

A SPECTRAL METHOD FOR A NONLINEAR EQUATION
ARISING IN FLUIDIZED BED MODELLING

L. Abia, J.M. Sanz-Serna
Departamento de Matemática Aplicada y Computación
Facultad de Ciencias
Universidad de Valladolid
Valladolid Spain

To appear in Numerical Treatment of Differential Equations, Proceedings
of the Fifth Seminar "NUMDIFF-5" held in Halle, 1989. Edited by Karl --
Strehmel, Teubner, Leipzig.

A SPECTRAL METHOD FOR A NONLINEAR EQUATION ARISING IN FLUIDIZED BED MODELLING

L. Abia, J. M. Sanz-Serna
Valladolid, Spain

1. Introduction

We consider the nonlinear, periodic initial-value problem

$$(1.1) \quad u_t + u_{xxx} + \beta(u^2)_x + (\gamma/2)(u^2)_{xx} + \epsilon u_{xx} - \delta u_{tx} = 0, \quad -\infty < x < \infty, \quad 0 < t \leq T < \infty,$$

$$(1.2) \quad u(x,t) = u(x+2\pi,t), \quad -\infty < x < \infty, \quad 0 \leq t \leq T,$$

$$(1.3) \quad u(x,0) = q(x).$$

where $\beta, \gamma, \epsilon, \delta$ are given real constants with $\epsilon, \delta > 0$, the unknown u is real-valued and the given function q is 2π -periodic. The problem (1.1)-(1.3) arises in the theory of flow in a gas-particle fluidized bed [7] with u representing the value of a spatially periodic small perturbation of the concentration of particles. Christie & Gans [3] have numerically studied (1.1)-(1.3) by means of finite-difference and modified-Galerkin methods. These authors discovered that the numerical integration of (1.1)-(1.3) is a difficult task, due to the delicate balance between the various terms in (1.1); a balance which is likely to be destroyed by the discretization procedure, resulting in an unstable scheme. In fact many 'reasonable' time-implicit schemes perform in an unexpectedly unstable manner, while the application of other implicit schemes leads to stable computations only if the time step is large enough relative to the mesh-size in space. In [1] Christie and the present authors have explained the numerical difficulties encountered in [3] and suggested a well-behaved finite-difference scheme.

Since the problem (1.1)-(1.3) is periodic, it is only natural to ask whether Fourier spectral and pseudospectral techniques can be successfully applied. In this paper we suggest a time-continuous pseudospectral scheme for (1.1)-(1.3) and prove that produces spectrally small errors. The analytical technique employed is similar to those in [4], [5]. In a forthcoming paper [2] we shall study a time-discrete version of the method presented here.

The new scheme is presented in section 2 and analyzed in section 3. The final section is devoted to some numerical illustrations.

2. Numerical method

We first need some notation. If J is a positive integer, we set $h=2\pi/(2J)$ and consider the grid-points $x_j = jh, j = 0, \pm 1, \pm 2, \dots$. We denote by Z_h the space of 2π -periodic real functions defined on the grid. Thus, each element $V \in Z_h$ is a real sequence $\{V_j\}_{j=0, \pm 1, \dots}$ such that $V_j = V_{j+2J}, j = 0, \pm 1, \dots$. The notation $[V]_p$ refers to the p -th discrete Fourier coefficient of the sequence V , i.e.

$$[\mathbf{V}]_p^\wedge = (1/2\pi) \sum''_{0 \leq |j| \leq 2J} h v_j \exp(-ipjh), \quad -J \leq p \leq J,$$

where the double prime in the summation means that the first and last terms are halved. To recover \mathbf{V} from its Fourier coefficients it is enough to evaluate at the grid points the trigonometric interpolant $V^*(x)$ of \mathbf{V} given by

$$V^*(x) = \sum''_{-J \leq p \leq J} [\mathbf{V}]_p^\wedge \exp(ipx), \quad -\infty < x < \infty.$$

On differentiating this identity and evaluating the result at the grid points we obtain the following definition of the spectral difference operator D , mapping \mathbb{Z}_h into itself

$$(2.1) \quad (D\mathbf{V})_j = \sum''_{-J \leq p \leq J} [\mathbf{V}]_p^\wedge (ip) \exp(ipjh), \quad \mathbf{V} \in \mathbb{Z}_h, \quad j = 0, 1, \dots$$

