Fourteenth International Conference on Domain Decomposition Methods
Editors: Ismael Herrera , David E. Keyes, Olof B. Widlund, Robert Yates

1. Nonlinearity, numerics and propagation of information

A. A. Aldama[1]

1. Introduction. In the study of evolution equations that describe the dynamics of natural and man-made systems, it is always useful to determine the way in which information is propagated by the said equations. In other words, the manner in which different scales present in the solution of an evolution equation travel and decay through space and time. The ideal tool to determine the propagation properties of (continuous or discrete) evolution equations is Fourier or harmonic analysis. In the case of continuous systems, the study of propagation properties allows the understanding of their stability. On the other hand, much insight regarding the behavior of discrete approximations of partial differential equations may be gained by comparing the propagation properties of a continuous equation and its corresponding discrete analogue. Thus, so-called amplitude and phase portraits that respectively depict the ratio of numerical and analytical amplification factor amplitudes and the difference between analytical and numerical phases, both as functions of wavenumber, may be developed (see, for example, Abbot [1] and Vichenevsky and Bowles [17]). These portraits show in a very objective way the effects of "numerical diffusion" and "numerical dispersion" associated to each wave number. Furthermore, the determination of the stability of numerical approximations may be viewed as a by-product of their amplitude propagation properties. Interestingly enough, a similar approach may be applied to study of the convergence properties of iterative schemes for the solution of systems of equations, a fact that has been exploited by the champions of the multigrid approach (see, for instance, [9]). The author and his collaborators have demonstrated the power of Fourier techniques in the study of the propagation properties of non-orthodox approximations of the linear transport equation, via least-squares collocation (Bentley et al., [10]) and the Eulerian-Lagrangian localized adjoint method (Aldama and Arroyo, [6]). Moreover, they have established the existence of an ordinary differential analogy that simplifies the determination of the stability conditions for high order time discretizations of the linear transport equation (Aldama, [3], and Aldama and Aparicio, [5]). Finally, they have studied the convergence properties of a semi-iterative scheme for the solution of a coupled diffusion-reaction system that describes the decay of argon in rocks and minerals (Lee and Aldama, [15]).

Unfortunately, the application of Fourier methods is limited to linear and constant coefficient equations, subject to periodic boundary conditions or to linear and constant coefficient pure initial value problems occurring in infinite spatial domains. The author has developed an approach that allows the use of Fourier techniques in finite spatial domains, variable coefficient or nonlinear problems. Such approach consists of an asymptotic approximation that is constructed by employing Taylor-Fréchet expansions of the differential operators arising in evolution equations, the method of multiple scales and local analysis. Numerical experiments have shown excellent results of the application of the said approach. This paper reviews the general theory on which the approach is based and presents a number of applications made by the author and his

[1] Mexican Institute of Water Technology, Mexican Academy of Engineering and School of Engineering and National Autonomous University of Mexico, aaldama@tlaloc.imta.mx

collaborators that have produced excellent results.

2. Nonlinear evolution problems. Let us consider the following nonlinear evolution problem for the components of the N-dimensional vector $\mathbf{U} = \mathbf{U}(\mathbf{x}, t) \equiv [U_1(\mathbf{x}, t), U_2(\mathbf{x}, t), ..., U_N(\mathbf{x}, t)]^T$, dependent on the three-dimensional position vector $\mathbf{x}$ and time t:

$$\frac{\partial U_i}{\partial t} - N_i\left(U_j\right) \ = \ 0, \ \ \mathbf{x} \in \Omega, \ \ t > 0; \ \ i = 1, 2, ..., N \tag{2.1}$$

$$B_k(U_j) \ = \ 0, \ \ \mathbf{x} \in \partial\Omega, \ \ t > 0; \ \ k = 1, 2, ..., M \tag{2.2}$$

$$U_i(\mathbf{x}, 0) \ = F_j(\mathbf{x}), \ \ \mathbf{x} \in \Omega; \ \ i = 1, 2, ..., N \tag{2.3}$$

where (2.1) represents a set of N evolution equations, involving a like number of *spatial* differential operators, $N_i(\cdot)$, acting upon the components of $\mathbf{U}$; Ω is the spatial domain of interest and $\partial\Omega$ its boundary; equation (2.2) represents a set of M boundary conditions involving a like number of differential operators, $B_k(\cdot)$; equation (2.3) represents a set of N initial conditions, where $F_j(\mathbf{x})$ stands for a like number of prescribed functions. The number M is determined by the order of the operators $N_i(\cdot)$ and by the number N.

Examples of evolution equations of the kind represented by equation (2.1) abound. Take, for example, the celebrated Navier-Stokes equations for incompressible flow:

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} \tag{2.4}$$

where u_i $(i = 1, 2, 3)$ are the components of the velocity vector, p is the dynamic pressure, ρ is the density, ν is the kinematic viscosity, t is time, and x_i $(i = 1, 2, 3)$ are the components of the position vector; or the shallow water equations:

$$\begin{array}{l} \frac{\partial h}{\partial t} + \frac{\partial U h}{\partial x} + \frac{\partial V h}{\partial y} = 0 \\ \frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} + V \frac{\partial U}{\partial y} - fV = -g \frac{\partial (z_b + h)}{\partial x} + \frac{1}{\rho h} \tau_{bx}(h, U, V) \\ \frac{\partial V}{\partial t} + U \frac{\partial V}{\partial x} + V \frac{\partial V}{\partial y} + fU = -g \frac{\partial (z_b + h)}{\partial y} + \frac{1}{\rho h} \tau_{by}(h, U, V) \end{array} \tag{2.5}$$

