

NUMERICAL ALTERNATIVE METHOD SCHEME FOR BURGERS' EQUATION

TITUS PETRILA and DAMIAN TRIF
(Cluj-Napoca)

1. Introduction

The aim of this paper is to give a numerical scheme based on the alternative method for the following quasi-linear parabolic equation

$$(1) \quad \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}$$

where $u = u(x, t)$ in some domain and ν is a parameter. Obviously some appropriate boundary-initial conditions will be joined to the equation in order to ensure the existence and the uniqueness of the (classical) solution.

Historically this equation first appears in a paper by H. Bateman [1] in 1915 when he gave a special solution for it. Besides some applications in the theory of stochastic processes, this equation having a structure roughly similar to that of Navier-Stokes equation, has a great importance in aerodynamics.

Precisely this is the equation given in J. Burgers' theory of a model of turbulence [2] when the relationship to the shock wave theory is also pointed out. A mathematical study of the general properties of this equation can be found in [3], [4].

In the last decade applied mathematicians have become increasingly interested in developing numerical stable schemes for the Burgers' equation considered as a "core" of the Navier-Stokes equation. Among all these researches the methods of spectral analysis seem to be of a very special interest [5], [6].

In what follows we shall give besides a classical scheme which uses the finite differences methods, a numerical scheme based upon the alternative method. Both the stationary and the non-stationary case are considered.

The alternative method divides the given problem into a fixed point problem for a contractive operator (which once iteratively solved supplies that part of the solution with superior harmonics) and a small-size algebraic or differential nonlinear system (according to the stationarity or the nonstationarity of the problem).

Some numerical tests finally prove the higher accuracy of the alternative method as well as some computational advantages for its use in the problem considered.

2. The alternative method

Let us consider an equation of the type

$$(2) \quad Lu = Nu$$

on a normed space S , where $L : D(L) \subset S \rightarrow S$ is a linear operator while $N : D(N) \subset S \rightarrow S$ is a nonlinear operator. The solutions of equation (2) are searched in $D(L) \cap D(N)$.

Let us consider a splitting $S = S_0 \oplus S_1$ of the S and $P : S \rightarrow S_0$ its projection onto S_0 . Let $H : S_1 \rightarrow S_1$ be a partial inverse for L , i.e. a linear application so that

$$(3) \quad H(I - P)Lu = (I - P)u \quad \forall u \in S$$

Applying now to the considered equation (2) the operators P and $H(I - P)$ respectively, it becomes equivalent with the system

$$PLu = PNu \quad (\text{bifurcation equation})$$

(4)

$$u = Pu + H(I - P)Nu \quad (\text{auxiliary equation})$$

Let us now fix $Pu = u^*$ (in the second equation). If the operator

$$(5) \quad Tu = u^* + H(I - P)Nu$$

is a contraction on a closed sphere from S then it admits a fixed point U which depends on u^* and which will be denoted by $U(u^*)$. This fixed point would be a solution of equation (2) only if it satisfies the equation of bifurcation (4) too

$$(6) \quad PLU(u^*) = PNU(u^*)$$

The solvability of equation (6) considered in the unknown u^* is thus equivalent with the solvability of the given equation (2) and, consequently, an approximate solution u^* for equation (6) generates, via the auxiliary equation (4), a fixed point $U(u^*)$ which represents as well an approximate solution for equation (2). Details related to this topic can be found in the papers of Cesari [7] and Trif [8].

For using this method to the numerical determining of the solutions of equation (2), we will suppose that S is a real, separable Hilbert space while L is a closed self-adjoint operator, whose domain $D(L)$ is dense into S and which has a kernel of finite size p . We will also suppose that L has the eigenvalues $\lambda_1 = \dots = \lambda_p = 0$, $\lambda_{p+1} > 0, \dots, \lambda_i \rightarrow \infty, i \rightarrow \infty$ and the corresponding eigenfunctions Φ_1, Φ_2, \dots which determine a complete orthonormed system in S .

We will also admit that there is a subspace S' of S (which contains $D(L)$ and $D(N)$), complete with regard to a norm μ in respect of the Fourier series of the elements $u \in D(L)$ converge too. Additionally, we admit that $\{\mu(\Phi_k)/\lambda_k\}$ $k > p$ is an l^2 sequence and that on S' we have $\|u\| \leq \alpha\mu(u)$.

