XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)
CIMNE, Barcelona 2010

A DESCRIPTION OF LINEAR AND NONLINEAR SOLVER FAILURES AND CURES FOR UNSATURATED FLOW CALCULATIONS

Fred T. Tracy[*]

[*] Engineer Research and Development Center
Information Technology Laboratory
DoD Supercomputing Resource Center
Vicksburg, MS USA 39180
e-mail: Fred.T.Tracy@usace.army.mil

Key words: Numerical methods, unsaturated flow, linear and nonlinear solvers

Summary. This breakdown of the linear system of equations observed in a three-dimensional (3-D) parallel finite element solution of unsaturated flow using Newton's method is illustrated using the one-dimensional (1-D) solution of the Green-Ampt infiltration problem. Causes and cures are then described.

1 INTRODUCTION

A recent investigation of linear and nonlinear solvers for difficult unsaturated flow problems [1] revealed that a BiCG-Stabilized solver [2] for Newton's method sometimes failed. This was in the context of a 3-D Galerkin finite element method discretization, where only tetrahedral elements were used. It was first thought that this was due to a weakness in the linear solver, so different linear solver options were considered using the Portable, Extensible Toolkit for Scientific Computation (PETSc) [3] library, as has recently been done by Nguyen et al. [4]. However, it was also observed that not only the 3-D solution failed, but also a two-dimensional (2-D) finite difference version of the test problem using a direct banded linear solver also failed. This indicated that something more fundamental might be at work. The original problem was essentially a 3-D version of the Green-Ampt problem [5] where a dry soil sample has a head applied at the top of the sample, thus generating a downward moving front of water. When the 1-D Green-Ampt problem was then tested using the finite difference method, the direct tridiagonal linear system of equations did not fail; but unless remedies were applied, the nonlinear solution process failed. This document will describe the cause and different cures of the problem.

2 PROBLEM DESCRIPTION

The Green-Ampt infiltration problem is typically 1-D unsaturated flow in a vertical column of soil that is initially dry until ponding occurs at the top (often pressure head is zero). A front then proceeds downward as the soil becomes saturated (see Figure 1). Node 1 is at $z = L$, and the bottom node is at $z = 0$.

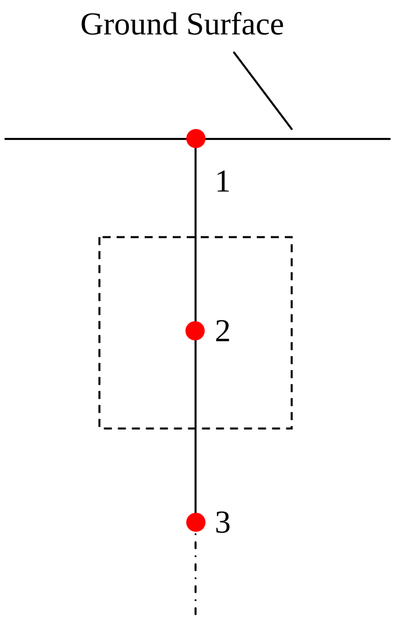

Figure 1. Finite difference representation of the first three nodes of the Green-Ampt problem

3 GOVERNING EQUATIONS

The 1-D version of Richards' equation was used as follows:

$$k_s \frac{\partial}{\partial z}\left(k_r \frac{\partial \phi}{\partial z}\right) = \frac{\partial \theta}{\partial t} \tag{1}$$

where

$$\phi = h + z \tag{2}$$

and k_r is the relative hydraulic conductivity, k_s is the constant saturated hydraulic conductivity, ϕ is the total head, θ is the moisture content, h is the pressure head, z is the positive upward z coordinate, and t is the time. The initial conditions are that the soil is dry, so the pressure head is a constant negative value, h_d ($h = h_d$ or $\phi = h_d + z$). The boundary conditions are $h = 0$ ($\phi = L$) at $z = L$, and $\phi = h = h_d$ at $z = 0$. For simplicity, use Gardner's equation [6],

$$k_r = e^{\alpha h} \tag{3}$$

where α is a positive parameter. Also, use

$$\frac{\theta - \theta_d}{\theta_s - \theta_d} = k_r \tag{4}$$

where θ_s is the moisture content when the soil is saturated, and θ_d is the moisture content when the soil is dry.