where U and V are the components of the velocity vector, h is the depth, ρ is the density, z_b is the bottom elevation, z_{bx} and z_{by} are the x and y components of the bottom shear stress, t is time, and x and y are the components of the position vector; or Richards equation:

$$S(\psi) \frac{\partial \psi}{\partial t} = \frac{\partial}{\partial x_j} \left[K(\psi) \frac{\partial}{\partial x_j} (\psi + z) \right] \tag{2.6}$$

where ψ is the pressure head, S is the specific moisture capacity, K is the unsaturated hydraulic conductivity, z is the vertical coordinate, t is time, and x_i $(i = 1, 2, 3)$ are the components of the position vector; or the two-species advection diffusion reaction system:

$$\begin{array}{l} \frac{\partial C_1}{\partial t} + V \frac{\partial C_1}{\partial x} = D \frac{\partial^2 C_1}{\partial x^2} - K_1(C_1) C_1 + f_1(C_2) \\ \frac{\partial C_2}{\partial t} + V \frac{\partial C_2}{\partial x} = D \frac{\partial^2 C_2}{\partial x^2} - K_2(C_2) C_2 + f_2(C_1) \end{array} \tag{2.7}$$

where C_1 and C_2 are the concentrations of species 1 and 2, V is the advective velocity, D is the diffusion/dispersion coefficient, $K_1(\cdot)$ and $K_2(\cdot)$ are nonlinear decay functions, $f_1(\cdot)$ and $f_2(\cdot)$are nonlinear source/sink functions, t is time, and x is the spatial coordinate.

Evidently, the problem (2.1)-(2.3) is continuous in space and time. Discrete analogues of such a problem may be developed through numerical approximations of the differential operators.

3. Taylor-Fréchet expansions of nonlinear operators. Let us decompose the dependent variable appearing in equation (2.1), U_i, as follows:

$$U_i = \bar{U}_i + u_i \tag{3.1}$$

where $\bar{U}_i$ represents a reference solution of problem (2.1)-(2.3) and u_i is a small perturbation around it, such that

$$\|u_i\| << \left\|\bar{U}_i\right\| \tag{3.2}$$

where $\|\cdot\|$ is a properly defined norm. The assumed nature of $\bar{U}_i$ implies that

$$\frac{\partial \bar{U}_i}{\partial t} - N_i\left(\bar{U}_j\right) \;=\; 0 \tag{3.3}$$

Substituting (3.1) in (2.1) yields:

$$\frac{\partial \bar{U}_i}{\partial t} + \frac{\partial u_i}{\partial t} - N_i\left(\bar{U}_j + u_j\right) \;=\; 0 \tag{3.4}$$

Employing a Taylor-Fréchet expansion (Milne, [16]) of the nonlinear operator that appears as the last term on the left hand side of the last equation results in:

$$N_i(\bar{U}_j + u_j) = N_j(\bar{U}_j) + \partial_{U_k} N_i(\bar{U}_j) \circ u_k + O(\,\|u_k u_k\|\,) \tag{3.5}$$

where $\partial_{U_k} N_i(\bar{U}_j) \circ (\cdot)$ stands for the first partial Fréchet derivative of the nonlinear differential operator $N_i(\cdot)$, which possesses a nonlinear parametric dependence on the reference solution and acts upon the perturbation u_k. It may be shown that first order Fréchet derivatives of nonlinear differential operators are themselves linear differential operators (Milne, [16]). Substituting (3.5) in (3.4) and accounting for (3.3) yields:

$$\frac{\partial u_i}{\partial t} - \partial_{U_k} N_i(\bar{U}_j) \circ u_k + O(\,\|u_k u_k\|\,) = 0 \tag{3.6}$$

As may be observed, to first order in u_i, equation (3.6) (for i=1,2,. . . ,N) is linear, a fact that will be exploited later on.

4. Multiple scale analysis. Let $(\mathbf{x}_0, t_0)$ be a fixed reference point in space and time, with x_{0i} representing the components of $\mathbf{x}_0$. Thus, let us define "slow" space and time variables as follows:

$$\begin{array}{l} \mathrm{X}_i = \frac{x_i - x_{io}}{L} \\ \mathrm{T} = \frac{t - t_o}{T} \end{array} \tag{4.1}$$

where L and T respectively represent characteristic "large" length and time scales present in U_i. Similarly, let us define "fast" space an time variables as follows:

$$\begin{array}{l} \chi_i = \frac{x_i - x_{io}}{\Lambda_x} \\ \tau = \frac{t - t_o}{\Lambda_t} \end{array} \tag{4.2}$$

where Λ_x and Λ_t respectively represent characteristic "small" length and time scales present in U_i. We will now assume that the following holds true:

$$\varepsilon = \frac{\Lambda_x}{L} = \frac{\Lambda_t}{T} << \; 1 \tag{4.3}$$

Now we are in position of introducing the *separation of scales hypothesis*:

$$\bar{U}_j = \bar{U}_j \left(\mathrm{X}_i, \mathrm{T} \right) \tag{4.4}$$

$$u_j = u_j \left(\chi_i, \tau \right) \tag{4.5}$$

Equations (4.4) and (4.5) express the assumption that the reference solution only depends on the slow variables, whereas the perturbation only depends on the fast variables. Hence the large and small length and time scales take on a more precise meaning. Indeed, L and T respectively represent the length and time scales characteristic of the reference solution, $\bar{U}_i$, and Λ_x and Λ_t respectively represent the length and time scales characteristic of the perturbation, u_i. As will be shown later, the separation of scales hypothesis (4.4)-(4.5) has proven to be valid in a number of specific cases. The reason for this is that it is very often the case that when stability or nonlinear iteration convergence are of interest, it is often the case that the most unstable or the most resistant modes correspond to small scale (i.e., high wavenumber) components of the solution, which through (4.4)-(4.5) may be isolated from a smoothly varying reference solution.