The relation (2.1) is of course equivalent to the following simple formula for the Fourier coefficients

$$(2.2) \quad [D\mathbf{V}]_p^\wedge = (ip) [\mathbf{V}]_p^\wedge, \quad -J \leq p \leq J.$$

The powers D^2, D^3, \dots of D are, by definition, the composite operators DD, DDD, \dots

With this notation, the time-continuous pseudospectral method for (2.1)-(2.3) consists of looking for a mapping $\mathbf{U} : [0, T] \rightarrow \mathbb{Z}_h$ such that $\mathbf{U}(0)$ is an approximation \mathbf{Q}_h to the (grid restriction of the) initial datum q and

$$(2.3) \quad (d/dt) \mathbf{U}(t) + D^3 \mathbf{U}(t) + \beta D \mathbf{U}^2(t) + (\gamma/2) D^2 \mathbf{U}^2(t) + \varepsilon D^2 \mathbf{U}(t) - \delta D (d/dt) \mathbf{U}(t) = \mathbf{0}, \quad 0 \leq t \leq T.$$

Here $\mathbf{U}(t)$ approximate the grid restriction of the solution $u(\cdot, t)$.

For implementation purposes it is best to transform (2.3) to obtain, on taking into account (2.2), the following system of ODE for the Fourier coefficients $[\mathbf{U}(t)]_p^\wedge, -J \leq p \leq J$, of $\mathbf{U}(t)$

$$(2.4) \quad (d/dt) [\mathbf{U}(t)]_p^\wedge + (ip)^3 [\mathbf{U}(t)]_p^\wedge + \beta (ip) [\mathbf{U}^2(t)]_p^\wedge + (\gamma/2) (ip)^2 [\mathbf{U}^2(t)]_p^\wedge \\ + \varepsilon (ip)^2 [\mathbf{U}(t)]_p^\wedge - \delta (ip) (d/dt) [\mathbf{U}(t)]_p^\wedge = \mathbf{0}, \quad -J \leq p \leq J, \quad 0 \leq t \leq T,$$

or

$$(2.5) \quad (d/dt) [\mathbf{U}(t)]_p^\wedge = -(1 - \delta ip)^{-1} \{ (ip)^3 [\mathbf{U}(t)]_p^\wedge + \beta (ip) [\mathbf{U}^2(t)]_p^\wedge + (\gamma/2) (ip)^2 [\mathbf{U}^2(t)]_p^\wedge \\ + \varepsilon (ip)^2 [\mathbf{U}(t)]_p^\wedge \}, \quad -J \leq p \leq J, \quad 0 \leq t \leq T.$$

On denoting by \mathbf{Y} the vector of unknown Fourier coefficient, the system (2.5) takes the form

$$(2.6) \quad (d/dt) \mathbf{Y} = \mathbf{f}(\mathbf{Y})$$

and can be advanced in time by means of any ODE solver, e.g. by means of an automatic package. Note that, according to (2.5), one evaluation of the right hand side of (2.6) requires an inverse

Fourier transform to recover \mathbf{U} from its Fourier coefficients and a Fourier transform to find the coefficients of \mathbf{U}^2 . Further comments on implementation will be given in the final section.

3. Analysis

We first construct the energy norm which will be used later in the stability and convergence analysis of (2.3).

Let D^{-1} represent the operator in Z_h defined by the relations

$$D(D^{-1}\mathbf{V}) = \mathbf{V} - \langle \mathbf{V} \rangle \mathbf{1},$$

$$\langle D^{-1}\mathbf{V} \rangle = \langle \mathbf{V} \rangle,$$

where $\langle \cdot \rangle$ denotes mean value (i.e. $\langle \mathbf{V} \rangle = [\mathbf{V}]_0$) and $\mathbf{1}$ represents the grid function which takes the value 1 at each grid point. In terms of Fourier coefficients, D^{-1} is defined by the formulas

$$[D^{-1}\mathbf{V}]_p = (ip)^{-1} [\mathbf{V}]_p, \quad p = \pm 1, \dots, \pm J,$$

