SIAM J. NUMER. ANAL.
Vol. 49, No. 6, pp. 2576–2597

FAST AND ROBUST NUMERICAL SOLUTION OF THE RICHARDS EQUATION IN HOMOGENEOUS SOIL*

HEIKO BERNINGER†, RALF KORNHUBER‡, AND OLIVER SANDER‡

Abstract. We derive and analyze a solver-friendly finite element discretization of a time discrete Richards equation based on Kirchhoff transformation. It can be interpreted as a classical finite element discretization in physical variables with nonstandard quadrature points. Our approach allows for nonlinear outflow or seepage boundary conditions of Signorini type. We show convergence of the saturation and, in the nondegenerate case, of the discrete physical pressure. The associated discrete algebraic problems can be formulated as discrete convex minimization problems and, therefore, can be solved efficiently by monotone multigrid methods. In numerical examples for two and three space dimensions we observe L^2-convergence rates of order $\mathcal{O}(h^2)$ and H^1-convergence rates of order $\mathcal{O}(h)$ as well as robust convergence behavior of the multigrid method with respect to extreme choices of soil parameters.

Key words. saturated-unsaturated porous media flow, Kirchhoff transformation, convex minimization, finite elements, monotone multigrid

AMS subject classifications. 65N30, 65N55, 76S05

DOI. 10.1137/100782887

1. Introduction.

The Richards equation [7, 16, 33] models saturated-unsaturated groundwater flow and reads

$$n\,\theta(p)_t + \operatorname{div}\mathbf{v}(p) = 0\,, \qquad \mathbf{v}(p) = -K_h kr(\theta(p))\nabla(p-z) \tag{1.1}$$

in case of a homogeneous soil. Here, p is the water or capillary pressure on $\Omega\times(0,T)$ for a time $T>0$ and a domain $\Omega\subset\mathbb{R}^3$ inhibited by the porous medium. It can be heterogeneous in the sense that the porosity and the hydraulic conductivity $n:\Omega\to(0,1)$ and $K_h:\Omega\to\mathbb{R}^+$, respectively, may vary in space. The coordinate in the direction of gravity is denoted by z. The saturation $\theta:\mathbb{R}\to[\theta_m,\theta_M]$ with $\theta_m,\theta_M\in[0,1]$ is an increasing function of p which is constant $\theta(p)=\theta_M$—the case of full saturation and ellipticity of (1.1)—if p is sufficiently large. The relative permeability $kr:[\theta_m,\theta_M]\to[0,1]$ is an increasing function of θ with $kr(\theta_M)=1$. It usually leads to a degeneracy in the elliptic-parabolic equation (1.1) by $kr(\theta)\to 0$ for $\theta\to\theta_m$ or even by $kr(\theta_m)=0$ whereby it becomes an ODE.

The soil in (1.1) is homogeneous inasmuch $\theta(\cdot)$ and $kr(\cdot)$ do not depend explicitly on $x\in\Omega$. Concrete forms of these parameter functions are given by Brooks and Corey [14] and van Genuchten [37]. We use the former which are constituted by the bubbling pressure $p_b<0$ and the pore size distribution factor $\lambda>0$ as the soil parameters.

It is a longstanding problem that "most discretization approaches for Richards' equation lead to nonlinear systems that are large and difficult to solve" [22] and that "poor iterative solver performance ... [is] often reported" [27]. Apart from the degeneracy of (1.1) this stems from the fact that the parameter functions degenerate to

*Received by the editors January 19, 2010; accepted for publication (in revised form) August 16, 2011; published electronically December 22, 2011. This work was supported by the BMBF–Programm "Mathematik für Innovationen in Industrie und Dienstleistungen."
http://www.siam.org/journals/sinum/49-6/78288.html
†Section de Mathématiques, Université de Genève, 2-4 rue du Lièvre, CH-1211 Genève, Switzerland (Heiko.Berninger@unige.ch).
‡Institut für Mathematik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany (kornhuber@math.fu-berlin.de, sander@math.fu-berlin.de).

step functions for extreme soil parameters. Therefore, it is necessary for robustness to refrain from linearizing the Richards equation (1.1) in the (iterative) solution process. To the best of our knowledge there are no numerical approaches to the Richards equation in the literature that meet this requirement. For example, although several different discretizations are used in Wagner et al. [38], Fuhrmann [24], Schneid, Knabner, and Radu [34], and Bastian et al. [6], all these authors apply Newton's method to the resulting finite-dimensional system.

In this paper we strive for robustness of the numerical solution with respect to soil parameters appearing in saturation $\theta(p)$ and relative permeability $kr(\theta)$, respectively, particularly of the algebraic solver for the arising large-scale, highly nonlinear spatial problems. Following Alt and Luckhaus [1] and Visintin [2], our approach is based on a Kirchhoff transformation of the physical pressure into a generalized pressure $u = \kappa(p)$. In this way the nonlinearity and the degeneracy are removed from the main part of the differential operator. Incorporating Signorini-type boundary conditions occurring, e.g., around seepage faces at the bank of a lake, the transformed problem can be formulated in a weak sense as a variational inequality involving the monotonically increasing nonlinearity $u \mapsto \theta(\kappa^{-1}(u))$.

By a time discretization in which only the gravitational term is treated explicitly, one obtains elliptic variational inequalities that are equivalent to strictly convex minimization problems. The spatial discretization is carried out by piecewise linear finite elements. Upwinding of the gravitational (i.e., convective) part guarantees stability for sufficiently small time steps (see Berninger [9, sec. 4.2] and the careful discussion in Forsyth and Kropinski [23]). We prove H^1-convergence of the finite element approximations u_j to the generalized pressure.

The discretization in generalized variables u_j can be reinterpreted as a standard finite element discretization of the original Richards equation (1.1) in physical pressure p_j with numerical integration based on particular (solution dependent) quadrature points. If the Richards equation is nondegenerate, we obtain H^1-convergence of $\kappa^{-1}(u_j)$ and L^2-convergence of its piecewise linear interpolation p_j to the physical solution of the time discrete problem. Similar convergence results are obtained for the discrete saturation $\theta(\kappa^{-1}(u_j))$.

Our new approach pays off in two regards. First, the ill-conditioning inherent in the degenerate problem is decoupled from the solution process and appears only in the inverse Kirchhoff transformation $u \mapsto p = \kappa^{-1}(u)$ after u has been determined. Second, the discretization is solver-friendly in the sense that there are fast and robust monotone multigrid solvers [26, 30] at hand for the large-scale algebraic problems occurring in each time step. Monotone multigrid methods are based on successive minimization rather than linearization. Therefore, they perform robustly even for nonsmooth nonlinearities. Moreover, these methods are fast in the sense that for good initial iterates, as obtained from the preceding time step or by nested iteration, they have a similar convergence speed as standard linear multigrid methods applied to linear self-adjoint problems. This is confirmed by asymptotic logarithmic upper bounds for the convergence rates established in [30].

By nonlinear domain decomposition techniques and monotone multigrid methods as local solvers our approach has been extended to heterogeneous soils that consist of different layers of homogeneous soil in the doctoral thesis of Berninger [9, Chapter 3]. For related work, we also refer to [11, 13] and [12].

Outline. In section 2 we first introduce the Brooks–Corey parameter functions and the Kirchhoff transformation. Then we give a weak formulation of a Signorini-type boundary value problem for the Richards equation as a variational inequality.

In section 3 we present our implicit-explicit time discretization and show that the resulting variational inequality is equivalent to a uniquely solvable convex minimization problem.

In section 4 we introduce a finite element discretization of the convex minimization problem that provides H^1-convergence. We also give a reinterpretation as a standard finite element discretization with numerical integration of the problem in physical pressure. Convergence results for the discrete saturation and the discrete physical pressure are derived.

Section 5 illustrates our theoretical reasoning by numerical experiments. First, we investigate the spatial discretization error. Both for transformed and physical variables we numerically obtain the order of convergence $\mathcal{O}(h^2)$ in the L^2-norm and $\mathcal{O}(h)$ in the H^1-norm. We both illustrate and analyze how the ill-conditioning of the inverse Kirchhoff transformation affects the numerical calculations. Then we consider a gravity dominated infiltration problem into almost dry soil in order to investigate the stability constraint on the time step as resulting from our explicit upwind discretization of the convective gravitational term. Finally, we apply our discretization to an infiltration problem for an almost dry dam in three space dimensions leading to large-scale spatial problems for each time step. For all the large-scale, highly nonlinear spatial problems occurring throughout the evolution we observe a similar convergence speed of our algebraic monotone multigrid solver as for the linear Darcy flow arising after full saturation. Moreover, fast convergence is preserved for a wide range of soil parameters confirming the robustness of our approach.

2. Signorini-type problem and variational inequality for the Richards equation. We give concrete forms of the parameter functions $p \mapsto \theta(p)$ and $\theta \mapsto kr(\theta)$ according to Brooks and Corey. Then we apply the Kirchhoff transformation to the Richards equation and introduce our boundary value problem in a strong form. Finally, we develop a weak formulation of a boundary value problem for the Richards equation with nonlinear outflow conditions of Signorini-type.

2.1. Brooks–Corey parameter functions. Let the residual and the maximal saturation $\theta_m, \theta_M \in [0,1]$, $\theta_m < \theta_M$, as well as the bubbling pressure $p_b < 0$ and the pore size distribution factor $\lambda > 0$ be given. Then, by Brooks and Corey [14] the saturation θ is given by

$$\theta(p) = \begin{cases} \theta_m + (\theta_M - \theta_m)\left(\frac{p}{p_b}\right)^{-\lambda} & \text{for } p \le p_b\,, \\ \theta_M & \text{for } p \ge p_b\,, \end{cases} \tag{2.1}$$

and (with results by Burdine [15]) the relative permeability kr reads

$$kr(\theta) = \left(\frac{\theta - \theta_m}{\theta_M - \theta_m}\right)^{3+\frac{2}{\lambda}}, \quad \theta \in [\theta_m, \theta_M]\,. \tag{2.2}$$

Typical shapes of these nonlinearities are depicted in Figures 2.1 and 2.2. Their essential properties are collected in the following lemma.