5. Localization. Now let us expand the reference solution, $\bar{U}_i(\mathbf{x}, t)$, and the perturbation, $u_i(\mathbf{x}, t)$, around the reference point $(\mathbf{x}_0, t_0)$, assuming the space and time displacements are of the same order of magnitude as Λ_x and Λ_t:

$$\begin{array}{l} \bar{U}_i(\mathbf{x}, t) = \bar{U}_i(\mathbf{x}_0, t_0) + (x_j - x_{0j}) \left. \frac{\partial \bar{U}_i}{\partial x_j} \right|_{(\mathbf{x_0}, t_0)} + (t - t_0) \left. \frac{\partial \bar{U}_i}{\partial t} \right|_{(\mathbf{x_0}, t_0)} + ... \\ u_i(\mathbf{x}, t) = u_i(\mathbf{x}_0, t_0) + (x_j - x_{0j}) \left. \frac{\partial u_i}{\partial x_j} \right|_{(\mathbf{x_0}, t_0)} + (t - t_0) \left. \frac{\partial u_i}{\partial t} \right|_{(\mathbf{x_0}, t_0)} + ... \end{array} \tag{5.1}$$

where

$$\frac{|\mathbf{x} - \mathbf{x}_0|}{\Lambda_x} = O(1) \tag{5.2}$$

$$\frac{|t - t_0|}{\Lambda_t} = O(1) \tag{5.3}$$

We may now introduce characteristic scales for the magnitudes of $\bar{U}_i$ and u_i:

$$\bar{U}_i = U \bar{U}_i^* \tag{5.4}$$

$$u_i = u u_i^* \tag{5.5}$$

where, say:

$$U = \left\|\bar{U}_i(\mathbf{x}_0, t_0)\right\| \tag{5.6}$$

$$u = \|u_i(\mathbf{x}_0, t_0)\| \tag{5.7}$$

and

$$\bar{U}_i^* = O(1) \tag{5.8}$$

$$u_i^* = O(1) \tag{5.9}$$

and $\bar{U}_i^* = O(1)$, $u_i^* = O(1)$. On account of equation (3.2), we may further assume that:

$$u = \varepsilon U \tag{5.10}$$

Now, from the separation of scales hypothesis (4.4)-(4.5), we get:

$$\begin{aligned} \frac{\partial \bar{U}_i}{\partial x_j} &= \frac{\partial \bar{U}_i}{\partial \mathrm{X}_k}\frac{\partial \mathrm{X}_k}{\partial x_j} = \frac{U}{L}\delta_{jk}\frac{\partial \bar{U}_i^*}{\partial \mathrm{X}_k} = \frac{U}{L}\frac{\partial \bar{U}_i^*}{\partial \mathrm{X}_j} \\ \frac{\partial \bar{U}_i}{\partial t} &= \frac{\partial \bar{U}_i}{\partial \mathrm{T}}\frac{\partial \mathrm{T}}{\partial t} = \frac{U}{T}\frac{\partial \bar{U}_j^*}{\partial \mathrm{T}} \end{aligned} \tag{5.11}$$

$$\begin{aligned} \frac{\partial u_i}{\partial x_j} &= \frac{\partial u_i}{\partial \chi_k}\frac{\partial \chi_k}{\partial x_j} = \frac{u}{\Lambda_x}\delta_{jk}\frac{\partial u_i^*}{\partial \chi_k} = \frac{u}{\Lambda_x}\frac{\partial u_i^*}{\partial \chi_j} \\ \frac{\partial u_i}{\partial t} &= \frac{\partial u_i}{\partial \tau}\frac{\partial \tau}{\partial t} = \frac{u}{\Lambda_t}\frac{\partial u_i^*}{\partial \tau} \end{aligned} \tag{5.12}$$

where (4.1), (4.2), (5.4) and (5.5) have been used. Employing now (5.2), (5.3), (5.8), (5.9), (5.11), and (5.12) in (5.1) it is readily shown that:

$$\bar{U}_i(\mathbf{x}, t) = \bar{U}_i(\mathbf{x}_0, t_0)\left[1 + O(\varepsilon)\right] \equiv U\left[1 + O(\varepsilon)\right] \tag{5.13}$$

$$u_i(\mathbf{x}, t) = u_i(\mathbf{x}_0, t_0)\left[1 + O(1)\right] \tag{5.14}$$

Equation (5.13) shows that whereas the reference solution, $\bar{U}_i$, may be localized in the neighborhood of the reference point $(\mathbf{x}_0, t_0)$ at space and time displacements commensurate with the small scales Λ_x and Λ_t, the perturbation, u_i, may not. In other words, an observer sensitive to the scales Λ_x and Λ_t, would only perceive the variations in the perturbation, and would view the reference solution as a constant.

