

Split-Step Spectral Schemes for Nonlinear Dirac Systems

J. DE FRUTOS AND J. M. SANZ-SERNA

*Departamento de Matemática Aplicada y Computación, Facultad de Ciencias,
Universidad de Valladolid, Valladolid, Spain*

Received March 22, 1988; revised September 6, 1988

The paper considers split-step spectral schemes for the numerical integration of nonlinear Dirac systems in $[1 + 1]$ -dimensions. Proofs of stability and convergence are given along with numerical experiments which clearly show the superiority of the suggested methods over standard and split-step finite-difference algorithms. © 1989 Academic Press, Inc.

1. INTRODUCTION

It is well known that nonlinear modifications of the (linear) time-dependent Schrödinger equation play an important role in the mathematical modelling of many phenomena. Often, the success of such modifications stems from the fact that the nonlinearity can oppose the dispersive behaviour of the linear terms, thus making it possible for solitary waves to exist [28]. The importance of the applications of the nonlinear Schrödinger equations has resulted in a rather large number of papers being devoted to their numerical solution (see, e.g., [14, 20, 27] and their references). The Dirac system, which, to some extent provides the relativistic counterpart of the Schrödinger equation, can also be subjected to useful nonlinear modifications. In the $(1 + 1)$ -dimensional case, nonlinear Dirac systems can be written in the form

$$u_t = Au_x + if(|u_1|^2 - |u_2|^2)Bu, \quad (1.1)$$

where $u = u(x, t)$ is the spinorial unknown, represented as a 2-dimensional complex vector $u = [u_1, u_2]^T$, i is the imaginary unit, $f(s)$ is a real valued function of a real variable s and A, B denote the matrices

$$A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Systems of the form (1.1) may give rise to *solitary waves* and in the physics literature have been suggested as models of extended particles (see [1] and

references therein). From the numerical point of view, Alvarez *et al.* [2] showed the convergence of a Crank–Nicolson scheme applied to the particular case

$$f(s) = m - 2\lambda s, \quad m, \lambda \text{ real constants.} \quad (1.2)$$

Their analysis explicitly uses the form of the nonlinearity (1.2) and cannot be readily extended to the case of a general f . However, it is now well known that, in the numerical integration of 1-dimensional waves, split-step spectral methods are likely to be more advantageous than finite-difference or finite-element methods (see, e.g., [24, 27]). Split-step spectral methods for wave computations were introduced by Hardin and Tappert [8] and Tappert [25].

The aim of the present paper is first to analyze and then to assess split-step spectral methods for Dirac systems. In this connection, we would like to mention that not many examples of analyses of split-step spectral methods are available in the literature (cf. the final remarks in [13]). Our analytical technique for the Dirac equations can be extended to cover discretizations of other nonlinear wave equations, such as nonlinear modifications of the Schrödinger and Klein–Gordon equations.

The organization of the paper is as follows. Section 2 describes a simple, first order in time, spectral-splitting scheme, whose analysis is presented in detail in Section 3. Section 4 deals with the second order time-splitting technique due to Strang [22]. The final section reports numerical tests of the split-step spectral schemes and comparisons with some standard and split-step finite-difference alternative schemes.

2. SIMPLE SPLITTING

We consider the periodic problem given by (1.1) and

$$u(x+1, t) = u(x, t), \quad -\infty < x < +\infty, \quad 0 \leq t \leq T < \infty, \quad (2.1)$$

$$u(x, 0) = q(x), \quad -\infty < x < +\infty, \quad (2.2)$$

with q a known 1-periodic function. For simplicity, we sometimes use the abbreviation

$$g(z) = if(|z_1|^2 - |z_2|^2)Bz, \quad (2.3)$$

if $z = [z_1, z_2]^T \in \mathbb{C}^2$. A simple split-step scheme can be described as follows:

Time Discretization

If $k > 0$ is the time step, we set $N = [T/k]$ (square brackets denote integer part) and denote by $u^n(\cdot)$, $0 \leq n \leq N$, the function $u(\cdot, t_n)$, with u the solution of (1.1), (2.1)–(2.2), and t_n , the time level $t_n = nk$. When a space-continuous approximation

\mathcal{U}^n to u^n , $n=0, 1, \dots, N-1$, has been computed, an intermediate approximation \mathcal{U}_*^{n+1} is obtained as the solution at $t=t_{n+1}$ of the problem

$$v_t(x, t) = g(v(x, t)), \quad -\infty < x < +\infty, \quad t_n < t \leq t_{n+1}, \quad (2.4a)$$

$$v(x, t_n) = \mathcal{U}^n(x), \quad -\infty < x < +\infty, \quad (2.4b)$$

where there is no evolution due to the linear term in (1.1). The approximation \mathcal{U}^{n+1} corresponding to the advanced time level t_{n+1} is then reached as the solution at time $t=t_{n+1}$ of the problem

$$w_t = Aw_x, \quad -\infty < x < +\infty, \quad t_n < t \leq t_{n+1}, \quad (2.5a)$$

$$w(x+1, t) = w(x, t), \quad -\infty < x < +\infty, \quad t_n < t \leq t_{n+1}, \quad (2.5b)$$

$$w(x, t_n) = \mathcal{U}_*^{n+1}(x), \quad -\infty < x < +\infty, \quad (2.5c)$$

where there is no evolution due to the nonlinear term in (1.1).

