G. Koeppe; Hans-Görg Roos; Lutz Tobiska

An enclosure generating modification of the method of discretization in time

Commentationes Mathematicae Universitatis Carolinae, Vol. 28 (1987), No. 3, 441--447

Persistent URL: http://dml.cz/dmlcz/106558

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE
28,3 (1987)

AN ENCLOSURE GENERATING MODIFICATION OF THE METHOD OF DISCRETIZATION IN TIME

G. KOEPPE, H.-G. ROOS, L. TOBISKA

Abstract: A modification of the method of discretization in time is proposed to generate upper and lower bounds for the solution of the original linear parabolic boundary value problem. It is proved that the modified Rothe function converges to the exact solution of the first order in the maximum norm if the step size tends to zero.

Key words: Parabolic boundary value problem, discretization in time, maximum principle, enclosure.

Classification: 65N59

---

1. Introduction. In the numerical solution of boundary value problems it is not only interesting to obtain a numerical approximation of the solution within a certain accuracy but it is also of practical importance to construct upper and lower bounds for the solution itself. Such inclusions generating discretizations for parabolic boundary value problems have been proposed in [1],[2],[7],[10].

The aim of the present paper consists in deriving a modification of the Rothe method (or the method of discretization in time) to generate upper and lower bounds for the solution of the original parabolic boundary value problem. In contrast to [9] we use maximum principles to prove the enclosing property in the n-dimensional case. Our technique allows us to consider more general boundary value problems in comparison to [9] and to omit any restriction with respect to the step size in time.

As in [9], it is possible in a second step to combine the modified Rothe method with the monotone discretization techniques proposed [4],[5], at least for the one-dimensional case.

2. The modified method of discretization in time. Let us consider the parabolic boundary value problem

$$\begin{aligned} \frac{\partial u}{\partial t} + Lu &= f \text{ in } \Omega \cup B_T \\ u &= 0 \text{ on } S_T \\ u &= u_0 \text{ in } B_0 \end{aligned} \tag{1}$$

in frame of the theory of classical solutions. In (1) Ω denotes a bounded domain of R^n with $C^{2+\alpha}$ boundary, $Q=\Omega\times(0,T)$, $S_T=\partial\Omega\times[0,T]$, $B_T=\Omega\times\{T\}$ and $B_0=\Omega\times\{0\}$. Let L be a linear, uniformly elliptic differential operator of second order

$$L=\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial x_i}-c(x)$$

with coefficients in $C^{2+\alpha}(\bar{\Omega})$, further we assume $f=f(x)$ in $C^{\alpha}(\bar{\Omega})$ and

(2) (i) u_0 satisfies the compatibility conditions

$$u_0=0,\ Lu_0=f \text{ on } \partial\Omega$$

(ii) $c(x)\geq c_0>0$ on $\bar{\Omega}$.

It is well known that the problem (1) admits exactly one classical solution in $C^{2+\alpha}(\Omega)$ (see for instance [3]) and that the elliptic operator L and the parabolic operator $\frac{\partial}{\partial t}+L$ satisfy a classical maximum principle, respectively.

Now we choose some $N\in\mathbb{N}$ and divide the t-interval into N subintervals $[t_{i-1},t_i]$ (i=1(1)N) by the definition $t_\mu=\tau\mu$ with $\tau=T/N$. Let us set

$$\varphi_i(t)=\begin{cases}(t-t_{i-1})/\tau & \text{for } t\in[t_{i-1},t_i]\\ (t_{i+1}-t)/\tau & \text{for } t\in[t_i,t_{i+1}]\\ 0 & \text{otherwise.}\end{cases}$$

The discretization in time is realized by means of the representation

$$u_\tau=\sum_{i=0}^{N} z_i\varphi_i(t) \tag{3}$$

for an approximate solution $u_\tau(x,t)$ of problem (1). The usual method of discretization in time consists in determining the functions $z_i(x)$ as solutions of

$$\frac{z_i-z_{i-1}}{\tau}+Lz_i=f \text{ in } \Omega\ (i\geq 1)$$
$$z_i=0 \text{ on } \partial\Omega$$

starting with $z_0=u_0$. In our modified method we choose a constant p and a function q of x and determine the functions $z_i(x)$ as solutions of