$$[D^{-1}\mathbf{V}]_0 = [\mathbf{V}]_0$$

The energy norm $\|\cdot\|_E$ in Z_h is then defined by

$$(3.1) \quad \|\mathbf{V}\|_E^2 = \|D^{-1}\mathbf{V}\|^2 + \delta^2 \|\mathbf{V}\|^2,$$

where $\|\cdot\|$ is the standard discrete L^2 -norm

$$\|\mathbf{V}\|^2 = \sum_{0 \leq j \leq 2J} h (V_j)^2.$$

Note that for each fixed δ the energy norm is equivalent to the discrete L^2 -norm uniformly in h .

Nonlinear stability

For any $t^* \in [0, T]$, let $\mathbf{V}, \mathbf{W} : [0, t^*] \rightarrow Z_h$ be Z_h -valued continuously differentiable functions and, for $0 \leq t \leq t^*$, define the residuals

$$(3.2) \quad F(t) = (d/dt) \mathbf{V}(t) + D^3 \mathbf{V}(t) + \beta D \mathbf{V}^2(t) + (\gamma/2) D^2 \mathbf{V}^2(t) + \varepsilon D^2 \mathbf{V}(t) - \delta D(d/dt) \mathbf{V}(t),$$

$$(3.3) \quad G(t) = (d/dt) \mathbf{W}(t) + D^3 \mathbf{W}(t) + \beta D \mathbf{W}^2(t) + (\gamma/2) D^2 \mathbf{W}^2(t) + \varepsilon D^2 \mathbf{W}(t) - \delta D(d/dt) \mathbf{W}(t).$$

Thus $\mathbf{V}(t), \mathbf{W}(t)$ can be viewed as solutions of (2.3) after perturbations, with $F(t), G(t)$ being the perturbations. The stability analysis attempts to estimate the distance between $\mathbf{V}(t)$ and $\mathbf{W}(t)$ in terms of the distance between $F(t)$ and $G(t)$. The following result holds:

Theorem 3.1 (Stability). Given a positive constant R , there exists a positive constant C , depending only on R, T, γ and β , such that for each $t^* \in [0, T]$ and for each pair of continuously differentiable

functions $V, W: [0, t^*] \rightarrow Z_h$ with

$$(3.4) \quad \|V(t) + W(t)\|_{\infty} \leq R, \quad 0 \leq t \leq t^*,$$

the following bound holds

$$(3.5) \quad \|V(t) - W(t)\|_E^2 \leq C \{ \|V(0) - W(0)\|_E^2 + \max_{0 \leq t \leq t^*} \|F(t) - G(t)\|_E^2 \}, \quad 0 \leq t \leq t^*,$$

where F y G denote the residuals (3.2)-(3.3) associated with V and W respectively.

Proof. Set $E(t) = V(t) - W(t)$, $L(t) = F(t) - G(t)$ and subtract (3.3) from (3.2) to arrive at

$$(3.6) \quad (d/dt) E + D^3 E + \beta D[V^2 - W^2] + (\gamma/2) D^2[V^2 - W^2] + \varepsilon D^2 E - \delta D(d/dt)E = L$$

We now apply the operator D^{-1} to (3.6). In doing so, it should be observed that, on forming the mean value of (3.6), $\langle L \rangle = \langle (d/dt)E \rangle$, while $D^2 E$, $D[V^2 - W^2]$ and DE have zero mean. Thus (3.5) implies

$$\begin{aligned} (d/dt) D^{-1} E + D^2 E + \beta [V^2 - W^2] - \beta \langle V^2 - W^2 \rangle + (\gamma/2) D[V^2 - W^2] \\ + \varepsilon DE - \delta (d/dt)E + \delta \langle L \rangle = D^{-1} L \end{aligned}$$

On taking the inner product of this formula with $D^{-1} E - \delta E$ and manipulating, we arrive at

$$\begin{aligned} (3.7) \quad ((d/dt)(D^{-1} E - \delta E), D^{-1} E - \delta E) + (D^2 E, D^{-1} E - \delta E) = -\beta \{ [(V^2 - W^2) - \langle V^2 - W^2 \rangle], D^{-1} E - \delta E \} \\ - (\gamma/2) (D[V^2 - W^2], D^{-1} E - \delta E) - \varepsilon (DE, D^{-1} E - \delta E) + (D^{-1} L - \delta \langle L \rangle, D^{-1} E - \delta E). \end{aligned}$$