6. Asymptotics. We now may seek an asymptotic solution to equation (3.6), of the form:

$$u_i = u_i^{(0)} + \varepsilon u_i^{(1)} + \varepsilon^2 u_i^{(2)} + ... \equiv \varepsilon U \left[u_i^{(0)*} + \varepsilon u_i^{(1)*} + \varepsilon^2 u_i^{(2)*} + ... \right] \qquad (6.1)$$

where $u_i^{(k)*}$ (k=0,1,2,. . .) are dimensionless and of $O(1)$, and (5.5) and (5.10) have been accounted for. Substituting (5.13) and (6.1) in (3.6) we get the following evolution system for the zeroth order approximation $u_i^{(0)}$ (i=1,2,. . . ,N):

$$\frac{\partial u_i^{(0)}}{\partial t} - \partial_{U_k} N_i(U_j) \circ u_i^{(0)} = 0 \ ; \ \ j = 1, 2, ..., N \qquad (6.2)$$

It must be noted that equation (6.2) is linear and with *constant coefficients* that parametrically depend (alas, nonlinearly) on the constants U_j (j=1,2,. . . ,N). Thus, equation (6.2) captures the dominant nonlinear behavior of equation (2.1) in the scales of Λ_x and Λ_t. Furthermore, the previously presented localization analysis was based on the assumption that:

$$|\chi_i| = |x_i - x_{io}| \, / \Lambda_x = |x_i - x_{io}| \, / \, (\, \varepsilon \, L) = O(1) \qquad (6.3)$$

Therefore, as $\varepsilon \downarrow 0$, the domain corresponding to the zeroth order approximation $u_i^{(0)}$ (i=1,2,. . . ,N) becomes unbounded.

7. Fourier analysis. In view of the above, the most general form of equation (6.2) may be written as follows in three-dimensional space:

$$\partial_t u_j^{(0)} = \sum_{r=1}^{N} \sum_{\mathbf{p} \in P} \alpha_{jr,\mathbf{p}} \partial_{\mathbf{p}} u_r^{(0)} \ ; \ j = 1, 2, ..., N; \ \text{ in } \mathbf{x} \in \Omega_\infty \qquad (7.1)$$

where $\Omega_\infty \equiv (-\infty, \infty)^3$; $\partial_t \equiv \partial / \partial t$; $\mathbf{p} \equiv (p_1, p_2, p_3)$ represents a multi-index; $P \equiv \{(p_1, p_2, p_3 \, | 0 \leq p_1 + p_2 + p_3 \leq R\}$, where R is the maximum order of the spatial derivatives present in (7.1); $\alpha_{jr,\mathbf{p}}$ are constant coefficients; $\partial_{\mathbf{p}}(\cdot) \equiv \frac{\partial^{p_1+p_2+p_3}(\cdot)}{\partial x_1^{p_1} \partial x_2^{p_2} \partial x_3^{p_3}}$, and the summation convention is understood in $\mathbf{p}$.

Now, assuming the functions prescribed in the initial conditions (2.3) are of the form

$$F_j = \bar{F}_j + f_j, \ \ f_j / \bar{F}_j = O(\varepsilon) \ ; \ \ j = 1, 2, ..., N \qquad (7.2)$$

it is consistent to write that the initial conditions that equation (7.1) is subject to, are:

$$u_j^{(0)} = f_j; \ \ j = 1, 2, ..., N \qquad (7.3)$$

Equations (7.1) and (7.3) constitute a *pure initial value problem*, that may be tackled via Fourier methods. With that purpose in mind, the following Fourier representation may be used (Champeney, [12]):

$$u_j^{(0)}(\mathbf{x}, t) = \frac{1}{(2\pi)^{3/2}} \int\limits_{\Omega_\infty} \hat{u}_j^{(0)}(\mathbf{k}, t) \exp(-\mathrm{i} \, \mathbf{k} \cdot \mathbf{x}) d\mathbf{k} \qquad (7.4)$$

where $\mathrm{i} \equiv \sqrt{-1}$, $\mathbf{k} \equiv (k_1, k_2, k_3)$ is the wavenumber vector, $d\mathbf{k} \equiv dk_1 dk_2 dk_3$, and the Fourier coefficients $\hat{u}_j^{(0)}$ are given by the following Fourier transforms:

$$\hat{u}_j^{(0)}(\mathbf{k}, t) \equiv \Im \left\{ u_j^{(0)}(\mathbf{x}, t) \right\} = \frac{1}{(2\pi)^{3/2}} \int\limits_{\Omega_\infty} u_j^{(0)}(\mathbf{x}, t) \exp(\mathrm{i}\, \mathbf{k} \cdot \mathbf{x}) d\mathbf{x} \tag{7.5}$$

Now, it may be shown that the Fourier coefficients $\hat{u}_j^{(0)}$ may be determined by employing the initial conditions (7.3). Nevertheless, when the propagation properties of the equation (7.1) and, in particular, its stability are of interest, the initial values of $u_j^{(0)}$ are inconsequential. In effect, the stability of equation (7.1) is determined by finding whether $\hat{u}_j^{(0)}(\mathbf{k}, t)$ grows or decays in time.

8. Discrete systems. An analysis similar to that presented earlier may be performed for discrete systems, that may correspond to numerical approximations of partial differential evolution equations, such as equation (2.1). In such a case, the only additional aspect of the analysis that must be considered is the determination of the *modified partial differential equations* that are satisfied when the discrete equations in terms of the perturbation quantities are solved. This consideration allows a local analysis such as the one presented for the continuous case. In addition, instead of using a continuous Fourier pair, like (7.4)-(7.5), a semidiscrete one must be used (i.e., an integral representation for the physical space variables and a Fourier series representation for the wavenumber space variables). Examples of the use of such a technique follow.