4 FINITE DIFFERENCE DISCRETIZATION FOR NODE 2

The finite difference equation for node 2 is

$$k_s \frac{\left[k_r(h)\right]_{1+1/2}^{n+1}\left(\frac{L-\phi_2^{n+1}}{\Delta z}\right) - \left[k_r(h)\right]_{2+1/2}^{n+1}\left(\frac{\phi_2^{n+1}-\phi_3^{n+1}}{\Delta z}\right)}{\Delta z} = \frac{\theta_2^{n+1}(h_2) - \theta_2^n}{\Delta t} \tag{5}$$

Here, k_r and θ are functions of h, making the discretization nonlinear. Two different ways of evaluating $k_r(h)$ will be described in detail below. Eq. 5 is rewritten as follows:

$$\left[k_r(h)\right]_{1+1/2}^{n+1}\left(\phi_2^{n+1} - L\right) + \left[k_r(h)\right]_{2+1/2}^{n+1}\left(\phi_2^{n+1} - \phi_3^{n+1}\right) + \frac{\Delta z^2}{k_s \Delta t}\left(\theta_2^{n+1}(h_2) - \theta_2^n\right) = 0 \tag{6}$$

where n is the time-step number, Δt is the time increment, $\left[k_r(h)\right]_{1+1/2}^{n+1}$ is the relative hydraulic conductivity evaluated halfway between nodes 1 and 2 at time-step, $n+1$, and Δz is the constant distance between nodes.

5 NEWTON LINEARIZATION FOR NODE 2

The nonlinear equation,

$$f_2 = \left[k_r(h)\right]_{1+1/2}\left(\phi_2 - L\right) + \left[k_r(h)\right]_{2+1/2}\left[\phi_2 - \phi_3\right] + \frac{\Delta z^2}{k_s \Delta t}\left[\theta_2(h_2) - \theta_2^n\right] = 0 \tag{7}$$

is to be solved using Newton's method. Thus,

$$\frac{\partial f_2^{n+1.k}}{\partial \phi_2}\Delta\phi_2^{n+1,k+1}+\frac{\partial f_2^{n+1.k}}{\partial \phi_3}\Delta\phi_3^{n+1,k+1}=-f_2^{n+1.k} \tag{8}$$

where k is the nonlinear iteration count number. For the first time-step and $k=0$, $\phi_2^{1,0}=\phi_2^0=h_d+L-\Delta z$, and $\phi_3^{1,0}=\phi_3^0=h_d+L-2\Delta z$.

6 LINEAR SYSTEM OF EQUATIONS

In general, with, for example, $N=7$ nodes, $\phi_1=L$ and $\phi_7=h_d$, thus giving the following tridiagonal system of equations to solve:

$$\begin{bmatrix} a_{22} & a_{23} & & & \\ a_{32} & a_{33} & a_{34} & & \\ & a_{34} & a_{44} & a_{54} & \\ & & a_{45} & a_{55} & a_{56} \\ & & & a_{65} & a_{66} \end{bmatrix}\begin{Bmatrix} \Delta\phi_2^{n+1,k+1} \\ \Delta\phi_3^{n+1,k+1} \\ \Delta\phi_4^{n+1,k+1} \\ \Delta\phi_5^{n+1,k+1} \\ \Delta\phi_6^{n+1,k+1} \end{Bmatrix}=\begin{Bmatrix} b_2 \\ 0 \\ 0 \\ 0 \\ b_6 \end{Bmatrix} \tag{9}$$

Then $\boldsymbol{\varphi}$ is updated by

$$\boldsymbol{\varphi}^{n+1,k+1}=\boldsymbol{\varphi}^{n+1,k}+\beta\Delta\boldsymbol{\varphi}^{n+1,k+1}, \quad 0<\beta\le 1 \tag{10}$$

where $\boldsymbol{\varphi}$ is the vector of unknown total heads at the nodes, and β is a parameter resulting from a line search.

