SIAM J. SCI. COMPUT.
Vol. 18, No. 5, pp. 1328–1354, September 1997

006

MONOTONICITY CONSIDERATIONS FOR SATURATED–UNSATURATED SUBSURFACE FLOW*

P. A. FORSYTH† AND M. C. KROPINSKI‡

Abstract. It is demonstrated that monotonicity is a sufficient condition to ensure that no new nonphysical local maxima and minima can be produced in the discrete nonlinear unsaturated flow equation. Monotonicity conditions are derived for various types of weighting for the mobility term. The basic discretization is of finite element type, but the results can be extended to finite volume discretizations. Central weightings are only conditionally monotone, while upstream weightings are unconditionally monotone. Sample computations are given for highly heterogeneous problems.

Key words. saturated–unsaturated flow, monotonicity

AMS subject classifications. 65N05, 76S05

PII. S1064827594265824

1. Introduction.

Simulation of trace contaminant migration in the subsurface requires knowledge of the flow of the aqueous phase in both the saturated and unsaturated zones. If the transport of the contaminant in the gas phase is considered unimportant, then the air phase pressure is assumed to be a constant. This results in a highly nonlinear parabolic equation [13, 24, 2, 16, 30, 20, 28, 9].

It is often of interest to simulate subsurface systems which have widely varying material properties. These different materials can have discontinuities in absolute permeabilities of several orders of magnitude. Also, the capillary pressure and relative permeability curves in different materials can be very different. These materials can be the naturally occurring soil or rock formations, or they can be part of an engineered design of a waste storage facility [10, 22, 11, 21, 29].

Although the unsaturated flow equation is parabolic, steep saturation fronts occur during infiltration into dry media. The appearance of these fronts is not surprising when it is recalled that the unsaturated flow equation is in fact an approximation to two-phase flow [5]. Standard finite element methods may produce oscillations near the wetting front [2].

Upstream weightings are commonly used in simulating true two-phase flow. In fact, for true two-phase flow, central weightings may converge to the incorrect solution [1, 25]. However, the saturated–unsaturated flow equation is a parabolic approximation to two-phase flow. For this equation, central-type weightings are commonly used [30]. Although it is well known that oscillations can occur in the discrete solution at coarse grid sizes, convergence to the incorrect solution, as the grid is refined, has never been reported.

The objective of this work is to develop a finite element-type discretization method for saturated–unsaturated flow which will not generate spurious local maxima and minima at coarse (i.e., practical) grid sizes. It will be shown that if certain mono-

*Received by the editors April 5, 1994; accepted for publication (in revised form) January 2, 1996. This work was supported by a grant from the National Sciences and Engineering Research Council of Canada and by the Information Technology Research Center funded by the Province of Ontario.
http://www.siam.org/journals/sisc/18-5/26582.html

†Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (paforsyth@yoho.uwaterloo.ca).

‡Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012 (kropinsk@acf4.nyu.edu).

tonicity conditions are satisfied, then spurious oscillations of a specific type cannot occur in the discrete solution. In order to ensure that these monotonicity conditions are satisfied, it may be necessary to use an upstream-like weighting for the relative permeability terms.

Note that the monotonicity properties of explicit discretizations of hyperbolic conservation law equations are described in [19]. However, the saturated–unsaturated subsurface flow equation is a parabolic equation, which is usually solved using an implicit discretization. This requires somewhat different methods of analysis compared to [19].

In [2] it was demonstrated that mass lumping was beneficial in eliminating nonphysical oscillations. However, [2] suggested the use of central weighting or geometric mean weighting for the relative permeability terms. As will be demonstrated in this paper, central weighting will not be monotone at nodes where the derivative of the capillary pressure as a function of saturation is small.

Computational results will be presented for some one- and two-dimensional problems having discontinuous material properties. It will be shown that use of central weighting may result in local maxima and minima which disappear or become smaller as the grid is refined. These local maxima and minima are not observed in the upstream weighted computations.

2. Formulation. Conservation of the water phase w implies

$$\frac{\partial}{\partial t}\left(\phi S_w \rho_w\right) = -\nabla\cdot\left(\rho_w \mathbf{V}_w\right) + q_w \rho_w, \tag{1}$$

where the velocity of phase w is given by

$$\mathbf{V}_w = -\mathbf{K}\cdot\lambda_w(\nabla P_w - \rho_w g \nabla D), \tag{2}$$

and where

$$\begin{aligned} S_w &= \text{ saturation of water,}\\ P_w &= \text{ pressure of water phase,}\\ \rho_w &= \text{ mass density of water phase,}\\ \mathbf{K} &= \text{ absolute permeability tensor,}\\ \mu_w &= \text{ viscosity of water phase,}\\ k_{rw} &= \text{ relative permeability of water phase,}\\ \lambda_w &= k_{rw}/\mu_w,\\ D &= \text{ depth,}\\ g &= \text{ gravitational acceleration,}\\ q_w &= \text{ source/sink term for water phase.} \end{aligned} \tag{3}$$

The water and air phase pressures are related by the capillary pressure P_{caw},

$$P_a = P_w + P_{caw}(S_w). \tag{4}$$

Assuming that the air phase pressure is constant, then the water phase pressure is given as a function of water saturation S_w from (4). Alternatively, S_w can be regarded as a function of P_w by inverting equation (4):

$$S_w = \min\left(1.0, P_{caw}^{-1}(P_a - P_w)\right). \tag{5}$$

Physically admissible P_{caw} and λ functions are such that

$$\begin{aligned} \frac{\partial S_w}{\partial P_w} &\geq 0 \\ &> 0 \quad \text{if } P_w < P_a, \\ \frac{\partial \lambda}{\partial S_w} &\geq 0, \\ \lambda(S_w = 1) &\neq 0. \end{aligned} \tag{6}$$

The water phase density and porosity are given by

$$\begin{aligned} \rho_w &= \rho_{w0}(1 + c_w(P_w - P_{w0})), \\ \phi &= \phi_0(1 + c_m(P_w - P_{w0})), \\ c_w &\geq 0, \\ c_m &\geq 0, \end{aligned} \tag{7}$$

where ρ_{w0} is the density at pressure P_{w0}, c_w is the water phase compressibility, and c_m is the compressibility of the porous media.

For future reference, we note that equations (6) and (7) imply

$$\begin{aligned} \lambda(P_w^*) &\geq \lambda(P_w^{**}), \\ S_w(P_w^*) &\geq S_w(P_w^{**}), \\ \rho_w(P_w^*) &\geq \rho_w(P_w^{**}), \\ \phi(P_w^*) &\geq \phi(P_w^{**}), \\ &\text{if } P_w^* > P_w^{**}. \end{aligned} \tag{8}$$

3. Discretization. The computational domain R is considered to be composed of a mesh of triangles (in two dimensions) or tetrahedra (in three dimensions). The mesh has n nodes, and linear Lagrange basis functions will be used.

In the case of heterogeneous systems, we assume that the nodal values of the permeability tensor are specified. If the relative permeability and capillary pressure functions are different functions at different points in the domain, then we also assume that these functions are specified at each node (as opposed to each element). For example, the notation

$$\lambda_{wi} = \{\lambda_w(P_{wi})\}_i \tag{9}$$

is to be interpreted as the value of the k_{rw}/μ function at the ith node, evaluated with the pressure value P_{wi}. Nodal representation of material properties results in considerable simplification in the form of the discrete equations. If element values are needed, they will be obtained by averaging the nodal values in the element.

Assume that N_i are the usual linear Lagrange polynomial C^0 basis functions defined on triangles or tetrahedra, where

$$\begin{aligned} N_i &= 1 \text{ at node } i \\ &= 0 \text{ at all other nodes,} \\ \sum_j N_j &= 1 \text{ everywhere in the solution domain } R, \end{aligned} \tag{10}$$

and assume that

$$P_w = \sum_j P_{wj} N_j,$$
$$D = \sum_j D_j N_j,$$
$$S_w = \sum_j S_{wj} N_j. \tag{11}$$

Then, noting that

$$\nabla \sum_j P_{wj} N_j = \sum_{j \neq i} (P_{wj} - P_{wi}) \nabla N_j,$$
$$\nabla \sum_j D_j N_j = \sum_{j \neq i} (D_j - D_i) \nabla N_j, \tag{12}$$

defining

$$V_i = \int_R N_i dv, \tag{13}$$

and using the usual Galerkin procedure (with mass lumping for the accumulation term) and backward Euler timestepping, we obtain

$$\left\{ \left[\phi S_w \rho_w \right]_i^{N+1} - \left[\phi S_w \rho_w \right]_i^{N} \right\} \frac{V_i}{\Delta t}$$
$$= - \sum_{j \in \eta_i} \int_R \left\{ (P_{wj}^{N+1} - P_{wi}^{N+1}) - \rho_w^{N+1} g (D_j - D_i) \right\} \left(\nabla N_j \cdot \mathbf{K} \cdot \nabla N_i \right) \lambda_w^{N+1} \rho_w^{N+1} \, dv$$
$$+ \left(q_w \rho_w \right)_i^{N+1} V_i. \tag{14}$$

Here η_i is the set of neighbor nodes of node i such that the integrals in equation (14) are nonzero.

All boundary conditions are modelled by suitable source/sink terms [6]. The source/sink terms $(q_w \rho_w)_i^{N+1}$ in equation (14) can be used to enforce constant rate injection/production boundary conditions with max/min pressure constraints, or constant pressure boundary conditions (see [4] for details). In order to bound discrete pressures, we must assume that the constant rate injection/production boundary conditions have max/min pressure constraints, and therefore, the worst case would activate these constraints. Consequently, since constant rate injector/producers reduce to constant pressure injector/producers in extreme cases, we can, without loss of generality, consider only the constant pressure injector/producer case in our analysis.

Therefore, all the source/sink terms will be assumed to be of the form (P_{si} is the specified pressure)

$$\left(\rho_w q_w \right)_i^{N+1} V_i = W_{pi}^{+} \rho_{wi}^{N+1} \max \left[\left(P_{si}^{N+1} - P_{wi}^{N+1} \right), 0 \right]$$
$$+ W_{pi}^{-} (\rho_w \lambda_w)_i^{N+1} \min \left[\left(P_{si}^{N+1} - P_{wi}^{N+1} \right), 0 \right]. \tag{15}$$

Here W_{pi}^{+}, W_{pi}^{-} are nonnegative geometric factors independent of P_{wi}, S_{wi} (see [5, 4, 18, 8, 6] for details). Note that if, for example, $W_{pi}^{+} = 0$, then (15) can be used to model a seepage face (interface between air and the porous media)[4].

