = \frac{1}{\Delta x^2} \begin{bmatrix}
(a^n_{1+1/2} + a^n_{1/2}) & -a^n_{1+1/2} & 0 \\
-a^n_{1+1/2} & \ddots & \ddots \\
0 & \ddots & -a^n_{M-1/2} \\
0 & \ddots & (a^n_{M+1/2} + a^n_{M-1/2}) & -a^n_{M-1/2}
\end{bmatrix}, \]

\[ B^n = \frac{1}{2\Delta x} \begin{bmatrix}
b^n_1 & 0 \\
0 & \ddots & 0 \\
0 & \ddots & b^n_{M-1} & 0 \\
0 & \ddots & (a^n_{M+1/2} + a^n_{M-1/2}) & -a^n_{M-1/2}
\end{bmatrix}, \]

and define the \( M \)-vectors \( \tau^n \), \( F^n(U) \), and \( G^n \) by

\[ \tau^n = \{ \tau^n_1, \tau^n_2, \ldots, \tau^n_M \}^T, \]

\[ F^n(U) = \{ f^n_1(u), f^n_2(u), \ldots, f^n_M(u) \}^T, \]

where

\[ f^n_i(u) = f(k\Delta x, n\Delta t, u^n_i) \]

and

\[ G^n = \frac{1}{\Delta x^2} \{ (a^n_{1+1/2} - \frac{1}{2}\Delta x b^n_1)\psi_1(n\Delta t), 0, 0, \ldots, 0, (a^n_{M+1/2} + \frac{1}{2}\Delta x b^n_M)\psi_2(n\Delta t) \}^T. \]

We may then write \((3.5)\), \((3.6)\) as

\[ \frac{v^{n+1} - v^{n-1}}{2\Delta t} = -L^n v^n - B^n v^n - F^n(U) + \tau^n + G^n \]

\[ + \frac{E^{n+1} - E^{n-1}}{2\Delta t} + B^n E^n + L^n E^n, \quad n = 1, 2, \ldots, N. \]

Some further definitions will enable us to write \((3.12)\) in "block" form. Define the \( MN \times MN \) block tridiagonal matrix \( P \) by (with \( \sigma = 1/2\Delta t \))

\[ P = \begin{bmatrix}
(L^1 + B^1) & \sigma I & \cdots & \cdots \\
-\sigma I & \ddots & \ddots & \cdots \\
\cdots & \cdots & -\sigma I & \ddots & \sigma I \\
0 & \cdots & (L^N + B^N) & \sigma I & \cdots
\end{bmatrix}. \]
For any real block vector $\xi$ define the $M \times M$ diagonal matrix $C(\xi)$ by

$$C(\xi) = \begin{bmatrix} f_u(\Delta x, n\Delta t, \xi_1^1) & 0 \\ \vdots & \ddots \\ 0 & f_u(\Delta x, n\Delta t, \xi_M^1) \end{bmatrix}$$

and let $C(\xi)$ be the block matrix

$$C(\xi) = \begin{bmatrix} C(\xi) & 0 \\ \vdots & \ddots \\ 0 & C(\xi) \end{bmatrix}.$$

Finally, define the block vectors $F$, $G^*$, $H$, and $\tau$ by

$$F = \{F_1, F_2, \ldots, F^N\}^T,$$

$$G^* = \left\{ G^1 + \frac{V^0}{2\Delta t}, G^2, \ldots, G^N - \frac{V^{N+1}}{2\Delta t} \right\}^T,$$

$$H = \left\{ \frac{E^2 - E^0}{2\Delta t} + (L^1 + B^1)E^1, \ldots, \frac{E^{N+1} - E^{N-1}}{2\Delta t} + (L^N + B^N)E^N \right\}^T,$$

$$\tau = \{\tau^1, \tau^2, \ldots, \tau^N\}^T.$$

With this notation we have from (3.12)

$$PV = -F(U) + G^* + \tau + H.$$

**Lemma 1.** Let $D$ be a diagonal matrix of order $MN$ with nonnegative real entries

$$D = P + D.$$

Let $b(x, t)$ in (1.1) satisfy

$$\frac{\partial b}{\partial x} \leq b_1 < 2a_0\pi^2, \quad \forall (x, t) \in \Omega T.$$