9. The one-dimensional Richards equation. Let us consider the one - dimensional analogue of equation (2.6):

$$S(\psi) \frac{\partial \psi}{\partial t} = \frac{\partial}{\partial z} \left[K(\psi) \frac{\partial(\psi)}{\partial z} \right] + \frac{\partial K(\psi)}{\partial z} \tag{9.1}$$

The θ-central difference or θ-lumped finite element (with constant element size) approximation of equation (9.1) is:

$$\begin{aligned} F(\psi_j^n) \equiv \overline{\theta}(S_j) \tfrac{\delta^{n+\frac{1}{2}} \psi_j}{\Delta t} - \overline{\theta} \left\{ \tfrac{K_{j+\frac{1}{2}} \delta_{j+\frac{1}{2}} \psi - K_{j-\frac{1}{2}} \delta_{j-\frac{1}{2}} \psi}{\Delta z^2} + \tfrac{(\delta_{j+\frac{1}{2}} - \delta_{j-\frac{1}{2}}) K}{2\Delta z} \right\} \\ = 0 \end{aligned} \tag{9.2}$$

where $\bar{\theta}(\phi) = \theta(\phi^{n+1}) + (1-\theta)(\phi^n)$, $\delta^{n+\frac{1}{2}}\phi = \phi^{n+1} - \phi^n$, $\delta_{j+\frac{1}{2}}\phi = \phi_{j+1} - \phi_j$ and the usual notation for discrete approximations in space and time is employed.

Now, since Richards' equation is a nonlinear diffusion (i.e., parabolic) equation, a simple frozen coefficient analysis yields unconditional stability for the Crank-Nicolson scheme ($\theta = 1/2$). This result is contradicted by computational evidence, which shows that the said scheme often becomes unstable. This led the author to believe that the explanation for the emergence of instabilities should lie on nonlinear effects. Thus, it is apparent that the theory presented herein may be of use.

The solution of equation (9.2) may be decomposed as follows:

$$\psi_j^n = \tilde{\psi}_j^n + \varepsilon_j^n \tag{9.3}$$

where $\tilde{\psi}_j^n$ is the exact solution of equation (9.2) and ε_j^n a roundoff error. Substituting (9.3) in (9.2), employing a Taylor-Fréchet expansion and localizing the result yields the following equation for the roundoff error:

$$\begin{aligned}
&S(\tilde{\psi}_0)\frac{\delta^{n+\frac{1}{2}}\varepsilon_j}{\Delta t} - K'(\tilde{\psi}_0)\left[2\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0 + 1\right]\overline{\theta}\left[\frac{(\delta_{j+\frac{1}{2}}+\delta_{j+\frac{1}{2}})\varepsilon}{2\Delta z}\right] = K(\tilde{\psi}_0)\times \\
&\times\overline{\theta}\left[\frac{(\delta_{j+\frac{1}{2}}-\delta_{j-\frac{1}{2}})\varepsilon}{\Delta z^2}\right] + K''(\tilde{\psi}_0)\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0\left[2\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0 + 1\right]\overline{\theta}\left[\frac{\varepsilon_{j+1}^{n+1}+\varepsilon_{j-1}^{n+1}}{2}\right] + \\
&+K'(\tilde{\psi}_0)\left(\frac{\partial^2\tilde{\psi}}{\partial z^2}\right)_0\overline{\theta}\left[\frac{\varepsilon_{j+\frac{1}{2}}+\varepsilon_{j-\frac{1}{2}}}{2}\right] - S'(\tilde{\psi}_0)\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0\overline{\theta}(\varepsilon_j)
\end{aligned} \tag{9.4}$$

Since equation (9.4) is linear and with constant coefficients, without the loss of generality, the behavior of a single (but arbitrary) Fourier mode may be studied. Thus let us employ the following Fourier representation:

$$\varepsilon_j^n = E_k\xi_k^n\exp(\mathrm{i}j\beta_k) \tag{9.5}$$

where E_k is the amplitude associated with the wavenumber k, ξ_k is the corresponding amplification factor and $\beta_k \equiv k\Delta x$ is a dimensionless wavenumber. Substituting (9.5) in (9.4) results in:

$$\xi_k = \frac{1+(1-\theta)\mu_k}{1-\theta\mu_k} \tag{9.6}$$

where $\mu_k = (\mu_k)_R + \mathrm{i}(\mu_k)_R$ and

$$\begin{aligned}
&(\mu_k)_R = \left\{\frac{K''(\tilde{\psi}_0)}{S(\tilde{\psi}_0)}\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0\left[\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0 + 1\right]\cos\beta_k + \frac{1}{2}\frac{K'(\tilde{\psi}_0)}{S(\tilde{\psi}_0)}\left(\frac{\partial^2\tilde{\psi}}{\partial z^2}\right)_0\times\right. \\
&\left.\times(1+\cos\beta_k) - \frac{S'(\tilde{\psi}_0)}{S(\tilde{\psi}_0)}\left(\frac{\partial\tilde{\psi}}{\partial z}\right)_0 - \frac{2}{\Delta z^2}\frac{K(\tilde{\psi}_0)}{S(\tilde{\psi}_0)}(1-\cos\beta_k)\right\}\Delta t \\
&(\mu_k)_I = \frac{K'(\tilde{\psi}_0)}{S(\tilde{\psi}_0)}\left[2\left(\frac{\partial\tilde{\psi}_0}{\partial z}\right)_0 + 1\right]\sin\beta_k\frac{\Delta t}{\Delta z}
\end{aligned} \tag{9.7}$$

The stability condition for Crank-Nicolson scheme $\theta = 1/2$is $(\mu_k)_R \leq 0,\ \forall k$. Aldama and Aparicio ([5]) have shown that this condition is often violated in the numerical solution of Richards' equation. This explains the computational evidence that indicates that the Crank-Nicolson scheme becomes unstable in the solution of Richards' equation.

