# AN OPTIMAL SWITCHING MECHANISM FOR A COMBINED PICARD-NEWTON METHOD IN THE SOLUTION OF RICHARDS' EQUATION

JAVIER APARICIO [1], ÁLVARO A. ALDAMA[1], CLAUDIO PANICONI [2] AND MARIO PUTTI [3]

[1]*Mexican Institute of Water Technology, Paseo Cuauhnáhuac 8532, Jiutepec 62550, Mor., México; japaricio@tlaloc.imta.mx, aaldama@tlaloc.imta.mx*
[2]*INRS-ETE, University of Quebec, 490 de la Couronne, G1K 9A9, Quebec City, Canada; claudio.paniconi@ete.inrs.ca*
[3]*Dept. Mathematical Methods and Models for Scientific Applications, University of Padova, via Belzoni 7 35131 Padova Italy; putti@dmsa.unipd.it*

## ABSTRACT

Richards' equation, describing flow in partially saturated porous media, contains strong nonlinearities arising from pressure head dependencies in soil moisture and hydraulic conductivity. Additionally, the time-dependent nature of boundary conditions can alter the nonlinear characteristics of this equation during a transient simulation. Various iterative methods are used for solving this nonlinear equation, most commonly the quadratically convergent Newton-Raphson technique and the simpler but only linearly convergent Picard method (successive approximations). The initial solution estimate can have a large influence on the behavior of these iterative schemes, and it has been observed through many applications of our numerical subsurface flow models that the Newton scheme is sometimes more sensitive to the initial solution than the Picard scheme. In this paper, the latter is used to calculate improved initial guesses for the Newton iteration. This scheme improves the global behavior of the iteration at less cost and complexity than alternative globalization techniques such as line search and trust region methods. In this work the combined Picard-Newton method is investigated via a theoretical analysis based on a Taylor-Fréchet expansion of the nonlinear Richards operator and a localization approach. In particular, we assess the convergence behavior of the individual schemes in order to determine an optimal criterion for switching from Picard to Newton iteration. This criterion, together with the overall convergence behavior of the Picard, Newton, and combined schemes, is illustrated in two numerical tests.

## 1. RICHARDS' EQUATION

In this paper we will consider the convergence properties of linearization schemes for the numerical solution of Richards' equation (Philip, 1969):

$$S(\psi)\frac{\partial \psi}{\partial t} = \frac{\partial}{\partial z}\left[ K(\psi)\left(\frac{\partial \psi}{\partial z} + 1\right)\right] \qquad (1)$$

where $\psi$ is the pressure head, *S* the specific moisture capacity, *K* the hydraulic conductivity, *t* time and *z* the space coordinate. Eq. (1) can be approximated by finite differences as

$$\left[\theta S_j^{n+1} + (1-\theta) S_j^n\right] \frac{\psi_j^{n+1} - \psi_j^n}{\Delta t} = \theta F_j^{n+1}(\psi) + (1-\theta) F_j^n(\psi) \qquad (2)$$

where

$$F_j^n = \frac{1}{\Delta z^2}\left[\left(\frac{K_{j+1}^n + K_j^n}{2}\left(\psi_{j+1}^n - \psi_j^n\right) - \frac{K_j^n + K_{j-1}^n}{2}\left(\psi_j^n - \psi_{j-1}^n\right)\right)\right] + \frac{K_{j+1}^n - K_{j-1}^n}{2\Delta z}$$

$\theta \in (0,1]$ is a time-weighting factor and $\psi_j^n$ represents a discrete approximation of $\psi(j\Delta z, n\Delta t)$, with *j*, *n*, $\Delta z$ and $\Delta t$ denoting nodal location, time level, gridsize and time step respectively.

## 2. ANALYSIS OF PICARD AND NEWTON CONVERGENCE

Eq. (2) is a nonlinear algebraic equation which can be solved by a number of numerical procedures, of which the Newton [*e.g.*, Dahlquist & Björck, 1974; Aparicio & Aldama, 2000] and Picard [*e.g.*, Huyakorn & Pinder, 1983; Aldama & Paniconi, 1992] algorithms are the most popular. Let, from eq. (2),

$$f(\psi_j) = \left[\theta S_j^{n+1} + (1-\theta) S_j^n\right] \frac{\psi_j^{n+1} - \psi_j^n}{\Delta t} - \theta F_j^{n+1}(\psi) - (1-\theta) F_j^n(\psi) = 0; \quad j = 1, 2, nz+1$$

Then, iterations based on the Newton algorithm can be expressed as

$$\overline{\overline{f}}'(\overline{\psi}^{n+1,m})(\overline{\psi}^{n+1,m+1} - \overline{\psi}^{n+1,m}) + \overline{f}(\overline{\psi}^{n+1,m}) = \overline{0} \qquad (3)$$

where $\overline{\overline{f}}'\left(\overline{\psi}^{n+1,m}\right)$ is the Jacobian matrix and *m* is the iteration number. The system of equations (3) is now linear in the pressure heads vector $\overline{\psi}^{n+1,m+1}$.