Fix $\epsilon > 0$ so that $a_0\pi^2 - b_1/2 - \epsilon \geq \omega > 0$. If $\Delta x \leq (12\epsilon/a_0\pi^2)^{1/2}$, $Q^{-1}$ exists and

$$\sup_{X \text{ real}, \|X\|_2 \leq 1} \|Q^{-1}X\|_2 \leq \frac{1}{\omega}.$$

Moreover, if $QW = Z$, where $Z$ is real we have

$$\|W\|_2 \leq \left( \frac{2\omega + b_1}{2a_0\pi^2} \right)^{1/2} \|Z\|_2.$$

**Proof.** See [3, Lemma 1].

**Remark.** If

$$D = \begin{bmatrix} \Lambda & 0 \\ \Lambda & \ddots \\ 0 & \ddots & \Lambda \end{bmatrix},$$

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
where $\Lambda$ is a diagonal $M \times M$ matrix with nonnegative real entries, and if $a(x, t), b(x, t)$ are independent of $t$, $Q^{-1}$ exists and remains bounded for all sufficiently small $\Delta x$ independently of hypothesis (3.22). This observation is relevant to the case where (1.1) is linear with time independent coefficients, i.e., $a = a(x), b = b(x)$, and $f(x, t, u) = c(x)u + h(x, t)$ with $c(x) \geq 0$. See [1, Lemma 1] and [2, Lemma 4.2].

4. A Posteriori Bounds. For each $n = 1, 2, \cdots, N + 1$, let $\hat{V}^n$ be the computed solution at $t = n\Delta t$, of the difference equations used to approximate (1.1) and consider the block vector

$$\hat{V} = \{\hat{V}^1, \hat{V}^2, \cdots, \hat{V}^N\}^T.$$  

Define $\hat{G}^*$ to be the block vector obtained from $G^*$ when $V^{N+1}$ is replaced by $\hat{V}^{N+1}$.

Compute the block vector $R$ given by

$$R = P\hat{V} + F(\hat{V}) - \hat{G}^*.$$  

Subtracting (4.2) from (3.20) we have

$$P(V - \hat{V}) = -F(U) + F(\hat{V}) + (G^* - \hat{G}^*) + \tau + H - R$$  

$$= -F(U) + F(V) + F(\hat{V}) - F(V) + (G^* - \hat{G}^*) + \tau + H - R$$  

$$= -C(\xi)(U - V) - C(\Psi)(V - \hat{V}) + (G^* - \hat{G}^*) + \tau + H - R,$$

for some intermediate real block vectors $\xi$ and $\Psi$ on using the mean value theorem. Hence,

$$[P + C(\Psi)](V - \hat{V}) = \tau + H - R + (G^* - \hat{G}^*) - C(\xi)(U - V).$$

Since $f_* \geq 0$, $C(\Psi)$ is a diagonal matrix with nonnegative real entries. By Lemma 1, we may estimate $\|V - \hat{V}\|_2, \|V_* - \hat{V}_*\|_2$, provided we can estimate the terms other than $R$ on the right-hand side of (4.4). We will make use of the a priori estimate (3.1).

Let $a^*, b^*$ be upper bounds for $a(x, t)$ and $|b(x, t)|$, respectively, in $\Omega_r$.

Since

$$(L^nE^n)_k = \frac{1}{\Delta x^2} a^*_{k-1/2}(e^n_k - e^{n-1}_k) + \frac{1}{\Delta x^2} a^*_{k+1/2}(e^n_k - e^{n+1}_k)$$

we have

$$\|L^nE^n\|_2 \leq \frac{a^*}{\Delta x} \left\{ \Delta x \sum_{k=1}^M \frac{(e^n_k - e^{n-1}_k)^2}{\Delta x^2} \right\}^{1/2} + \frac{a^*}{\Delta x} \left\{ \Delta x \sum_{k=1}^M \frac{(e^n_{k+1} - e^n_k)^2}{\Delta x^2} \right\}^{1/2}$$

$$\leq \frac{2a^*}{\Delta x} \|E^n\|_2 \leq 2a^* K(T) \left( \frac{\Delta x^{r+1}}{\Delta x} + \Delta x^r \right).$$

Similarly,

$$\|B^nE^n\|_2 \leq b^* K(T)(\Delta x^{r+1} + \Delta t^{r+1})$$

and we have

$$\frac{1}{2\Delta t} \|E^{n+1} - E^n\|_2 \leq K(T) \left( \frac{\Delta t^r + \Delta x^{r+1}}{\Delta t} \right).$$

Hence, we can estimate $\|H\|_2$ by
SOLUTION OF MILDLY NONLINEAR PARABOLIC EQUATIONS