LEMMA 2.1. *The Brooks–Corey functions θ and kr in* (2.1) *and* (2.2) *are nonnegative, bounded, monotonically increasing, and continuous.*

2.2. Kirchhoff transformation. In the following we assume $n = K_h = 1$ for simplicity and thus deal with the Richards equation in the form

$$\theta(p)_t - \operatorname{div}\Big(kr(\theta(p))\nabla\,(p - z)\Big) = 0\,. \tag{2.3}$$

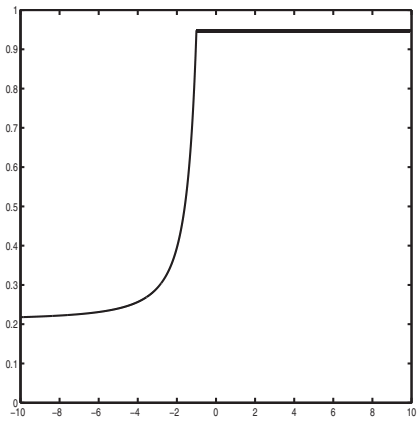

FIG. 2.1. $p \mapsto \theta(p)$.

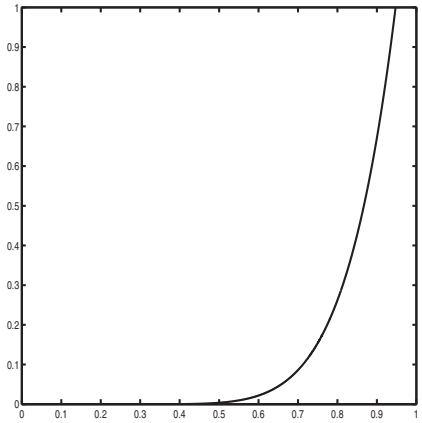

FIG. 2.2. $\theta \mapsto kr(\theta)$.

An essential well-known tool for our approach is Kirchhoff's transformation; see, for example, Alt, Luckhaus, and Visintin [2] or Eymard, Gutnic, and Hilhorst [21]. It is defined by

$$\kappa : p \mapsto u := \int_0^p kr(\theta(q))\, dq\,, \tag{2.4}$$

where u shall be called *generalized pressure*. If we take the chain rule

$$\nabla u = kr(\theta(p))\nabla p \tag{2.5}$$

into account and define

$$M(u) := \theta(\kappa^{-1}(u)) \tag{2.6}$$

and $e_z := \nabla z$, the transformed Richards equation (2.3) reads

$$M(u)_t - \operatorname{div}\Big(\nabla u - kr(M(u))e_z\Big) = 0\,. \tag{2.7}$$

Hence we obtain a semilinear equation from the quasilinear equation (2.3). In case of the Brooks–Corey parametrization, M has unbounded derivatives and κ^{-1} is ill-conditioned around a critical pressure u_c; compare Figures 2.3 and 2.4.

LEMMA 2.2. *Let θ and kr satisfy the properties in Lemma* 2.1. *Then M defined by* (2.6) *is nonnegative, bounded, monotonically increasing, and continuous. Furthermore, $\kappa : \mathbb{R} \to \mathbb{R}$ is monotonically increasing and in $C^1(\mathbb{R})$.*

Let θ and kr be chosen according to (2.1) *and* (2.2). *Then we have $u = p$ for $p \geq p_b$ and $p \leq 0 \Leftrightarrow u \leq 0$. Furthermore, $\lim_{p\to-\infty} \kappa(p) =: u_c < 0$ exists and, with $M(u_c) := \theta_m$, the function M is defined on $[u_c, \infty)$ and Hölder continuous.*

Let $kr \in L^\infty(\theta(\mathbb{R}))$ in the nondegenerate case

$$kr(\cdot) \geq c \quad \text{for a } c > 0\,. \tag{2.8}$$

Then both κ and κ^{-1} are Lipschitz continuous functions on $\mathbb{R}$, and if, in addition, θ is Lipschitz continuous on $\mathbb{R}$ (as in (2.1)*), so is M.*

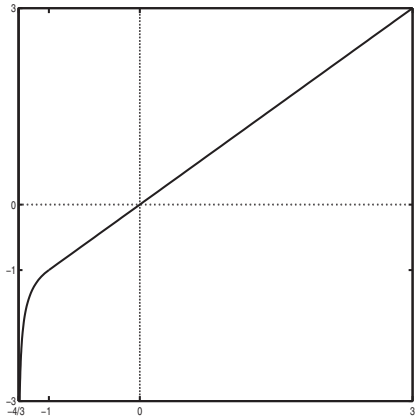

FIG. 2.3. $u \mapsto \kappa^{-1}(u)$.

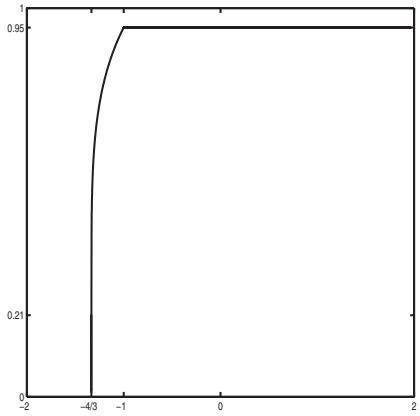

FIG. 2.4. $u \mapsto M(u)$.

Now we give our minimal set of global assumptions for this paper, which by Lemma 2.2 are satisfied in the Brooks–Corey case. However, all following considerations apply analogously if $[u_c, \infty)$ is replaced by $\mathbb{R}$.

Assumption 2.3. Let $kr : M(\mathbb{R}) \to \mathbb{R}^+$ be a bounded Borel function and $M : [u_c, \infty) \to \mathbb{R}$, $u_c \in \mathbb{R}$, be bounded, monotonically increasing, and continuous.

2.3. Signorini-type boundary value problem. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. For a time $t \in (0, T]$ we assume that a decomposition of $\partial\Omega$ into nonempty submanifolds $\gamma_D(t)$, $\gamma_N(t)$, and $\gamma_S(t)$ as well as functions $u_D(t) \in H^{1/2}(\gamma_D(t))$ and $f_N(t) \in L^2(\gamma_N(t))$ are given. Then for u and the flux

$$\mathbf{v} = -(\nabla u - kr(M(u))e_z) = -kr(\theta(p))\nabla\,(p - z)$$

we consider the boundary conditions

$$u = u_D(t) \quad \text{on } \gamma_D(t)\,, \tag{2.9}$$

$$\mathbf{v} \cdot \mathbf{n} = f_N(t) \quad \text{on } \gamma_N(t)\,, \tag{2.10}$$

$$u \le 0\,, \quad \mathbf{v} \cdot \mathbf{n} \ \ge\ 0\,, \quad u \cdot (\mathbf{v} \cdot \mathbf{n}) = 0 \quad \text{on } \gamma_S(t)\,. \tag{2.11}$$

The complementarity conditions (2.11) are sometimes called outflow conditions [35] or seepage face conditions [8, 18]. They lead to a free boundary value problem as they model possible unrestricted outflow on and close to seepage faces, which can be found, e.g., at the bank around lakes or in dam problems; see, e.g., [17]. Since they are known from Signorini problems [28], we call them *Signorini-type boundary conditions* (cf. [36, 39]) and refer to the corresponding boundary value problem (2.7), (2.9)–(2.11) as a *Signorini-type problem* for the (Kirchhoff-transformed) Richards equation.

2.4. Variational inequality. As a result of (2.11) we obtain a variational inequality on a convex subset of the space $H^1(\Omega)$ as a weak formulation of (2.7), (2.9)–(2.11). For a justification we refer to an equivalence result that holds for smooth functions [9, Prop. 1.5.3].

We introduce some notation. Let $\gamma \subset \partial\Omega$ be a nonempty submanifold. We call $tr_\gamma : H^1(\Omega) \to H^{1/2}(\gamma)$ the corresponding trace operator. With the decomposition of $\partial\Omega$ and the functions $u_D(t)$ and $f_N(t)$ given above we set

$$\mathcal{K}(t) := \{v \in H^1(\Omega) : v \ge u_c \wedge tr_{\gamma_D(t)} v = u_D(t) \wedge tr_{\gamma_S(t)} v \le 0\}\,, \tag{2.12}$$

which is a nonempty, closed, and convex subset of $H^1(\Omega)$ if $u_D(t)$ is chosen to be compatible with the other conditions constituting $\mathcal{K}(t)$; see [9, Prop. 1.5.5].

Given Assumption 2.3, we say that $u \in L^2(0,T;H^1(\Omega))$, with the property $M(u)_t \in L^2(\Omega)$ a.e. on $(0,T]$, is a weak solution of (2.7), (2.9)–(2.11) at the time $t \in (0,T]$ if

(2.13)
$$u(t) \in \mathcal{K}(t): \int_\Omega M(u(t))_t\,(v-u(t))\,dx + \int_\Omega \nabla u(t)\,\nabla(v-u(t))\,dx \geq \int_\Omega kr(M(u(t)))e_z\nabla(v-u(t))\,dx - \int_{\gamma_N(t)} f_N(t)\,(v-u(t))\,d\sigma \quad \forall v \in \mathcal{K}(t)\,.$$

It is possible to relate this variational inequality to a corresponding one given in the physical pressure $p(t)$ for the original Richards equation (2.3) with the boundary values (2.9)–(2.11) retransformed in physical variables. More concretely, $u(t)$ solves (2.13) if $p(t)$ solves the corresponding variational inequality, and, in case (2.8) and $\gamma_S(t) = \emptyset$, both formulations are equivalent (see [9, sec. 1.5.4] or [10]).

3. Implicit-explicit time discretization and convex minimization. In the following we give our implicit-explicit time discretization of the variational inequality (2.13). Our aim in this section is to derive an equivalent uniquely solvable convex minimization problem from the resulting variational inequality.