If $v(x, t) = [v_1(x, t)^T, v_2(x, t)^T]^T$ satisfies (2.4a), then, taking into account (2.3), we can write, for $v = 1, 2$,

$$\begin{aligned} (d/dt) |v_v(x, t)|^2 &= 2 \operatorname{Re}\{[(d/dt) v_v(x, t)] v_v^*(x, t)\} \\ &= 2 \operatorname{Re}\{(-1)^v \operatorname{if}(|v_1(x, t)|^2 - |v_2(x, t)|^2) |v_v(x, t)|^2\} = 0, \end{aligned}$$

so that $|v_v(x, t)|$ is independent of t . Therefore, recalling (2.3) once more, the ODE (2.4a) can be rewritten as

$$v_t = \operatorname{if}(|v_1(x, t_n)|^2 - |v_2(x, t_n)|^2) Bv, \quad -\infty < x < +\infty, \quad t_n < t \leq t_{n+1}, \quad (2.4a')$$

whence the initial value problem (2.4) can readily be integrated in closed form to yield

$$\mathcal{U}_*^{n+1}(x) = \exp\{\operatorname{if}(|\mathcal{U}_1^n(x)|^2 - |\mathcal{U}_2^n(x)|^2) kB\} \mathcal{U}^n(x), \quad -\infty < x < +\infty, \quad (2.6)$$

where $\mathcal{U}_1^n, \mathcal{U}_2^n$ are the components of \mathcal{U}^n . We shall denote by \mathcal{N}_k the nonlinear operator defined by $\mathcal{N}_k \mathcal{U}^n(x) = \mathcal{U}_*^{n+1}(x)$ with \mathcal{U}_*^{n+1} the function given in (2.6).

Turning now to the problem (2.5) the method of separation of variables shows that

$$\mathcal{U}^{n+1}(x) = \sum_{p \in \mathbb{Z}} \exp\{2\pi ipkA\} [\mathcal{U}_*^{n+1}]_p^\wedge e^{2\pi ipx}, \quad (2.7)$$

where the symbol $[\mathcal{U}_*^{n+1}]_p^\wedge$ refers to the p th Fourier coefficient of the 1-periodic function \mathcal{U}_*^{n+1} . We denote by \mathcal{L}_k the linear operator defined by $\mathcal{L}_k \mathcal{U}_*^{n+1} = \mathcal{U}^{n+1}$, with \mathcal{U}_*^{n+1} given in (2.7).

Space Discretization

If J is a positive integer number, we set $h = 1/(2J)$ and consider the mesh-points $x_j = jh$, $0 \leq j \leq 2J$. We denote by $\mathbb{C}_p^{2(2J+1)}$ the subspace of all vectors $\mathbf{V} =$

$[V_0^T, V_1^T, \dots, V_{2J}^T]^T$ in $\mathbb{C}^{2(2J+1)}$ with $V_0 = V_{2J}$. If $\mathbf{U}^n = [U_0^{nT}, U_1^{nT}, \dots, U_{2J}^{nT}]^T \in \mathbb{C}_p^{2(2J+1)}$ is a vector containing approximations U_j^n to $\mathcal{Q}^n(x_j)$, $j=0, 1, \dots, 2J$, we obtain approximations U_{*j}^{n+1} to $\mathcal{Q}_{*j}^{n+1}(x_j)$ by setting

$$U_{*j}^{n+1} = \exp\{if(|U_{j1}^n|^2 - |U_{j2}^n|^2)kB\} U_j^n, \quad 0 \leq j \leq 2J. \quad (2.8)$$

We introduce the operator N_k defined through $\mathbf{U}_{*j}^{n+1} = N_k \mathbf{U}^n$, where \mathbf{U}_{*j}^{n+1} is the vector in $\mathbb{C}_p^{2(2J+1)}$ with components U_{*j}^{n+1} given in (2.8). Note that N_k is a discrete version of \mathcal{N}_k and that, after (2.6)–(2.8) there is no local truncation error involved in the substitution of \mathcal{N}_k by N_k , i.e., if $U_j^n = \mathcal{Q}^n(x_j)$, $j=0, 1, \dots, 2J$, then $[N_k \mathbf{U}^n]_j = U_{*j}^{n+1} = \mathcal{Q}_{*j}^{n+1}(x_j) = \mathcal{N}_k \mathcal{Q}^n(x_j)$, $j=0, 1, \dots, 2J$.

Once \mathbf{U}_{*j}^{n+1} has been formed, we use (2.7) to compute approximations U_j^{n+1} to $\mathcal{Q}^{n+1}(x_j)$, according to the formula

$$U_j^{n+1} = \sum_{|p| \leq J}'' \exp\{2\pi i p k A\} [\mathbf{U}_{*j}^{n+1}]_p^\wedge e^{2\pi i p j h}, \quad 0 \leq j \leq 2J, \quad (2.9)$$

where the double prime in the summation means that the terms corresponding to $p = \pm J$ are halved and $[\mathbf{U}_{*j}^{n+1}]_p^\wedge$ is the p th discrete Fourier coefficient of the vector \mathbf{U}_{*j}^{n+1} , i.e.,

$$[\mathbf{U}_{*j}^{n+1}]_p^\wedge = (1/(2J)) \sum_{0 \leq j \leq 2J} U_{*j}^{n+1} e^{-2\pi i p j h}, \quad -J \leq p \leq J. \quad (2.10)$$

We introduce the operator L_k , the space-discrete version of \mathcal{L}_k given by $\mathbf{U}^{n+1} = L_k \mathbf{U}_{*j}^n$. In practice, the computations for $L_k \mathbf{U}_{*j}^n$, i.e., the summations in (2.9)–(2.10), are, of course, best performed by FFT techniques.

To sum up, the numerical method consists of a recursion

$$\mathbf{U}^{n+1} = L_k N_k \mathbf{U}^n, \quad n = 0, 1, \dots, N-1, \quad (2.11)$$

with L_k, N_k the linear and nonlinear operators defined in (2.9)–(2.10) and (2.8), respectively. The initial vector \mathbf{U}^0 for (2.11) is chosen as approximation \mathbf{q} to the vector

$$\mathbf{r}q = [q(x_0)^T, q(x_1)^T, \dots, q(x_{2J})^T]^T, \quad (2.12)$$

where q is the initial function in (2.2). Here and later, \mathbf{r} denotes restriction to the spatial grid.

3. STABILITY AND CONVERGENCE ANALYSIS

To investigate the stability and convergence of the scheme (2.11) we employ a general analytical framework, introduced by López-Marcos and Sanz-Serna [15, 10–12]. The use of this framework makes it possible to avoid the need for a priori estimates for nonlinear problems [6]. To facilitate the readability of the subsequent

analysis, we first present a very brief summary of the general definitions and main result of [10–12]. This is followed by a study of the stability, consistency, and convergence of (2.11).