The definition in (3.1) shows that the first term in the left hand side of (3.7) equals $(1/2)(d/dt)\|E\|_E^2$. The second term equals $\delta \|DE\|^2$, since the operator D is skew-symmetric. We now successively consider each of the inner products in the right hand side of (3.7). The first of them can be bounded using the Cauchy-Schwartz inequality and (3.4)

$$\begin{aligned} (3.8) \quad |-\beta \{ [(V^2 - W^2) - \langle V^2 - W^2 \rangle], D^{-1} E - \delta E \}| \leq \beta \|V^2 - W^2\| \|D^{-1} E - \delta E\| \\ \leq \beta \|V + W\|_{\infty} \|E\| (\|D^{-1} E\| + \delta \|E\|) \leq C \|E\|_E^2. \end{aligned}$$

Here and later C represents a constant depending only on the allowed quantities and not necessarily the same at each occurrence. Turning now to the second inner product in the right hand side of (3.7), we have

$$(3.9) \quad (\gamma/2) (D[V^2 - W^2], D^{-1} E - \delta E) = -(\gamma/2) (V^2 - W^2, E - \langle E \rangle) + (\gamma\delta/2) (V^2 - W^2, DE).$$

The first term in the right hand side of (3.9) can be bounded as in (3.8); for the second we write

$$\begin{aligned}
|(\gamma\delta/2) (V^2 - W^2, DE)| &\leq (\gamma\delta/2) \|V + W\|_{\infty} \|V - W\| \|DE\| \\
&\leq (\gamma\delta/2) R [(\eta/2) \|E\|^2 + (1/(2\eta)) \|DE\|^2]
\end{aligned}$$

where $\eta = \eta(\gamma, R)$ is chosen so large that $(\gamma R)/(4\eta) < 1$. This concludes the treatment of the γ term in (3.7). The term with ε satisfies

$$\varepsilon (DE, D^{-1}E - \delta E) = \varepsilon (DE, D^{-1}E) - \varepsilon (E, DD^{-1}E) = -\varepsilon \|E - \langle E \rangle\|^2 \leq 0.$$

Finally

$$(D^{-1}L - \delta \langle L \rangle, D^{-1}E - \delta E) \leq \|L\|_E \|E\|_E \leq (1/2) (\|L\|_E^2 + \|E\|_E^2).$$

Substitution of all the estimates in (3.7) leads to

$$(1/2) (d/dt) \|E\|_E^2 \leq C (\|E\|_E^2 + \|L\|_E^2)$$

and the Gronwall inequality yields (3.5). Q. E. D.

Remark. It should be noted that the previous theorem is only local due to the hypothesis (3.4). This sort of local stability result is typically the strongest that can be proved for nonlinear partial differential equations, see [6], [8], [9].

Consistency and convergence

By definition, the truncation error \mathfrak{F} is the residual associated with the grid-restriction $r_h u$ of the theoretical solution, as in (3.2)-(3.3)

$$\mathfrak{F}(t) = (d/dt) r_h u(\cdot, t) + D^3 r_h u(\cdot, t) + \beta D r_h u^2(\cdot, t) + (\gamma/2) D^2 r_h u^2(\cdot, t) + \varepsilon D^2 r_h u(\cdot, t) - \delta D(d/dt) r_h u(\cdot, t).$$

By (1.1), we can write

$$\begin{aligned}
(3.10) \quad \mathfrak{F}(t) &= [D^3 r_h u - r_h u_{xxx}] + \beta [D r_h u^2 - r_h (u^2)_x] + (\gamma/2) [D^2 r_h u^2 - r_h (u^2)_{xx}] \\
&\quad + \varepsilon [D^2 r_h u - r_h u_{xx}] - \delta [D r_h u_t - r_h u_{tx}].
\end{aligned}$$

The terms in brackets in (3.10) are easily bounded by using e.g. the lemma 2.2 of [10]. If, for some $s > 1/2$, $u \in H^{s+3}$, $u^2 \in H^{s+2}$, $u_t \in H^{s+1}$, uniformly in t (here H^p denotes the standard periodic Sobolev space), then

$$(3.11) \quad \max_{0 \leq t \leq T} \|\mathfrak{F}(t)\|_E = O(h^s), \quad h \rightarrow 0.$$

It is important to note that the exponent in (3.11) (i.e. the order of consistency) depends only on the smoothness of u . In particular, if u is indefinitely differentiable, the truncation error tends to 0 faster

than any power of h . Furthermore the truncation error may even be exponentially small [10].