In order to ensure that the pressure drop is physically correct in hydrostatic situations, the density term in brace brackets under the integral in equation (14) is approximated as the average of the values at nodes i and j, so that

$$\begin{aligned}
&\left\{ \left[\phi S_w \rho_w\right]_i^{N+1} - \left[\phi S_w \rho_w\right]_i^{N} \right\} \frac{V_i}{\Delta t} \\
&= -\sum_{j \in \eta_i} \left((P_{wj}^{N+1} - P_{wi}^{N+1}) - \rho_{w,ij+1/2}^{N+1} g(D_j - D_i) \right) \int_R \left(\nabla N_j \cdot \mathbf{K} \cdot \nabla N_i \right) \lambda_w^{N+1} \rho_w^{N+1} \, dv \\
&\qquad + \left(\rho_w q_w\right)_i^{N+1} V_i \,,
\end{aligned} \tag{16}$$

where

$$\rho_{w,ij+1/2}^{N+1} = \left(\rho_{w,i} + \rho_{w,j}\right)/2. \tag{17}$$

It is computationally efficient to approximate the nonlinear terms in the integrals in equation (16) by an influence coefficient method [15]. There are various ways to do this. The simplest method is to use a single-point quadrature rule using values for $\lambda_w^{N+1}\rho_w^{N+1}$ only on the line joining nodes i and j.

$$\begin{aligned}
-\int_R \left(\nabla N_j \cdot \mathbf{K} \cdot \nabla N_i \right) \lambda_w^{N+1} \rho_w^{N+1} \, dv &\simeq \left(\lambda_w^{N+1}\rho_w^{N+1}\right)_{f(i,j)} \gamma_{ij}, \\
\gamma_{ij} &= -\int_R \nabla N_i \cdot \mathbf{K} \cdot \nabla N_j dv.
\end{aligned} \tag{18}$$

Some common choices for $\left(\lambda_w^{N+1}\rho_w^{N+1}\right)_{f(i,j)}$ are central weighting (*cent*), upstream weighting (*ups*), and geometric [12] weighting. For brevity, we will not discuss geometric weighting in this work. Geometric weighting is discussed in detail in [7]. For notational convenience, we define

$$\psi_j^{N+1} - \psi_i^{N+1} = \left((P_{wj}^{N+1} - P_{wi}^{N+1}) - \rho_{w,ij+1/2}^{N+1} g(D_j - D_i) \right); \tag{19}$$

then the various weightings are given by

$$\begin{aligned}
\left(\lambda_w^{N+1}\rho_w^{N+1}\right)_{ups(i,j)} &= \left(\lambda_w^{N+1}\rho_w^{N+1}\right)_i \text{ if } \gamma_{ij}\left(\psi_j^{N+1} - \psi_i^{N+1}\right) < 0 \\
&= \left(\lambda_w^{N+1}\rho_w^{N+1}\right)_j \text{ if } \gamma_{ij}\left(\psi_j^{N+1} - \psi_i^{N+1}\right) \geq 0, \\
\left(\lambda_w^{N+1}\rho_w^{N+1}\right)_{cent(i,j)} &= \left\{ \left(\lambda_w^{N+1}\rho_w^{N+1}\right)_i + \left(\lambda_w^{N+1}\rho_w^{N+1}\right)_j \right\}/2.
\end{aligned} \tag{20}$$

The final form of the discrete equations, if a single-point quadrature rule as in equation (16) is used, is then

$$\begin{aligned}
&\left\{ \left[\phi S_w \rho_w \ \right]_i^{N+1} - \left[\phi S_w \rho_w\right]_i^{N} \right\} \frac{V_i}{\Delta t} \\
&= \sum_{j \in \eta_i} \left(\rho_w^{N+1}\lambda_w^{N+1}\right)_{f(i,j)} \gamma_{ij} \left(\psi_j^{N+1} - \psi_i^{N+1}\right) \\
&\qquad + \left(\rho_w q_w\right)_i^{N+1} V_i,
\end{aligned} \tag{21}$$

where $f(i,j)$ refers to either *ups(i,j)* or *cent(i,j)*.

For later purposes, it is convenient to write equation (21) as

$$g_i(P_{wi}^{N+1}, P_{wi}^{N}, P_{wj}^{N+1}, P_{si}) = 0, \tag{22}$$

where

$$\begin{aligned} g_i(P_{wi}^{N+1}, P_{wi}^{N}, P_{wj}^{N+1}, P_{si}) = & -\left\{ \left[\ \phi S_w \rho_w \ \right]_i^{N+1} - \left[\ \phi S_w \rho_w \ \right]_i^{N} \right\} \frac{V_i}{\Delta t} \\ & + \sum_{j \in \eta_i} \left(\rho_w^{N+1} \lambda_w^{N+1}\right)_{f(i,j)} \gamma_{ij} \left(\psi_j^{N+1} - \psi_i^{N+1}\right) + \left(\rho_w q_w\right)_i^{N+1} V_i \\ & = 0. \end{aligned} \tag{23}$$

Note that equation (23) has the same form as a finite volume discretization. In this case, V_i would be the cell volume and γ_{ij} would be the interface area between node i and node j divided by the distance between node i and node j.

Another common method for approximating the integral in equation (18) is to use the centroid of each triangle (tetrahedra) which contributes to the integral [14]. Let e_{ij} be the set of elements in the intersection of the support of N_i and N_j. In the usual finite element manner, the integral in equation (16) is evaluated as the sum of integrals over elements E_k in e_{ij}.

$$-\int_R \left(\nabla N_j \cdot \mathbf{K} \cdot \nabla N_i\right) \lambda_w^{N+1} \rho_w^{N+1} \, dv \simeq \sum_{k \in e_{ij}} \gamma_{ij}^{E_k} \left\{ \left(\rho_w \lambda_w\right)_{E_k}^{N+1} \right\}_{centroid},$$

$$\gamma_{ij}^{E_k} = -\int_{E_k} \nabla N_i \cdot \mathbf{K} \cdot \nabla N_j dv, \tag{24}$$

where

$$\left\{ \left(\rho_w \lambda_w\right)_{E_k}^{N+1} \right\}_{centroid} = \sum_{j \in E_k} \left(\rho_w \lambda_w\right)_j / N_{elm}, \tag{25}$$

and where N_{elm} is the number of nodes in the element (three for triangles, four for tetrahedra). It also follows from equations (18) and (24) that

$$\gamma_{ij} = \sum_{k \in e_{ij}} \gamma_{ij}^{E_k}. \tag{26}$$

If the centroidal method (equation (24)) is used to approximate the integral in equation (16), then the final discretized equations can be written in an analogous way to equation (23):

$$\begin{aligned} g_i(P_{wi}^{N+1}, P_{wi}^{N}, P_{wj}^{N+1}, P_{si}) = & -\left\{ \left[\phi S_w \rho_w\right]_i^{N+1} - \left[\phi S_w \rho_w\right]_i^{N} \right\} \frac{V_i}{\Delta t} \\ & + \sum_{j \in \eta_i} \left(\sum_{k \in e_{ij}} \left\{ \left(\rho_w \lambda_w\right)_{E_k}^{N+1} \right\}_{centroid} \gamma_{ij}^{E_k} \right) \left(\psi_j^{N+1} - \psi_i^{N+1}\right) \\ & + \left(\rho_w q_w\right)_i^{N+1} V_i \\ & = 0. \end{aligned} \tag{27}$$

For either discretization (23) or (27), the values of $(P_{wi}^{N+1}, P_{wj}^{N+1})$ are determined by solving the simultaneous equations (22) for all nodes i. For future reference, let

$$\begin{aligned} P_i^{\min} &= \min(P_{wi}^{N}, P_{si}, P_{wj}^{N+1}) \quad \forall j \in \eta_i, \\ P_i^{\max} &= \max(P_{wi}^{N}, P_{si}, P_{wj}^{N+1}) \quad \forall j \in \eta_i. \end{aligned} \tag{28}$$

4. Nonoscillatory solutions. In the following, we will bound the values of P_{wi}^{N+1} (solutions of equations (23) and (27)) in terms of the values of $P_{wi}^{N}, P_{wj}^{N+1}, P_{si}$. This will suffice to determine if new local maxima and minima are possible solutions of the discrete equations. Consequently, for this purpose, we can regard equations (23) and (27) as an implicit equation for P_{wi}^{N+1} with $P_{wi}^{N}, P_{wj}^{N+1}, P_{si}$ as independent variables.

Before proceeding further, it will be convenient to define some terms.

DEFINITION 1 (homogeneous node). *Let the set of nodes which are members of elements having node i as a vertex be denoted by η_i (η_i excludes node i itself). Then node i is a homogeneous node if*

1.

$$\mathbf{K}_j = \mathbf{K}_i \quad \forall j \in \eta_i. \tag{29}$$

2. *All nodes in η_i have the same constitutive functions for relative permeability and capillary pressure as at node i; i.e.,*

$$\begin{aligned} \lambda_j(S_w) &= \lambda_i(S_w), \\ (P_{caw}(S_w))_j &= (P_{caw}(S_w))_i, \\ &\quad \textit{for } S_w \textit{ such that the functions are defined} \\ &\quad \forall j \in \eta_i. \end{aligned} \tag{30}$$

DEFINITION 2 (interior node). *A node is an interior node in the mesh if there exists a (small) sphere surrounding the node such that all points on the surface of the sphere belong to some element in the mesh. To avoid trivial situations, we also require that an interior node i satisfy*

1. $\mathbf{K}_i \neq 0$,
2. $\exists \mathbf{K}_j \neq 0\ ,\ j \in \eta_i$.

An interior node is simply one that is not on the boundary of the domain. The two extra conditions in the above definition simply ensure that the node is not isolated from other nodes if all $\mathbf{K} \equiv 0$.