Concerning the nonlinear application N , we suppose that it is locally Lipschitzian, i.e. for every $R > 0$ there is a $L_R > 0$ so that for

any $u, v \in D(N)$ with $\mu(u) \leq R, \mu(v) \leq R$ the inequality

$$(7) \quad \mu(Nu - Nv) \leq L_R\mu(u - v)$$

Let now $m \geq p, S_m = sp\{\Phi_1, \dots, \Phi_p, \Phi_{p+1}, \dots, \Phi_m\}, S_0 = \{0\}$ and let us define, for every $u \in S$

$$P_m : S \rightarrow S_m, P_m u = \sum_{k=1}^m (u, \Phi_k) \Phi_k$$

$$(8) \quad H_m : S \rightarrow S, H_m u = \sum_{k=m+1}^{\infty} \frac{(u, \Phi_k)}{\lambda_k} \Phi_k$$

where $u = \sum_{k=1}^{\infty} (u, \Phi_k) \Phi_k$. It proves at once that for every $u \in S$ we have $H_m u \in D(L), LH_m u = (I - P_m)u, P_m H_m u = 0$ and, respectively, for every $u \in D(L), H_m Lu = (I - P_m)u, P_m Lu = LP_m u$. At the same time, it can show (cf. [8]) that

$$\|H_m\| = \frac{1}{\lambda_{m+1}}, \mu(H_m) \leq \alpha\sigma(m)$$

where

$$(9) \quad \sigma(m) = \left[\sum_{k=m+1}^{\infty} \left(\frac{\mu(\Phi_k)}{\lambda_k} \right)^2 \right]^{1/2}$$

which implies that $\mu(H_m) \rightarrow 0, \|H_m\| \rightarrow 0$ when $m \rightarrow \infty$.

Under all these circumstances it proves that for a sufficiently large m , the operator T given by (5) becomes a contraction on a metric space (a bounded closed set from S'). According to the Banach fixed point theorem, the operator T admits a unique fixed point which can be got by the method of successive approximations, namely by

$$u^0 = u^*, u^{s+1} = u^* + H_m Nu^s \quad s = 0, 1, \dots$$

and which depends continuously on u^* (with respect to μ). Taking into account the assumptions made so far, this fixed point called also the *associate element* for $u^* \in S_m$ fulfils the auxiliary equation (4).

The bifurcation equation (4) becomes then

$$Lu^* = P_m NU(u^*)$$

or, if $u^* = \sum_{k=1}^m c_k^* \Phi_k$, on components

$$(10) \quad \lambda_k c_k^* - (NU(u^*), \Phi_k) = 0 \quad k = 1, \dots, m$$

which represents the so-called *system of determining equations*.

Summarizing, under above hypotheses, for m sufficiently large, equation (2) admits a solution \bar{u} if and only if equations (10) admit a solution u^* and then $\bar{u} = U(u^*)$. For details connected with the proofs and the evaluation of approximation errors one can use [8] and its references.

3. The approximation of the solutions of Burgers' equation. Stationary case

To compare and to test the numerical solutions got through the alternative method with those obtained using other algorithms or even with exact solution, we shall consider the following nonhomogeneous Burgers equation with joined boundary conditions

$$(11) \quad u_{xx} = uu_x - f \quad u(-1) = u(1) = 0$$

where $x \in (-1, 1)$, $f(x) = 2x^3 - 2x + 2$. The unique exact solution of this problem is $u(x) = 1 - x^2$.

In this case $Lu = u_{xx}$, $Nu = uu_x - f$, $S = L^2(-1, 1)$, $D(L) = D(N) = \{u \in C^2(-1, 1) \cap C[-1, 1], u(-1) = u(1) = 0\}$ endowed with the uniform norm μ . The spectral problem

$$(12) \quad u_{xx} = \lambda u, \quad u(-1) = u(1) = 0$$

admits the eigenfunctions $\Phi_k(x) = \sin \frac{k\pi}{2}(x+1)$ and the eigenvalues $\lambda_k = -(k\pi/2)^2$ satisfying the conditions made in the previous paragraph.