We can now prove:

Theorem 3.2 (convergence). Assume that: (i) The problem (1.1)-(1.3) possesses a unique classical solution u for which (3.11) holds with $s > 1/2$. (ii) The starting vectors Q_h satisfy

$$(3.12) \quad \|Q_h - r_h q\|_E = O(h^s), \quad h \rightarrow 0.$$

Then, for h sufficiently small, the Cauchy problem given by (2.3) and $U(0) = Q_h$ possesses a unique solution U defined for $0 \leq t \leq T$ and

$$(3.13) \quad \max_{0 \leq t \leq T} \|U(t) - r_h u(\cdot, t)\|_E = O(h^s), \quad h \rightarrow 0.$$

Proof. A solution U exists which is defined, at least, for small values of t , $0 \leq t \leq t_{\max}(h) > 0$. Set $R = 1 + 2M$ with

$$M = \max_{0 \leq t \leq T} \{ \|r_h u(\cdot, t)\|_{\infty} \}$$

and note that

$$(3.14) \quad \|U(t) - r_h u(\cdot, t)\|_{\infty} < 1$$

implies

$$\|U(t) + r_h u(\cdot, t)\|_{\infty} < R.$$

The assumption (3.12) implies that at $t = 0$, (3.14) holds, provided that h is sufficiently small. Denote by $t(h)$ the largest number $S \leq T$ such that $U(t)$ exists and satisfies (3.14) for $0 \leq t < S$. By continuity, either $t(h) = T$ or the left hand side of (3.14) evaluated at $t(h)$ equals 1. On applying Theorem 3.1 with $V = U$, $W = r_h u$, $t^* = t(h)$ and noticing that F vanishes identically while G represents the truncation error, we conclude that

$$\max_{0 \leq t \leq t(h)} \|U(t) - r_h u(\cdot, t)\|_E = O(h^s), \quad h \rightarrow 0.$$

Therefore the left hand side of (3.14) at $t = t(h)$ is $O(h^{s-1/2}) = o(1)$ ($h \rightarrow 0$), so that for h small

$$\|U(t(h)) - r_h u(\cdot, t(h))\|_{\infty} < 1$$

and as a consequence $t(h) = T$. Q. E. D.

4. Numerical examples

As in [3] and [1], we consider the nonlinear problem (1.1)-(1.3) with the parameter values $\beta = -0.45000$, $\gamma = 0.37947$, $\delta = 0.04216$, $\varepsilon = 0.09487$ and the initial condition $q(x) = 0.1 \sin(x)$.

To advance in time the system of ODEs (2.3)-(2.5) we have employed a combination of the trapezoidal rule and the two-step Adams-Bashforth algorithm. More precisely, if Y_p^n, Z_p^n denote the approximation to the Fourier coefficients $[U(t_n)]_p, [U^2(t_n)]_p, t_n = nk, k > 0$, we discretize (2.5) as

$$(4.1) \quad (Y_p^{n+1} - Y_p^n) / k = -(1 - \delta ip)^{-1} \{ (ip)^3 [(Y_p^{n+1} + Y_p^n) / 2] + \beta (ip) [(3 Z_p^n - Z_p^{n-1}) / 2] \\ + (\gamma/2) (ip)^2 [(3 Z_p^n - Z_p^{n-1}) / 2] + \epsilon (ip)^2 [(3 Y_p^n - Y_p^{n-1}) / 2] \}, \quad -J \leq p \leq J, \quad n = 1, 2, \dots, [T/k]$$

Thus the nonlinear terms and the term with ϵ are treated explicitly, while the stiffest term arising from u_{xxx} in (1.1) is dealt with in an implicit manner. Therefore there is no nonlinear equations to be solved and each time-step requires a Fourier transform and an inverse Fourier transform.