Discretization Framework

Consider a fixed, given problem concerning a differential or integral equation. Let u be a solution of this problem. We denote by U_h the numerical approximation to u . The subscript h shows that U_h depends on a small parameter h , such as a mesh-size. We assume that h takes values in a set H of positive numbers with $\inf H = 0$. The numerical approximation U_h is obtained, for each fixed h in H , by solving a *discrete problem*

$$\Phi_h(U_h) = 0, \quad (3.1)$$

where Φ_h is a mapping with domain $D_h \subset X_h$ and values in Y_h . Here X_h and Y_h are normed spaces, both real or both complex, with the same finite dimension.

To investigate how close U_h is to u , we choose, for each h in H , an element u_h in D_h . This element is a suitable discrete representation of u . Typically, in a difference method, u_h will be a set of nodal values of u . The *global discretization error* is defined to be the vector $e_h = u_h - U_h$ and the *local discretization error* is given by $l_h = \Phi_h(u_h)$. We say that the discretization (3.1) is convergent if there exists $h_0 > 0$, such that for h in H , $h \leq h_0$, (3.1) has a solution U_h and, as $h \rightarrow 0$, $\lim \|u_h - U_h\| = 0$. The convergence is of order p , if $\|u_h - U_h\| = \mathcal{O}(h^p)$. The discretization (3.1) is consistent (respectively consistent of order p) if as $h \rightarrow 0$, $\|\Phi_h(u_h)\| = o(1)$ (resp. $\mathcal{O}(h^p)$).

Assume that for each h in H , R_h is a value with $0 < R_h \leq +\infty$. We say that (3.1) is *stable restricted to the thresholds* R_h , if there exist two positive constants h_0 and S such that for any h in H , $h \leq h_0$, the open ball $B(u_h, R_h)$ is contained in the domain D_h and for any V_h, W_h in that ball

$$\|V_h - W_h\| \leq S \|\Phi_h(V_h) - \Phi_h(W_h)\|. \quad (3.2)$$

It should be emphasized that the stability bound (3.2) need to be proved not for arbitrary V_h and W_h , but only for vectors V_h and W_h “near” the theoretical solution, near in the sense that $\|V_h - u_h\| < R_h$, $\|W_h - u_h\| < R_h$. Thus, this notion of stability is weaker than others used [12]. However, stability and consistency still imply convergence, namely:

THEOREM 3.1. *Assume that (3.1) is consistent and stable with thresholds R_h . If Φ_h is continuous in $B(u_h, R_h)$ and $\|l_h\| = o(R_h)$ as $h \rightarrow 0$, then:*

(i) *For h small enough, the discrete equations (3.1) possess a unique solution in $B(u_h, R_h)$.*

(ii) *As $h \rightarrow 0$ the solutions in (i) converge with an order of convergence not smaller than the order of consistency.*

We write the scheme (2.11) within the previous abstract framework as follows:

(i) First of all, since only one discretization parameter is allowed in the abstract framework, a relation between k and h needs to be imposed. At this stage, we only assume that $k = \sigma(h)$, where σ is an increasing, continuous function with $\sigma(0) = 0$. Hereafter the subindex h , used so far in the formalism, will often be omitted; for instance, we shall write X, Y, \mathbf{u} rather than X_h, Y_h, \mathbf{u}_h .

(ii) We take $X = Y = (\mathbb{C}_p^{2(2J+1)})^{N+1}$. In $\mathbb{C}_p^{2(2J+1)}$ we use the discrete L^2 and L^∞ -norms

$$\|\mathbf{Z}\| = \left[h \sum_{0 \leq j \leq 2J} |Z_j|^2 \right]^{1/2},$$

$$\|\mathbf{Z}\|_\infty = \max_{0 \leq j \leq 2J} |Z_j|,$$

where $\mathbf{Z} = [Z_0^T, Z_1^T, \dots, Z_{2J}^T]^T \in \mathbb{C}_p^{2(2J+1)}$ and $|\cdot|$ denotes the standard Euclidean norm in \mathbb{C}^2 . In X we use a maximum norm

$$\|\mathbf{V}\|_X = \max\{\|\mathbf{V}^n\| : 0 \leq n \leq N\}, \quad \mathbf{V} = [\mathbf{V}^{0T}, \mathbf{V}^{1T}, \dots, \mathbf{V}^{NT}]^T \in X,$$

and in Y we employ an L^1 -norm

$$\|\mathbf{F}\|_Y = \|\mathbf{F}^0\| + k \sum_{1 \leq n \leq N} \|\mathbf{F}^n\|, \quad \mathbf{F} = [\mathbf{F}^{0T}, \mathbf{F}^{1T}, \dots, \mathbf{F}^{NT}]^T \in Y.$$

The relevance of this choice of norms is made clear by the fact that for linear initial value problems, $L^\infty - L^1$ stability is equivalent to the familiar Lax stability [17, 15].

(iii) On defining the mapping Φ given by $\Phi(\mathbf{V}) = \mathbf{F}$ with

$$\begin{aligned} \mathbf{F}^{n+1} &= k^{-1}(\mathbf{V}^{n+1} - L_k N_k \mathbf{V}^n), & 0 \leq n \leq N-1, \\ \mathbf{F}^0 &= \mathbf{V}^0 - \mathbf{q}, \end{aligned} \tag{3.3}$$

the recursion (2.11) with initial vector \mathbf{q} adopts the abstract from (3.1). Each of the $N+1$ components of Φ corresponds to the computation of a time level.

(iv) Finally the representation of the theoretical solution u is given by the obvious grid-restriction choice

$$\mathbf{u} = [\mathbf{r}u^{0T}, \mathbf{r}u^{1T}, \dots, \mathbf{r}u^{NT}]^T.$$

Stability

We need some preliminary results.