The solution of problem (11) will be looked for in a truncated Fourier series of the type

$$(13) \quad u(x) = \sum_{k=1}^n c_k \sin \frac{k\pi}{2}(x+1)$$

Let $1 \leq m < n$. Then

$$(14) \quad u^* = P_m u = \sum_{k=1}^m c_k \sin \frac{k\pi}{2}(x+1)$$

$$H_m u = - \sum_{k=m+1}^{\infty} c_k \frac{4}{k^2 \pi^2} \sin \frac{k\pi}{2}(x+1)$$

The coefficients of the development of f using the corresponding system are

$$f_k = \int_{-1}^1 (2x^3 - 2x + 2) \sin \frac{k\pi}{2}(x+1) dx = \begin{cases} \frac{192}{k^3 \pi^3} & k \text{ even} \\ \frac{8}{k\pi} & k \text{ odd} \end{cases}$$

In this case

$$(15) \quad Nu = \sum_{i,j=1}^n c_i \sin \frac{i\pi}{2}(x+1) c_j \frac{j\pi}{2} \cos \frac{j\pi}{2}(x+1) - \sum_{k=1}^n f_k \sin \frac{\pi k}{2}(x+1)$$

from where the Fourier coefficients for Nu can be obtained at once

$$(16) \quad C_k = \sum_{i,j=1}^n c_i c_j \frac{j\pi}{2} \int_{-1}^1 \sin \frac{i\pi}{2}(x+1) \cos \frac{j\pi}{2}(x+1) \sin \frac{k\pi}{2}(x+1) dx - f_k = \\ = \frac{\pi}{4} \left(\sum_{j=1}^{k-1} j c_j c_{k-j} - k \sum_{j=1}^{n-k} c_j c_{k+j} \right) -$$

Consequently the iterations which lead to the associate element for u^* are

$$(17) \quad c_k^{s+1} = - \frac{1}{k^2 \pi^2} \left(\sum_{j=1}^{k-1} j c_j^s c_{k-j}^s - k \sum_{j=1}^{n-k} c_j^s c_{k+j}^s \right) + \frac{4f_k}{k^2 \pi^2}, \quad k = m+1, \dots, n$$

with c_1, \dots, c_m being fixed.

We remark that in the case that the iterative process converges (in the frame of the computer accuracy), for m sufficiently great, at S^{st} step

$$u^{s+1} = \sum_{k=1}^m c_k^* \Phi_k + \sum_{k=m+1}^n C_k^s \Phi_k$$

represents an approximation of the associated function $U(u^*)$. The determining equations become

$$- c_k^* \left(\frac{k\pi}{2} \right)^2 = - \frac{k^2 \pi^2}{4} C_k^{s+1} \quad k = 1, \dots, m$$

that means

$$(18) \quad g_k \equiv c_k^* - C_k^{s+1} = 0 \quad k = 1, \dots, m$$

When equation (11) has a solution for m sufficiently great, equation (18) will also have a solution which can be approximated by an arbitrary procedure. Such a procedure using the data c_1, \dots, c_m computes g_1, \dots, g_m (by the iterative process (17)) and on the base of the results obtained it will improve the initial data c_1, \dots, c_m , the cycle being retaken until the requested accuracy is achieved. The function associated with this approximate solution of the system (18), got by the iterations (17), represents an approximation of the solution of problem (11).

4. The approximation of the solutions of Burgers' equations. Nonstationary case

In the sequel we will consider the problem

$$(19) \quad \begin{aligned} u_{xx} &= u_t + uu_x - f & x \in (-1, 1), t > 0. \\ u(x, 0) &= u_0(x) & x \in (-1, 1) \\ u(-1, t) &= u(1, t) = 0 & t > 0 \end{aligned}$$

were, for numerical computations, we will take

$$f(x, t) = (1 - x^2) \cos t + 2 \sin t - 2x \sin^2 t + 2x^3 \sin^2 t$$

This problem has the exact (unique) solution $u(x, t) = (1 - x^2) \sin t$.