DEFINITION 3 (monotone discretization). *Let $P_i^{\min}$ and $P_i^{\max}$ be given from equation* (28)*; then the discretization* (23) *or* (27) *is monotone if, for all $(P_{wi}^{N}, P_{wj}^{N+1}, P_{si})$ such that each of $(P_{wi}^{N}, P_{wj}^{N+1}, P_{si})$ is in the interval $[P_i^{\min}, P_i^{\max}]$,*

$$\begin{aligned} &\frac{\partial g_i}{\partial P_{wi}^{N+1}} \neq 0, \\ &\frac{\partial P_{wi}^{N+1}}{\partial P_{wi}^{N}} \geq 0, \\ &\frac{\partial P_{wi}^{N+1}}{\partial P_{wj}^{N+1}} \geq 0 \quad \forall j \in \eta_i, \end{aligned} \tag{31}$$

$$\frac{\partial P_{wi}^{N+1}}{\partial P_{si}} \geq 0. \tag{32}$$

Note that condition (31) *is required by the implicit function theorem to ensure that P_{wi}^{N+1} can be regarded as a differentiable function of $P_{wi}^{N}, P_{wj}^{N+1}, P_{si}$. There is a slight complication in that the above derivatives may not exist at all points (if upstream weighting is used, for example). Let X represent P_{wi}^{N} , P_{wj}^{N+1}, or P_{si}. If at $X = X^*$,*

$$\frac{\partial P_{wi}^{N+1}}{\partial X} \text{ does not exist,}$$

then we interpret the monotonicity condition at $X = X^*$ *as*

$$\lim_{\substack{X \to X^* \\ X > X^*}} \frac{\partial P_{wi}^{N+1}}{\partial X} \geq 0,$$

$$\lim_{\substack{X \to X^* \\ X < X^*}} \frac{\partial P_{wi}^{N+1}}{\partial X} \geq 0. \tag{33}$$

We will show that, for a monotone discretization, no new local maxima or minima of a certain type can appear at interior, homogeneous nodes (i.e. no physically incorrect oscillations in the solution).

More formally, we can state the following theorem.

THEOREM 1 (nonoscillatory solutions). *If* P_{wi}^{N+1} *are determined as solutions of the discrete equations* (27) *or* (23)*, and if node* i *is an interior, homogeneous node where the monotonicity conditions* (32) *are satisfied, then*

$$P_i^{\min} \leq P_{wi}^{N+1} \leq P_i^{\max}. \tag{34}$$

Proof. Since either equation (23) or (27) can be written as

$$g_i(P_{wi}^{N+1}, P_{wi}^{N}, P_{wj}^{N+1}, P_{si}) = 0, \quad j \in \eta_i, \tag{35}$$

and since P_{wi}^{N+1} is a monotone function of $\{P_{wi}^{N}, P_{wj}^{N+1}, P_{si}\}$, then

$$\max\{P_i^{N+1}\} \leq P_i^{upper}, \tag{36}$$

where P_i^{upper} is the solution of

$$g_i(P_i^{upper}, P^*, P^*, P^*) = 0, \qquad P^* = P_i^{\max}, \tag{37}$$

and similarly,

$$\min\{P_{wi}^{N+1}\} \geq P_i^{lower}, \tag{38}$$

where P_i^{lower} is the solution of

$$g_i(P_i^{lower}, P^*, P^*, P^*) = 0, \qquad P^* = P_i^{\min}. \tag{39}$$

We will now demonstrate that

$$g_i(P^*, P^*, P^*, P^*) = 0. \tag{40}$$

Setting

$$P_{wi}^{N+1} = P_{wi}^{N} = P_{wj}^{N+1} = P_{si} = P^* \tag{41}$$

in either equation (23) or (27) and noting that node i is a homogeneous node (equations (29) and (30)), we obtain

$$\begin{aligned} g_i &= -(\rho\lambda)^* g \rho^* \sum_{j \in \eta_i} \gamma_{ij}(D_j - D_i), \\ \rho^* &= \rho(P^*), \\ (\rho\lambda)^* &= \rho(P^*)\lambda(P^*), \end{aligned} \tag{42}$$

all other terms being trivially zero. Now

$$\begin{aligned}\sum_{j\in\eta_i}\gamma_{ij}(D_j-D_i) &= -\sum_{j\in\eta_i}\int_R \nabla N_i\cdot\mathbf{K}\cdot\nabla N_j(D_j-D_i)\,dv\\ &= -\int_R \nabla N_i\cdot\mathbf{K}\cdot\nabla\left(\sum_j D_jN_j\right)\,dv\\ &= -\int_R \nabla N_i\cdot\mathbf{K}\cdot\nabla D\,dv, \qquad (43)\end{aligned}$$

where $\mathbf{K}=\mathbf{K}_i=\mathbf{K}_j$ (since i is a homogeneous node). Recall that D (the depth) is exactly a linear function. Therefore,

$$\begin{aligned}\nabla\cdot(N_i\mathbf{K}\cdot\nabla D) &= \nabla N_i\cdot\mathbf{K}\cdot\nabla D + N_i\nabla\cdot(\mathbf{K}\cdot\nabla D)\\ &= \nabla N_i\cdot\mathbf{K}\cdot\nabla D + 0. \qquad (44)\end{aligned}$$

Let $\mathrm{supp}(N_i)$ be the set of elements such that $N_i\neq 0$, and $\partial\mathrm{supp}(N_i)$ be the surface of $\mathrm{supp}(N_i)$. Then, equation (43) becomes (using equation (44))

$$\begin{aligned}-\int_R \nabla N_i\cdot\mathbf{K}\cdot\nabla D\,dv &= -\int_{\mathrm{supp}(N_i)}\nabla\cdot(N_i\cdot\mathbf{K}\cdot\nabla D)\,dv\\ &= -\int_{\partial\,\mathrm{supp}(N_i)} N_i\cdot\mathbf{K}\cdot\nabla D\cdot\,\vec{ds}\\ &= 0, \qquad (45)\end{aligned}$$

which follows from the fact that node i is an interior node, so that

$$N_i = 0 \ \text{ on } \ \partial\,\mathrm{supp}(N_i). \qquad (46)$$

It then follows from equations (42), (43), (44), and (45) that

$$P_i^{upper} = P^* = P_i^{\max} \qquad (47)$$

is a solution of equation (37), and

$$P_i^{lower} = P^* = P_i^{\min} \qquad (48)$$

is a solution of equation (39). It remains to show that these solutions of equations (37) and (39) are unique. This can be done as follows. Suppose that there exists a solution to

$$g_i(P^{**},P^*,P^*,P^*) = 0 \qquad (49)$$

so that

$$P^{**} > P^*. \qquad (50)$$

Then we can show that in fact

$$g_i(P^{**},P^*,P^*,P^*) \neq 0, \qquad (51)$$

and that therefore equation (49) is false. Similarly, suppose that a solution to equation (49) exists with

$$P^{**} < P^*; \tag{52}$$

then we can show that

$$g_i(P^{**}, P^*, P^*, P^*) \neq 0, \tag{53}$$

and that therefore, equation (49) is false.

This demonstration of uniqueness requires a detailed examination of several different cases. We give the detailed proof for only a few cases. The other cases can be proved in a similar fashion.

Suppose that central weighting is used as in equation (21), or the centroidal method is used as in equation (27). We will consider upstream weighting in a separate case. Suppose

$$P^{**} > P^*; \tag{54}$$

then

$$g_i(P^{**}, P^*, P^*, P^*) = \text{Term1} + \text{Term2} + \text{Term2}, \tag{55}$$

where

$$\text{Term1} = -\left\{ \left[\ \phi S_w \rho_w \ \right]^{**} - \left[\ \phi S_w \rho_w \ \right]^{*} \right\} \frac{V_i}{\Delta t},$$

$$\text{Term2} = w\left[(\rho_w \lambda_w)^*, (\rho_w \lambda_w)^{**}\right] \times$$

$$\left\{ \sum_{j \in \eta_i} \gamma_{ij}(P^* - P^{**}) - (\rho_w^* + \rho_w^{**})/2 \sum_{j \in \eta_i} \gamma_{ij}(D_j - D_i) \right\},$$

$$\text{Term3} = W_{pi}^{-} \left(\rho_w \lambda_w\right)^{**} \left(P^* - P^{**}\right), \tag{56}$$

and where $w(\cdot, \cdot)$ depends on the precise form of the weighting used. Note that

$$w\left[(\rho_w \lambda_w)^*, (\rho_w \lambda_w)^{**}\right] > 0 \quad \text{if} \quad (\rho_w \lambda_w)^* > 0, \quad (\rho_w \lambda_w)^{**} > 0. \tag{57}$$

Since $P^{**} > P^*$, it follows from equations (8) and (15) that

$$\begin{aligned} \text{Term1} &\leq 0, \\ \text{Term3} &\leq 0. \end{aligned} \tag{58}$$

Recall from equations (43) and (45) that

$$\sum_{j \in \eta_i} \gamma_{ij}(D_j - D_i) = 0, \tag{59}$$

so that

$$\text{Term2} = w\left[(\rho_w \lambda_w)^*, (\rho_w \lambda_w)^{**}\right] (P^* - P^{**}) \sum_{j \in \eta_i} \gamma_{ij}. \tag{60}$$

From the definition of the basis functions (10) and the definition of γ_{ij} (18), it follows that

$$\begin{aligned} \sum_{j\in\eta_i} \gamma_{ij} + \gamma_{ii} &= 0, \\ \gamma_{ii} &< 0 \\ \Rightarrow \sum_{j\in\eta_i} \gamma_{ij} &> 0, \end{aligned} \tag{61}$$

where strict inequality follows from the definition of an interior node. Consequently,

$$\text{Term2} \le 0, \tag{62}$$

which means that

$$g_i(P^{**}, P^*, P^*, P^*) \le 0, \quad P^{**} > P^*. \tag{63}$$

To prove strict inequality, we note that there are two cases. If $S_w^* = S_w(P^*) < 1$, then from equation (7), Term1 < 0 in equation (56). If $S_w^* = 1$, then from equations (60) and (61), Term2 < 0. Therefore,

$$g_i(P^{**}, P^*, P^*, P^*) < 0, \quad P^{**} > P^*, \tag{64}$$

and thus there is no solution to $g_i(P^{**}, P^*, P^*, P^*) = 0$ with $P^{**} > P^*$. A similar argument shows that there is no solution to $g_i(P^{**}, P^*, P^*, P^*) = 0$ with $P^{**} < P^*$.