On the other hand, Picard linearization leads to:

$$\left[\theta S_j^{n+1,m} + (1-\theta) S_j^n\right] \frac{\psi_j^{n+1,m+1} - \psi_j^n}{\Delta t} = \theta F_j^{n+1,m+1}(\psi) + (1-\theta) F_j^n(\psi) \qquad (4)$$

where

$$F_j^{n+1,m+1}(\psi)=\frac{1}{\Delta z^2}\left[\left(\frac{K_{j+1}^{n+1,m}+K_j^{n+1,m}}{2}\left(\psi_{j+1}^{n+1,m+1}-\psi_j^{n+1,m+1}\right)-\right.\right.$$
$$\left.\left.-\frac{K_j^{n+1,m}+K_{j-1}^{n+1,m}}{2}\left(\psi_j^{n+1,m+1}-\psi_{j-1}^{n+1,m+1}\right)\right)\right]+\frac{K_{j+1}^{n+1,m}-K_{j-1}^{n+1,m}}{2\Delta z}$$

Picard iteration tends to work better when relaxation is used (Paniconi & Putti, 1994), whereby the solution $\psi_j^{n+1,m+1}$ obtained from solving eq. (4) is updated as $\psi_j^{*n+1,m+1}$ by the relationship

$$\psi_j^{*n+1,m+1}=\psi_j^{n+1,m}+\alpha_r\left(\psi_j^{n+1,m+1}-\psi_j^{n+1,m}\right) \tag{5}$$

where $\alpha_r$ is the relaxation parameter.

## 3. CONVERGENCE ANALYSIS

### 3.1 Picard iterations

Aldama & Paniconi (1992), through a “frozen coefficients” approach and a Fourier analysis of the discretization error, found that the condition for iteration convergence

$$|\xi_k|<1 \quad \forall k \tag{6}$$

$\xi_k$ being the amplification factor for the *k*th Fourier mode, can be written as

$$\xi_k=\frac{\mu_{2,k}}{1+\mu_{1,k}} \tag{7}$$

where $\mu_{1,k}$ and $\mu_{2,k}$ are complex numbers. Denoting with subindices *R* and *I* respectively the real and imaginary parts, from (7) it follows that

$$\mu_{2R,k}^2+\mu_{2I,k}^2<1+2\mu_{1R,k}+\mu_{1R,k}^2+\mu_{1I,k}^2 \quad \forall k \tag{8}$$

in which

$$\mu_{1R,k} = 2\theta \frac{K(\psi_0)}{S(\psi_0)}(1-\cos\beta_k)\frac{\Delta t}{\Delta z^2} + \theta \frac{S'(\psi_0)}{S(\psi_0)}\left(\frac{\partial \psi}{\partial t}\right)_0 \Delta t \tag{9}$$

$$\mu_{1I,k} = \theta \frac{K'(\psi_0)}{S(\psi_0)}\left(\frac{\partial \psi}{\partial z}\right)_0 \frac{\Delta t}{\Delta z}\sin\beta_k \tag{10}$$

$$\begin{aligned}\mu_{2R,k} = {} & \theta \frac{K''(\psi_0)}{S(\psi_0)}\left(\frac{\partial \psi}{\partial z}\right)_0 \left[\left(\frac{\partial \psi}{\partial z}\right)_0 + 1\right]\cos\beta_k \Delta t + \\ & + \frac{1}{2}\theta \frac{K'(\psi_0)}{S(\psi_0)}\left(\frac{\partial^2 \psi}{\partial z^2}\right)_0 (1+\cos\beta_k)\Delta t - \theta \frac{S'(\psi_0)}{S(\psi_0)}\left(\frac{\partial \psi}{\partial t}\right)_0 \Delta t\end{aligned} \tag{11}$$

$$\mu_{2I,k} = -\mu_{1I,k}\left[1 + \left(\frac{\partial \psi}{\partial z}\right)_0^{-1}\right] \tag{12}$$