3.1. Time discretization. Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of $[0,T]$ with the time step sizes $\tau_n := t_n - t_{n-1}$, $n \in \{1,\dots,N\}$, and set $u^0 = u(0) \in H^1(\Omega)$ as the given initial condition for (2.7). Without loss of generality we set $f_N(t) = 0$. Then, successively for $n = 1,\dots,N$, our time discretized version of (2.13) reads

(3.1)
$$u^n \in \mathcal{K}(t_n): \int_\Omega M(u^n)\,(v-u^n)\,dx + \tau_n \int_\Omega \nabla u^n \nabla(v-u^n)\,dx \geq \int_\Omega M(u^{n-1})\,(v-u^n)\,dx + \tau_n \int_\Omega kr(M(u^{n-1}))e_z\nabla(v-u^n)\,dx \quad \forall v \in \mathcal{K}(t_n)\,.$$

We proceed with some notation and abbreviations. For a given $n \in \{1,\dots,N\}$ we set $\mathcal{K} := \mathcal{K}(t_n)$ and also $\gamma_D := \gamma_D(t_n)$, $\gamma_S := \gamma_S(t_n)$ and $\gamma_N := \gamma_N(t_n)$ as well as $u_D := u_D(t_n)$, and we denote $u = u^n$. We abbreviate the norm in the Sobolev space $H^1(\Omega)$ by $\|\cdot\|_1$ and define the subspace

$$H^1_{\gamma_D}(\Omega) := \{v \in H^1(\Omega) : tr_{\gamma_D} v = 0\}\,.$$

The left-hand side in (3.1) is given by a continuous linear functional ℓ on $\mathcal{K} \subset H^1(\Omega)$ defined as

$$\ell(v) := \int_\Omega M(u^{n-1})\,v\,dx + \tau_n \int_\Omega kr(M(u^{n-1}))e_z\nabla v\,dx \quad \forall v \in H^1(\Omega)\,. \tag{3.2}$$

Since $\gamma_D \subset \partial\Omega$ is a nonempty submanifold, the continuous symmetric bilinear form $a(\cdot,\cdot)$ on $H^1(\Omega)$ given by

$$a(v,w) := \tau_n \int_\Omega \nabla v \nabla w\,dx \quad \forall v,w \in H^1(\Omega) \tag{3.3}$$

is coercive on $H^1_{\gamma_D}(\Omega)$. With this notation we can write (3.1) more compactly as the variational inequality

$$u \in \mathcal{K} : \int_\Omega M(u)(v-u)\,dx + a(u, v-u) - \ell(v-u) \geq 0 \quad \forall v \in \mathcal{K}\,. \tag{3.4}$$

3.2. Convex minimization. The variational inequality (3.4) is equivalent to a convex minimization problem. In what is to come we sketch the reasoning to derive this fact and refer to [9, sec. 2.3.2–2.3.4] for proofs and further details. We start with a primitive $\Phi : [u_c, \infty) \to \mathbb{R}$ of M defined as

$$\Phi(z) := \int_0^z M(s)\,ds \quad \forall z \in [u_c, \infty) \tag{3.5}$$

which gives rise to a functional $\phi : \mathcal{K} \to \mathbb{R}$ by

$$\phi(v) := \int_\Omega \Phi(v(x))\,dx \quad \forall v \in \mathcal{K}\,. \tag{3.6}$$

Since M is monotonically increasing, Φ is convex, and since M is bounded, Φ is Lipschitz continuous. Therefore, ϕ is a well-defined convex and Lipschitz continuous functional with an affine lower bound

$$\phi(v) \geq -c_1\|v\|_1 - c_2 \quad \forall v \in \mathcal{K}$$

for c_1, $c_2 > 0$. Furthermore, since M is continuous, Φ is differentiable with $\Phi' = M$.

Recall that for a function $F : S \to \mathbb{R}$ on a subset $S \subset V$ of a normed space V the one-sided limit

$$\partial_v F(u) := \lim_{h \downarrow 0} \frac{F(u+hv) - F(u)}{h}\,, \quad u, u+hv \in S\,,$$

if it exists, is the directional derivative of F at u in the direction of $v \in V$.

Since Φ is convex and differentiable, one can interchange differentiation with the integral in (3.6) and obtain the following result.

LEMMA 3.1. *For any $u, v \in \mathcal{K}$ the directional derivative $\partial_{v-u}\phi(u)$ exists and can be written as*

$$\partial_{v-u}\phi(u) = \int_\Omega \Phi'(u(x))(v(x)-u(x))\,dx = \int_\Omega M(u(x))(v(x)-u(x))\,dx\,.$$

It is well known that the quadratic functional $\mathcal{J} : H^1_{\gamma_D}(\Omega) \to \mathbb{R}$ defined by

$$\mathcal{J}(v) := \frac{1}{2}a(v,v) - \ell(v) \quad \forall v \in H^1_{\gamma_D}(\Omega) \tag{3.7}$$

is strictly convex, continuous, and coercive. Moreover, $\mathcal{J}$ is Fréchet-differentiable in $u \in H^1_{\gamma_D}(\Omega)$ with the derivative

$$\mathcal{J}'(u)(v) = \partial_v \mathcal{J}(u) = a(u,v) - \ell(v) \quad \forall v \in H^1_{\gamma_D}(\Omega)\,.$$

Consequently, the functional $F : \mathcal{K} \to \mathbb{R}$ defined by

$$F(v) := \phi(v) + \mathcal{J}(v) \quad \forall v \in \mathcal{K} \tag{3.8}$$

(and extended by $+\infty$ on $H^1_{\gamma_D}(\Omega)\backslash\mathcal{K}$) is strictly convex, proper, continuous, and coercive, and $\partial_{v-u}F(u)$ exists for any $u, v \in \mathcal{K}$. Altogether, we conclude that (3.4) has the form

$$u \in \mathcal{K} : \quad \partial_{v-u}F(u) \geq 0 \quad \forall v \in \mathcal{K}\,.$$

The next result provides the link to convex minimization.

LEMMA 3.2. *Let V be a real vector space, let $K \subset V$ be a convex set, and let $F : K \to \mathbb{R}$ be a convex functional whose directional derivative $\partial_{v-u}F(u)$ exists for all $u, v \in K$. Then*

$$u \in K : \quad \partial_{v-u}F(u) \geq 0 \quad \forall v \in K$$

is equivalent to

$$u \in K : \quad F(u) \leq F(v) \quad \forall v \in K\,.$$

Now we can apply a well-known existence and uniqueness result for convex minimization problems (see, e.g., [20, p. 35]) to obtain the main result of this section.

THEOREM 3.3. *Let $\mathcal{K} \subset H^1(\Omega)$, $a(\cdot,\cdot)$, and $\ell(\cdot)$ be defined as in* (2.12), (3.3), *and* (3.2), *respectively. Then, with Assumption* 2.3, *the variational inequality* (3.4) *has a unique solution. More specifically, it is equivalent to the minimization problem*

$$u \in \mathcal{K} : \quad \mathcal{J}(u) + \phi(u) \leq \mathcal{J}(v) + \phi(v) \quad \forall v \in \mathcal{K} \tag{3.9}$$

with $\mathcal{J}$ and ϕ as defined in (3.7) *and* (3.6), *respectively.*

Theorem 3.3 can be generalized to the case of nonnegative and bounded porosity $n(\cdot)$ and hydraulic conductivity $K_h(\cdot)$ satisfying

$$c \leq K_h(\cdot) \leq C \quad \text{with some } c, C > 0\,. \tag{3.10}$$

4. Finite element discretization. In this section we present a finite element discretization of (3.9), which extends the results in [29, pp. 36–43] to our more general boundary conditions. We give a reinterpretation as a certain finite element discretization of the problem in physical variables, thus making clear that our discretization in the transformed variables is not artificial. We obtain convergence of the discrete generalized solutions u_j to the continuous solution in the H^1-norm, which entails H^1-convergence of the corresponding saturation $M(u_j)$ and L^2-convergence of its piecewise linear interpolation. In the nondegenerate case (2.8) we can also prove H^1-convergence of the retransformed pressure $\kappa^{-1}(u_j)$ as well as L^2-convergence of its piecewise linear interpolation.

4.1. Discretized problem in generalized variables. For the sake of presentation we consider the case of a polygonal domain $\Omega \subset \mathbb{R}^2$. Let $\mathcal{T}_j$, $j \in \mathbb{N}_0$, be a conforming triangulation of Ω. The set of all vertices of the triangles in $\mathcal{T}_j$ is denoted by $\mathcal{N}_j$. We require that each intersection point of two closures of γ_D, γ_N, and γ_S is contained in $\mathcal{N}_j$ and define $\mathcal{N}_j^D := \mathcal{N}_j \cap \gamma_D$ and $\mathcal{N}_j^S := \mathcal{N}_j \cap \gamma_S$.

We choose the finite element space $\mathcal{S}_j \subset H^1(\Omega)$ as the subspace of all continuous functions in $H^1(\Omega)$, which are linear on each triangle $t \in \mathcal{T}_j$. Analogously, we define $\mathcal{S}_j^D \subset H^1_{\gamma_D}(\Omega)$. The nodal basis function corresponding to $q \in \mathcal{N}_j$ is denoted by $\lambda_q^{(j)}$. For the finite dimensional analogue of $\mathcal{K}$ we assume that u_D is continuous in each $q \in \mathcal{N}_j^D$, $j \in \mathbb{N}_0$, so that writing $u_D(q)$ makes sense in these nodes. Then we define the nonempty, closed, and convex set $\mathcal{K}_j \subset \mathcal{S}_j$ by

$$\mathcal{K}_j := \left\{ v \in \mathcal{S}_j : v(q) \geq u_c \,\forall q \in \mathcal{N}_j \,\wedge\, v(q) = u_D(q) \,\forall q \in \mathcal{N}_j^D \,\wedge\, v(q) \leq 0 \,\forall q \in \mathcal{N}_j^S \right\}. \tag{4.1}$$

We discretize the convex functional in (3.6) by $\mathcal{S}_j$-interpolation of the integrand $\Phi(v)$ arriving at $\phi_j : \mathcal{S}_j \to \mathbb{R} \cup \{+\infty\}$ given by

$$\phi_j(v) := \sum_{q \in \mathcal{N}_j} \Phi(v(q))\, h_q \quad \forall v \in \mathcal{S}_j\,, \quad h_q := \int_\Omega \lambda_q^{(j)}(x)\, dx\,. \tag{4.2}$$

The properties of ϕ_j are inherited by ϕ. Concretely, ϕ_j, $j \geq 0$, are convex, proper, Lipschitz continuous, lower semicontinuous, and admit affine lower bounds. The constants are independent of $j \geq 0$. Moreover, for $v_j \in \mathcal{S}_j$, $j \geq 0$, and $v \in H^1(\Omega)$ we have

$$v_j \rightharpoonup v,\, j \to \infty \implies \liminf_{j\to\infty} \phi_j(v_j) \geq \phi(v)\,,$$

where $v_j \rightharpoonup v$ denotes the weak convergence of v_j to v in $H^1(\Omega)$.