Now consider the case if upstream weighting is used. Assume that there exists a solution to equation (49) with

$$P^{**} < P^*; \tag{65}$$

then equation (23) becomes

$$g_i(P^{**}, P^*, P^*, P^*) = \text{Term1} + \text{Term2} + \text{Term3}, \tag{66}$$

where

$$\begin{aligned} \text{Term1} &= -\left\{ \left[\ \phi S_w \rho_w \ \right]^{**} - \left[\ \phi S_w \rho_w \ \right]^{*} \right\} \frac{V_i}{\Delta t}, \\ \text{Term2} &= \sum_{j\in\eta_i} \left(\rho_w \lambda_w\right)_{ups(i,j)} \gamma_{ij} \left[(P^* - P^{**}) - \frac{(\rho_w^* + \rho_w^{**})}{2} (D_j - D_i) \right], \\ \text{Term3} &= W_{pi}^{+} \left(\rho_w\right)^{**} (P^* - P^{**}). \end{aligned} \tag{67}$$

Since $P^{**} < P^*$, it follows from equations (7) and (15) that

$$\begin{aligned} \text{Term1} &\ge 0, \\ \text{Term3} &\ge 0. \end{aligned} \tag{68}$$

Now consider the expression

$$\begin{aligned} \beta &= -\sum_{j\in\eta_i} \left(\rho_w \lambda_w\right)^{**} \gamma_{ij} \left[(P^* - P^{**}) - \frac{(\rho_w^* + \rho_w^{**})}{2} (D_j - D_i) \right] \\ &= -\left(\rho_w \lambda_w\right)^{**} \left[(P^* - P^{**}) \right] \sum_{j\in\eta_i} \gamma_{ij} \\ &\le 0 \quad \text{if} \ \ P^{**} < P^*, \end{aligned} \tag{69}$$

which follows from equations (59) and (61). Thus

$$\begin{aligned}
&\text{Term2} + \beta \\
&= \sum_{j\in\eta_i} \left\{ (\rho_w\lambda_w)_{ups(i,j)} - (\rho_w\lambda_w)^{**} \right\} \gamma_{ij} \left[(P^* - P^{**}) - \frac{(\rho_w^* + \rho_w^{**})}{2}(D_j - D_i) \right] \\
&= \sum_{j\in\alpha_i} \left\{ (\rho_w\lambda_w)^{*} - (\rho_w\lambda_w)^{**} \right\} \gamma_{ij} \left[(P^* - P^{**}) - \frac{(\rho_w^* + \rho_w^{**})}{2}(D_j - D_i) \right], \qquad (70)
\end{aligned}$$

where α_i is defined as $j \in \eta_i$ such that $j = ups(i,j)$. Recall from the definition of upstream weighting (20) that

$$\gamma_{ij} \left[(P^* - P^{**}) + \frac{(\rho_w^* - \rho_w^{**})}{2}(D_j - D_i) \right] > 0, \quad j = ups(i,j),$$

so that (using equation (8))

$$\begin{aligned}
&\beta + \text{Term2} \geq 0 \\
&\Rightarrow \text{Term2} \geq 0. \qquad (71)
\end{aligned}$$

To prove strict inequality, we again note that there are two cases. If $S_w^{**} < 1$, then from equation (7), Term1 > 0 in equation (67). If $S_w^{**} = 1$, then trivially Term2 > 0 in equation (67). Consequently,

$$g_i(P^{**}, P^*, P^*, P^*) > 0, \quad P^{**} < P^*, \qquad (72)$$

and therefore no solution to $g_i(P^{**}, P^*, P^*, P^*) = 0$ with $P^{**} < P^*$ exists. By a similar argument, no solution exists to $g_i(P^{**}, P^*, P^*, P^*) = 0$ with $P^{**} > P^*$. ∎

5. Summary of nonoscillatory solution result. Theorem 1 essentially states that if the discretization is monotone, then at all interior, homogeneous nodes, no new local maxima and minima which violate condition (34) can be generated as solutions of the discrete algebraic equations.

The requirement that the node be interior and homogeneous is necessary, since new local maxima or minima can be generated at nodes which are on boundaries, or adjacent to material discontinuities. A simple example of this would be a wetting front which encounters an impermeable layer. This would result in water pooling on top of the layer, and hence produce a local maxima in the saturation.

Consequently, if a monotone discretization is used for heterogeneous media, then Theorem 1 implies that any new local maxima/minima which violate condition (34) which appear in the discrete solution can only occur at nodes which lie on the boundary of the computational domain, or at nodes which are adjacent to discontinuities (more precisely, nodes which are not both interior and homogeneous). Assuming that monotone schemes converge to the correct solution of the unsaturated-saturated flow equation as the grid is refined, then new local maxima/minima which appear at interior homogeneous nodes are clearly numerical artifacts, and are not present in the exact solution. In the following, we will refer to these spurious new local maxima/minima as nonphysical.

DEFINITION 4 (nonphysical maxima and minima). *A local maximum or minimum in the discrete solution at timestep $N+1$ which appears at an interior, homogeneous node, which violates condition* (34)*, and was not a local maximum or minimum at timestep N, is a nonphysical maximum or minimum.*

All the above results are also valid for finite volume discretizations that are similar in form to equation (23).

6. Monotonicity conditions. Before determining the monotonicity conditions for various types of discretizations, we shall make some simplifying assumptions. For example, the compressibility of the porous media and the water phase are normally very small, i.e.,

$$\begin{aligned} c_w &= 3.0 \times 10^{-6} (1/\text{kpa}), \\ c_m &= 1.0 \times 10^{-7} (1/\text{kpa}), \end{aligned} \tag{73}$$

and so are frequently ignored. It is possible to continue the analysis with nonzero c_m and c_w, but the algebra becomes involved, and any restrictions on the monotonicity are trivial. For example, a finite compressibility in the gravity term (equation (17)) results in an additional restriction (for monotonicity) that the difference in depth between two nodes must be less than 500 km, which is clearly not very restrictive [17].

Consequently, for ease of exposition, we will assume that

$$\begin{aligned} c_w &= 0, \\ c_m &= 0 \end{aligned}$$

in the following.

6.1. Single-point quadrature discretizations. We first consider the discretizations which use a single-point quadrature rule to evaluate the nonlinear integral term in equation (18), which results in discrete equations having the form (23). It will be convenient to note the following derivatives:

$$\begin{aligned}
\frac{\partial g_i}{\partial P_{wi}^N} &= \left(\rho_w \phi \frac{V_i}{\Delta t} \right) \frac{\partial S_{wi}^N}{\partial P_{wi}^N}, \\
\frac{\partial g_i}{\partial P_{si}} &= \begin{cases} W_{pi}^+ \rho_w & \text{if } P_{si} > P_{wi}^{N+1}, \\ W_{pi}^- \rho_w \lambda_{wi}^{N+1} & \text{if } P_{si} < P_{wi}^{N+1}, \end{cases} \\
\frac{\partial g_i}{\partial P_{wj}^{N+1}} &= \rho_w \gamma_{ij} \left[(P_{wj}^{N+1} - P_{wi}^{N+1}) - \rho_w g (D_j - D_i) \right] \frac{\partial (\lambda_w)_{f(ij)}^{N+1}}{\partial P_{wj}^{N+1}} \\
&\quad + \rho_w \gamma_{ij} (\lambda_w)_{f(ij)}^{N+1}, \\
\frac{\partial g_i}{\partial P_{wi}^{N+1}} &= - \left(\rho_w \phi \frac{V_i}{\Delta t} \right) \frac{\partial S_{wi}^{N+1}}{\partial P_{wi}^{N+1}} - \sum_{j \in \eta_i} \rho_w \gamma_{ij} (\lambda_w)_{f(ij)}^{N+1} \\
&\quad + \sum_{j \in \eta_i} \rho_w \gamma_{ij} \left[(P_{wj}^{N+1} - P_{wi}^{N+1}) - \rho_w g (D_j - D_i) \right] \frac{\partial (\lambda_w)_{f(ij)}^{N+1}}{\partial P_{wi}^{N+1}} \\
&\quad + \begin{cases} -\rho_w W_{pi}^+ & \text{if } P_{si} > P_{wi}^{N+1}, \\ (P_{si} - P_{wi}^{N+1}) \rho_w W_{pi}^- \dfrac{\partial (\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}} & \\ \qquad - \rho_w W_{pi}^- (\lambda_w)_i^{N+1} & \text{if } P_{si} < P_{wi}^{N+1}. \end{cases}
\end{aligned} \tag{74}$$

Since

$$\begin{aligned} \frac{\partial g_i}{\partial P_{wi}^N} &\geq 0, \\ \frac{\partial g_i}{\partial P_{si}} &\geq 0, \end{aligned}$$

then necessary and sufficient conditions for monotonicity are, for all $(P_{wi}^N, P_{wj}^{N+1}, P_{si})$ such that each of $(P_{wi}^N, P_{wj}^{N+1}, P_{si})$ is in the interval $[P_i^{\min}, P_i^{\max}]$,

$$\frac{\partial g_i}{\partial P_{wj}^{N+1}} \geq 0,$$

$$\frac{\partial g_i}{\partial P_{wi}^{N+1}} < 0. \tag{75}$$

Note that if $S_w = 1$ at nodes i, $j \in \eta_i$, then conditions (75) immediately imply that

$$\gamma_{ij} \geq 0, \tag{76}$$

which is simply the M-matrix condition. For a problem with a constant permeability tensor, this is equivalent (in two dimensions) to a Delauney triangulation in the transformed plane where $\mathbf{K} = \mathbf{I}$. For three-dimensional problems, a sufficient condition for tetrahedra which satisfy (76) is that all interior angles in the tetrahedra are nonobtuse (in the transformed plane where $\mathbf{K} = \mathbf{I}$ [6]). Of course, finite volume discretizations satisfy (76) by definition. In the following, we will assume that the mesh satisfies condition (76).

6.1.1. Upstream weighting. If upstream weighting is used, then

$$\begin{aligned}\frac{\partial g_i}{\partial P_{wj}^{N+1}} &= \rho_w \gamma_{ij} \max(\psi_j^{N+1} - \psi_i^{N+1}, 0) \frac{\partial (\lambda_w)_j^{N+1}}{\partial P_{wj}^{N+1}} \\ &\quad + \rho_w \gamma_{ij} \left(\lambda_w^{N+1}\right)_{ups(i,j)}\end{aligned} \tag{77}$$

is always nonnegative. Assuming that node $\mathbf{K}_i \neq 0$ and that one of $\mathbf{K}_j \neq 0$, $j \in \eta_i$, and that condition (76) holds, then

$$\begin{aligned}\frac{\partial g_i}{\partial P_{wi}^{N+1}} &= -\left(\rho_w \phi \frac{V_i}{\Delta t}\right) \frac{\partial S_{wi}^{N+1}}{\partial P_{wi}^{N+1}} - \sum_{j \in \eta_i} \rho_w \gamma_{ij} (\lambda_w)_{ups(i,j)}^{N+1} \\ &\quad + \sum_{j \in \eta_i} \rho_w \gamma_{ij} \frac{\partial (\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}} \min(0, \psi_j^{N+1} - \psi_i^{N+1}) \\ &\quad + \begin{cases} -\rho_w W_{pi}^+ & \text{if } P_{si} > P_{wi}^{N+1}, \\ (P_{si} - P_{wi}^{N+1}) \rho_w W_{pi}^- \dfrac{\partial (\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}} & \\ \quad - \rho_w W_{pi}^- (\lambda_w)_i^{N+1} & \text{if } P_{si} < P_{wi}^{N+1} \end{cases}\end{aligned} \tag{78}$$

is always negative.

Therefore, if upstream weighting (as defined in (20)) is used, then the discretization is unconditionally monotone.