Subindex 0 represents a "frozen" value and $\beta_k \in (-\pi, \pi]$ is the dimensionless wave number corresponding to the *k*th Fourier mode. Inequality (8) implies that there is a set of values of $\Delta t$ for which convergence is attained. Therefore, this analysis can be used to guide the choice of $\Delta t$ to ensure Picard convergence at the start of the iteration procedure. Substituting eqs. (9) to (12) in (8), the following convergence condition in terms of $\Delta t$ is obtained:

$$A_p \Delta t^2 + B_p \Delta t + C_p < 0 \tag{13}$$

where

$$\begin{aligned}A_p = {} & \left[-S_0'\psi_{t0} + K_0''\psi_{z0}\cos\beta_k(\psi_{z0}+1) + \frac{1}{2}K_0'\psi_{zz0}(1+\cos\beta_k)\right]^2 + \\ & + \frac{1}{\Delta z^2}\left[K_0'\psi_{z0}\sin\beta_k\right]^2\left[2\psi_{z0}^{-1} + \left(\psi_{z0}^{-1}\right)^2\right] - \left[S_0'\psi_{t0} + \frac{2}{\Delta z^2}K_0(1-\cos\beta_k)\right]^2\end{aligned}$$

$$B_p = -\frac{2S_0}{\theta}\left[S_0'\psi_{t0} + \frac{2}{\Delta z^2}K_0(1-\cos\beta_k)\right]$$

$$C_p = -\frac{S_0^2}{\theta^2}$$

where, to simplify the notation with respect to the previous equations, we use $\psi_{z0} = \left(\frac{\partial \psi}{\partial z}\right)_0, \psi_{t0} = \left(\frac{\partial \psi}{\partial t}\right)_0, K_0 = K(\psi_0)$ and so on.

### 3.2 Newton iterations

Aparicio & Aldama (2000), using a similar methodology, found the following equation for the amplification factor $\xi$ of the Newton scheme:

$$\frac{\theta\Delta t\xi}{2\Delta z^2}\left\{2\left(K_0+\Delta tK_{t0}\right)\left(\cos\beta_k-1\right)+i\Delta z\left[K_{z0}+K'_0\left(\psi_{z0}+1\right)\sin\beta_k\right]\right\}- \\ -\xi\left(S_0+\Delta tS_{t0}+\theta\Delta tS'_0\psi_{t0}\right)=0 \tag{14}$$

Let

$$D\equiv\frac{\theta\Delta t}{\Delta z^2}\left(K_0+\Delta tK_{t0}\right)\left(\cos\beta_k-1\right)-\left(S_0+\Delta tS_{t0}+\theta\Delta tS'_0\psi_{t0}\right) \tag{15}$$

and

$$B\equiv\frac{\theta\Delta t}{2\Delta z}\left(K_{z0}+K'_0\left(\psi_{z0}+1\right)\sin\beta_k\right) \tag{16}$$

Eq. (14) implies that either $\xi=0$ —that is, we have unconditional convergence— or that the coefficient of $\xi$ is null, meaning that $\xi$ is not necessarily equal to zero, and thus also that $|\xi|<1$ does not necessarily hold. To avoid this uncertain case, we wish to use the Newton scheme only when the modulus of $D+Bi$

$$M=\sqrt{D^2+B^2} \tag{17}$$

is *sufficiently* large. In this case it most follow that for (14) to hold, we must have $\xi=0$ — thus $|\xi|<1$ holds— and Newton will converge.

## 4. SWITCHING MECHANISM

Due to the fact that *M* varies along the solution domain, the following norm will be used to evaluate the convergence and switching criterion:

$$\|M\|\equiv\frac{1}{(N_z)}\sum_{j=1}^{N_z}M_j \tag{18}$$

where $N_z$ is the number of nodes. The switching criterion from Picard to Newton iterations will thus be

$$\|M\| > \varepsilon \tag{19}$$

where $\varepsilon$ is a suitable tolerance.