With the definitions from above, our discrete version of (3.9) reads

$$u_j \in \mathcal{K}_j : \;\; \mathcal{J}(u_j) + \phi_j(u_j) \leq \mathcal{J}(v) + \phi_j(v) \quad \forall v \in \mathcal{K}_j\,. \tag{4.3}$$

Since $\mathcal{K}_j$, $\mathcal{J}$, and ϕ_j have the same properties as $\mathcal{K}$, $\mathcal{J}$, and ϕ in Theorem 3.3, now in the subspace $\mathcal{S}_j$ of the Hilbert space $H^1(\Omega)$, we obtain the following result.

THEOREM 4.1. *The discrete minimization problem* (4.3) *has a unique solution.*

In order to ensure stability of the explicit time discretization of the convective gravity term, we use a standard upwind technique based on artificial viscosity. See [9, sec. 4.2] for details. This does not affect the properties of $\mathcal{J}$ that are relevant for our analysis. The stability properties of the resulting discretization will be illustrated by numerical experiments to be reported in subsection 5.2.

The discretization presented in the preceding two sections is solver-friendly in the sense that the resulting spatial problems can be solved by monotone multigrid methods [26, 30]. These methods can be regarded as multilevel descent methods and thus rely on convex minimization rather than linearization. Asymptotic logarithmic bounds for the convergence rates are available and fast and robust convergence has been observed for model problems [30]. The convergence behavior for a problem in three space dimensions and a wide range of soil parameters will be reported in section 5.3.

4.2. Interpretation in physical space: Discrete Kirchhoff transformation. Now we give a reinterpretation of (4.3) in terms of discrete physical variables. It turns out that (4.3) can be understood as a finite element discretization of problem (2.3), written in physical variables, where a particular quadrature rule with quadrature points for $kr(\theta(p))$ depending on $kr \circ \theta$ is applied.

By Lemma 3.2 the discrete minimization problem (4.3) is equivalent to the variational inequality

$$u_j \in \mathcal{K}_j : \sum_{q\in\mathcal{N}_j} M(u_j(q))\,(v(q) - u_j(q))\,h_q + a(u_j, v - u_j) - \ell(v - u_j) \geq 0 \quad \forall v \in \mathcal{K}_j\,, \tag{4.4}$$

which can be regarded as the corresponding discretization of the original variational inequality (3.4).

It is clear that in case of $u_j(q) = u_c$ for a $q \in \mathcal{N}_j$ we have $\kappa^{-1}(u_j(q)) = -\infty$, which is a physically unrealistic situation. Note that the somewhat unnatural condition $v \geq u_c$ instead of $v > u_c$ in (2.12) and, correspondingly, in (4.1) is necessary to guarantee the existence of a solution to the minimization problem by the closedness of the convex sets $\mathcal{K}$ and $\mathcal{K}_j$, respectively, and does not occur in the original physical problem. Therefore, we assume

$$u_j(q) > u_c \quad \forall q \in \mathcal{N}_j \tag{4.5}$$

from now on, which entails real-valuedness of $\kappa^{-1}(u_j)$ and allows the following definition.

DEFINITION 4.2. *Let* $I_{\mathcal{S}_j} : H^1(\Omega) \cap C(\overline{\Omega}) \to \mathcal{S}_j$ *be the piecewise linear interpolation operator defined by* $(I_{\mathcal{S}_j} v)(q) = v(q)\ \forall q \in \mathcal{N}_j$ *for* $v \in H^1(\Omega) \cap C(\overline{\Omega})$. *With assumption* (4.5) *we call*

$$I_{\mathcal{S}_j}\kappa : \mathcal{S}_j \to \mathcal{S}_j$$

the discrete Kirchhoff transformation *on* $\mathcal{S}_j$ *and*

$$p_j := I_{\mathcal{S}_j}\kappa^{-1}(u_j)$$

the discrete physical pressure *corresponding to problem* (4.3).

We are now going to investigate what kind of discretization of the untransformed problem corresponds to the discrete pressure variable p_j. To this end we impose the condition

$$\kappa \in C^1(\mathbb{R})$$

on the Kirchhoff transformation (2.4) which means that $kr \circ \theta$ is continuous. The latter is satisfied for the Brooks–Corey parameter functions in (2.1) and (2.2).

First, by (2.6) we clearly have

$$M(u_j(q)) = \theta(p_j(q)) \quad \forall q \in \mathcal{N}_j\,.$$

Accordingly, the linear term $\ell(\cdot)$ arising from the solution of the previous time step on the right-hand side in (2.13) is retransformed in discrete physical variables. The remaining problem is to see how the bilinear form

$$a(u_j, w) = \int_\Omega \nabla u_j \nabla w \, dx\,, \quad w = v - u_j\,,\ \ v \in \mathcal{K}_j\,, \tag{4.6}$$

looks in physical variables. For the continuous problem (2.3) the reformulation is provided by the chain rule (2.5) in a weak sense; consult [9, sec. 1.5.4] or [10]. For the discrete problem we need a discrete counterpart of (2.5) and argue as follows with the help of the mean value theorem.

First, we consider the integral in (4.6) only on a triangle $t \in \mathcal{T}_j$. Recall that the transformation from the reference triangle

$$T \subset \mathbb{R}^2 \quad \text{with the vertices} \quad a = (0,0)\,,\ \ b = (1,0)\,,\ \ c = (0,1) \tag{4.7}$$

onto the triangle t is given by an affine map

$$G_t : \xi \mapsto x = B_t\xi + b_t$$

acting on $\mathbb{R}^2$ with a nonsingular matrix $B_t \in \mathbb{R}^{2\times 2}$ and a vector $b_t \in \mathbb{R}^2$. Transformed functions on the reference element shall be denoted by

$$\hat{v}(\xi) := v(G_t(\xi)) = v(x) \quad \forall x \in t\,,\ \ \forall v \in H^1(\Omega) \cap C(\overline{\Omega})\,.$$

By the chain rule we can write

$$\nabla_\xi\, \hat{v}(\xi) = \nabla_x\, v(x)\, B_t\,, \quad \forall x = G_t(\xi) \in t\,,\ \ \forall v \in H^1(\Omega) \cap C(\overline{\Omega})\,.$$

Without loss of generality we assume $\hat{u}_j(b) \neq \hat{u}_j(a)$. Then, with the Eucledian norm $|\cdot|$ in $\mathbb{R}^2$ and (4.7), the first component in $\nabla_\xi\, \hat{u}_j$ is given by

$$\left(\nabla_\xi\, \hat{u}_j\right)_1 = \frac{\hat{u}_j(b) - \hat{u}_j(a)}{|b-a|} = \frac{\kappa(\hat{p}_j(b)) - \kappa(\hat{p}_j(a))}{\hat{p}_j(b) - \hat{p}_j(a)} \cdot \frac{\hat{p}_j(b) - \hat{p}_j(a)}{|b-a|}\,.$$

The range of the affine function $\hat{p}_j$ on the edge between a and b is the interval with the endpoints $\hat{p}_j(a)$ and $\hat{p}_j(b)$. Since κ is bijective and continuously differentiable on this interval, there exists a unique point $\bar{\xi}_1$ on the edge between a and b with the property

$$\left(\nabla_\xi \hat{u}_j\right)_1 = \kappa'(\hat{p}_j(\bar{\xi}_1)) \, \frac{\hat{p}_j(b) - \hat{p}_j(a)}{|b-a|} = kr(\theta(\hat{p}_j(\bar{\xi}_1))) \, \left(\nabla_x \hat{p}_j\right)_1 . \tag{4.8}$$

Analogously, we can find a point $\bar{\xi}_2$ on the edge of T between the vertices a and c with the corresponding property. Altogether, with the transformation onto the reference triangle, the reformulation in physical variables and the transformation back onto t, we obtain

$$\nabla u_j = D_t(p_j) \nabla p_j \quad \text{on } t \tag{4.9}$$

with the diagonal matrix

$$D_t(p_j) = \begin{pmatrix} kr(\theta(p_j(\bar{x}_1))) & 0 \\ 0 & kr(\theta(p_j(\bar{x}_2))) \end{pmatrix}$$

and points

$$\bar{x}_1 = G_t(\bar{\xi}_1) \quad \text{and} \quad \bar{x}_2 = G_t(\bar{\xi}_2) \tag{4.10}$$

situated on edges of t.

Since p_j is affine and $\kappa : \mathbb{R} \to (u_c, \infty)$ is bijective, the points $\bar{x}_1$ and $\bar{x}_2$ are uniquely defined by the properties (4.8) and (4.10). Therefore, one can interpret (4.9) as the discrete counterpart of the chain rule (2.5) for the discrete Kirchhoff transformation in $\mathcal{S}_j$.

Now we introduce the nonlinear form

$$b(p_j, v) := \sum_{t \in \mathcal{T}_j} \int_t D_t(p_j) \nabla p_j \nabla v \, dx \, , \quad p_j, v \in \mathcal{S}_j \, . \tag{4.11}$$

Then, with discrete Dirichlet boundary data p_D on $\mathcal{N}_j^D$ and the closed and convex set

$$\mathcal{K}_j^0 := \left\{ v \in \mathcal{S}_j : v(q) = p_D(q) \ \ \forall q \in \mathcal{N}_j^D \wedge \ v(q) \le 0 \ \ \forall q \in \mathcal{N}_j^S \right\}$$

we consider the discrete problem

$$p_j \in \mathcal{K}_j^0 : \sum_{q \in \mathcal{N}_j} \theta(p_j(q)) \, (v(q) - p_j(q)) \, h_q + b(p_j, v - p_j) - \ell(v - p_j) \ge 0 \quad \forall v \in \mathcal{K}_j^0 \tag{4.12}$$

in physical variables. Note that in case of $\gamma_S = \emptyset$ we have

$$\mathcal{K}_j - u_j = \left\{ v \in \mathcal{S}_j : v(q) \ge -\varepsilon \ \ \forall q \in \mathcal{N}_j \ \wedge \ v(q) = 0 \ \ \forall q \in \mathcal{N}_j^D \right\}$$

with an $\varepsilon > 0$ due to (4.5), so that by linearity the corresponding set of test functions $v - u_j$ in (4.4) can be chosen as the space $\mathcal{S}_j^D$ which is equal to $\mathcal{K}_j^0 - p_j$. On the other hand, with the assumption $kr(\theta(\mathbb{R})) \subset (0, 1]$ we have $p \le \kappa(p) \ \ \forall p \in \mathbb{R}$ and, therefore,

$$\mathcal{K}_j - u_j \subset \mathcal{K}_j^0 - p_j \, .$$

In general, these sets of test functions in (4.4) and (4.12), respectively, are not equal. However, with these ingredients one can prove the following discrete counterpart of Theorem 1.5.18 in [9], with arguments as given there for the continuous case.