The scheme (4.1) is analysed in [2]. The missing starting level $n = 1$ is computed by a step of the standard forward Euler scheme applied to (2.5), so that the overall algorithm is second order accurate in time. While the results quoted here correspond to a constant time-step, variable time-step implementations are easily formulated.

The following table gives the absolute errors at $x = \pi, t = 20$ (the true solution is $u(\pi, 20) = -0.258126$).

	2J = 4	2J = 8
k = 0.1	0.005 827	0.000 794
k = 0.05	0.006 215	0.000 183
k = 0.025	0.006 295	0.000 047
k = 0.0125	0.006 315	0.000 012

Note that the errors in the first column are roughly independent of k . This shows that for $2J = 4$ the spatial errors dominate. For $2J = 8$, the picture is reversed and the errors show an $O(k^2)$ behaviour, so that the spatial error is negligible. Such a drastic error reduction when doubling the number of grid points is typical of pseudospectral methods and cannot be found when using finite differences or finite elements.

As a comparison we have implemented the method suggested in [1], with the Adams-Bashforth/trapezoidal rule time-stepping used in (4.1).

The errors are as follows:

	2J = 32	2J = 64	2J = 128
k = 0.1	0.030588	0.006865	0.001118

Errors corresponding to smaller values of k are not given, as a decrease in k increases the computational costs without reducing the error. It is clear that the pseudospectral scheme is far more accurate. As far as computational costs go, the most expensive run in Table 1 ($2J = 8, k = 0.0125$) took 7 seconds CPU time in a VAX 11/760 with the Fast Fourier Transform coded by us in FORTRAN.

The most expensive run in Table 2 took 13 seconds. (The CPU time of the remaining runs can be found from those we have just given, as the cost increases linearly with J and $1/k$.) When both accuracy and cost are taken into account the superiority of the pseudospectral method is perfectly clear.

References

- [1] Abia, L., I. Christie and J. M. Sanz-Serna, *Stability of schemes for the numerical treatment of an equation modelling fluidized beds*, RAIROM²AN 23(1989), pp. 125-138.
- [2] Abia, L. and J. M. Sanz-Serna, *The spectral accuracy of a fully-discrete scheme for a nonlinear third order equation*, (in preparation).
- [3] Christie, I. and G. H. Ganser, *A numerical study of nonlinear waves arising in a one-dimensional model of a fluidized bed*, J. Comput. Phys. 81(1989), pp. 300-318.
- [4] de Frutos, J., T. Ortega and J. M. Sanz-Serna, *A pseudo-spectral method for the "good Boussinesq equation*, (submitted).
- [5] de Frutos, J. and J. M. Sanz-Serna, *Split-step spectral schemes for nonlinear Dirac systems*, J. Comput. Phys. 83(1989) (in the press).
- [6] de Frutos, J. and J. M. Sanz-Serna, *h-dependent stability thresholds avoid the need for a priori bounds in nonlinear convergence proofs*, in *Computational Mathematics III, Proceedings of the Third International Conference on Numerical Analysis and its Applications, Benin City, Nigeria, 1987*, edited by Simeon Ola Fatunla, (in the press).
- [7] Ganser, G. H. and D. A. Drew, *Nonlinear periodic waves in a two-phase flow model*, SIAM J. Appl. Math. 47 (1987), pp. 726-736.
- [8] López-Marcos, J. C. and J. M. Sanz-Serna, *A definition of stability for nonlinear problems*, in *Numerical Treatment of Differential Equations, Proceedings of the Fourth Seminar NUMDIFF-4 held in Halle 1987*, edited by K. Strehmel, Teubner-Texte zur Mathematik, Leipzig, 1988, pp.216-226.
- [9] López-Marcos, J. C. and J. M. Sanz-Serna, *Stability and convergence in numerical analysis III, linear investigation of nonlinear stability*, IMA J. Numer. Anal. 7(1988), pp.71-84.
- [10] Tadmor, E., *The exponential accuracy of Fourier and Chebyshev schemes*, SIAM J. Numer. Anal. 23 (1986), pp.1-10.