6.1.2. Central weighting. For central weighting, the conditions (75) are

$$\begin{aligned}\frac{\partial g_i}{\partial P_{wj}^{N+1}} &= \rho_w \gamma_{ij} (\psi_j^{N+1} - \psi_i^{N+1}) \frac{1}{2} \frac{\partial (\lambda_w)_j^{N+1}}{\partial P_{wj}^{N+1}} \\ &\quad + \rho_w \gamma_{ij} \left(\lambda_w^{N+1}\right)_{cent(i,j)} \\ &\geq 0\end{aligned} \tag{79}$$

and

$$\begin{aligned}\frac{\partial g_i}{\partial P_{wi}^{N+1}} = &-\left(\rho_w\phi\frac{V_i}{\Delta t}\right)\frac{\partial S_{wi}^{N+1}}{\partial P_{wi}^{N+1}} - \sum_{j\in\eta_i}\rho_w\gamma_{ij}(\lambda_w)_{cent(i,j)}^{N+1}\\ &+\sum_{j\in\eta_i}\rho_w\gamma_{ij}\frac{1}{2}\frac{\partial(\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}}(\psi_j^{N+1}-\psi_i^{N+1})\\ &+\begin{cases}-\rho_w W_{pi}^{+} & \text{if } P_{si} > P_{wi}^{N+1},\\ (P_{si}-P_{wi}^{N+1})\rho_w W_{pi}^{-}\dfrac{\partial(\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}} & \\ \qquad -\rho_w W_{pi}^{-}(\lambda_w)_i^{N+1} & \text{if } P_{si} < P_{wi}^{N+1}\end{cases}\\ &<0.\end{aligned} \tag{80}$$

Equations (79) and (80) are only conditionally satisfied. For example, if $(\psi_j^{N+1} - \psi_i^{N+1}) < 0$; then the first term in equation (79) is negative. Let $\Delta x = \|\mathbf{x}_i - \mathbf{x}_j\|$; then condition (79) is satisfied if

$$(\lambda_w)_{cent(i,j)}^{N+1} \geq \frac{\Delta x}{2}\left|\frac{(\psi_j^{N+1}-\psi_i^{N+1})}{\Delta x}\right|\left(\frac{\partial(\lambda_w)_j^{N+1}}{\partial S_{wj}^{N+1}}\right)\left(\frac{\partial P_{wj}^{N+1}}{\partial S_{wj}^{N+1}}\right)^{-1}. \tag{81}$$

Assuming that $(\psi_j^{N+1} - \psi_i^{N+1})/\Delta x$ approaches a finite limit as the grid is refined, then condition (81) is satisfied for sufficiently small Δx (for nonzero $(\lambda_w)_{cent(i,j)}^{N+1}$). Note that if the derivative of the capillary pressure as a function of saturation is very small, then the diffusion forces are very small, and hence central weighting will be nonmonotone. Over a wide range of saturation values, gravel is an example of a material which has a flat capillary pressure saturation curve [21].

6.2. Centroidal approximation. As for the single-point quadrature methods,

$$\frac{\partial g_i}{\partial P_{wi}^{N}} \geq 0,$$

$$\frac{\partial g_i}{\partial P_{si}} \geq 0.$$

Therefore, if the centroidal approximation is used (equation (27)), then the monotonicity conditions are

$$\begin{aligned}\frac{\partial g_i}{\partial P_{wi}^{N+1}} = &-\left(\rho_w\phi\frac{V_i}{\Delta t}\right)\frac{\partial S_{wi}^{N+1}}{\partial P_{wi}^{N+1}}\\ &-\sum_{j\in\eta_i}\rho_w\left(\sum_{k\in e_{ij}}\gamma_{ij}^{E_k}\sum_{l\in E_k}\frac{(\lambda_w)_l^{N+1}}{N_{elm}}\right)\\ &+\sum_{j\in\eta_i}\rho_w\left(\sum_{k\in e_{ij}}\gamma_{ij}^{E_k}\frac{1}{N_{elm}}\frac{\partial(\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}}\right)(\psi_j^{N+1}-\psi_i^{N+1})\\ &+\begin{cases}-\rho_w W_{pi}^{+} & \text{if } P_{si} > P_{wi}^{N+1},\\ (P_{si}-P_{wi}^{N+1})\rho_w W_{pi}^{-}\dfrac{\partial(\lambda_w)_i^{N+1}}{\partial P_{wi}^{N+1}} & \\ \qquad -\rho_w W_{pi}^{-}(\lambda_w)_i^{N+1} & \text{if } P_{si} < P_{wi}^{N+1}\end{cases}\\ &<0\end{aligned} \tag{82}$$

and

$$\frac{\partial g_i}{\partial P_{wj}^{N+1}} = \rho_w \left(\sum_{k \in e_{ij}} \gamma_{ij}^{E_k} \sum_{l \in E_k} \frac{(\lambda_w)_l^{N+1}}{N_{elm}} \right) + \sum_{\alpha \in \eta_i} \rho_w \left(\sum_{k \in e_{i\alpha}} \gamma_{i\alpha}^{E_k} \sum_{l \in E_k} \frac{\delta_{jl}}{N_{elm}} \frac{\partial (\lambda_w)_l^{N+1}}{\partial P_{wj}^{N+1}} \right) (\psi_\alpha^{N+1} - \psi_i^{N+1}) \geq 0, \tag{83}$$

where δ_{jl} is the Kronecker delta. Note that the first term in equation (83) would be identically zero if all angles opposite the edge from node i to node j are equal to $\pi/2$ (in the transformed plane where $\mathbf{K} = \mathbf{I}$). In this case, it is possible that the second term in equation (83) could be nonzero and negative. In this situation, the discretization would never be monotone, no matter how fine the grid, although the size of the negative derivative would be $O(\Delta x)$. This suggests that a centroidal approximation is inappropriate if bricks or rectangles are broken up into tetrahedra or triangles having angles of size $\pi/2$.

6.3. Summary of monotonicity conditions. The monotonicity results for the various discretizations can be summarized as follows.

- Upstream weighting is unconditionally monotone, regardless of mesh size or timestep size.
- Central weighting is monotone if the grid is sufficiently fine (see condition (81)).
- For certain grid configurations, centroidal weighting is never monotone, no matter how fine the grid.

7. Practical significance of the monotonicity conditions. In the remainder of this paper, we will focus on upstream weighting, which is unconditionally monotone, and one of the conditionally monotone weightings. Since central weighting is commonly used, and can be employed with either finite volume or finite element discretizations, we will consider central weighting as a prototypical, conditionally monotone weighting.

Since upstream weighting is guaranteed to be monotone, we can be certain that an upstream weighted discretization will not produce nonphysical oscillations (i.e., any new local maxima or minima cannot appear at interior homogeneous nodes). However, since monotonicity is a sufficient, not necessary, condition for nonoscillatory solutions, a discretization which is nonmonotone does not necessarily produce nonphysical local maxima and minima. However, we can certainly expect that problems may occur (for nonmonotone discretizations) at nodes near discontinuities.

The previous analysis does not permit us to make any general statements about local maxima and minima which appear at nodes adjacent (or on) material boundaries. In some cases, these local maxima or minima are undoubtedly correct (i.e., water encountering an impermeable layer). However, it is usually considered prudent, when dealing with complex nonlinear PDEs to use monotone schemes, which more closely reflect the underlying physical transport process [23]. If the monotone schemes do not exhibit local maxima at nodes adjacent to material boundaries, while nonmonotone discretizations have local maxima/minima at these nodes, then the nonmonotone results must certainly be viewed with suspicion.

Central-type weightings are widely used for the mobility term in saturated–unsaturated flow [2, 24]. The usual argument for using central weightings is that

upstream-like methods are overly diffusive. However, the mobility term is a highly nonlinear function of saturation, and therefore the wetting front tends to be self-sharpening.

For simple homogeneous problems, where very fine grids are used, there may not be much advantage in using upstream weightings. However, for highly heterogeneous systems, which are typical of layered cover designs for waste facilities, central-type weightings may result in nonphysical oscillations, for practical grid sizes.

Consequently, we would advocate the use of pure upstream weighted discretizations, which are guaranteed to produce physically meaningful results at all grid sizes, for simulating highly heterogeneous systems. For such problems, the risk of using central weighting seems unwarranted. The following examples, where we compare results using central and upstream weightings, will illustrate this point.

8. Numerical examples.

8.1. Computational details. The nonlinear algebraic equations were solved using full Newton iteration, with variable substitution [9]. The material balance errors [2] in all cases were less than 10^{-5} at the end of the runs. The Jacobians were solved using the CGSTAB algorithm [26] with a reduced-system level-one incomplete factorization preconditioning [3].

8.2. Problem 1. Figure 1 shows the domain for this one-dimensional example. The region is 1 m by 1 m by 15 m. The initial pressure was 90 kpa everywhere, and water was injected at a rate of .25 m^3/day at the top. The bottom surface was held at a constant pressure of 100 kpa. Table 1 gives the material properties for this problem. Note that the capillary pressure curve has a small derivative in range $S_w \in [.4, .9]$.

This problem has a simple heterogeneity. In Zone 1 (see Fig. 1), the absolute permeability is $10^{-12}m^2$, while in Zone 2, the permeability is $10^{-13}m^2$. This is actually a very mild form of heterogeneity, compared to those encountered in the modelling of waste cover systems (see Problem 3).

This problem was run using 31, 61, and 121 nodes (one-dimensional linear elements were used). Figures 2 and 3 show the results for the various grid sizes, at a simulated time of 4 days, for both central and upstream weightings.

The results show clearly that the discontinuity in permeability at a height of 7.4 m causes spurious oscillations to appear upstream of the discontinuity for central weighting. Eventually, this oscillation disappears as the grid is refined, as is expected from equation (81). No nonphysical local maxima or minima are generated by the upstream weighted solutions, regardless of grid size.

The coarse grid centrally weighted solutions are clearly unacceptable, while the coarse grid upstream weighted solutions at least capture the qualitative behavior of the solution. At the finest grid level, where the central weighted solution has no oscillations, both methods are in close agreement.

8.3. Problem 2. Figure 4 shows the domain for a two-dimensional problem of infiltration into an initially dry region. This problem is similar to the example used in [9], which was based on a Las Cruces trench field experiment [28]. The initial pressure was -880.0 kpa, and no flow conditions were imposed on the left, right, and bottom boundaries. Infiltration occurred along the top surface as indicated in Figure 4.