## 5. NUMERICAL TESTS

The numerical tests are based on a one-dimensional infiltration example, with $\Delta t = 0.2$ h, a uniform grid spacing $\Delta z$ = 0.1 m and $\psi = 0$ at $z = 0$. The boundary condition at the top of the soil $z = (N_z + 1)\Delta z = L$ is a known infiltration flux $k_H$:

$$K(\psi_L)\left(\frac{\partial \psi_L}{\partial t} + 1\right) = k_H$$

where, in this case, $N_z = 100$ and $k_H = 0.0001$ m/s. We used saturated conductivity $K_s = 0.01$ m/s and van Genuchten characteristic equations (van Genuchten and Nielsen, 1985), with $\beta$ = 5, $\theta_s - \theta_r = 0.37$, and $\psi_s = -3$ m. A weighting factor $\theta = 1$ (i.e., a backward Euler scheme for eq. (2)) was set. Two linear pressure head distributions were used as initial conditions: $\psi = 0$ at $z = 0$ to $\psi = -1$ m at $z = L$ (BC1) and $\psi = 0$ at $z = 0$ to $\psi = -4$ m at $z = L$ (BC2). In test case BC2 the Newton procedure converges, whereas it fails to converge for BC1. On the other hand, the Picard procedure converges in both cases and condition (13) is always satisfied, with $A_p$, $B_p$ and $C_p$ negative or zero for every node and value of $\beta_k$.

Modulus *M* (eq. (17)) was computed for each node during the calculation in the first iteration. These values remained essentially unchanged during the rest of the iterations for the convergent case BC2, while they were not considered valid for iterations 2 and higher in the nonconvergent case BC1, due to the erroneous solution being propagated in this case. It was also observed that the minimum values for each node were always attained when $\beta_k = -\pi$. Figure 1 shows values of the module *M* for both initial conditions. It can be seen that for BC2 *M* stays well above a suitable tolerance for most of the calculation domain, whereas BC1 yields modules below or near the tolerance. Therefore, for test case BC1 eq. (14) is satisfied for values of $\xi$ not necessarily equal to zero and this could explain the nonconvergent behavior of the Newton algorithm in this case. Conversely, the high *M* values for test case BC2 suggest that $\xi$ must be zero, condition (6) is satisfied and thus Newton converges.

Other experiments were done for different initial conditions and it was found that, for this case, an appropriate limit for convergence is $\varepsilon \cong 1.5 \times 10^{-2}$.

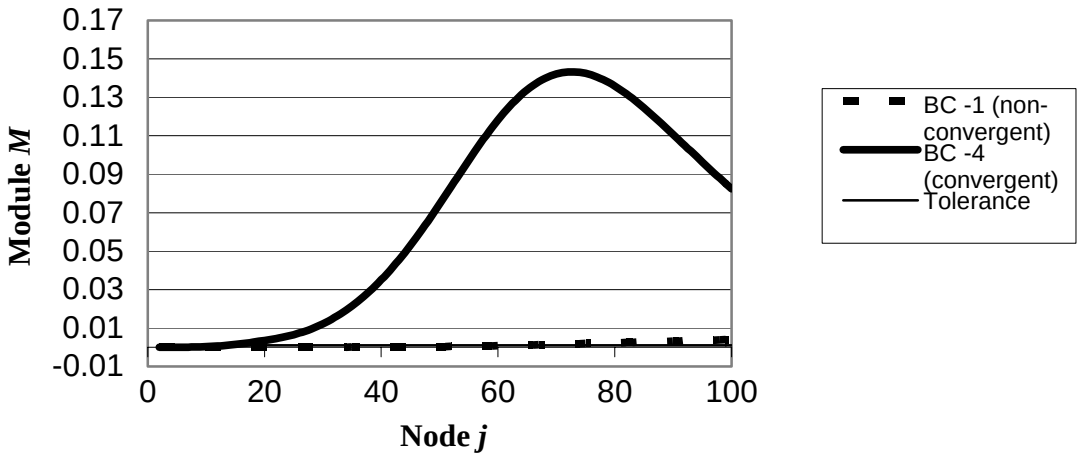


FIGURE 1. Moduli for test cases BC1 and BC2.