Since Richards equation (9.1) is nonlinear, its discrete analogue (9.2) generates an algebraic system of equations that is nonlinear as well. Thus, equation (9.2) must be solved in practice via an iterative scheme. The Picard or successive approximation iterative scheme for equation (9.2) may be written as follows:

$$\begin{array}{l}\left[\theta\, S_j^{n+1,m} + (1-\theta)\, S_j^n\right] \frac{\psi_j^{n+1,m+1} - \psi_j^n}{\Delta t} \\ -\theta \left\{ \frac{1}{2\Delta z^2} \left[\left(K_{j+1}^{n+1,m} + K_j^{n+1,m} \right) \left(\psi_{j+1}^{n+1,m+1} - \psi_j^{n+1,m+1} \right) \right.\right. \\ \left. - \left(K_j^{n+1,m} + K_{j-1}^{n+1,m} \right) \left(\psi_j^{n+1,m+1} - \psi_{j-1}^{n+1,m+1} \right) \right] \\ \left. + \frac{K_{j+1}^{n+1,m} - K_{j-1}^{n+1,m}}{2\Delta z} \right\} - (1-\theta) \left\{ \frac{1}{2\Delta z^2} \left[\left(K_{j+1}^n + K_j^n \right) \left(\psi_{j+1}^n - \psi_j^n \right) \right.\right. \\ \left.\left. - \left(K_j^n + K_{j-1}^n \right) \left(\psi_j^n - \psi_{j-1}^n \right) \right] + \frac{K_{j+1}^n - K_{j-1}^n}{2\Delta z} \right\} = 0 \end{array} \tag{9.8}$$

where the superindex m refers to iteration number. Now, a frozen coefficients analysis predicts unconditional convergence for scheme (9.8). This is not consistent with the observations of Huyarkon et al ([13]) and Celia et al ([11]), who have reported that the Picard scheme (9.8) sometimes diverges. In particular, it has been observed that it converges for small values of the time step, Δt, diverges for intermediate values and converges again for large values. This behavior would not be expected were the equation under study a linear one and, thus, may be attributed to nonlinearity.

In order to properly characterize the behavior of the Picard scheme applied to the solution of the discrete Richards equation, the theory presented in this paper may be applied. With that purpose in mind, let us express the $(m+1)$th iterate in equation (9.8) as follows:

$$\psi_j^{n+1,m+1} = \tilde{\psi}_j^{n+1} + \delta_j^{m+1} \tag{9.9}$$

where, as before, $\tilde{\psi}_j^{n+1}$ represents the exact solution of equation (9.2) and δ_j^{m+1}, the error corresponding to iteration $m+1$. Substituting (9.9) in equation (9.8), performing a Taylor-Fréchet expansion and localizing the result yields:

$$\begin{array}{l} S(\psi_0) \frac{\delta_j^{n+1}}{\Delta t} - \theta K'(\psi_0) \left(\frac{\partial \psi}{\partial z} \right)_0 \left(\frac{\delta_{j+1}^{m+1} - \delta_{j-1}^{m+1}}{2\Delta z} + \frac{\delta_{j+1}^m - \delta_{j-1}^m}{2\Delta z} \right) \\ -\theta K'(\psi_0) \frac{\delta_{j+1}^m - \delta_{j-1}^m}{2\Delta z} \theta K(\psi_0) \frac{\delta_{j+1}^{m+1} - \delta_j^{m+1} + \delta_{j-1}^{m+1}}{\Delta z^2} \\ +K''(\psi_0) \left(\frac{\partial \psi}{\partial z} \right)_0 \left[\left(\frac{\partial \psi}{\partial z} \right)_0 + 1 \right] + \frac{\delta_{j+1}^m + \delta_{j-1}^m}{2} \\ +K'(\psi_0) \left(\frac{\partial^2 \psi}{\partial z^2} \right)_0 \frac{\delta_{j+1}^m + 2\delta_j^m + \delta_{j-1}^m}{4} - +\theta S'(\psi_0) \left(\frac{\partial \psi}{\partial z} \right)_0 \left(\delta_j^{m+1} - \delta_j^m \right) \end{array} \tag{9.10}$$

Let us now study the behavior of a single (but arbitrary) Fourier mode in the solution of equation (9.10), by employing the following representation for the iteration error:

$$\delta_j^m = \Delta_k \xi_k^m \exp(\mathrm{i} j \beta_k) \tag{9.11}$$

where Δ_k is the amplitude associated with the wavenumber k, ξ_k is the corresponding amplification factor and $\beta_k \equiv k\Delta x$ is a dimensionless wavenumber. Substituting (9.5) in (9.4) results in:

$$\xi_k = \frac{\mu_{2,k}}{1 + \mu_{1,k}} \tag{9.12}$$

where $\mu_{1,k} = \mu_{1R,k} + \mathrm{i}\mu_{1I,k}$, $\mu_{2,k} = \mu_{2R,k} + \mathrm{i}\mu_{2I,k}$ and:

$$\begin{aligned}
&\mu_{1R,k} = 2\theta \tfrac{K(\psi_0)}{S(\psi_0)} \left(1 - \cos\beta_k\right) \tfrac{\Delta t}{\Delta z^2} + \theta \tfrac{S'(\psi_0)}{S(\psi_0)} \left(\tfrac{\partial\psi}{\partial t}\right)_0 \Delta t \\
&\mu_{1I,k} = -\theta \tfrac{K'(\psi_0)}{S(\psi_0)} \left(\tfrac{\partial\psi}{\partial z}\right)_0 \tfrac{\Delta t}{\Delta z^2} \sin\beta_k \Delta t \\
&\mu_{2R,k} = \theta \tfrac{K''(\psi_0)}{S(\psi_0)} \left(\tfrac{\partial\psi}{\partial z}\right)_0 \left[\left(\tfrac{\partial\psi}{\partial z}\right)_0 + 1\right] \cos\beta_k \Delta t + \tfrac{1}{2}\theta \tfrac{K'(\psi_0)}{S(\psi_0)} \\
&\times \left(+1 \cos\beta_k\right) \Delta t \left(\tfrac{\partial^2\psi}{\partial z^2}\right)_0 - \theta \tfrac{S'(\psi_0)}{S(\psi_0)} \left(\tfrac{\partial\psi}{\partial t}\right)_0 \Delta t \\
&\mu_{2I,k} = -\mu_{1I,k} \left[1 + \left(\tfrac{\partial\psi}{\partial z}\right)_0^{-1}\right]
\end{aligned} \tag{9.13}$$

The convergence condition for the Picard iterative scheme may be written as follows:

$$|\, \xi_k| \ < 1 \ \ \forall\, k \tag{9.14}$$

It may be shown that the above inequality leads to a quadratic inequality in Δt, which explains the observation that Picard iterations are sometimes convergent for "small" values of Δt, divergent for "intermediate" values, and convergent again for "large" values. Numerical experiments performed by Aldama and Paniconi ([8]) have validated such theoretical considerations.

10. The Saint-Venant equations. Another nonlinear evolution system that commonly arises in applications is the one constituted by the Saint-Venant equations that govern nonuniform, transient open channel flow:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \tag{10.1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + gA\frac{\partial h}{\partial x} + gA\frac{\partial z}{\partial x} + gS_f = 0 \tag{10.2}$$

where equation (10.1) expresses the conservation of mass principle and equation (10.2), the momentum principle. There, A represents the hydraulic area; Q, the discharge; h, the depth; z, the bottom elevation; S_f, the frictional slope; g, the acceleration of gravity; x, the spatial coordinate along the channel, and t, time. When Manning's formula is employed, the frictional slope may be expressed as follows:

$$S_f = \alpha \left(\frac{k_s}{R}\right)^{1/3} \frac{Q\,|Q|}{A^2 R} \tag{10.3}$$

where $\alpha \cong 17/100$ (Aldama and Ocón, [7]); k_s is Nikuradse's equivalent roughness and R is the hydraulic radius.

The so-called generalized Preismann scheme ([1]) for the numerical solution of the Saint-Venant system (10.1)-(10.2) may be written as follows:

$$\frac{A_{j+1}^{n+1} - A_{j+1}^{n}}{\Delta t} + (1-\theta)\,\frac{Q_{j+1}^{n} - Q_{j}^{n}}{\Delta x} + \theta\frac{Q_{j+1}^{n+1} - Q_{j}^{n+1}}{\Delta x} = 0 \tag{10.4}$$

$$\begin{array}{l}
(1-\psi)\,\frac{Q_j^{n+1}-Q_j^n}{\Delta t}+\psi\frac{Q_{j+1}^{n+1}-Q_{j+1}^n}{\Delta t}+(1-\theta)\,\frac{\left(\frac{Q^2}{A}\right)_{j+1}^n-\left(\frac{Q^2}{A}\right)_j^n}{\Delta x}+\theta\frac{\left(\frac{Q^2}{A}\right)_{j+1}^{n+1}-\left(\frac{Q^2}{A}\right)_j^{n+1}}{\Delta x}+\\
\quad +g\left\{(1-\theta)\,\left[(1-\psi)\,A_j^n+\psi A_{j+1}^n\right]+\theta\left[(1-\psi)\,A_j^{n+1}+\psi A_{j+1}^{n+1}\right]\,\right\}\\
\left[(1-\theta)\,\frac{h_{j+1}^n-h_j^n}{\Delta x}+\theta\frac{h_{j+1}^{n+1}-h_j^{n+1}}{\Delta x}+\qquad\frac{z_{j+1}-z_j}{\Delta x}\right]+(1-\theta)\,\left[(1-\psi)\,A_j^nS_{fj}^n+\psi A_{j+1}^nS_{fj+1}^n\right]\\
+\theta\left[(1-\psi)\,A_j^{n+1}S_{fj}^{n+1}+\psi A_{j+1}^{n+1}S_{fj+1}^{n+1}\right]=0
\end{array} \tag{10.5}$$

where $\psi \in [0\,,\,1]$ is a space weighting factor and $\theta \in [0\,,\,1]$ is a time weighting factor.

By applying the theory presented herein, it may be shown that the stability conditions for the generalized Preismann scheme (10.4)-(10.5) are:

$$|\,V_e\,|\ \le\ 1,\quad \psi\ =\ 0.5,\quad \theta\ \ge\ 0.5 \tag{10.6}$$

where V_e is the Vedernikov number. The validity of the conditions (10.6) has been assessed via numerical experimentation (Aguilar, [2]).