7 RELATIVE HYDRAULIC CONDUCTIVITY CHOICES

Two ways of computing relative hydraulic conductivity are considered. It is important to note here that this is not a comprehensive study of different ways to average relative hydraulic conductivity and the errors associated with each method as others have done [7]. Rather, the focus of this work is to use the two selected methods of averaging to illustrate linear and nonlinear solver difficulties and fixes. That is, robustness, not accuracy, is the primary focus.

7.1 Average pressure head at the midpoint

The first choice is

$$\bar{h}_{1+1/2}=\frac{0+h_2}{2}=\frac{h_2}{2}, \quad \bar{h}_{2+1/2}=\frac{1}{2}\left(h_2+h_3\right) \tag{11}$$

$$[k_r(h)]_{1+1/2} = k_r(\bar{h}_{1+1/2}), \quad [k_r(h)]_{2+1/2} = k_r(\bar{h}_{2+1/2}) \tag{12}$$

Eq. 8 for the first nonlinear iteration for the first time-step becomes

$$\begin{aligned}
&k_r(\bar{h}^{1,0}_{1+1/2})\Delta\phi_2^{1,1} + \frac{1}{2}\frac{dk_r}{dh}(\bar{h}^{1,0}_{1+1/2})[\phi_2^{1,0} - L]\Delta\phi_2^{1,1} + k_r(\bar{h}^{1,0}_{2+1/2})\Delta\phi_2^{1,1} + \\
&\frac{1}{2}\frac{dk_r}{dh}(\bar{h}^{1,0}_{2+1/2})[\phi_2^{1,0} - \phi_3^{1,0}]\Delta\phi_2^{1,1} + \left(\frac{\Delta z^2}{k_s \Delta t}\right)\frac{d\theta}{dh}(h_2^{1,0})\Delta\phi_2^{1,1} - \\
&k_r(\bar{h}^{1,0}_{2+1/2})\Delta\phi_3^{1,1} + \frac{1}{2}\frac{dk_r}{dh}(\bar{h}^{1,0}_{2+1/2})[\phi_2^{1,0} - \phi_3^{1,0}]\Delta\phi_3^{1,1} = \\
&-k_r(\bar{h}^{1,0}_{1+1/2})[\phi_2^{1,0} - L] - k_r(\bar{h}^{1,0}_{2+1/2})[\phi_2^{1,0} - \phi_3^{1,0}] - \frac{\Delta z^2}{k_s \Delta t}\left[\theta_2(h_2^{1,0}) - \theta_2^0\right]
\end{aligned} \tag{13}$$

Putting Eq. 3 in the above equation gives

$$\begin{aligned}
&\left[e^{\alpha\frac{h_d}{2}} + \frac{\alpha}{2}e^{\alpha\frac{h_d}{2}}(h_d - \Delta z) + e^{\alpha h_d} + \frac{\alpha\Delta z}{2}e^{\alpha h_d} + \frac{\alpha\Delta z^2}{k_s\Delta t}(\theta_s - \theta_d)e^{\alpha h_d}\right]\Delta\phi_2^{1,1} \\
&- e^{\alpha h_d}\left(1 - \frac{\alpha\Delta z}{2}\right)\Delta\phi_3^{1,1} = -e^{\alpha\frac{h_d}{2}}(h_d - \Delta z) - e^{\alpha h_d}\Delta z
\end{aligned} \tag{14}$$

Using $L = 50\ cm$, $\Delta z = 0.25\ cm$, $\Delta t = 0.1\ hr$, $\alpha = 0.2\ cm^{-1}$, $h_d = -50\ cm$, $\theta_s = 0.45$, $\theta_d = 0.15$, and $k_s = 0.1\ cm/hr$, the coefficient of $\Delta\phi_2^{1,1}$ in Eq. 13 becomes

$$\begin{aligned}
&e^{\alpha\frac{h_d}{2}} + \frac{\alpha}{2}e^{\alpha\frac{h_d}{2}}(h_d - \Delta z) + e^{\alpha h_d} + \frac{\alpha\Delta z}{2}e^{\alpha h_d} + \frac{\alpha\Delta z^2}{k_s\Delta t}(\theta_s - \theta_d)e^{\alpha h_d} = \\
&0.006738 - 0.033858 + 0.000045 + 0.000001 + 0.000017 = \\
&-0.027057
\end{aligned} \tag{15}$$