PROPOSITION 3.2. *The operator L_k defined after (2.9)–(2.10) is L^2 -isometric in $\mathbb{C}_p^{2(2J+1)}$.*

Proof. This follows from the Parseval theorem, since $\exp\{2\pi ipkA\}$ is an isometry in \mathbb{C}^2 . ■

PROPOSITION 3.3. *Assume that the function f in (1.1) is continuously differentiable. If D is a bounded subset of \mathbb{C}^2 , then there exists a positive constant $L = L(D, f)$ such that for $x = [x_1, x_2]^T, y = [y_1, y_2]^T$ in D and for $k > 0$,*

$$|\exp\{ikf(|x_1|^2 - |x_2|^2)B\}x - \exp\{ikf(|y_1|^2 - |y_2|^2)B\}y| \leq (1 + kL)|x - y|.$$

Proof. We can write

$$\begin{aligned} & |\exp\{ikf(|x_1|^2 - |x_2|^2)B\}x - \exp\{ikf(|y_1|^2 - |y_2|^2)B\}y| \\ & \leq |[\exp\{ikf(|x_1|^2 - |x_2|^2)B\} - \exp\{ikf(|y_1|^2 - |y_2|^2)B\}]x| \\ & \quad + |\exp\{ikf(|y_1|^2 - |y_2|^2)B\}(x - y)|. \end{aligned}$$

The second term in the right-hand side equals $|x - y|$. It is easy to see that the first can be bounded by $kL_1v|x - y|$, where L_1 is a Lipschitz constant for $f(|x_1|^2 - |x_2|^2)$ in D and $v = \sup\{|x| : x \in D\}$. ■

THEOREM 3.4. *Assume that the function f is continuously differentiable and that u is a classical solution of the problem (1.1), (2.1), (2.2). Then to each $R > 0$, there corresponds a positive constant S , which only depends on R, T, f , and $M = \max\{|u(x, t)| : 0 \leq x \leq 1, 0 \leq t \leq T\}$, such that for any \mathbf{V} and \mathbf{W} in X with*

$$\max_{0 \leq n \leq N} \|\mathbf{V}^n - r\mathbf{u}^n\|_\infty < R, \quad \max_{0 \leq n \leq N} \|\mathbf{W}^n - r\mathbf{u}^n\|_\infty < R, \tag{3.4}$$

the following bound holds:

$$\|\mathbf{V} - \mathbf{W}\|_X \leq S \|\Phi(\mathbf{V}) - \Phi(\mathbf{W})\|_Y. \tag{3.5}$$

Proof. Let \mathbf{V} and \mathbf{W} be elements in X fulfilling (3.4) and set $\mathbf{F} = \Phi(\mathbf{V}), \mathbf{G} = \Phi(\mathbf{W})$. By the definition of Φ given in (3.3)

$$\mathbf{F}^{n+1} - \mathbf{G}^{n+1} = k^{-1}(\mathbf{V}^{n+1} - \mathbf{W}^{n+1}) - k^{-1}(L_k N_k \mathbf{V}^n - L_k N_k \mathbf{W}^n), \quad 0 \leq n \leq N - 1,$$

and therefore

$$\|\mathbf{V}^{n+1} - \mathbf{W}^{n+1}\| \leq \|L_k\| \|N_k \mathbf{V}^n - N_k \mathbf{W}^n\| + k \|\mathbf{F}^{n+1} - \mathbf{G}^{n+1}\|. \tag{3.6}$$

Now, by Proposition 3.2, $\|L_k\| = 1$. On the other hand, (3.4) implies that the components $V_j^n, W_j^n, 0 \leq j \leq 2J, 0 \leq n \leq N$ of \mathbf{V} and \mathbf{W} belong to the ball $D \subset \mathbb{C}^2$ centered at the origin and having radius $R + M$. If $L = L(D, f)$ is the constant from Proposition 3.3

$$\|N_k \mathbf{V}^n - N_k \mathbf{W}^n\| \leq (1 + kL) \|\mathbf{V}^n - \mathbf{W}^n\|.$$

Summing up, (3.6) implies that

$$\|\mathbf{V}^{n+1} - \mathbf{W}^{n+1}\| \leq (1 + kL) \|\mathbf{V}^n - \mathbf{W}^n\| + k \|\mathbf{F}^{n+1} - \mathbf{G}^{n+1}\|, \quad 0 \leq n \leq N - 1,$$

and a simple recursion leads to (3.5). ■

COROLLARY 3.5. *In the hypotheses of the theorem and for any positive R , the scheme (2.11) is stable with thresholds $R_h = Rh^{1/2}$.*

Proof. The relations $\|\mathbf{V} - \mathbf{u}\|_x \leq R_h$, $\|\mathbf{W} - \mathbf{u}\|_x \leq R_h$, imply (3.4) and, according to the theorem, lead to (3.5). ■

Consistency

The $(n + 1)$ -component, $0 \leq n \leq N - 1$, of the local truncation error $\mathbf{l} = \Phi(\mathbf{u})$ is given by

$$\begin{aligned} \mathbf{l}^{n+1} &= k^{-1}(\mathbf{r}u^{n+1} - L_k N_k \mathbf{r}u^n) \\ &= k^{-1}\{\mathbf{r}(u^{n+1} - \mathcal{L}_k \mathcal{N}_k u^n) + [\mathbf{r}(\mathcal{L}_k \mathcal{N}_k u^n) - L_k \mathbf{r}(\mathcal{N}_k u^n)] \\ &\quad + [L_k \mathbf{r}(\mathcal{N}_k u^n) - L_k N_k \mathbf{r}u^n]\}. \end{aligned} \tag{3.7}$$

Thus, the local error appears as the sum of the local splitting error and the local errors in the integration of the fractional steps (2.4)–(2.5) (cf. [9]).

Note that

$$L_k \mathbf{r}(\mathcal{N}_k u^n) - L_k N_k \mathbf{r}u^n = \mathbf{0}, \tag{3.8}$$

because, as noted before, $\mathbf{r}\mathcal{N}_k = N_k \mathbf{r}$. For the local splitting error, we have the following result.

PROPOSITION 3.6. *Assume that the function f in (1.1) is twice continuously differentiable and that the solution u of (1.1), (2.1), (2.2) possesses bounded second derivatives in $[0, 1] \times [0, T]$. Then*

$$\|\mathbf{r}u^{n+1} - \mathbf{r}\mathcal{L}_k \mathcal{N}_k u^n\| \leq C_1 k^2, \quad 0 \leq n \leq N - 1, \tag{3.9}$$

where C_1 is a positive constant independent of n , k , and h .