(4) (i) $$\frac{z_1-z_0}{\tau}+Lz_0=f+\tau q \text{ in } \Omega$$
$$z_0=0 \quad \text{on } \partial\Omega$$

(ii) $$\frac{z_i-z_{i-1}}{\tau}+Lz_i=f+\tau p \text{ in } \Omega\ (i=1(1)N$$
$$z_i=0 \quad \text{on } \partial\Omega .$$

In the sequel we will show that it is possible to choose p, q in such a way that, for instance, u_τ is an upper solution of our original problem (1) that means $u_\tau(x,t) \geq u(x,t)$ in $\overline{Q}$. Finally we will show that the function $u_\tau(x,t)$ converges to $u(x,t)$ of the first order in the maximum norm if τ tends to zero.

3. Analysis of the modified method of discretization in time. We denote by $C^{k,\ell}$ the set of functions being k-times and ℓ-times continuously differentiable with respect to x and t, respectively. Furthermore, let $Q_j = \Omega \times (t_{j-1}, t_j)$ and $B_j = \Omega \times \{t_j\}$, $j=1(1)N$. Our analysis is based on a weak maximum principle of the following type:

Lemma 1: Let the functions v and w be in $C(\overline{Q}) \cap \bigcap_{j=1}^{N} C^{2,1}(Q_j \cup B_j)$ and satisfy

(i) $\frac{\partial v}{\partial t} + Lv \leq \frac{\partial w}{\partial t} + Lw$ on $Q_j \cup B_j$, $j=1,\dots,N$

(ii) $v \leq w$ on $S_T \cup B_0$.

Then, it follows $v \leq w$ on $\overline{Q}$.

The validity of Lemma 1 follows from the successive application of the classical maximum principle.
Now we discuss the solvability of the problems (4),(i),(ii). Subtracting (4),(i) and (4),(ii) for i=1 we obtain

$$(5)\quad \begin{aligned} L(z_1-z_0) &= (p-q)\tau \quad \text{in } \Omega \\ z_1-z_0 &= 0 \quad \text{on } \partial\Omega . \end{aligned}$$

Hence, the function $\tilde{z}_1 = (z_1-z_0)/\tau$ is uniquely determined and z_0 satisfies

$$(6)\quad \begin{aligned} Lz_0 &= f+\tau q-\tilde{z}_1 \quad \text{in } \Omega \\ z_0 &= 0 \quad \text{on } \partial\Omega . \end{aligned}$$

Introducing the operator L_τ by $L = \tau L + I$ (I represents the identity), the functions $z_i(x)$ for $i \geq 1$ fulfil

$$(7)\quad \begin{aligned} L_\tau z_i &= z_{i-1}+f\tau+p\tau^2 \quad \text{in } \Omega \\ z_i &= 0 \quad \text{on } \partial\Omega . \end{aligned}$$

Therefore, in the representation (3) all coefficients of our approximate solution are uniquely determined provided that p, q are known.

Now we proceed to choose the parameters p, q in a suitable way to generate an upper solution. According to Lemma 1 we compute the defect of the approximate solution u_τ on $Q_j \cup B_j$. One obtains

$$\frac{\partial u_\tau}{\partial t} + Lu_\tau = \frac{z_1 - z_0}{\tau} + \varphi_0 Lz_0 + \varphi_1 Lz_1 \text{ on } Q_1 \cup B_1$$

and

$$\frac{\partial u_\tau}{\partial t} + Lu_\tau = f + \tau p + \varphi_{i-1}(t)\, \frac{z_i - 2z_{i-1} + z_{i-2}}{\tau} \text{ on } Q_i \cup B_i \quad (i=2(1)N).$$

Taking into account (4)(i),(ii) one gets

$$\frac{\partial u_\tau}{\partial t} + Lu_\tau = f + \tau(p\,\varphi_1 + q\,\varphi_0) \text{ on } Q_1 \cup B_1.$$

Let us introduce the notation

(8) $$s_i = \frac{z_{i+1} - 2z_i + z_{i-1}}{\tau^2} \quad (i=1)(N-1).$$

Applying Lemma 1, our considerations with respect to the defect on every subinterval result in

Lemma 2: Let us suppose

(9) (i) $p, q \geq 0$,

(ii) $p + s_i \geq 0$, $(i=1(1)N-1)$

(iii) $z_0 \geq u_0$.