The main diagonal term for node 2 is indeed negative, and Δt must be made substantially smaller to create a positive main diagonal term. This phenomenon is, in fact, the crucial cause of the failure of the linear solver and thus the nonlinear solver in the 1-D, 2-D, and 3-D cases. The resulting linear system often has an extremely high condition number because negative main diagonal terms near the front cause near-zero main diagonal terms a bit away from the front.

7.2 Average relative hydraulic conductivity at the midpoint

The second choice is

$$[k_r(h)]_{1+1/2} = \frac{1}{2}[1 + k_r(h_2)], \quad [k_r(h)]_{2+1/2} = \frac{1}{2}[k_r(h_2) + k_r(h_3)] \tag{16}$$

Eq. 8 for the first nonlinear iteration for the first time-step now becomes

$$\begin{aligned}
&\frac{1}{2}\left[1 + k_r\left(h_2^{1,0}\right)\right]\Delta\phi_2^{1,1} + \frac{1}{2}\frac{dk_r}{dh}\left(h_2^{1,0}\right)\left[\phi_2^{1,0} - L\right]\Delta\phi_2^{1,1} + \frac{1}{2}\left[k_r\left(h_2^{1,0}\right) + k_r\left(h_3^{1,0}\right)\right]\Delta\phi_2^{1,1} \\
&+ \frac{1}{2}\frac{dk_r}{dh}\left(h_2^{1,0}\right)\left[\phi_2^{1,0} - \phi_3^{1,0}\right]\Delta\phi_2^{1,1} + \left(\frac{\Delta z^2}{k_s\Delta t}\right)\frac{d\theta}{dh}\left(h_2^{1,0}\right)\Delta\phi_2^{1,1} - \frac{1}{2}\left[k_r\left(h_2^{1,0}\right) + k_r\left(h_3^{1,0}\right)\right]\Delta\phi_3^{1,1} \\
&+ \frac{1}{2}\frac{dk_r}{dh}\left(h_3^{1,0}\right)\left[\phi_2^{1,0} - \phi_3^{1,0}\right]\Delta\phi_3^{1,1} = -\frac{1}{2}\left[1 + k_r\left(h_2^{1,0}\right)\right]\left(\phi_2^{1,0} - L\right) \\
&- \frac{1}{2}\left[k_r\left(h_2^{1,0}\right) + k_r\left(h_3^{1,0}\right)\right]\left(\phi_2^{1,0} - \phi_3^{1,0}\right) - \frac{\Delta z^2}{k_s\Delta t}\left[\theta_2\left(h_2^{1,0}\right) - \theta_2^0\right]
\end{aligned} \tag{17}$$

Putting Eq. 3 in the above equation as before gives

$$\begin{aligned}
&\left[\frac{1}{2}\left(1 + e^{\alpha h_d}\right) + \frac{\alpha}{2}e^{\alpha h_d}\left(h_d - \Delta z\right) + e^{\alpha h_d} + \frac{\alpha\Delta z}{2}e^{\alpha h_d} + \frac{\alpha\Delta z^2}{k_s\Delta t}\left(\theta_s - \theta_d\right)e^{\alpha h_d}\right]\Delta\phi_2^{1,1} \\
&- e^{\alpha h_d}\left(1 - \frac{\alpha\Delta z}{2}\right)\Delta\phi_3^{1,1} = -\frac{1}{2}\left(1 + e^{\alpha h_d}\right)\left(h_d - \Delta z\right) - e^{\alpha h_d}\Delta z
\end{aligned} \tag{18}$$