Proof. The definition of \mathcal{N}_k in (2.6) shows that $\mathcal{N}_k u^n$ possesses second derivatives with respect to x and k , bounded uniformly in n . The method of characteristics reveals that the solution of (2.5) has second bounded derivatives if the initial condition has a bounded second derivative with respect to x . Then a Taylor expansion and (2.5a) imply that

$$\mathcal{L}_k \mathcal{N}_k u^n(x) = \mathcal{N}_k u^n(x) + kA \partial_x \mathcal{N}_k u^n(x) + R_1(x),$$

where $R_1(x) = \mathcal{O}(k^2)$, uniformly in x and n . Analogously

$$\mathcal{N}_k u^n(x) = u^n(x) + kg(u^n(x)) + R_2(x),$$

with $R_2(x) = \mathcal{O}(k^2)$, uniformly in x and n , and

$$\partial_x \mathcal{N}_k u^n(x) = \partial_x u^n(x) + R_3(x),$$

where $R_3(x) = \mathcal{O}(k)$ uniformly in x and n . Summing up,

$$\mathcal{L}_k \mathcal{N}_k u^n(x) = (u^n(x) + k(A \partial_x u^n(x) + g(u^n(x))) + (R_1(x) + R_2(x) + kAR_3(x)),$$

which leads immediately to (3.9). ■

To analyze the local error of the linear fractional step we need the Sobolev space H^s , $s > 0$, whose elements are the 1-periodic functions $w(\cdot)$ with finite s -norm

$$\|w\|_{H^s} = \left(\sum_{p \in \mathbb{Z}} (1 + |p|)^{2s} |[w]_p^\wedge|^2 \right)^{1/2} < +\infty.$$

PROPOSITION 3.7. *If $w \in H^{s+1}$, $s > \frac{1}{2}$, then there exists a positive constant C_2 , depending on w and s but not on h and k , such that*

$$\|\mathbf{r}\mathcal{L}_k w - L_k \mathbf{r}w\| \leq C_2 kh^s \|w\|_{H^{s+1}}. \tag{3.10}$$

Proof. We need the following well-known relation [7], [23] between the Fourier coefficients $[v]_p^\wedge$ of a 1-periodic function and the discrete Fourier coefficients $[\mathbf{r}v]_p^\wedge$ of its grid restriction

$$[\mathbf{r}v]_p^\wedge = \sum_{j \in \mathbb{Z}} [v]_{p+2jJ}^\wedge, \quad v \in H^s, \quad s > \frac{1}{2}, \quad |p| \leq J \tag{3.11}$$

(the series converges absolutely). The identity (3.11) shows that the p th discrete Fourier coefficients of $L_k \mathbf{r}w$ and $\mathbf{r}\mathcal{L}_k w$ are given, respectively, by

$$\exp\{2\pi i p k A\} [\mathbf{r}w]_p^\wedge = \sum_{j \in \mathbb{Z}} \exp\{2\pi i p k A\} [w]_{p+2jJ}^\wedge, \quad |p| \leq J,$$

and

$$\sum_{j \in \mathbb{Z}} [\mathcal{L}_k w]_{p+2jJ}^\wedge = \sum_{j \in \mathbb{Z}} \exp\{2\pi i (p + 2jJ) k A\} [w]_{p+2jJ}^\wedge, \quad |p| \leq J.$$

Therefore, Parseval's theorem yields

$$\begin{aligned} & \|\mathbf{r}\mathcal{L}_k w - L_k \mathbf{r}w\|^2 \\ &= \sum_{|p| \leq J} \left| \sum_{j \neq 0} (\exp\{2\pi i (p + 2jJ) k A\} - \exp\{2\pi i p k A\}) [w]_{p+2jJ}^\wedge \right|^2. \end{aligned} \tag{3.12}$$

The application of the Cauchy–Schwarz inequality to each term in the summation leads to

$$\begin{aligned} & \left| \sum_{j \neq 0}'' (\exp\{2\pi i(p + 2jJ)kA\} - \exp\{2\pi ipkA\}) [w]_{\rho + 2jJ}^\wedge \right|^2 \\ & \leq \left\{ \sum_{j \neq 0} (1 + |p + 2jJ|)^{-2s} \right\} \\ & \quad \times \left\{ \sum_{j \neq 0} (1 + |p + 2jJ|)^{2s} |(\exp\{2\pi i(p + 2jJ)kA\} - \exp\{2\pi ipkA\}) [w]_{\rho + 2jJ}^\wedge|^2 \right\}. \end{aligned}$$

It is easy to show that

$$\|\exp\{2\pi i(p + 2jJ)kA\} - \exp\{2\pi ipkA\}\| \leq 8\pi k |j| J.$$

and therefore (3.12) implies

$$\begin{aligned} \|\mathbf{r}_k \mathcal{L}_k w - L_k \mathbf{r} w\|^2 & \leq \sum_{|p| \leq J}'' \left\{ \left(\sum_{j \neq 0} (1 + |p + 2jJ|)^{-2s} \right) \right. \\ & \quad \times \left. \left(\sum_{j \neq 0} (1 + |p + 2jJ|)^{2s} (8\pi k |j| J)^2 |[w]_{\rho + 2jJ}^\wedge|^2 \right) \right\} \\ & \leq \sum_{|p| \leq J}'' \left\{ \left(J^{-2s} \sum_{j \neq 0} (2|j| - 1)^{-2s} \right) \right. \\ & \quad \times \left. (8\pi k)^2 \sum_{j \neq 0} (1 + |p + 2jJ|)^{2s+2} |[w]_{\rho + 2jJ}^\wedge|^2 \right\}, \end{aligned}$$

where we have taken into account that for $|p| \leq J$,

$$\begin{aligned} |j|J & \leq 1 + |p + 2jJ| \\ J(2|j| - 1) & \leq 1 + |p + 2jJ|. \end{aligned}$$

It is now easily concluded that (3.10) holds with

$$C_2 = \left[(8\pi)^2 2^{2s} \sum_{1 \leq j \leq \infty} 2/(2j - 1)^{2s} \right]^{1/2}. \quad \blacksquare$$

Finally we can formulate the following consistency result.