Here and in Problem 3, we assume that the capillary pressures and relative permeabilities are of the van Genuchten form [27]:

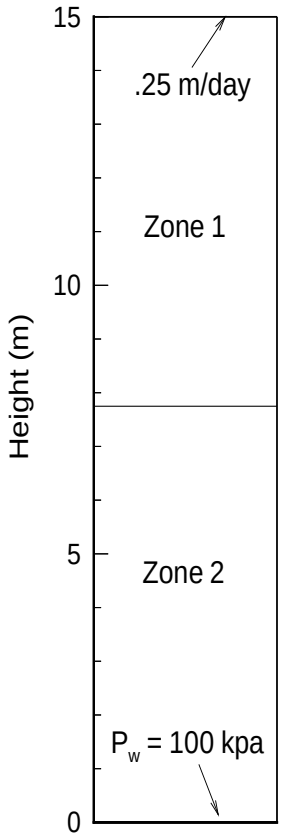

FIG. 1. *Domain for one-dimensional test problem* 1.

TABLE 1
Material properties for Problem 1.

S_w	K_{rw}	P_{caw} (kpa)
.18	0.0	100.0
.2	.0059	50.0
.3	.024	9.9
.4	.071	9.8
.5	.152	9.7
.6	.26	9.6
.7	.40	9.5
.8	.57	9.4
.9	.77	9.3
1.0	1.0	0.0

$$
\begin{aligned}
P_{caw} &= \frac{\rho_w g}{\alpha}\left((\bar{S_w})^{-1/\gamma} - 1\right)^{1/\beta}, \\
\bar{S_w} &= \frac{S_w - S_{wr}}{1 - S_{wr}}, \qquad (84) \\
\gamma &= 1 - \frac{1}{\beta}.
\end{aligned}
$$

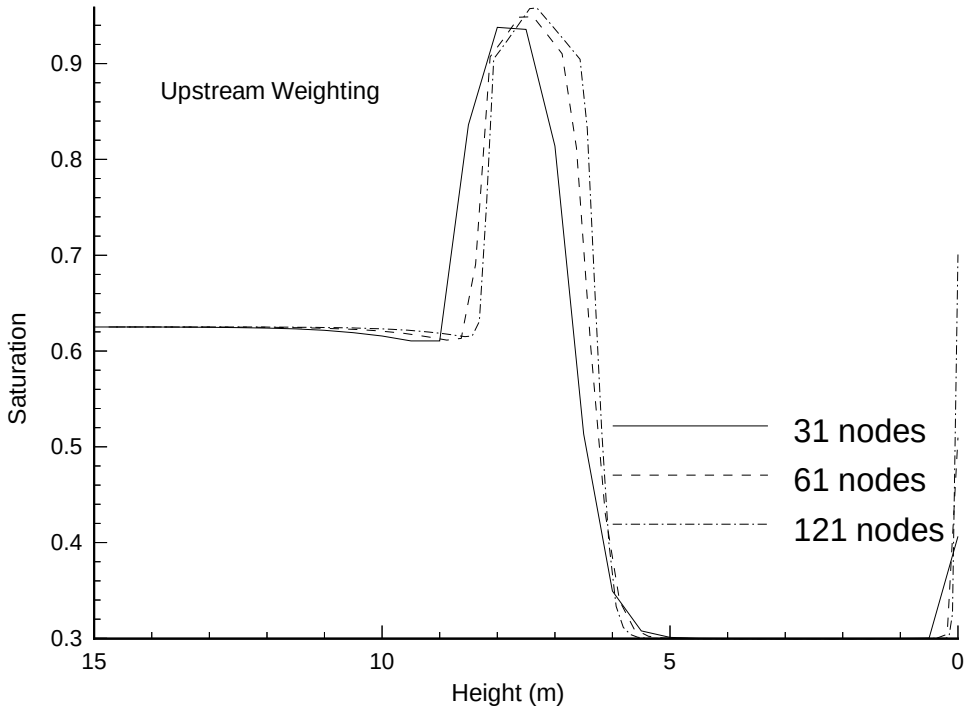

FIG. 2. *Saturation profile for Problem* 1*, upstream weighting, four days.*

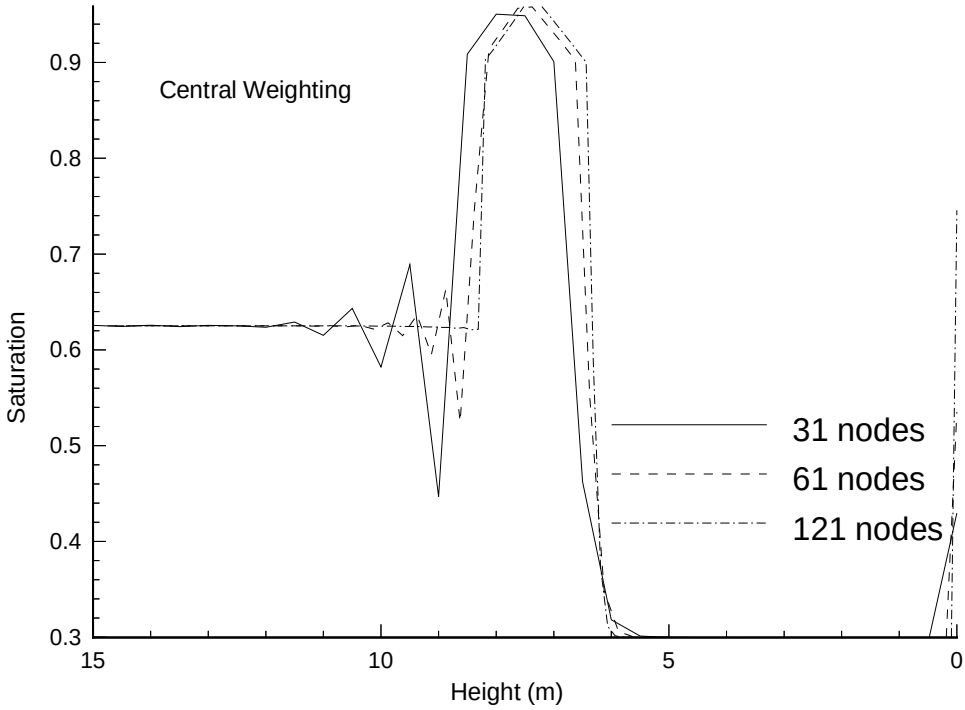

FIG. 3. *Saturation profile for Problem* 1*, central weighting, four days.*

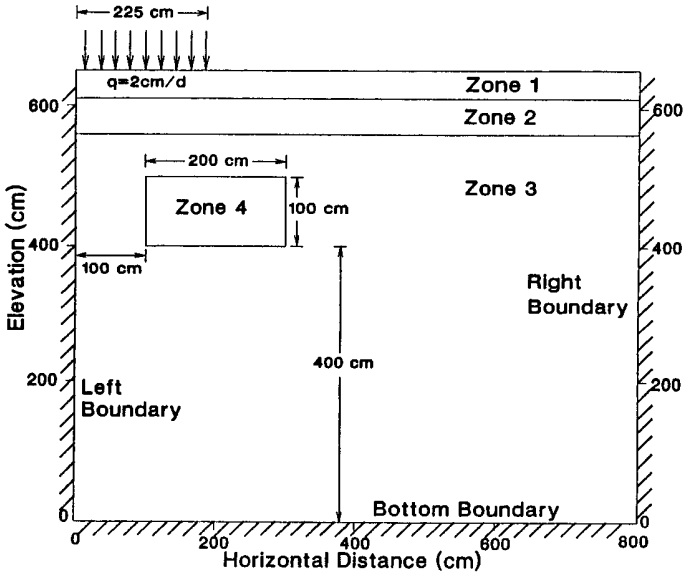

FIG. 4. *Domain for Problem* 2.

TABLE 2
Material properties for Problem 2.

Zone	$K_x = K_y$ (m^2)	ϕ	S_{wr}	α (1/cm)	β
1	9.33×10^{-12}	.3680	.2771	.0334	1.982
2	5.55×10^{-12}	.3510	.2806	.0363	1.632
3	4.898×10^{-12}	.3250	.2643	.0345	5.0
4	4.898×10^{-11}	.3250	.2643	.0345	5.0

The relative permeabilities are given by

$$k_{rw} = (\bar{S_w})^{1/2} \left\{ 1 - \left[1 - (\bar{S_w})^{1/\gamma} \right]^{\gamma} \right\}^2 . \tag{85}$$

The material properties are given in Table 2.

Note that the difference between this problem and the original problem posed in [9, 28] is that we have increased the value of β in Zones 3 and 4. Large values of β make the capillary pressure curve flat at intermediate values of the saturation, and hence we can expect this to cause some difficulties for central weightings.

Figures 5 and 6 show the saturation contours for central weightings for coarse and fine grids (at a simulated time of 30 days). Similar contour maps are given in Figs. 7 and 8 for upstream weighting.

Examination of Fig. 5 demonstrates a striking phenomenon. Figure 5 shows local minima at ($x = 1.0$ m, $z = 1.0$ m) and ($x = 1.0$ m, $z = 2.0$ m). These local minima disappear on the finest grid in Fig. 6, and hence we can conclude that the coarse grid central weighted solutions contain spurious local minima.

Note that these local minima are entirely absent from the upstream weighted results. Qualitatively correct solutions are generated at all grid sizes. For fine grids where the central weightings do not show spurious oscillations, both central and upstream contour maps are in close agreement.