THEOREM 4.3. *Let $\theta : \mathbb{R} \to \mathbb{R}$ and let $kr : \theta(\mathbb{R}) \to (0,1]$ be bounded, monotonically increasing, and continuous, while $\kappa : \mathbb{R} \to \mathbb{R}$ is defined by* (2.4). *In addition, let $p_D = \kappa^{-1}(u_D)$ on $\mathcal{N}_j^D$. Then $u_j = I_{\mathcal{S}_j}\kappa(p_j)$ solves* (4.4) *if p_j solves* (4.12). *Conversely, $p_j = I_{\mathcal{S}_j}\kappa^{-1}(u_j)$ solves* (4.12) *if u_j solves* (4.4) *with* (4.5) *in case of $\gamma_S = \emptyset$. If* (2.8) *and $\gamma_S = \emptyset$ hold, then* (4.4) *and* (4.12) *are equivalent in the sense that u_j satisfies* (4.4) *if and only if $p_j = I_{\mathcal{S}_j}\kappa^{-1}(u_j)$ satisfies* (4.12).

Our discretization (4.12) of problem (3.4), retransformed in physical variables, involves a quadrature formula with special quadrature points for the term

$$\int_\Omega kr(\theta(p))\nabla p \nabla(v-p)\,dx \tag{4.13}$$

which is given by (4.11). This quadrature is uniquely defined by the given functions kr and θ. Even though one would not use it in practical calculations, one would certainly be forced to use some quadrature for (4.13). At the end of this section we will prove that the quadrature (4.11) is as good as any appropriately chosen quadrature in the sense that it leads to a convergent discretization; see Theorem 4.10.

4.3. Convergence of the generalized pressure. Now we address the convergence of our finite element solutions from (4.3) to the solution of the continuous problem (3.9). The derivation of the results is based on the arguments in [29, pp. 38–42] for the case of homogeneous Dirichlet boundary conditions on all of $\partial\Omega$. Therefore, we only state the assumptions used for the inhomogeneous case and refer to [9, sec. 2.5.2] for details.

Assumption 4.4. Let the sequence of triangulations $(\mathcal{T}_j)_{j\geq 0}$ be shape regular with

$$h_j := \max_{t\in\mathcal{T}_j} \operatorname{diam} t \to 0 \quad \text{for } j \to \infty\,. \tag{4.14}$$

Let $u_D = tr_{\gamma_D} w$ for a $w \in H^1(\Omega) \cap C(\overline{\Omega})$ satisfying

$$w_j := I_{\mathcal{S}_j} w \to w \quad \text{for } j \to \infty \text{ in } H^1(\Omega)\,. \tag{4.15}$$

For the sets $C^\infty_{\gamma_D}(\overline{\Omega}) := \{v \in C^\infty(\overline{\Omega}) : v = 0 \text{ in a neighborhood of } \gamma_D\}$ and $\mathcal{K}_{\gamma_D} := \mathcal{K} - w = \{v \in H^1_{\gamma_D}(\Omega) : v \geq u_c - w \wedge tr_{\gamma_S} v \leq -tr_{\gamma_S} w\}$ we require that

$$C^\infty_{\gamma_D}(\overline{\Omega}) \cap \mathcal{K}_{\gamma_D} \text{ is dense in } \mathcal{K}_{\gamma_D}\,. \tag{4.16}$$

By Ciarlet [19, pp. 122–124] one could replace (4.15) by $w \in H^2(\Omega)$ or a corresponding condition for $d > 2$. Assumption (4.16) holds for $\gamma_S = \emptyset$ if γ_D is sufficiently smooth and, given that $[u_c, \infty)$ is replaced by $\mathbb{R}$ in Assumption 2.3, it is also true for $\gamma_S = \partial\Omega$; consult [25, pp. 36–39, 61].

With Assumption 4.4 one can prove the consistency of the discrete functionals ϕ_j. Concretely, for $v \in w + C^\infty(\overline{\Omega})$ and $v_j = I_{\mathcal{S}_j} v$, $j \geq 0$, one has

$$v_j \to v \text{ in } H^1(\Omega) \quad \text{and} \quad \phi_j(v_j) \to \phi(v) \quad \text{for } j \to \infty\,.$$

Together with the properties of ϕ_j named in subsection 4.1 one can now prove the following theorem.

THEOREM 4.5. *Let Assumptions* 2.3 *and* 4.4 *with possibly discontinuous M be given. Then the solutions u_j of the discrete minimization problem* (4.3) *converge to the solution u of* (3.9) *in the sense that*

$$u_j \to u \text{ in } H^1(\Omega) \quad \textit{and} \quad \phi_j(u_j) \to \phi(u) \quad \textit{for } j \to \infty\,.$$

Theorem 4.5 also holds in the case (3.10) of space-dependent hydraulic conductivity $K_h(\cdot)$ and in case of a positive and bounded porosity $n(\cdot)$ in (1.1) if the discretization of the resulting ϕ in (3.6) is adapted accordingly in (4.2).

4.4. Convergence of the saturation and the physical pressure. The last part of this section is devoted to what can be inferred from Theorem 4.5 on the behavior of the saturation $M(u_j)$ and the physical pressure $\kappa^{-1}(u_j)$ as well as their piecewise linear interpolations, discrete saturation $\theta_j(p_j) := I_{\mathcal{S}_j}\theta(p_j) = I_{\mathcal{S}_j}M(u_j)$, and discrete pressure $p_j = I_{\mathcal{S}_j}\kappa^{-1}(u_j)$, for $j \to \infty$.

The L^2-convergence results for the saturation hold in quite general situations including the Brooks–Corey model. For the physical variables we obtain only convergence results in case of uniformly bounded p_j, $j \geq 0$, which is reasonable in realistic situations and is guaranteed in case of nondegeneracy (2.8). Although we prove only L^2-convergence of $I_{\mathcal{S}_j}M(u_j)$ and p_j for $j \to \infty$, we show H^1-convergence for the iterates $M(u_j)$ and $\kappa^{-1}(u_j)$, which can also be evaluated on a discrete level.

Recall that the real functions M and κ^{-1} induce superposition operators by composition $M \circ u$ and $\kappa^{-1} \circ u$. In order to deduce $M(u_j) \to M(u)$ in $L^2(\Omega)$ by $u_j \to u$ for $j \to \infty$ with the help of Theorem 4.5, we note the following lemma (cf. [3] and [9, pp. 90/91]). In particular, the second assertion holds for the Brooks–Corey case.

LEMMA 4.6. *Let $\Omega \subset \mathbb{R}^d$ be bounded. If $M : \mathbb{R} \to \mathbb{R}$ is continuous and bounded, it induces a continuous superposition operator on $L^2(\Omega)$. If $M : \mathbb{R} \to \mathbb{R}$ is α-Hölder continuous w.r.t. $\alpha \in (0, 1]$, it induces an α-Hölder continuous superposition operator on $L^2(\Omega)$.*

The situation is more convenient in the nondegenerate case (2.8) since here the convergence properties of the generalized pressure are inherited by the saturation and the retransformed pressure.

THEOREM 4.7. *In the nondegenerate case* (2.8) *and with Assumptions* 2.3 *and* 4.4 *we have the convergence*

$$M(u_j) \to M(u) \quad \textit{and} \quad \kappa^{-1}(u_j) \to \kappa^{-1}(u) \quad \textit{in } H^1(\Omega) \textit{ for } j \to \infty\,.$$

For the proof we can use the following result. Remarkably, its converse is also true for $d \geq 2$ even without imposing continuity of the superposition operator [31].

LEMMA 4.8. *If $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, then the corresponding superposition operator acts on $H^1(\Omega)$ and is continuous.*

With respect to discrete solutions, one will certainly be interested in the convergence behavior of the $\mathcal{S}_j$-interpolations of $M(u_j)$ and $\kappa^{-1}(u_j)$, in particular, since the latter is the discrete physical pressure from the finite element discretization (4.12).

LEMMA 4.9. *Let $f : \mathbb{R} \to \mathbb{R}$ be α-Hölder continuous w.r.t. $\alpha \in (0, 1]$. Then, for $u_j \in \mathcal{S}_j$, $j \geq 0$, with $u_j \to u$ in $H^1(\Omega)$ for $j \to \infty$, we have*

$$f(u_j) - I_{\mathcal{S}_j} f(u_j) \to 0 \quad \textit{in } L^2(\Omega) \textit{ for } j \to \infty\,. \tag{4.17}$$

Proof. For any point x contained in a triangle $t \in \mathcal{T}_j$ with the vertices q_1, q_2, q_3 there are $\vartheta_i \in [0, 1]$, $i = 1, 2, 3$, with $\sum_{i=1}^3 \vartheta_i = 1$ such that

$$I_{\mathcal{S}_j} f(u_j)(x) = \sum_{i=1}^{3} \vartheta_i\, f(u_j(q_i))\,.$$

Therefore, using binomial formulas and the Hölder continuity of f with the Hölder

constant C_α, we can estimate

$$\begin{aligned} |f(u_j(x)) - I_{\mathcal{S}_j} f(u_j)(x)|^2 &\le \left(\sum_{i=1}^{3} \vartheta_i \, |f(u_j(x)) - f(u_j(q_i))| \right)^2 \\ &\le 3 \sum_{i=1}^{3} |f(u_j(x)) - f(u_j(q_i))|^2 \\ &\le 3 \, C_\alpha^2 \sum_{i=1}^{3} |u_j(x) - u_j(q_i)|^{2\alpha} \, . \end{aligned} \tag{4.18}$$