The main difference with respect to the previous case consists in the structure of the operator N where the term u_t is involved now. Supposing calculated the solution u_j at the level of time t_j the auxiliary equation (4) becomes at the time level t_{j+1}

$$(20) \quad u_{j+1}^{s+1} = u_{j+1}^* + H_m \left(N u_{j+1}^s + \frac{u_{j+1}^s - u_j}{\delta t} \right) \quad s = 0, 1, \dots, u_{j+1}^0 = u_j$$

where δt is a time step and $(u_{j+1}^s - u_j)/\delta t$ is an approximation of u_t at the level t_{j+1} . If m is sufficiently large, the iterations (20) converge towards the associated function $U(u_{j+1}^*)$. This would be a solution of equation (19) if

$$(21) \quad \frac{\partial^2}{\partial x^2} u_{j+1}^* = P_m N U(u_{j+1}^*) + \frac{\partial}{\partial t} u^* |_{t=t_{j+1}}$$

In this case the coefficients of $f(x, t)$ according to the structure of N are

$$(22) \quad f_k(t) = \begin{cases} \frac{32}{k^3 \pi^3} \cos t + \frac{8}{k\pi} \sin t & k \text{ odd} \\ \frac{192}{k^3 \pi^3} \sin^2 t & k \text{ even} \end{cases}$$

As $u^*(x, t) = \sum_{k=1}^m c_k^*(t) \sin \frac{k\pi}{2} (x+1)$, equations (21) represent a system of differential equations with respect to the unknown functions $c_1(t), \dots, c_m(t)$. These will be approximated at different levels of time, being known at the prior level. For the envisaged numerical example, $u(x, 0) = 0$ and hence $c_1^*(0) = \dots = c_m^*(0) = 0$.

To system (21) of the shape $u' = F(t, u)$, one could apply different numerical procedures in order to get an approximate solution. For instance a predictor-corrector procedure involves

$$(23) \quad \bar{u}_{j+1} = u_j + \delta t F(t_j, u_j) \quad (\text{the predictor})$$

$$u_{j+1} = u_j + \frac{\delta t}{2} [F(t_j, u_j) + F(t_{j+1}, \bar{u}_{j+1})] \quad (\text{the corrector})$$

where, of course, the corrector can be retaken.

The result of the numerical integration represents u^* at the level of time t_{j+1} . The associated function for u_{j+1}^* is then an approximation of the solution of problem (19) at that time level. The algorithm of this procedure is then the following:

One knows the approximate solution at the time level t_j , its coefficients being B_1, \dots, B_n .

1. It evaluates

$$F_k(t_j, B_1, \dots, B_m) = - \left(\frac{k\pi}{2} \right)^2 B_k - \int_{-1}^1 u u_x \Phi_k dx + f_k(t_j)$$

2. It calculates the predictor

$$\bar{c}_k = B_k + \delta t F_k(t_j, B_1, \dots, B_m) \quad k = 1, \dots, m$$

3. It calculates the associated function for $\bar{c}_1, \dots, \bar{c}_m$, $U(\bar{u}^*)$, as the limit of the sequence

$$u^{s+1} = \bar{u}^* + H_m (N u^s - f(t_{j+1})) - \sum_{k=m+1}^n \frac{4}{k^2 \pi^2} \frac{\bar{c}_k - B_k}{\delta t} \Phi_k$$

where the iterations stop at a convenient rank S .

4. It evaluates

$$F_k(t_{j+1}, \bar{c}_1, \dots, \bar{c}_m) = - \left(\frac{k\pi}{2} \right)^2 \bar{c}_k - \int_{-1}^1 U U_x \Phi_k dx + f_k(t_{j+1})$$

for $k = 1, \dots, m$

5. It calculates the "corrected" c_1, \dots, c_m

$$c_k = B_k + \frac{\delta t}{2} [F_k(t_j, B) + F_k(t_{j+1}, \bar{c})] \quad k = 1, \dots, m$$

Steps 3, 4, 5 are resumed (if necessary)

6. It calculates the associated function for c_1, \dots, c_m as the limit of the sequence

$$u^{s+1} = u^* + H_m (N u^s - f(t_{j+1})) - \sum_{k=m+1}^n \frac{4}{k^2 \pi^2} \frac{c_k - B_k}{\delta t} \Phi_k$$

where the iterations are stopped at a convenient rank S .