We estimate \( ||G^* - \hat{G}^*||_2 \) as follows: First,

\[
||G^* - \hat{G}^*||_2 = \frac{1}{2\Delta t^{1/2}} ||\hat{\nu}^{N+1} - \nu^{N+1}||_2.
\]

If \( U^* \) is the \( M \)-vector consisting of the steady state solution, we have from (3.1)

\[
||G^* - \hat{G}^*||_2 \leq \frac{1}{2\Delta t^{1/2}} \{ ||\hat{\nu}^{N+1} - U^*||_2 + ||U^{N+1} - U^*||_2 \}
\]

\[
+ \frac{K(T)}{\Delta t} \left( \frac{\Delta t^{r+1} + \Delta x^{*+1}}{2\Delta t^{1/2}} \right).
\]

Since we assume \( U^* \) is known, and that \( \{u(x, t) - u^*(x)\} \) can be estimated as a function of \( t \), the right-hand side of (4.10) can be estimated.

We may estimate \( ||C(\xi)||_2 \) by using the a priori estimate (3.1), since \( \xi \) is an intermediate value, and since \( f_u(x, t, w) \) is bounded if \( w \) is bounded. This means we can find a constant \( K_3 \) such that

\[
||C(\xi)(V - u)||_2 \leq K_3 K(T)T^{1/2}(\Delta t^{r+1} + \Delta x^{*+1}).
\]

Finally, we assume a bound is known for the derivatives of \( u \) occurring in (3.4) so that

\[
||\tau||_2 \leq T^{1/2} K_2(\Delta t^2 + \Delta x^2), \quad \text{for some constant } K_3.
\]

Using Lemma 1 and (4.8), (4.10), (4.11) and (4.12) we have

**Theorem.** Let \( b(x, t) \) in (1.1) satisfy \( |\partial b/\partial x| \leq b_1 < 2a_0\pi^2 \) and fix \( \epsilon > 0 \) so that \( a_0\pi^2 - \frac{1}{2} b_1 - \epsilon \geq \omega > 0 \). Let

\[
\Delta x \leq \left( \frac{12\epsilon}{a_0\pi^2} \right)^{1/2}.
\]

Let \( V = \{V^n\} \) and \( \hat{V} = \{\hat{V}^n\} \) be, respectively, the exact and computed solution of a difference approximation for (1.1) satisfying (3.1). Finally, let \( R \) be defined by (4.2). Then

\[
||V - \hat{V}||_2 \leq \frac{1}{2\omega\Delta t^{1/2}} \{ ||\hat{\nu}^{N+1} - U^*||_2 + ||U^{N+1} - U^*||_2 \} + \frac{||R||_2}{\omega}
\]

\[
+ T^{1/2} K_3(\Delta t^2 + \Delta x^2) \frac{1}{\omega} \left\{ K_1 + \frac{1}{2(T\Delta t)^{1/2}} + b^* + \frac{2a^*}{\Delta x} + \frac{1}{\Delta t} \right\}
\]

\[
||V_x - \hat{V}_x||_2 \leq \left( \frac{2\omega + b_1}{2a_0\omega^2} \right)^{1/2} \left\{ \frac{1}{2\Delta t^{1/2}} \{ ||\hat{\nu}^{N+1} - U^*||_2 + ||U^{N+1} - U^*||_2 \} + ||R||_2 + K_2 T^{1/2}(\Delta t^2 + \Delta x^2) + K(T)T^{1/2}(\Delta t^{r+1} + \Delta x^{*+1}) \right\}
\]

\[
\left\{ K_1 + \frac{1}{2(T\Delta t)^{1/2}} + b^* + \frac{2a^*}{\Delta x} + \frac{1}{\Delta t} \right\}.
\]
1. A. CARASSO, “Finite-difference methods and the eigenvalue problem for nonselfadjoint
Numer. Math. (To appear.)
4. J. DOUGLAS, Jr., A Survey of Numerical Methods for Parabolic Differential Equations,
5. A. FRIEDMAN, Partial Differential Equations of Parabolic Type, Prentice-Hall, Englewood
6. F. JOHN, “On integration of parabolic equations by difference methods. I: Linear and
MR 13, 947.
7. H. O. KREISS & O. B. WIDLUND, Difference Approximations for Initial Value Problems for
Partial Differential Equations, Department of Computer Sciences, Report NR 7, Upsala University,
1967.
#7356.