Test case BC1 was run using the above mentioned parameters, with $\Delta t = 3\,\mathrm{h}$ and a tolerance of $10^{-5}$. The Picard procedure with relaxation ($\alpha_r = 0.5$) took 173 iterations to converge in the first time step, while unrelaxed Newton diverged. The combined Picard-Newton method with switching criterion (19) (using $\varepsilon \cong 1.5\times10^{-2}$, switching occurs at the sixth iteration) converged much faster. In all Picard iterations, $A_p$, $B_p$ and $C_p$ were negative, thus condition (13) was always satisfied. The norm (eq. (18)) calculated at the sixth iteration was $1.6302\times10^{-2}$. This relatively high value (higher than $\varepsilon$) indicate that a switch to Newton at this stage will likely lead to convergence and at a faster rate than Picard, as indeed we were able to verify. Figure 2 shows the convergence behavior of the combined method in terms of the convergence error norm

$$\left\|\psi_{diff}\right\| \equiv \sqrt{\sum_{j=1}^{N_z}\left(\psi_j^{n+1,m+1} - \psi_j^{n+1,m}\right)^2} \qquad (20)$$

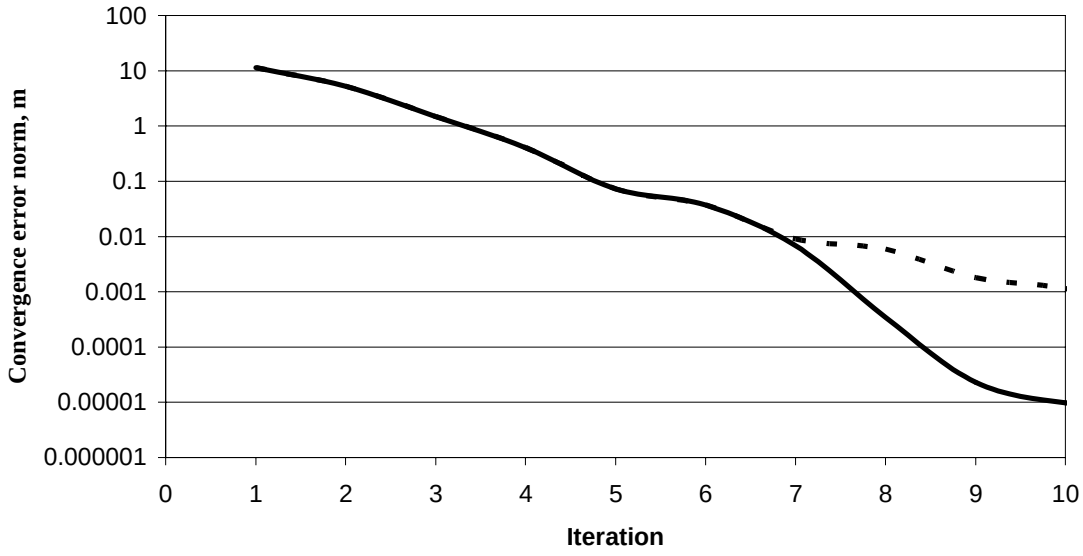


FIGURE 2. Convergence behavior for the Picard scheme (dashed line) and the combined Picard-Newton method (continous line).

## 6. CONCLUSIONS

The convergence properties of the Picard and Newton iterative methods to solve Richards equation have been assessed through the analysis of previously published results. The behavior of the Newton algorithm, which may become nonconvergent for certain initial conditions, was explained and the region of convergence of the Picard algorithm was derived. An *optimal* criterion was obtained for switching from Picard to Newton iteration during the solution procedure based on the analysis of the convergence condition. Numerical experiments confirmed the above results and the good performance of the proposed combined Picard-Newton method.

## REFERENCES

Aldama, A.A. and C. Paniconi (1992), An analysis of the convergence of Picard iterations for implicit approximations of Richards' equation, *Proc. IX Int. Conf. Comp. Meth. Water Res. 2*: 521-528. Comp. Mech. Publ., Southampton

Aparicio, J. and A.A. Aldama, (2000), Convergence of Newton iterations for the solution of Richard's equation, *Proc. XIII Int. Conf. Comp. Meth. Water Res.*, 107-111. Balkema, Rotterdam

Dahlquist, G. and Å. Björck, (1974), *Numerical Methods*. Prentice-Hall, Englewood Cliffs

Philip, J.R. (1969). Theory of infiltration, *Adv. In Hydrosc.*, 5:215-296

Huyakorn, P.S. and G.F. Pinder, (1983), *Computational methods in subsurface flow*, Academic Press, New York

Paniconi, C., and M.Putti, (1994). A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems, *Water Resour. Res.*, 30(12), 3357-3374

van Genuchten, M.T. and D.R. Nielsen (1985), On describing and predicting the hydraulic properties of unsaturated soils, *Ann. Geophys.* 3(5):615-628