In c_1, \dots, c_m we have now the coefficients of the approximate solution of problem (19) at the level of time t_{j+1} .

7. The values got through new c_k come into B_k , $k = 1, \dots, n$ and step 1 is resumed for a new level of time.

5. Numerical results

(a) In the stationary case, with the same prescribed data, we have also used for our problem for finite difference method, i.e. the following scheme

$$\frac{u_{j+1} - 2u_j + u_{j-1}}{\delta x^2} - u_j \frac{u_{j+1} - u_j}{\delta x} + f_j = 0 \quad j = 1, \dots, 15$$

where $u_0 = u_{16} = 0$, $\delta x = 1/8$ ($n = 16$). Here u_j represents the approximation of the exact solution u taken in the nodal point $x_j = -1 + 8/j$, $j = 1, \dots, 15$. The resulting nonlinear system has been solved by

Newton method, obtaining after five iterations an error of the replacement into the system of magnitude 6×10^{-8} , the norm of the corrections on the solution being 1.8×10^{-9} .

The same problem has been numerically treated with the alternative method too. For $m = 1$ iterations (17) are convergent with a contraction constant evaluated at $\alpha = 1/2$ on the respective orbit. System (18) leads then to only one nonlinear equation which has been solved by the method of bisection. The number of iterations for determining the associate function has varied from 9 at the beginning up to 1 at the end (for getting a difference between two successive approximations less than 10^{-6}), the computed coefficient c_1 having the same first 8 digits as the exact coefficient $c_{1ex} = 1.0320491$. The even coefficients have vanished (vis-a-vis the computer accuracy), $c_3 - c_{11}$ have been computed with errors less than 5×10^{-9} with regard to the exact solution, c_{13} with an error of 8×10^{-9} and c_{15} with an error of 10^{-7} .

The values of the solution taken in nodal points $x_j = -1 + j/8$, by the above two methods, with the corresponding errors, are the following :

j	finite difference	error $\times 10^{-2}$	alternative method	error $\times 10^{-4}$
0	0.	0.	0.	0.
1	0.242105	0.8	0.234749	4.
2	0.412932	1.6	0.437232	-3.
3	0.632417	2.3	0.609579	2.
4	0.780453	3.	0.749833	-2.
5	0.896877	3.7	0.859520	2.
6	0.981464	4.4	0.937368	-2.
7	1.033911	5.	0.984500	2.
8	1.053839	5.	0.999877	-2.

The values on the other nodal points $x_9 \dots x_{16}$ are almost symmetrical, their errors being of the same order of magnitude.

According to the above results, the use of the alternative method with $n = 16$ has led to errors with two orders of magnitude less than in the case of finite differences method. At the same time the volume of calculations has been smaller than that in the case of the use of a spectral method, the solving of only one nonlinear equation even with some supplementary iterations being a task much easier than the solving of a system of 15 nonlinear equations for the same order of accuracy.

(b) In the evolution case, we used (to compare the numerical results) the implicit finite differences Crank-Nicolson method for u_{xx} and the explicit Euler method for uu_x . The corresponding system

$$\frac{u_k^{j+1}}{\delta t} - \frac{u_{k+1}^{j+1}}{2\delta x^2} + \frac{u_k^{j+1}}{\delta x^2} - \frac{u_{k-1}^{j+1}}{2\delta x^2} = \frac{u_{k+1}^j - 2u_k^j + u_{k-1}^j}{2\delta x^2} - u_k^j \frac{u_{k+1}^j - u_k^j}{\delta x} + f_k^j,$$

which is linear with tridiagonal matrix, allows the calculation of the numerical solution at a new time level using its previous values. Choosing $\delta t = \pi/32$, $\delta x = 1/8$ it remarks that the maximum of the error is reached at $x = 0$.

The alternative method using the Euler technique for equation (21), i.e.

$$\frac{\partial u^*}{\partial t} \Big|_{t=t_{j+1}} \approx \frac{u_{j+1}^* - u_j^*}{\delta t}$$

becomes now an implicit method. In this case, $m = 1$ leads to a contraction constant for iterations (20) of about 0.9, which implies a very slow convergence. But the value $m = 2$ leads to $\alpha \approx 1/2$ for a time step $\delta t = \pi/32$, $n = 16$.