11. The shallow water equations. The one-dimensional version of the shallow water equations may be written as follows:

$$\begin{array}{l}
M_a(h,U)\equiv\frac{\partial h}{\partial t}+\frac{\partial Uh}{\partial x}=0\\
M_0(h,U)\equiv\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial x}+g\frac{\partial(z_b+h)}{\partial x}+gS_f=0
\end{array} \tag{11.1}$$

where $M_a(\cdot,\cdot)$ is the mass conservation operator and $M_o(\cdot,\cdot)$ is the momentum operator. The Generalized Wave Continuity Equation (GWCE) formulation was introduced in order to eliminate the spurious oscillations that arise in the numerical solution of the shallow water equations, in their primitive formulation (11.1), when collocated grids are used (see, for example Kinmark, [14]). The GWCE formulation introduces the following equation, which is derived from (11.1):

$$W(h,U)\equiv\frac{\partial M_a(h,U)}{\partial t}-\frac{\partial M_o(h,U)}{\partial x}+GM_a(h,U)=0 \tag{11.2}$$

where $W(\cdot,\cdot)$ is the so-called GWCE operator. The GWCE formulation consists of solving the coupled equations and $M_o(h,U)=0$. As is apparent, when $G\to\infty$, the GWCE formulation approaches the primitive formulation, and when $G\to 0$, the equation $W(h,U)=0$ approaches a nonlinear wave equation.

A number of investigators have become concerned with the fact that, apparently, the GWCE formulation does not possess good mass conservation properties (see Aldama et al., [4], for details). It may be shown, by applying the theory presented in this paper that such formulation does not satisfies the continuity equation and that the error is larger for high wavenumbers. This theoretical result is consistent with observations that indicate that relatively large mass conservation errors arise in refined grids.

12. Conclusions. A theory that consists of the Taylor-Fréchet expansion of nonlinear operators, multiple scale analysis, localization and asymptotic analysis has been presented in order to include dominant nonlinear effects in the study of the propagation properties (stability, amplitude and phase portraits, nonlinear iteration convergence) of nonlinear evolution systems. The theory presented has been tested via a number of applications, a few of which are presented in this paper, with excellent results.

REFERENCES

[1] M. B. Abbot. *Computational hydraulics. Elements of the theory of free surface flows.* Pitman. London, 1979.

[2] A. Aguilar. *Propagation properties of numerical schemes for free flow simulation.* PhD thesis, UNAM. Mexico, 2002.

[3] A. Aldama. Stability analysis of discrete approximations of the advection diffusion equation through the use of an ordinary differential equation analogy. *Developments in Water Science*, (5):3–8, 1988.

[4] A. Aldama, A. Aguilar, J. Westerink, and R. Kolar. A mass conservation analysis of the GWCE formulation. In B. et al, editor, *XIII International Conference on Computational Methods in Water Resources*, pages 597–601 907–912. Balkema. Rotterdam, 2000.

[5] A. Aldama and J. Aparicio. The effect of nonlinearities in the stability of numerical solutions of Richards' equation. In B. et al, editor, *XII International Conference on Computational Methods in Water Resources. Volume I*, pages 289–296. Computational Mechanics Publications. Southampton, 1998.

[6] A. Aldama and V. Arroyo. Propagation properties of Eulerian Lagrangian Localized Adjoint Methods. In B. et al, editor, *XIII International Conference on Computational Methods in Water Resources*, pages 597–601. Vol 2. Balkema. Rotterdam, 2000.

[7] A. Aldama and A. Ocon. Flow resistance in open channels and Manning's formula limits of applicability (in spanish). *Ingenieria Hidraulica en Mexico*, XVII:107–115, Jan–Mar 2002.

[8] A. Aldama and C. Paniconi. An analysis of the convergence of picard iterations for implicit approximations of Richards' equation. In R. et al, editor, *IX International Conference on Computational Methods in Water Resources*, pages 521–528. Computational Mechanics Publications and Elsevier. Southampton and London., 1992.

[9] J. Aparicio and A. Aldama. On the efficient determination of stability properties for higher order approximations of the transport equation. In A. et al, editor, *XI International Conference on Computational Methods in Water Resources*, pages 29–36. Computational Mechanics Publications. Southampton, 1996.

[10] L. Bentley, A. Aldama, and G. Pinder. Fourier analysis of the Eulerian-Lagrangian-least-squares-collocation method. *International Journal for Numerical Methods in Fluids*, (11):427–444, 1990.

[11] M. Celia, E. T. Bouloutas, and R. L. Zarba. A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resources Research*, (23):1483–1496, 1990.

[12] D. C. Champeney. *A Handbook of Fourier Theorems.* Cambridge University Press, 1989.

[13] P. Huyakon, S. Thomas, and B. Thompson. Techniques for making finite elements competitive in modelling flow in variably saturated porous media. *Water Resources Research*, (20):1099–1115, 1984.

[14] I. Kinmark. *The Shallow Water Wave Equations: Formulation, Analysis and Application.* Springer-Verlag. Berlin-Heidelberg, 1986.

[15] J. Lee and A. Aldama. Multipath diffusion: A general numerical model. *Computers and Geosciences*, (18):531–555, 1992.

[16] R. D. Milne. *Applied Functional Analysis.* Pitman. London, 1980.

[17] R. Vichnevetsky and J. B. Bowles. *Fourier Analysis of Numerical Approximations of Hyperbolic Equations.* SIAM, Philadelphia, 1982.