THEOREM 3.8. *Assume that the hypotheses of Proposition 3.6 hold and that there exist constants $s > \frac{1}{2}$ and $C_3 > 0$ such that*

$$\|\mathcal{N}_k u^n\|_{H^{s+1}} \leq C_3, \quad 0 \leq n \leq N - 1, \quad k > 0. \tag{3.13}$$

Then, as the mesh is refined,

$$\|\Phi(\mathbf{u})\|_Y = \mathcal{O}(k + h^s), \tag{3.14}$$

provided that the starting vectors \mathbf{q} satisfy

$$\|\mathbf{r}\mathbf{u}^0 - \mathbf{q}\| = \mathcal{O}(h^s).$$

Proof. The theorem follows from (3.7)–(3.10). ■

Note that according to (2.6) the hypothesis (3.13) can be enforced by demanding regularity in f and u . In particular, under the hypotheses of Proposition 3.6 (f, u twice continuously differentiable), the bound (3.13) holds with $s=1$. Additional regularity in f and u makes it possible to increase the value of s . When f and u are C^∞ , s can be chosen arbitrarily high and therefore the truncation error in space has bounds of the form $C_s h^s$ with s arbitrarily large. The fact that the order of accuracy of spectral discretizations may be infinite was perhaps first pointed out by Fornberg [4] (cf. [23]).

Convergence

The abstract Theorem 3.1 leads now to:

THEOREM 3.9. *Assume that the hypotheses of the consistency theorem 3.8 are satisfied and that k and h are subject to the relation $k = rh^\nu$, for constants $r > 0$ and $\nu > \frac{1}{2}$. Then, there exists a positive constant C , independent of h and k , such that*

$$\max_{0 \leq n \leq N} \|\mathbf{U}^n - \mathbf{r}\mathbf{u}^n\| \leq C(k + h^s). \tag{3.15}$$

4. STRANG’S SPLITTING

The order of consistency/convergence in time of the simple scheme (2.11) is only 1. A well-known modification introduced by Strang [22] leads to $\mathcal{O}(k^2)$ errors. With our notation, Strang’s splitting leads to the recursion

$$\mathbf{U}^{n+1} = N_{k/2} L_k N_{k/2} \mathbf{U}^n. \tag{4.1}$$

Sheng [21] has recently proved that, for a class of initial value problems, Strang’s splitting is optimal. Note that in view of the relation

$$(N_{k/2} L_k N_{k/2})^n = N_{k/2} (L_k N_k)^{n-1} L_k N_{k/2}$$

the cost of method (4.1) is virtually the same as that of (2.11).

The recursion (4.1) can be analyzed along the lines of the previous section (see [5]). Bounds $\mathcal{O}(k^2 + h^s)$ for the global error $\|\mathbf{U}^n - \mathbf{r}\mathbf{u}^n\|$ can be derived, uniformly in n , under the following hypotheses: (i) The starting vectors \mathbf{q} have errors $\mathcal{O}(h^s)$.

(ii) f is three times continuously differentiable and u possesses bounded third derivatives in $[0, 1] \times [0, T]$. (iii) $\|\mathcal{N}_k u^n\|_{H^{r+1}} \leq C_3$, $0 \leq n \leq N-1$, $k > 0$ for constants $C_3 > 0$ and $s > \frac{1}{2}$. (iv) The mesh refinement is performed according to a rule $k = rh^v$, $r > 0$, $v > \frac{1}{4}$.

5. NUMERICAL EXPERIMENTS

The split-step spectral schemes (2.11) and (4.1) have been implemented in single precision complex arithmetic on a VAX 11/780 machine with a VAX-11 FORTRAN compiler. The Fourier transforms were carried out by the Cooley-Tukey algorithm [3] implemented by us, also in FORTRAN. (No doubt a better coding of the FFT would have increased the efficiency of the spectral schemes, but we preferred not to give them this advantage in the numerical tests.)

For comparison, we also implemented, in the same environment, a leap-frog, a Crank-Nicolson [2], and a Crank-Nicolson split-step finite-difference scheme for (1.1), (2.1), (2.2). All of them used the standard centered difference approximation of ∂_x .

The Crank-Nicolson equations take the form

$$(I - (k/2)L_h)U^{n+1} = (I + (k/2)L_h)U^n + kG((U^{n+1} + U^n)/2), \quad 0 \leq n \leq N-1, \quad (5.1)$$

where L_h is a skew-symmetric matrix and G is a nonlinear mapping. At each time level (5.1) is solved by first computing a prediction

$$U^* = (I + kL_h)U^n + kG(U^n)$$

and then performing a fix-point iteration [20]

$$(I - (k/2)L_h)U_{[r+1]} = (I + (k/2)L_h)U^n + kG((U_{[r]} + U^n)/2), \quad r = 0, 1, \dots \quad (5.2)$$

Thus the only matrix to factorize is $I - (k/2)L_h$. This factorization was performed, prior to each time integration, by a Gaussian elimination subroutine which takes full advantage of the structure of L_h . Also, (5.2) was implemented in the efficient form [20]

$$(I - (k/2)L_h)U^{**} = U^n + (k/2)G((U_{[r]} + U^n)/2), \quad r = 0, 1, \dots$$

$$U_{[r+1]} = 2U^{**} - U^n$$

which avoids the computation of $(I + (k/2)L_h)U^n$.

The Crank-Nicolson split-step scheme (cf. [20]) is similar to the spectral method (4.1), the only difference being that now the linear operator L_k is replaced by the approximation of (2.5a) based on standard central differences in space along with a Crank-Nicolson time discretization. With this method, as with (5.2), there is only

one matrix to factorize per time integration. Furthermore, there are no nonlinear systems to be solved iteratively.

As a theoretical solution we employ the standing wave [1, 2]

$$\psi_A(x, t) = [M(x), iN(x)]^T e^{-iAt}, \tag{5.3}$$

$$M(x) = 2^{1/2}(1 - A^2)^{1/2} (1 + A)^{1/2} \frac{\text{ch}((1 - A^2)^{1/2} x)}{1 + A \text{ch}(2(1 - A^2)^{1/2} x)}$$

$$N(x) = 2^{1/2}(1 - A^2)^{1/2} (1 - A)^{1/2} \frac{\text{sh}((1 - A^2)^{1/2} x)}{1 + A \text{ch}(2(1 - A^2)^{1/2} x)}$$

with frequency $A = 0.75$. All schemes were implemented in $-16 \leq x \leq 16$, $0 \leq t \leq 8$ with periodic boundary conditions. While ψ_A is obviously nonperiodic, there is virtually no error in assuming periodicity, since ψ_A is exponentially small away from $x = 0$.