The alternative method has been also used together with the predictor-corrector method (23), the corrector being applied only once. To get the predictor, we need about 12 iterations while the corrections have involved in (20) 1-5 iterations. The results at different time levels (the numerical values of the solution at $x = 0$, with the corresponding errors with respect to the exact solution are :

t $x\pi/32$	finite difference	error $\times 10^{-2}$	alternative Euler	error $\times 10^{-3}$	alternative pred.-corr.	error $\times 10^{-4}$
1	0.107132	0.9	0.097747	0.27	0.097908	-1.1
2	0.212022	1.7	0.194213	0.88	0.194837	-2.5
3	0.313310	2.3	0.288539	1.6	0.289862	-4.2
4	0.410377	2.7	0.379861	2.8	0.382071	-6.1
5	0.502842	3.1	0.467363	4.0	0.470582	-8.2
6	0.590110	3.5	0.550219	5.3	0.554546	-10.

Going on with the numerical solving with new time steps up to $t = 2\pi$ (the length of a period), the error by using the predictor-corrector method for $m = 2$, $\delta t = \pi/32$ has an oscillating variation with a maximum amplitude 2.7×10^{-3} . When the period is over the error is 5.5×10^{-4} by comparison to its beginning which confirm the stability of the method with these parameters.

While growing m to 3, 4, 6, 8 in order to accelerate the convergence of iterations (20), the phenomenon of the instability of the numerical solutions is recorded, the step $\delta t = \pi/32$ being now too large.

On the other hand, the step $\delta t = \pi/64$, $m = 3$ leads to a contraction constant $\alpha \approx 1/3$, the procedure becoming again stable and the oscillations of the error have an amplitude less than 6.9×10^{-4} while at the end of a period the total error does not exceed 1.25×10^{-4} . As we could expect, the diminishing of the step time has been joined to the growing up of m in order to ensure the upper boundness of the contraction constant less than 1/2 (which seems to be numerically convenient).

The growing of n did not lead to some spectacular effects, the error diminishing with only about 30%. This shows that the most important error source is the numerical integration and not the space spectral discretisation.

Retaking many times the corrector (23), at $m = 2$, $n = 16$, $\delta t = \pi/32$ the stability of the method is kept up, the maximum of the error amplitude being under 1.8×10^{-4} .

6. Final remarks

The alternative method seems to be a real way to improve the accuracy of the numerical solutions, to reduce the computing effort and it could be easily combined with the spectral methods both for stationary and nonstationary case. A detailed study of the stability of the method and of the time step δt dependence on m and n for different procedures of temporary numerical integration will be the topic of a future work.

The authors thank prof. G. Labrosse from the University Paris-Orsay for his helpful remarks and suggestions made while he conducted the seminar of spectral methods at the University of Cluj-Napoca.

REFERENCES

1. H. Bateman, *Some recent researches on the motion of fluids*. Mon. Weather Rev. 43, 163—170, 1915.
2. Burgers, J. M., *A mathematical model illustrating theory of turbulence*. Adv. Appl. Mech. 1, 171—199, 1948.
3. E. Hopf, *The partial differential equation $u_t + uu_x = u_{xx}$* , Comm. on Pure and Appl. Math., vol. III, 3, 201—230, 1950
4. J. Cole, *On a quasilinear parabolic equation occurring in aerodynamics*. Quart. of applied math., vol. IX, 3, 225—236, 1951
5. N. Bressan, A. Quarteroni, *An implicit/explicit spectral method for Burgers' equation*, Calcolo, XXIII, fasc. III, 265—284, 1986
6. C. Canuto, M. Y. Hussaini, A. Quarteroni, Th. A. Zang, *Spectral methods in fluid dynamics* Springer Verlag, 1988.
7. L. Cesari, *Functional analysis and Galerkin's method*, Mich. Math. J., 11, 3, 1964, 383—414.
8. D. Trif, *La méthode de l'alternative et les solutions numériques des équations $L_y = N_y$* . Preprint 7, 70—89, 1984, Univ. of Cluj-Napoca.

Received 1.IX. 1992

Universitatea din Cluj-Napoca
 Facultatea de Matematică
 Str. Kogălniceanu 1
 3400 Cluj-Napoca
 România