Then, $u_\tau(x,t)$ is an upper solution of our original problem, that means $u_\tau(x,t) \geq u(x,t)$ for all $(x,t) \in \overline{Q}$.

In the next step we analyze the validity of condition (9),(ii). From the identities

$$\frac{z_1 - z_0}{\tau} + Lz_0 = f + \tau q, \quad \frac{z_1 - z_0}{\tau} + Lz_1 = f + \tau p, \quad \frac{z_2 - z_1}{\tau} + Lz_2 = f + \tau p,$$

it follows immediately

$$\frac{z_2 - 2z_1 + z_0}{\tau} + L(z_2 - 2z_1 + z_0) = \tau(q-p),$$

thus

(10) (i) $L_\tau s_1 = q - p$

Adding the identities

$$\frac{z_i - z_{i-1}}{\tau} + Lz_i = f + \tau p, \qquad -2\,\frac{z_{i+1} - z_i}{\tau} - 2Lz_{i+1} = -2f - 2\tau p,$$

$$\frac{z_{i+2} - z_{i+1}}{\tau} + Lz_{i+2} = f + \tau p,$$

one obtains similarly

(10) (ii) $L_\tau s_{i+1} = s_i \quad (i \geq 1)$.

Remembering p being a constant we conclude

(11) $L_\tau(s_1+p)=q+\tau c(x)p,\ L_\tau(s_{i+1}+p)=s_i+p+\tau c(x)p \quad (i\geq 1).$

We want to derive advantage from the inverse-monotonicity of the operator $A_\tau=(L_\tau,R)$ where R denotes the restriction of functions on $\partial\Omega$. Because of (9),(i) s_i+p $(i=1(1)N-1)$ is nonnegative on $\partial\Omega$ such that with (2),(ii) it follows $s_i+p\geq 0$ successively and condition (9),(ii) is automatically fulfilled.

To generate an upper solution u_τ it is only necessary to guarantee the conditions $p,q\geq 0$ and $z_0\geq u_0$. However, the function z_0 is defined in a not so simple way - one has to solve (5),(6) - therefore it is essential to find a practical criterion for the parameters p, q in order to safeguard the inequality $z_0\geq u_0$.

Taking into consideration the inverse-monotonicity of the operator $A=(L,R)$, the inequality $z_0\geq u_0$ is fulfilled provided that $Lz_0\geq Lu_0$. The inequality

$$\tilde{z}_1\leq f-Lu_0$$

is sufficient for $Lz_0\geq Lu_0$ and valid on $\partial\Omega$ because of the compability condition for u_0. Consequently, the inequality $z_0\geq u_0$ is fulfilled if

$$p-q\leq L(f-Lu_0)$$

and we have the following result:

<u>Theorem 1:</u> Let us additionally assume that $f-Lu_0$ belongs to $C^2(\bar{\Omega})$ and $p,q\geq 0$ are chosen such that the inequality

$$q\geq p-L(f-Lu_0) \tag{12}$$

holds. Then, the modified method of discretization in time (3),(4) generates an upper solution for the original problem for all $\tau>0$.

Now we are going to prove the convergence of our modified method or discretization in time in the maximum norm. For this we choose

$$p=\max(0,\ \max_{x\in\bar{\Omega}} L(f-Lu_0)) \tag{13}$$

$$q=p-L(f-Lu_0) \tag{14}$$

such that the assumptions of Theorem 1 are fulfilled.

Because of (5) and $f-Lu_0=0$ on $\partial\Omega$ we have

$$\tilde{z}_1=f-Lu_0$$

that means we start with the solution z_0 of

$$Lz_0 = Lu_0 + \tau q \text{ in } \Omega$$
$$z_0 = 0 \qquad \text{on } \partial\Omega .$$

From the barrier function technique it follows

$$\max_{x \in \bar{\Omega}} |z_0(x) - u_0(x)| \leq M\tau . \tag{15}$$

For the difference of the exact solution $u(x,t)$ and the approximate solution we obtain

$$(16) \qquad \begin{aligned} &(\tfrac{\partial}{\partial t} + L)(u_\tau - u) = \tau(p\varphi_1 + q\varphi_0) \text{ on } Q_1 \cup B_1 \\ &(\tfrac{\partial}{\partial t} + L)(u_\tau - u) = \tau(p + \varphi_{i-1}s_{i-1}) \text{ on } Q_i \cup B_i \quad (i=2(1)N). \end{aligned}$$