Using the same input data values as before, the main diagonal term for node 2 is

$$\begin{aligned}
&\frac{1}{2}\left(1 + e^{\alpha h_d}\right) + \frac{\alpha}{2}e^{\alpha h_d}\left(h_d - \Delta z\right) + e^{\alpha h_d} + \frac{\alpha\Delta z}{2}e^{\alpha h_d} + \frac{\alpha\Delta z^2}{k_s\Delta t}\left(\theta_s - \theta_d\right)e^{\alpha h_d} = \\
&0.500023 - 0.000228 + 0.000045 + 0.000001 + 0.000017 = \\
&0.499858
\end{aligned} \tag{19}$$

Here, the first term is much bigger, and the negative term is much smaller than the first option for k_r. It is also important to note that if the $\frac{dk_r}{dh}$ terms are not used (k_r lagged to the previous nonlinear iteration (Picard iteration [8])), all the negative contributions to the main diagonal vanish.

8 COMPUTATIONAL RESULTS

The finite difference solution was run with the two options for relative hydraulic conductivity using combinations of Picard and Newton iterations. The number of nonlinear iterations for the first time-step is presented in Table 1. Here, any Picard iterations are done first and then any Newton iterations are done after that. The linear system of equations was done by a direct solution without any pivoting.

k_r	**Picard**	**Newton**	**Total**
$(k_r)_{i+1/2} = k_r\left(\frac{h_i + h_{i+1}}{2}\right)$	0	Not converged after 20,000	-
	5	Blew up after 4	
	10	Blew up after 2	
	20	58	78
	72	0	72
$(k_r)_{i+1/2} = \frac{1}{2}\left[k_r(h_i) + k_r(h_{i+1})\right]$	0	76	76
	71	0	71

Table 1. Nonlinear iteration count for the first time-step

9 CONCLUSIONS

- For the Newton iteration, a negative contribution to the main diagonal of every internal node, m, exists in this problem because flow is downward. This negative value is because $\phi_m - \phi_{m-1} < 0$ (nodes numbered consecutively from top to bottom) and $\frac{dk_r}{dh} > 0$ (see the second term in Eqs. 15 and 19).
- This negative contribution is largest near the front.
- For option 1 (equivalent to a constant k_r inside each element when using the finite element method), when this negative term is big enough, the main diagonal term becomes negative, and the numerical solution breaks down. The use of a full direct linear solver with pivoting might help, but the more fundamental issue is the potentially ill-conditioned linear system of equations that are generated when negative diagonal terms exist. This can be remedied by decreasing Δt or doing some Picard iterations before Newton iterations are done.
- The most robust way to compute k_r is to average respective node values (option 2). This is equivalent in the finite element equations to having k_r vary linearly inside the element. This option does not usually need initial Picard iterations to get the solution close to the final result since the negative contribution is much smaller and therefore the main diagonal terms remain positive.

10 ACKNOWLEDGMENT

This work was supported in part by a grant of computer time from the DoD High Performance Computing Modernization Program.

REFERENCES

[1] F.T. Tracy, "Testing line search techniques for finite element discretizations for unsaturated flow", *Proc. of the 9th Int. Conf. in Computational Science*, Baton Rouge, LA, May 25-27, (2009).

[2] Y. Saad, *Iterative methods for sparse linear systems*, Second Edition, SIAM, pp. 231-234, (2003).

[3] PETSc Development Team, *Portable, Extensible Toolkit for Scientific Computation*, http://www.mcs.anl.gov/petsc/petsc-as/, (2009).

[4] H.V. Nguyen, J.C. Cheng, and R.S. Maier, "Study of parallel linear solvers for three-dimensional subsurface flow problems", *Proc. of the 9th Int. Conf. in Computational Science*, Baton Rouge, LA, May 25-27, (2009).

[5] A.W. Warrick, *Soil Water Dynamics*, Oxford University Press, 168-170, (2003).

[6] W. Gardner, "Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table", *Soil Sci.*, **85**, 228-232, (1958).

[7] A. Szymkiewicz, "Approximation of intermodal conductivities in numerical simulation of one-dimensional infiltration, drainage, and capillary rise in unsaturated soils", *Water Resour. Res.*, **45**, W10403, doi:10.1029/2008WR007654, (2009).

[8] S. Mehl, "Use of Picard and Newton iteration for solving nonlinear ground water flow equations", *Ground Water*, **44**, 583-594, (2006).