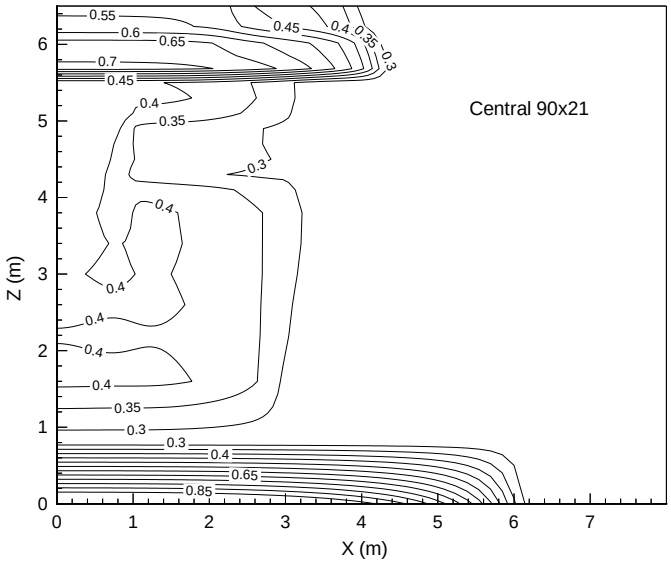

FIG. 5. *Saturation profile for Problem* 2*, central weighting,* 90×21 *grid.*

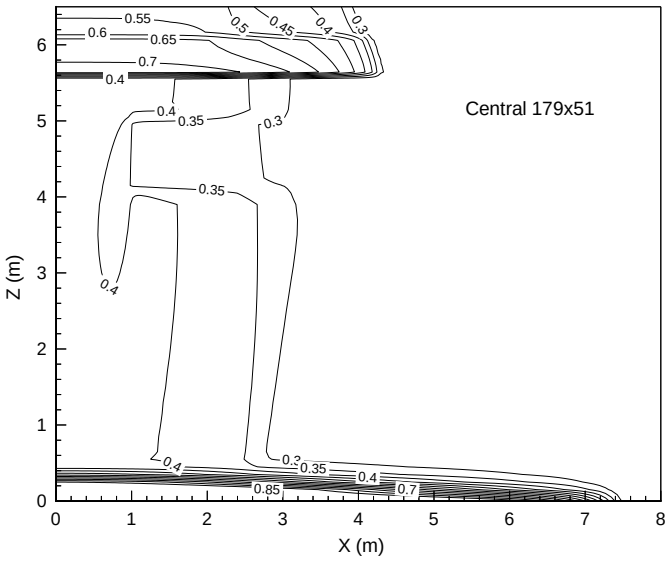

FIG. 6. *Saturation profile for Problem* 2*, central weighting,* 179×51 *grid.*

8.4. Problem 3. Figure 9 shows a two-dimensional model of a capillary barrier waste cover design [10, 22, 11, 21, 29]. The usual EPA-recommended waste design [11] uses a layer of clay to divert water away from the waste. However, clay barriers are difficult to install, cracking can cause the effective permeability of the clay to increase, and the barrier can fail. An alternative is to utilize the differing capillary pressure curves in gravel and sand [22, 11, 21, 29]. Since the capillary pressure in gravel is much lower (at moderate saturations) than the capillary pressure in sand, this will divert water along the gravel–sand interface.

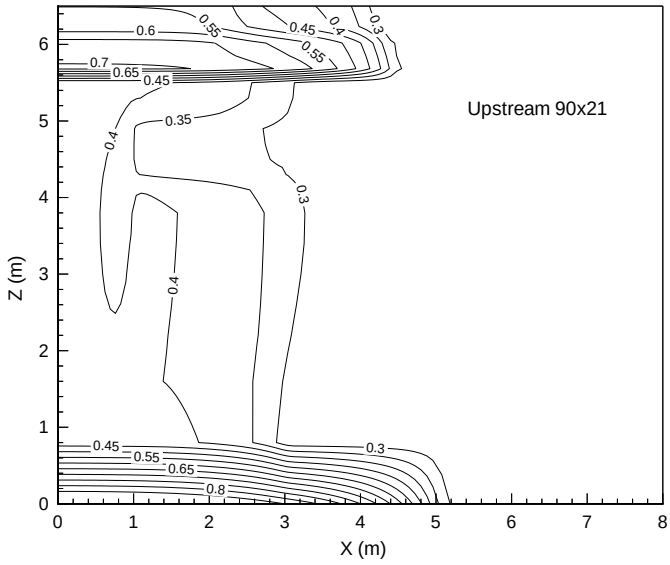

FIG. 7. *Saturation profile for Problem* 2*, upstream weighting,* 90×21 *grid.*

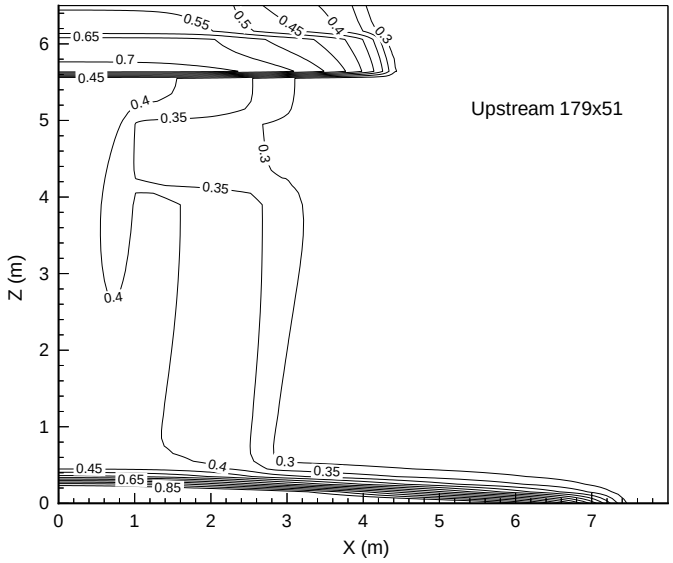

FIG. 8. *Saturation profile for Problem* 2*, upstream weighting,* 179×51 *grid.*

The material properties for this problem are given in Table 3. These properties were obtained from various published sources [11, 21].

No flow boundaries are imposed on the left boundary, constant infiltration is imposed on the top surface, and the right-hand side is no-flow except for a drain (seepage point), as shown in Fig. 9. The actual computational domain is extended (in the vertical direction) much further than is shown in Fig. 9, so that the bottom boundary is effectively at infinity. Initial conditions are water pressure $-1 \times 10^{+6}$ kpa in the loam and sand, -6×10^{10} kpa in the tuff, and 70 kpa in the gravel.

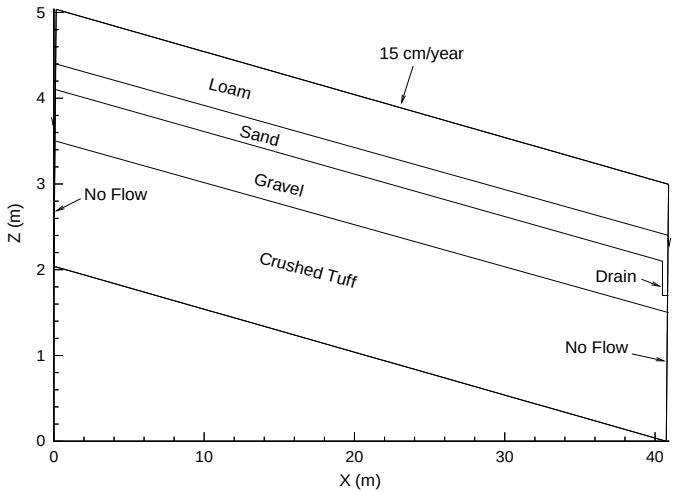

FIG. 9. *Domain for Problem* 3.

TABLE 3
Material properties for Problem 3.

Zone	$K_x = K_y$ (m^2)	ϕ	S_{wr}	α (1/cm)	β
Loam	$.17 \times 10^{-11}$	.452	.0752	.043	1.246
Sand	$.67 \times 10^{-11}$	.345	.046	.0634	1.53
Gravel	$.357 \times 10^{-9}$	.419	.074	4.69	2.57
Crushed tuff	$.283 \times 10^{-12}$	.345	.032	.0143	1.506

This problem was run on two different grid sizes, a coarse 52×46 grid, and a fine 102×92 grid, for both central and upstream weighting. The grid size was highly variable in the z-direction. At the sand–gravel interface, for example, the z-node spacing on the finest grid was .005 m.

To see the differences between upstream and central weighting clearly, the saturation profiles are plotted for only three contour levels. Figures 10, 11, 12, and 13 show the saturation contours at thirty years. Note that in this case, the capillary barrier has clearly failed at about 10 m along the slope.

Figure 12 shows a jagged profile at $x = 16$ m, $z = 3.5$ m. The jagged profile disappears as the grid is refined in Fig. 13. This effect is absent from both coarse and fine upstream results (Figs. 10 and 11).

Also, Fig. 12 shows a local minima at nodes adjacent to the gravel–tuff interface. As the grid is refined in Fig. 13, this area of local minima becomes smaller. Note that this local minima is not seen in the upstream profiles (Figs. 10 and 11).

Since this minima (for central weighting) appears at nodes adjacent to the discontinuity, Theorem 1 does not apply, since we can only say for sure that new local maxima and minima cannot occur for monotone discretizations at interior, homogeneous nodes (see equation (30)). Nevertheless, the fact that the jagged profile in Fig. 12 disappears as the grid is refined, and the observation that the local minima in Figs. 12 and 13 are becoming smaller in size as the grid is refined, while the coarse and fine upstream weighted profiles are qualitatively similar, would appear to indicate that the centrally weighted results are producing spurious local maxima and minima.

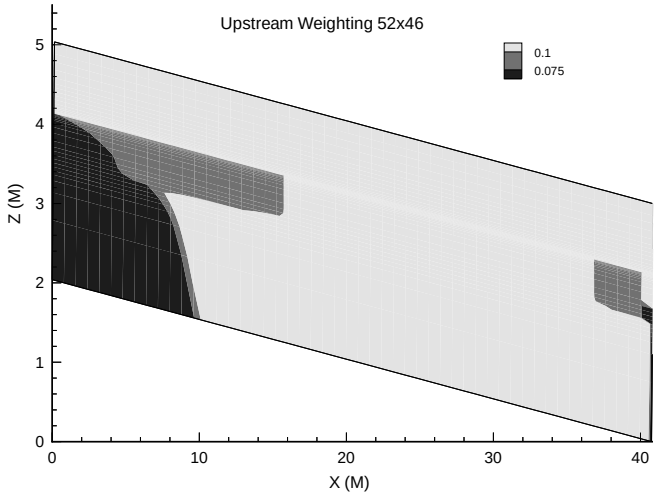

FIG. 10. *Saturation profile, Problem* 3*, upstream weighting* 52×46 *grid.*

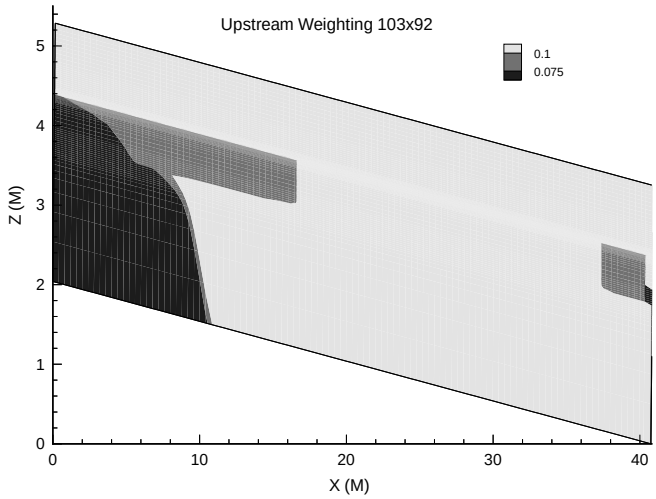

FIG. 11. *Saturation profile, Problem* 3*, upstream weighting* 103×92 *grid.*

From a practical point of view, the performance of the capillary barrier can be summarized in a single number: the fraction of infiltrating water that crosses the gravel–tuff interface. If this number is near zero, then the barrier is very effective, while a number near 1 indicates barrier failure.