Using the mean value theorem

$$|u_j(x) - u_j(q_i)| \le |\nabla u_j| |x - q_i|$$

on the triangle t (with the Euclidean norm $|\cdot|$ on $\mathbb{R}^d$) while considering that $|\nabla u_j|$ is constant on t, we can go on estimating the last term in (4.18) to obtain

$$|f(u_j(x)) - I_{\mathcal{S}_j} f(u_j)(x)|^2 \le 9 \, C_\alpha^2 \, |\nabla u_j|^{2\alpha} \, h_j^{2\alpha}$$

with h_j as in (4.14). Now, integration over Ω provides

$$\begin{aligned} \int_\Omega |f(u_j(x)) - I_{\mathcal{S}_j} f(u_j)(x)|^2 \, dx &\le \sum_{t \in \mathcal{T}_j} \int_t |f(u_j(x)) - I_{\mathcal{S}_j} f(u_j)(x)|^2 \, dx \\ &\le 9 \, C_\alpha^2 \, h_j^{2\alpha} \int_\Omega (|\nabla u_j|^2 + 1) \, dx \, . \end{aligned}$$

Since $(u_j)_{j \ge 0}$ converges in $H^1(\Omega)$, the last integral is uniformly bounded and, therefore, this whole last term tends to 0 as $j \to \infty$ due to (4.14). ∎

We remark that due to the Sobolev embedding theorem, Lemmas 4.6 and 4.9 also hold in one space dimension if $L^2(\Omega)$ is replaced by $(C(\overline{\Omega}), \|\cdot\|_\infty)$. As a consequence of Lemmas 4.6, 4.8, and 4.9 we obtain the following convergence results.

THEOREM 4.10. *Let Assumptions* 2.3 *and* 4.4 *be satisfied. Then we have*

$$\theta_j(p_j) = I_{\mathcal{S}_j} M(u_j) \to M(u) = \theta(p) \quad \text{in } L^2(\Omega) \text{ for } j \to \infty \, .$$

In the nondegenerate case (2.8) *we also have*

$$p_j = I_{\mathcal{S}_j} \kappa^{-1}(u_j) \to p = \kappa^{-1}(u) \quad \text{in } L^2(\Omega) \text{ for } j \to \infty \, .$$

Note that in the proof of Lemma 4.9 we also obtained the order of convergence $\mathcal{O}(h_j^\alpha)$ for (4.17). Therefore, altogether we can prove that the convergence $p_j \to p$ and $I_{\mathcal{S}_j} \theta(p_j) \to \theta(p)$ in $L^2(\Omega)$ is of order $\mathcal{O}(h_j)$ and $\mathcal{O}(h_j^\alpha)$, respectively, if the convergence $u_j \to u$ in $L^2(\Omega)$ has the order $\mathcal{O}(h_j)$. Section 5.1 reveals that numerically one can observe much more, even in the (degenerate!) Brooks–Corey case.

5. Numerical results. We now concentrate on the numerical properties of the discretization suggested above and on the efficiency and robustness of the associated

monotone multigrid method. The implementation has been performed in the numerics environment DUNE [5] using the grid manager from UG [4].

5.1. Spatial discretization error. This subsection is devoted to adding a quantitative flavor to Theorems 4.5, 4.7, and 4.10 by determining numerically the order of convergence of $u_j \to u$ and $p_j \to p$ as $j \to \infty$ for an example in two space dimensions. We consider the function $(x, y) \mapsto \tilde{p}(x, y) = 0.1 - 10\,(x^2 + y^2)$ on $\overline{\Omega} = [0, 2] \times [0, 1]$ (with pressure and length unit $[m]$) and set

$$f := n\,\theta(\tilde{p}) - \operatorname{div}\left(K_h kr(\theta(\tilde{p}))\nabla\tilde{p}\right).$$

One can regard $\tilde{p}$ as a stationary solution of a corresponding time-discretized Richards equation (1.1) without gravity with the time step size $\tau = 1\,[s]$. We use the Brooks–Corey model. The soil parameters given in Table 5.1 are in a realistic range of sandy soils (see [32]).

We approximate $\tilde{p}$ by solving the discretized equation

$$n\,\theta(p) - \operatorname{div}\left(K_h kr(\theta(p))\nabla p\right) = f$$

in the finite element space $\mathcal{S}_j$ as described above and determine discrete solutions u_j and p_j for $j = 1, \ldots, 11$ with monotone multigrid. We choose Dirichlet boundary conditions on $\partial\Omega$ and $I_{\mathcal{S}_j}\tilde{p}$ as the initial iterate. We start with a uniform coarse triangular grid for $j = 1$ with 15 nodes and obtain the higher levels by uniform refinement. This leads to $8,394,753$ nodes on the finest level.

The exact solution is a paraboloid directed downwards, and we have full saturation $\theta(\tilde{p}) = \theta_M$ on a disc around the origin with the radius $\sqrt{0.02} \approx 0.14$ only, so that a large part of the domain is dominated by the nonlinear nature of the problem. Besides, note that the problem is not radially symmetric.

Figures 5.1 and 5.2 show an order of convergence $\mathcal{O}(h_j^2)$ for both $u_j \to \tilde{u} = \kappa(\tilde{p})$ and $p_j \to \tilde{p}$ as $j \to \infty$ in the L^2-norm. Figures 5.3 and 5.4 show that with the H^1-norm we obtain only an order of convergence $\mathcal{O}(h_j)$, which one might expect from the result for the L^2-norm, and which is optimal even for linear problems.

The anomalous behavior of the curves corresponding to the physical pressure for small mesh sizes can be explained by the ill-conditioning of the inverse Kirchhoff transformation $\kappa^{-1} : (u_c, \infty) \to \mathbb{R}$ around u_c. The following estimates illuminate this effect and even confirm its order of magnitude.

Concretely, if $\bar{u}$ is an approximation of $\tilde{u}$ up to the numerical accuracy of

$$|\bar{u}(x, y) - \tilde{u}(x, y)| = 10^{-16} \quad \text{on } \Omega\,,$$

the square of this error is only given up to an accuracy of

$$0.01 \int_\Omega |\kappa^{-1}(\bar{u}(x, y)) - \kappa^{-1}(\tilde{u}(x, y))|^2\,dx\,dy = 10^{-34} \int_\Omega |(\kappa^{-1})'(u(x, y))|^2\,dx\,dy \tag{5.1}$$

TABLE 5.1
Soil parameters of sandy soil.

n	θ_m	θ_M	λ	p_b	K_h
0.38	0.21	0.95	1.0	$-0.1\,[m]$	$2 \cdot 10^{-3}\,[m/s]$

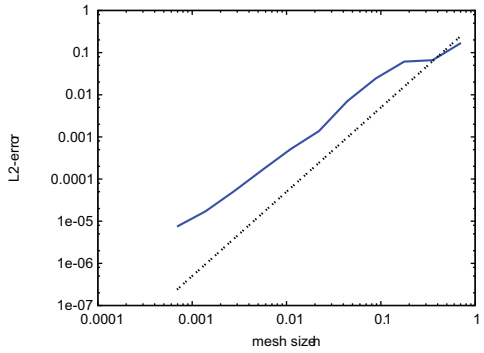

FIG. 5.1. *L^2-error in u (dotted line: $\mathcal{O}(h^2)$).*

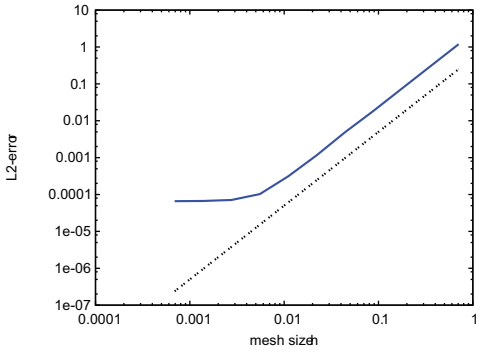

FIG. 5.2. *L^2-error in p (dotted line: $\mathcal{O}(h^2)$).*

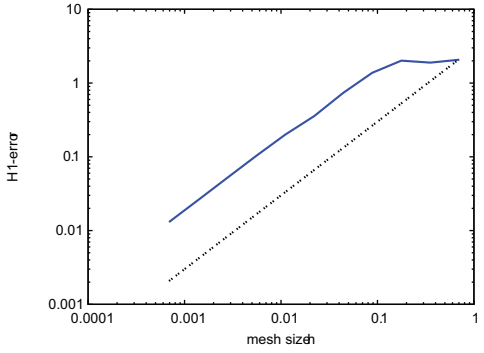

FIG. 5.3. *H^1-error in u (dotted line: $\mathcal{O}(h)$).*

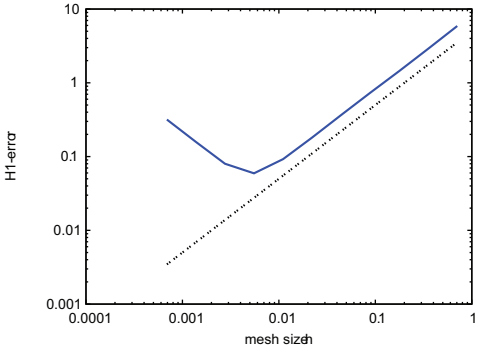

FIG. 5.4. *H^1-error in p (dotted line: $\mathcal{O}(h)$).*

with suitable $u(x,y)$ between $\bar{u}(x,y)$ and $\tilde{u}(x,y)$. (The factor 0.01 enters (5.1) because we treat u in the unit $|p_b|$, whereas the unit for p is given by $[m]$.) Now, we have

$$(\kappa^{-1})'(u) = \frac{1}{\kappa'(\kappa^{-1}(u))} = \frac{1}{\kappa'(p)} = \frac{1}{kr(\theta(p))} = (-10\,p)^5$$

for $p \le p_b = -0.1$ due to (2.1) and (2.2) and the choice of $\lambda = 1$. If we insert this into (5.1) for $p = \tilde{p}$, we can get an estimation of the numerical accuracy for the square of the L^2-error in p by considering the integral only on the right half of the quadrilateral Ω where we have $x^2 + y^2 \ge 1$. Therefore, we obtain the estimate

$$10^{-34} \int_0^1 \int_1^2 (100\,(x^2+y^2) - 1)^{10}\,dx\,dy \\ \ge 10^{-34}\,99^{10} \int_0^1 \int_1^2 (x^2+y^2)^{10}\,dx\,dy \approx 5 \cdot 10^{-5}$$

for the numerical accuracy that we can expect for the L^2-error in p. In fact, the L^2-error in p on levels 9, 10, and 11 is already around $7 \cdot 10^{-5}$ as one can see in Figure 5.2.