The terms in brackets on the right hand sides of (16) are uniformly bounded on $\bar{\Omega}$ because (10),(i) and (10),(ii) imply

$$\max_{\bar{\Omega}} |s_{i+1}(x)| \leq \max_{\bar{\Omega}} |s_i(x)| \leq \max_{\bar{\Omega}} |q(x) - p| .$$

Thus, applying Lemma 1 we summarize

<u>Theorem 2:</u> Let $f - Lu_0$ belong to $C^2(\bar{\Omega})$ and $p, q(x)$ satisfy (13),(14). Then, the modified method of discretization in time (3),(4) generates an upper solution u_τ for the original problem. Moreover, there exists a constant $M = M(p,q)$ such that the error estimate

$$\max_{\bar{\Omega}} |u(x,t) - u_\tau(x,t)| \leq M\tau$$

holds.

<u>Remark:</u> We considered classical solutions of our parabolic boundary value problem (1). It is possible to extend Theorem 1 to weak solutions; to be more precise, to $W_2^1([0,T], H_0^1(\Omega), L^2(\Omega))$ - type solutions. Of course, it is necessary to specify some order relations and to use generalized maximum principles for parabolic and elliptic problems, for instance, instead Lemma 1. It is possible to choose for p, q appropriate elements from the dual space $(H_0^1(\Omega))^*$. But in practice one will use simple functions to simplify the realization of the method - therefore we restricted ourselves to choose p as a constant parameter in our classical framework.

In general, for weak solutions an error estimation in the maximum norm does not hold. But, similarly as in [8] one can prove the $O(\tau)$ convergence in the L^2-norm and the $O(\tau^{1/2})$ convergence in the H^1-norm (compare [9]). Details on the modified method of discretization in time for weak solutions are due to G. Koeppe and can be found in [6].

References

[1] ADAMS E., SPREUER H.: Konvergente numerische Schrankenkonstruktionen mit Spline-Funktionen für nichtlineare gewöhnliche bzw. lineare parabolische Randwertaufgaben, Intern. Math . (ed. K. Nickel), Springer Verlag, Berlin 1975.

[2] FAASS E.: Beliebig genaue numerische Schranken für die Lösung parabolischer Randwertaufgaben, Diss., Karlsruhe 1975.

[3] FRIEDMAN A.: Partial Differential Equations of Parabolic Type, Prentice-Hall, Inc., Englewood Cliffs, N.J. 1964.

[4] GROSSMANN Ch.: KRÄTZSCHMAR, M., ROOS H.-G.: Gleichmässig einschliessende Diskretisierungsverfahren für schwach nichtlineare Randwertaufgaben, Numerische Mathematik 49(1986), 95-110.

[5] GROSSMANN Ch., KRÄTZSCHMAR M., ROOS H.-G.: Uniformly enclosing discretization methods of high order for boundary value problems, Math. of Comput. (to appear).

[6] KOEPPE G.: Eine monoton einschliessende Variante des Rothe-Verfahrens für lineare parabolische Randwertprobleme, Diplomarbeit, TH Magdeburg 1986.

[7] MEYN K.H.: Monotonieaussagen für elliptische und parabolische Randwertaufgaben und Anwendung auf Finite-Elemente-Funktionen, Diss., Hamburg 1979.

[8] REKTORYS K.: The method of discretization in time and partial differential equations, Dordrecht/Boston 1982.

[9] ROOS H.-G.: The Rothe method and monotone discretization for parabolic equations, To appear in:"Discretization in differential equations and enclosure" (ed. Adams, Ansorge, Grossmann, Roos).

[10] SCHEU G.: Schrankenkonstruktionen für die Randwertaufgabe der linearen Wärmeleitungsgleichung, Diss., Karlsruhe 1973.

G. Koeppe, L. Tobiska: Technische Hochschule, Magdeburg 3040,

H.-G. Roos: Technische Universität Dresden, Mommsenstrasse 13, Dresden 8027, GDR

(Oblatum 2.2. 1987)