Table 4 shows the fraction of water entering the tuff at thirty years for both central and upstream weighting, for both fine and coarse grids. Clearly, the centrally weighted solutions are not converging (at least for this parameter) any faster than the upstream weighted ones. In fact, it would appear that the fine grid centrally weighted result yields similar accuracy (for this parameter) as the coarse grid upstream computation. Moreover, the centrally weighted computations require 3–4 times the number of Newton iterations (and hence 3–4 times the CPU cost) compared to the upstream simulations.

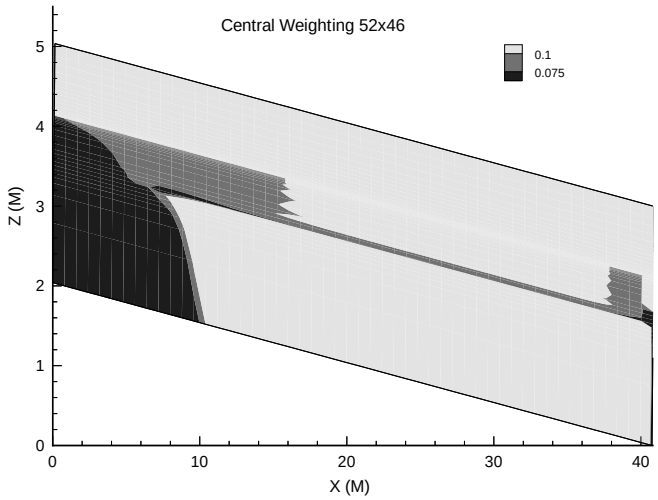

FIG. 12. *Saturation profile, Problem* 3*, central weighting* 52×46 *grid.*

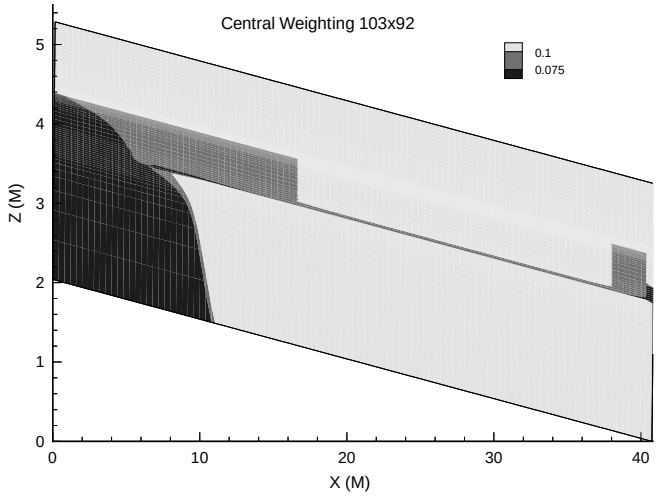

FIG. 13. *Saturation profile, Problem* 3*, central weighting* 103×92 *grid.*

TABLE 4
Comparison of central and upstream weighting, Problem 3.

	Fraction entering tuff		Newton iterations		CPU (sec)	
Grid size	Upstream	Central	Upstream	Central	Upstream	Central
52×46	.461	.488	342	1243	109	380
103×92	.454	.470	788	2463	1057	3137

9. Conclusion. Provided that the finite element mesh has certain properties (i.e., a Delauney triangulation in two dimensions) then an upstream weighting for the mobility term results in a finite element discretization which is unconditionally monotone.

Since the type of finite element discretization used in this work (if a single-point quadrature method is used, as in equation (23)) has the same form as a finite volume discretization, the above result also holds for an unstructured grid, finite volume method as well.

Monotone discretizations can be shown to have the property that no new local maxima or minima of a certain type can appear in the discrete solution at interior, homogeneous nodes, for any mesh size or timestep size.

Central weightings are monotone only for sufficiently fine grids, which implies that under some conditions, nonphysical (see Definition 4) local maxima and minima can occur at finite mesh sizes. A centroidal weighting method, which is natural for finite element discretizations, was also analyzed. There exist certain grid configurations which result in centroidal weighting being never monotone, no matter how fine the mesh.

Example computations have demonstrated that use of central weighting for the mobility for heterogeneous problems often results in spurious local maxima and minima. This effect is particularly severe if the capillary pressure saturation curve contains a range of saturations where the curve is very flat.

Sample computation of a model of a capillary barrier waste cover design, using realistic constitutive data, was also carried out. The centrally weighted computation produces local minima which are not observed in the upstream results. These local minima become smaller (in area) as the grid is refined.

From a practical point of view, the performance of the capillary barrier can be summarized in a single parameter: the fraction of inflowing water that crosses the tuff-gravel interface. In terms of this parameter, the centrally weighted results appear to converge (as the grid is refined) more slowly than the upstream weighted computation. Moreover, the centrally weighted computations require 3–4 times more CPU time than the upstream weighted computation. Although we have given only a single detailed example of this effect in this work, we have seen essentially the same phenomenon many times in dealing with heterogeneous systems.

Consequently, it is our conclusion that monotone schemes (i.e., upstream weightings at coarse grid sizes) should be routinely used when simulating heterogeneous systems. Nonmonotone schemes may have slightly less front smearing compared to monotone methods, but can produce spurious local maxima and minima in the solution.

REFERENCES

[1] K. Aziz and A. Settari, *Petroleum Reservoir Simulation*, Academic Press, New York, 1979.

[2] M. Celia, E. Boultoulas, and R. Zarba, *A general mass-conservative numerical solution method for the unsaturated flow equation*, Water Res. Research, 26 (1990), pp. 1483–1496.

[3] E. D'Azevedo, P. Forsyth, and W. Tang, *Ordering methods for preconditioned conjugate gradient methods applied to unstructured grid problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 944–961.

[4] P. Forsyth, *Comparison of single phase and two phase numerical model formulation for saturated–unsaturated groundwater flow*, Comput. Methods Appl. Mech. Engrg., 69 (1988), pp. 243–259.

[5] P. Forsyth, *Simulation of nonaqueous phase groundwater contamination*, Adv. Water Res., 11 (1988), pp. 74–83.

[6] P. Forsyth, *A control volume finite element approach to NAPL groundwater contamination*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 1029–1057.

[7] P. Forsyth and M. Kropinski, *Monotonicity Considerations for Saturated–Unsaturated Subsurface Flow*, Tech. Rep. CS-94-17, Department of Computer Science, University of Waterloo, Ontario, Canada, 1994.

[8] P. Forsyth and B. Shao, *Numerical simulation of gas venting for NAPL site remediation*, Adv. Water Res., 14 (1991), pp. 354–367.
[9] P. Forsyth, Y. Wu, and K. Pruess, *Robust numerical methods for saturated–unsaturated flow with dry initial conditions in heterogeneous media*, Adv. Water Res., 18 (1995), pp. 25–38.
[10] E. Frind, R. Gilham, and J. Pickens, *Application of unsaturated flow properties in the design of geologic environments for radioactive waste storage facilities*, in Finite Elements in Water Resources, W.G. Gray, G.F. Pinder, C.A. Brebbia, eds., Pentech, London, 1977, pp 3.133–3.163.
[11] T. Hakonson, K. Bostick, G. Trujillo, K. Manies, R. Warren, L. Lane, J. Kent, and W. Wilson, *Hydrologic Evaluation of Four Landfill Cover Designs at Hill Air Force Base, Utah*, Los Alamos National Laboratory Report LA-UR-93-4469, Los Alamos, NM, 1993.
[12] R. Haverkamp and M. Vauclin, *A note on estimating finite difference interblock conductivities for transient unsaturated flow problems*, Water Res. Research, 15 (1979), pp. 181–187.
[13] R. Hills, I. Porro, D. Hudson, and P. Wierenga, *Modeling of one dimensional infiltration into very dry soils* 1. *Model development and evaluation*, Water Res. Research, 25 (1989), pp. 1259–1269.
[14] P. Huyakorn and G. Pinder, *Computational Methods in Subsurface Flow*, Academic Press, New York, 1983.
[15] P. Huyakorn, S. Thompson, and B. Thompson, *Techniques for making finite element methods competitive in modelling flow in variably saturated porous media*, Water Res. Research, 20 (1984), pp. 1099–1115.
[16] M. Kirkland, R. Hills, and P. J. Wieranga, *Algorithms for solving Richards' equation for variably saturated soils*, Water Res. Research, 28 (1992), pp. 2049–2058.
[17] M. Kropinski, *Numerical techniques for saturated unsaturated flow*, Master's Thesis, Dept. of Applied Mathematics, University of Waterloo, Ontario, Canada, 1990.
[18] F. Letniowski and P. Forsyth, *A control volume finite element method for three dimensional NAPL groundwater contamination*, Internat. J. Numer. Methods Fluids, 13 (1991), pp. 955–970.
[19] R. J. LeVeque, *Numerical Methods for Conservation Laws*, Birkhauser, Basel, 1990.
[20] C. Li, *A simplified Newton method with linear finite elements for transient unsaturated flow*, Water Res. Research, 29 (1993), pp. 965–971.
[21] P. Meyer, *Application of an infiltration evaluation methodology to a hypothetical low-level waste disposal facility*, U.S. Nuclear Regulatory Commission Report NUREG/CR-6114, PNL-8842, Washington, DC, 1993.
[22] C. M. Oldenburg and K. Pruess, *On numerical modelling of capillary barriers*, Water Res. Research, 29 (1993), pp. 1045–1056.
[23] S. Patankar, *Numerical Heat Transfer and Fluid Flow*, Hemisphere, Washington, DC, 1980.
[24] P. J. Ross, *Efficient numerical methods for infiltration using Richards' equation*, Water Res. Research, 26 (1990), pp. 279–290.
[25] B. Rubin and P. Sammon, *Practical control of timestep selection in thermal simulation*, SPE12268, in Proc. of 1983 Society of Petroleum Engineers Reservoir Simulation Symposium, San Francisco.
[26] H. van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–645.
[27] M. T. van Genuchten, *A closed form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil Sc. Soc. Am. J., 44 (1980), pp. 892–898.
[28] Y. Wu, J. Kool, and J. McCord, *An evaluation of alternative numerical formulations for two-phase air water flow in unsaturated soils*, in Proc. American Geophysical Union Spring Meeting, Montreal, 1992.
[29] T.-C. Yeh, A. Guxman, R. Srivastava, and P. Gagnard, *Numerical simulation of the wicking effect in liner systems*, Groundwater, 32 (1994), pp. 2–11.
[30] J. Zaidel and D. Russo, *Estimation of finite difference interblock conductivities for simulation of infiltration into initially dry soils*, Water Res. Research, 28 (1992), pp. 2285–2295.