The results for the leap-frog scheme are given in Table I, where not all entries can be computed since, for stability, $k < h$. The numbers in brackets represent CPU times in hundredths of a second and the entries without brackets give the L^2 -errors. An $\mathcal{O}(k^2 + h^2)$ behaviour is clearly seen when refining with $k/h = \text{constant}$.

Table II corresponds to the Crank–Nicolson scheme. In the runs, the inner iteration (5.2) was stopped when two consecutive iterants were found whose difference was less than 10^{-5} in the L^2 -norm. The underlined quantities represent the number of inner iterations per step. The comparison of the Tables I and II reveals that for low accuracies the leap-frog scheme, whose coding is trivial, is slightly more efficient than the Crank–Nicolson, because the larger time steps that can be taken by the implicit scheme do not make up for the larger work per step required by the solution of the nonlinear equations (5.1). Unfortunately, the leap-frog scheme was observed to lead to nonlinear blow-up (cf. [16, 18, 19, 26]) in long run integrations such as those necessary in the study of wave interactions. For such problems the Crank–Nicolson scheme could provide a more reliable choice than the leap-frog

TABLE I

$k \backslash h$	0.2 ($J = 160$)	0.1 ($J = 320$)	0.05 ($J = 640$)
0.1	0.2342E - 01 (399)		
0.05	0.2120E - 01 (802)	0.5770E - 02 (1635)	
0.025	0.2073E - 01 (1618)	0.5224E - 02 (2262)	0.1442E - 02 (6588)
0.0125	0.2062E - 01 (3260)	0.5107E - 02 (6582)	0.1306E - 02 (13122)

TABLE II

$k \backslash h$	0.2 ($J = 160$)	0.1 ($J = 320$)	0.05 ($J = 640$)
0.4	0.7199E-02 (1327) <u>6</u>	0.1819E-01 (2560) <u>6</u>	0.2166E-01 (5146) <u>6</u>
0.2	0.1493E-01 (1766) <u>4</u>	0.1829E-02 (3629) <u>4</u>	0.4686E-02 (7052) <u>4</u>
0.1	0.1910E-01 (2813) <u>3</u>	0.3691E-02 (5496) <u>3</u>	0.4809E-03 (10990) <u>3</u>
0.05	0.2021E-01 (5501) <u>3</u>	0.4673E-02 (10925) <u>3</u>	0.9737E-03 (21947) <u>3</u>
0.025	0.2049E-01 <u>2</u>	0.4980E-02 <u>2</u>	0.1099E-02 <u>2</u>

algorithm. Alternatively, the leap-frog scheme could be supplemented by filtering or odd-even averaging in order to avoid the occurrence of the nonlinear blow-ups.

Table III gives the results for the split-step finite-difference method. A comparison with Table II shows that, for given k and h , the effect of the splitting is a decrease in CPU time and an increase in the size of the error. The latter indicates that, in the truncation error, the negative effect of the splitting does not compensate for the fact that now the nonlinear term is integrated exactly in closed form.

The results for the spectral split-step schemes are given in the Tables IV (simple splitting) and V (Strang's splitting). In both tables it is apparent that reducing h does not lead to more accuracy, a clear indication of the fact that the error stems,

TABLE III

$k \backslash h$	0.2 ($J = 160$)	0.1 ($J = 320$)	0.05 ($J = 640$)
0.4	0.7188E-01 (145)	0.5668E-01 (275)	0.5298E-01 (550)
0.2	0.3321E-01 (271)	0.1783E-01 (566)	0.1414E-01 (1066)
0.1	0.2368E-01 (517)	0.8239E-02 (1019)	0.4432E-02 (2165)
0.05	0.2136E-01 (1016)	0.5805E-02 (2061)	0.2106E-02 (4194)
0.025	0.2078E-01 (2026)	0.5264E-02 (4125)	0.1379E-02 (8334)

TABLE IV

$k \backslash h$	0.5 ($J = 64$)	0.25 ($J = 128$)	0.125 ($J = 256$)
0.5	0.2242E+00 (128)	0.2242E+00 (283)	0.2242E+00 (616)
0.25	0.1120E+00 (254)	0.1120E+00 (561)	0.1120E+00 (1209)
0.125	0.5614E-01 (501)	0.5615E-01 (1111)	0.5614E-01 (2424)
0.0625	0.2813E-01 (993)	0.2814E-01 (2192)	0.2814E-01 (4796)

almost exclusively, from the time-integration (recall that the smoothness of (5.3) leads to high spatial accuracy of the spectral technique). The $\mathcal{O}(k)$ behaviour of (2.11) and the $\mathcal{O}(k^2)$ behaviour of (5.1) are clearly borne out by the tables. No blow-up problems were encountered when using the split-step schemes, which revealed themselves to be very reliable.

In order to facilitate a comparison of the methods, we have summarized in Fig. 1 some of the information of the Tables I–III and V. The simple splitting spectral method is not considered, as it is clearly less competitive than (4.1). For the leap-frog scheme we have depicted the runs with $r = k/h = 0.5$, the most favourable mesh ratio of those considered in Table I. For the Crank–Nicolson and split-step Crank–Nicolson methods we show the runs with $r = k/h = 2$, the most favourable value of those considered in Tables II and III. For the spectral scheme (4.1) we display the runs with $h = 0.5$. Thus for each method we have chosen the combination of h and k that, for a given computational effort, results in smaller errors.

From the figure it is clear that the performances of the leap-frog and split-step Crank–Nicolson schemes are virtually the same. For both methods the error is

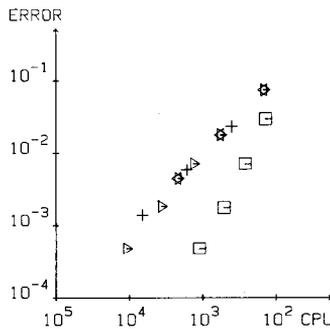


FIG. 1. Error against CPU time in hundredths of a second for the methods considered in the text: + leap-frog; ▷ Crank–Nicolson; ☆ split-step Crank–Nicolson; □ split-step spectral.