Consequently, the H^1-error in p raises from level 9 to 10 and from level 10 to 11 by a factor of 2, since the numerical accuracy of these terms is given by the numerical accuracy of the L^2-error in p divided by the horizontal mesh size $h_j/\sqrt{2}$. For example, with $h_{11}/\sqrt{2} = 2^{-11}$ and the numerical accuracy of $7 \cdot 10^{-5}$ for the L^2-error we obtain

0.14 as an estimate for the numerical accuracy of the H^1-error in p on level 11. In fact, here we obtain the H^1-error 0.32 as one can see in Figure 5.4.

We point out that this ill-conditioning is part of the problem, i.e., a measure for the degeneracy of the Richards equation (1.1), and has to be dealt with in one way or the other within any solution process. The advantage of our approach is the separation of this ill-conditioning from the solution process.

5.2. Stability of time discretization. In the time discretization as described in section 4.1, the convective gravitational term is treated by explicit upwinding. On the one hand, explicit discretizations are usually more accurate than implicit schemes which tend to smear out sharp fronts, but on the other hand, explicit methods give rise to stability constraints on the time step. The stability properties of our implicit–explicit upwind discretization are illustrated by the following numerical experiments.

We consider the computational domain $\Omega = (0,1) \times (0,1)$ with unit length $[m]$ and soil parameters for sand as obtained from [32] and listed in Table 5.2. We assume that the upper surface $\Gamma = [0,1] \times \{1\}$ of Ω is covered by water with a constant height of $2\,[m]$ and no flow conditions are imposed at the rest of the boundary.

Starting from an initial pressure $p_0 = -10\,[m]$, which by Table 5.2 corresponds to almost dry sand with saturation $\theta_0 = 0.0771$, a horizontal saturation front driven by gravity propagates through Ω from above to below. It reaches the bottom and thus leads to full saturation after $1126\,[s]$. The problem in convection dominated in the sense that the Reynolds number $R(u) = kr'(u)M'(u)$ of the transformed Richards equation (2.7) peaks at $Re(u) \approx 50$ for $u \lessapprox p_b$ directly before the front and is almost zero or even zero elsewhere.

We use a uniform triangular grid with mesh size $h = 2^{-6}\,[m]$ and vary the time step from $\tau = 1\,[s]$ to $\tau = 1000\,[s]$. Here, no instability occurs. It seems that due to the one-dimensional character of the problem, dominating convection at the saturation front is compensated by dominating diffusion arising directly after the front.

This is no longer the case for a curved saturation front which, for example, is obtained by restricting the constant water table of $2\,[m]$ to $\Gamma_0 = [0.75,1] \times \{1\}$ and assume no flow conditions at $\Gamma \setminus \Gamma_0$. We now observe instabilities for time steps larger than $\tau = 175\,[s]$ occurring in the upper left corner at just before full saturation is reached after $2700\,[s]$. For $\tau = 200\,[s]$ this is shown in Figure 5.5. As expected, we found larger (smaller) stability constraints for coarser (finer) spatial meshes.

TABLE 5.2
Soil parameters of sand.

n	θ_m	θ_M	λ	p_b	K_h
0.437	0.0458	1.0	0.694	$-0.0726\,[m]$	$6.54 \cdot 10^{-5}\,[m/s]$

FIG. 5.5. *Evolution of the saturation front at time* $t = 400\,s,\ 800\,s,\ 1600\,s,\ 2600\,s$.

5.3. A dam problem in 3D. In this last example we illustrate that our discretization is solver-friendly in the sense that existing monotone multigrid methods [30] exhibit fast convergence when applied to the discrete minimization problems (4.3). Moreover, the convergence speed turns out to be robust with respect to soil parameters.

5.3.1. Fast multigrid solution. We consider a dam described by the coarse grid (consisting of prisms and hexahedra) as depicted in Figure 5.6. The dam has a constant width on the bottom and a constant maximal height, both equal to $9.81\,[m]$. Its length is four times this value. We assume that the dam consists of sand with material parameters listed in Table 5.2 and, as in section 5.2, we select the constant pressure $p_0 = -10\,[m]$, or, equivalently, the very low saturation $\theta_0 = 0.0771$ as initial condition for the Richards equation with gravity. As to the boundary conditions, we assume a constant sea level of the maximal height $9.81\,[m]$ of the dam on the front side (left in Figure 5.6) leading to Dirichlet conditions by hydrostatic pressure. On the small faces of the dam as well as its bottom side we impose homogeneous Neumann conditions. Finally, on the back side water may (and eventually will) flow out so that we have a Signorini-type condition (2.11) there.

As a result, water infiltrates until a fully saturated dam with an overall nonnegative pressure, i.e., a stationary state, is reached. With the time step size $\tau = 2.5\,[s]$ this takes 106 time steps. See Figures 5.7–5.10 for the evolution of the wetting front ($p = p_b$) on the left and color plots of the physical pressure (between -10 and 9.81) on a vertical cut through the dam (situated at about a third of the dam length from the left small face).

The space discretization is carried out by first order Lagrangian finite elements (compare Figures 5.6 and 5.10 for the coarse grid). We have four refinement levels with $216,849$ nodes on the finest level. The monotone multigrid starts with the function obtained by nested iteration and stops as soon as the relative distance of succeeding iterates u^{k-1}, u^k in the H^1-seminorm $|\cdot|_1$ satisfies

$$\frac{|u^k - u^{k-1}|_1}{|u^{k-1}|_1} < 10^{-13}\,. \tag{5.2}$$

Let u^n be the last iterate. Then for each time step we calculate the multigrid convergence rate as the geometric mean of the rates

$$\frac{|u^k - u^n|_1}{|u^{k-1} - u^n|_1}\,, \quad k = 1, \ldots, n-1\,, \tag{5.3}$$

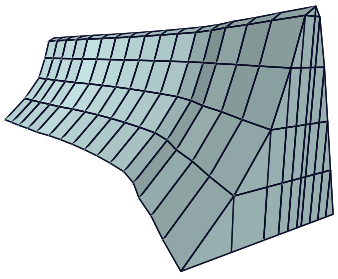

FIG. 5.6. *Coarse (prism) grid of the dam.*

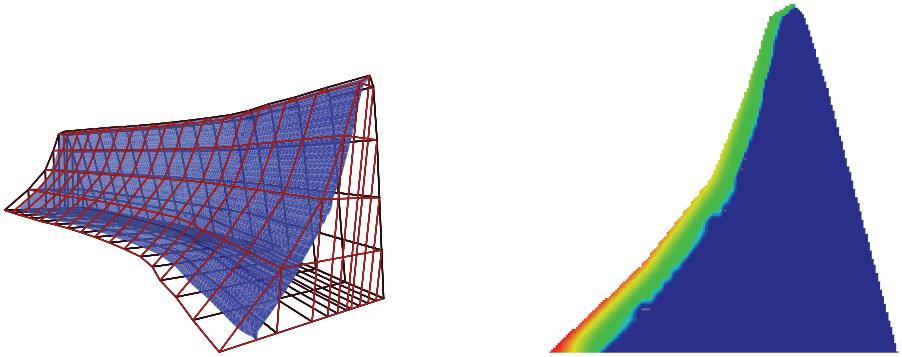

FIG. 5.7. $t = \tau = 2.5\,s$.

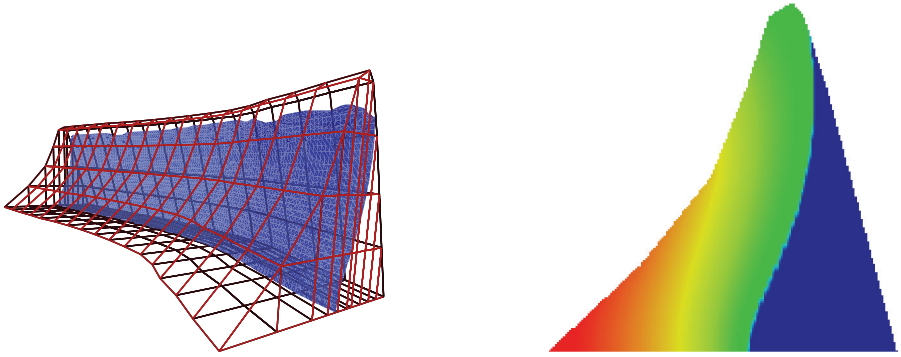

FIG. 5.8. $t = 20\,\tau = 50\,s$.

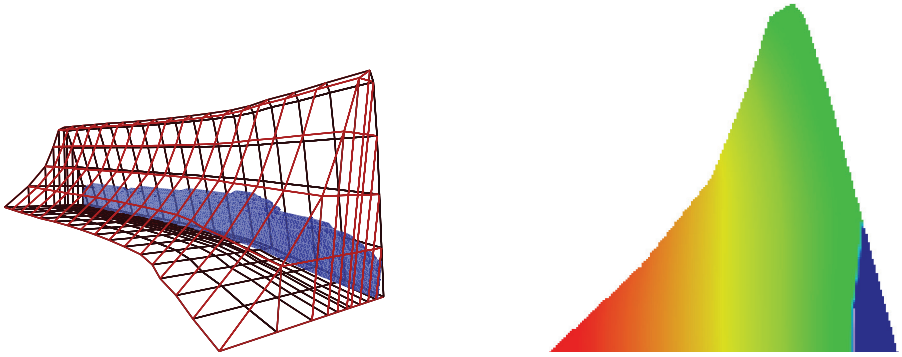

FIG. 5.9. $t = 60\,\tau = 150\,s$.