TABLE V

$k \backslash h$	0.5 ($J=64$)	0.25 ($J=128$)	0.125 ($J=256$)
0.5	0.2917E-01 (138)	0.2917E-01 (295)	0.2917E-01 (633)
0.25	0.7066E-02 (264)	0.7067E-02 (576)	0.7068E-02 (1214)
0.125	0.1767E-02 (510)	0.1770E-02 (1127)	0.1772E-02 (2410)
0.0625	0.4816E-03 (1108)	0.4874E-03 (2212)	0.4927E-03 (4761)

proportional to h^2 , i.e., to $(\text{CPU time})^{-1}$ and this is borne out by the figure where the corresponding symbols are located on a straight line with slope 1.

The slope of the Crank–Nicolson symbols is slightly larger than 1, due to the fact that for smaller values of k fewer inner iterations per step are needed. As a consequence, the Crank–Nicolson scheme should be preferred to the leap-frog and split-step finite-difference schemes except for very low accuracies.

However, none of the finite-difference schemes is competitive with the spectral method (4.1), for which the error is proportional to k^2 and thus to $(\text{CPU time})^{-2}$. For an accuracy of 10^{-3} , the split-step scheme (4.1) demands a computer time approximately *an order of magnitude lower* than that necessitated by the finite-difference algorithms. This advantage in efficiency would be even larger if more demanding accuracies were required. (However, note that (5.3) is very smooth and that, in the comparison above, this tends to benefit the spectral schemes more than the finite-difference algorithms.) We conclude that the scheme (4.1) is reliable and much more efficient than the finite-difference methods used for comparison.

ACKNOWLEDGMENTS

This work is part of the Project PB-86-0313 supported by “Fondo Nacional para el Desarrollo de la Investigación Científica y Técnica.” We are thankful to T. Ortega for his assistance.

REFERENCES

1. A. ALVAREZ AND B. CARRERAS, *Phys. Lett. A* **86**, 327 (1981).
2. A. ALVAREZ, P. KUO, AND L. VÁZQUEZ, *Appl. Math. Comput.* **13**, 1 (1983).
3. J. W. COOLEY AND T. W. TUKEY, *Math. Comput.* **19**, 297 (1965).
4. B. FORNBERG, *SIAM J. Numer. Anal.* **12**, 509 (1975).
5. J. DE FRUTOS, Ph.D. thesis, Universidad de Valladolid, 1987 (unpublished).

6. J. DE FRUTOS AND J. M. SANZ-SERNA, "h-dependent stability thresholds avoid the need for a priori bounds in nonlinear convergence proofs," in *Computacional Mathematics III, Proceedings of the Third International Conference on Numerical Analysis and its Applications, Benin City, Nigeria 1987*, edited by S. O. Fatunla (Boole, Dublin, in press).
7. D. GOTTLIEB, M. Y. HUSSAINI, AND S. A. ORSZAG, in *Spectral Methods for Partial Differential Equations*, edited by R. G. Voigt, D. Gottlieb and M. Y. Hussaini (SIAM, Philadelphia, 1984), p. 1.
8. R. H. HARDIN AND F. D. TAPPERT, *SIAM Rev. Chronicle* **15**, 423 (1973).
9. R. J. LEVEQUE AND J. OLIGER, *Math. Comput.* **40**, 469 (1983).
10. J. C. LÓPEZ-MARCOS, Ph.D. thesis, Universidad de Valladolid, 1985 (unpublished).
11. J. C. LÓPEZ-MARCOS AND J. M. SANZ-SERNA, *IMA J. Numer. Anal.* **8**, 71 (1988).
12. J. C. LÓPEZ-MARCOS AND J. M. SANZ-SERNA, A definition of stability for nonlinear problems, in *Numerical Treatment of Differential Equations, Proceedings of the Fourth Seminar Held in Halle*, (Teubner-Texte, Leipzig, 1988), p. 216.
13. J. E. PASCIAK, in *Spectral Methods for Partial Differential Equations*, edited by R. G. Voigt, D. Gottlieb, and M. Y. Hussaini (SIAM, Philadelphia, 1984), p. 217.
14. J. M. SANZ-SERNA, *Math. Comput.* **43**, 21 (1984).
15. J. M. SANZ-SERNA, in *Nonlinear Differential Equations and Applications*, edited by J. K. Hale and P. Martinez-Amores (Pitman, Boston, 1985), p. 64.
16. J. M. SANZ-SERNA, *SIAM J. Sci. Stat. Comput.* **6**, 923 (1985).
17. J. M. SANZ-SERNA AND C. PALENCIA, *Math. Comput.* **45**, 143 (1985).
18. J. M. SANZ-SERNA AND F. VADILLO, in *Pitman Research Notes in Mathematics Vol. 140*, edited by D. F. Griffiths and G. A. Watson (Longman Scientific & Technical, London, 1986), p. 187.
19. J. M. SANZ-SERNA AND F. VADILLO, *SIAM J. Appl. Math.* **47**, 92 (1987).
20. J. M. SANZ-SERNA AND J. G. VERWER, *IMA J. Numer. Anal.* **6**, 25 (1986).
21. Q. SHENG, Solving linear partial differential equations by exponential splitting, preprint, 1987 (unpublished).
22. G. STRANG, *SIAM J. Numer. Anal.* **5**, 506 (1968).
23. E. TADMOR, *SIAM J. Numer. Anal.* **23**, 1 (1986).
24. T. R. TAHA AND M. J. ABLOWITZ, *J. Comput. Phys.* **55**, 203 (1984).
25. F. D. TAPPERT, *Lect. Appl. Math. Amer. Math. Soc.* **15**, 215 (1974).
26. F. VADILLO AND J. M. SANZ-SERNA, *J. Comput. Phys.* **66**, 225 (1986).
27. J. A. C. WEIDEMAN AND B. M. HERBST, *SIAM J. Numer. Anal.* **23**, 485 (1986).
28. G. B. WHITHAM, *Linear and Nonlinear Waves* (Wiley-Interscience, New York, 1974).