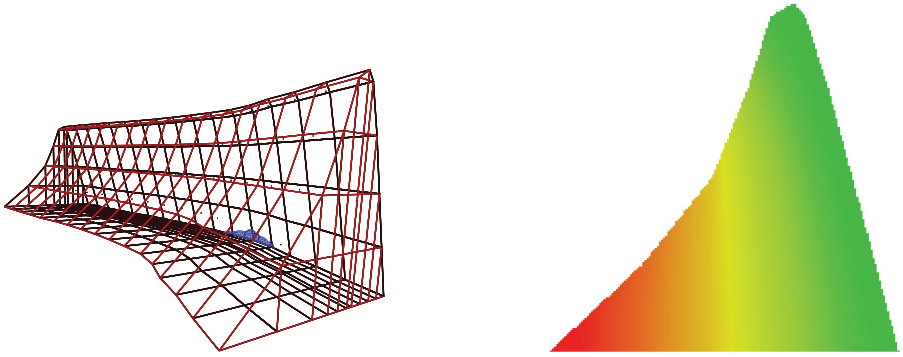

FIG. 5.10. $t = 100\,\tau = 250\,s$.

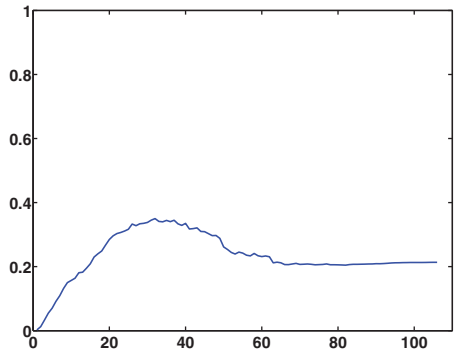

FIG. 5.11. *Multigrid convergence rates for the spatial problems over the time steps* $j = 1, \ldots, 106$.

setting the rate equal to 0 if $n \leq 2$. Figure 5.11 shows the multigrid convergence rates for the different spatial problems over the time steps $j = 1, \ldots, 106$. As a result, the maximal rate for all time steps is $\rho_{\max} = 0.35$ and the average rate is about $\rho_{av} = 0.23$ in this example, which we find to be a quite good performance of the multigrid solver. Note that these rates do not differ much from the rate $\rho_{lin} = 0.21$ in the linear case, which is a Darcy problem that has to be solved at the end of the evolution.

5.3.2. Robustness with respect to soil parameters. Now we illustrate the robust behavior of the multigrid solver with respect to the soil parameters p_b and λ which enter the nonlinearities. Concretely, we fix the time step size $\tau = 2.5\,[s]$ and the initial condition $\theta_0 = 0.0771$ as well as the parameters in Table 5.2 apart from p_b or λ. We vary $-p_b$ or λ, respectively, within a large range of the decimal powers between 10^{-10} and 10^{10} and, in addition, on the intervals $[0.01, 0.1]$, $[0.1, 1]$, and $[1, 10]$, each subdivided in 10 subintervals with equal length, which represent a hydrologically realistic range (compare [32, Table 5.3.2]). We have computed the evolution for each case until a stationary state with a fully saturated dam has been reached. This takes between 2 and 115 time steps. For each time step we calculated the multigrid convergence rates according to (5.2) and (5.3) as above. Then we determined the maximum $\rho_{\max}$ and the average ρ_{av} of these rates for each evolution.

The saturation $\theta(p)$ as a function $M(u) = \theta(\kappa^{-1}(u))$ of u degenerates to step functions for $\lambda \to 0$, $\lambda \to \infty$, or $p_b \to 0$. As a consequence, variation of λ and $-p_b$ over 20 orders of magnitude requires considerable care to obtain a numerically stable implementation of $M(u)$. For example, already for $\lambda = 10^{-4}$ the interval $|u - u_c| < 10^{-200}$ covers $0 - 95\%$ of full saturation.

Figures 5.12 and 5.13 show the maximal and, as a dashed line, the average convergence rates $\rho_{\max}$ and ρ_{av} per evolution for varying λ and $-p_b$, respectively. In light of the preceding remarks, we cannot rule out that the oscillations occurring in Figure 5.12 for (completely unphysical) values $\lambda < 10^{-4}$ are due to numerical instabilities. In Figure 5.13 one can see a peak with unusually big convergence rates of about 0.9 for (unphysical) values $-p_b \approx 10^2\,[m]$. It seems that for these cases nested iteration does not provide an initial iterate which is accurate enough to enter the fast asymptotic regime of monotone multigrid convergence immediately (cf. [30]). Nevertheless, our extensive numerical experiments reveal that for a wide variation of soil parameters λ and p_b the monotone multigrid solver exhibits good convergence rates which are often comparable to the linear self-adjoint case.

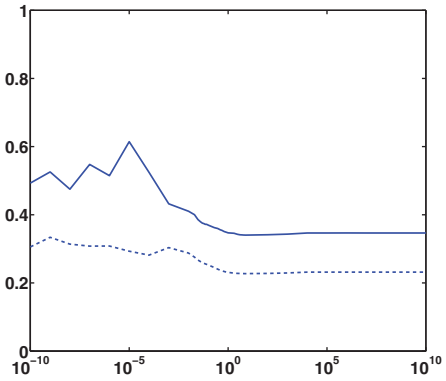

FIG. 5.12. $\rho_{\max}$ *and* ρ_{av} *over* λ.

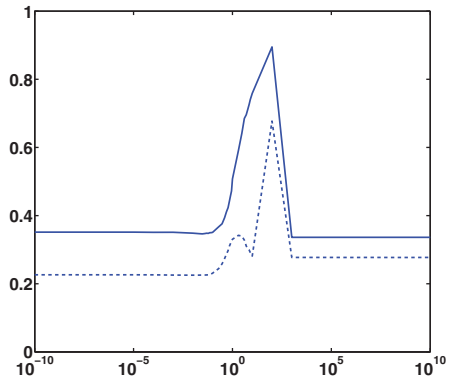

FIG. 5.13. $\rho_{\max}$ *and* ρ_{av} *over* $-p_b$.

REFERENCES

[1] H. ALT AND S. LUCKHAUS, *Quasilinear elliptic-parabolic differential equations*, Math. Z., 183 (1983), pp. 311–341.

[2] H. ALT, S. LUCKHAUS, AND A. VISINTIN, *On nonstationary flow through porous media*, Ann. Mat. Pura Appl. (4), 136 (1984), pp. 303–316.

[3] J. APPELL AND P. ZABREJKO, *Nonlinear Superposition Operators*, Cambridge University Press, Cambridge, UK, 1990.

[4] P. BASTIAN, K. BIRKEN, K. JOHANNSEN, S. LANG, N. NEUSS, H. RENTZ–REICHERT, AND C. WIENERS, *UG – A flexible software toolbox for solving partial differential equations*, Comput. Vis. Sci., 1 (1997), pp. 27–40.

[5] P. BASTIAN, M. BLATT, A. DEDNER, C. ENGWER, R. KLÖFKORN, R. KORNHUBER, M. OHLBERGER, AND O. SANDER, *A generic grid interface for parallel and adaptive scientific computing. Part* II: *Implementation and tests in DUNE*, Computing, 82 (2008), pp. 121–138.

[6] P. BASTIAN, O. IPPISCH, F. REZANEZHAD, H. VOGEL, AND K. ROTH, *Numerical simulation and experimental studies of unsaturated water flow in heterogeneous systems*, in Reactive Flows, Diffusion and Transport, W. Jäger, R. Rannacher, and J. Warnatz, eds., Springer, Berlin, 2007, pp. 579–597.

[7] J. BEAR, *Dynamics of Fluids in Porous Media*, Dover Publications, New York, 1988.

[8] H. BEAUGENDRE, A. ERN, T. ESCLAFFER, E. GAUME, I. GINZBURG, AND C. KAO, *A seepage face model for the interaction of shallow water tables with the ground surface: Application of the obstacle-type method*, J. Hydrol., 329 (2006), pp. 258–273.

[9] H. BERNINGER, *Domain Decomposition Methods for Elliptic Problems with Jumping Nonlinearities and Application to the Richards Equation*, Ph.D. thesis, FU Berlin, Berlin, Germany, 2007.

[10] H. BERNINGER, *Non-overlapping domain decomposition for the Richards equation via superposition operators*, in Domain Decomposition Methods in Science and Engineering XVIII, Lecture Notes in Comput. Sci. Eng. 70, Springer, Berlin, 2009, pp. 169–176.

[11] H. BERNINGER, R. KORNHUBER, AND O. SANDER, *Convergence Behaviour of Dirichlet–Neumann and Robin Methods for a Nonlinear Transmission Problem*, in Domain Decomposition Methods in Science and Engineering XIX, Lecture Notes in Comput. Sci. Eng., Springer, Berlin, 2011.

[12] H. BERNINGER, R. KORNHUBER, AND O. SANDER, *A new solver for the Richards equation in heterogeneous soil*, to appear.

[13] H. BERNINGER AND O. SANDER, *Substructuring of a Signorini-type problem and Robin's method for the Richards equation in heterogeneous soil*, Comput. Vis. Sci., 13 (2010), pp. 187–205.

[14] R. BROOKS AND A. COREY, *Hydraulic Properties of Porous Media*, Technical report Hydrology Paper No. 3, Civil Engineering Department, Colorado State University, Fort Collins, CO, 1964.

[15] N. BURDINE, *Relative permeability calculations from pore-size distribution data*, Petr. Trans., Am. Inst. Mining Metall. Eng., 198 (1953), pp. 71–77.

[16] G. Chavent and J. Jaffré, *Dynamics of Fluids in Porous Media*, Elsevier Science, Amsterdam, 1986.
[17] M. Chipot and A. Lyaghfouri, *The dam problem for non-linear Darcy's laws and non-linear leaky boundary conditions*, Math. Methods Appl. Sci., 20 (1997), pp. 1045–1068.
[18] T. Chui and D. Freyberg, *The use of COMSOL for integrated hydrological modeling*, in Proceedings of the COMSOL Conference 2007, Boston, pp. 217–223.
[19] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North–Holland, Amsterdam, 1978.
[20] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North–Holland, Amsterdam, 1976.
[21] R. Eymard, M. Gutnic, and D. Hilhorst, *The finite volume method for Richards equation*, Comput. Geosci., 3 (1999), pp. 259–294.
[22] M. Farthing, C. Kees, T. Coffey, C. Kelley, and C. Miller, *Efficient steady-state solution techniques for variably saturated groundwater flow*, Adv. Water Resour., 26 (2003), pp. 833–849.
[23] P. Forsyth and M. Kropinski, *Monotonicity considerations for saturated-unsaturated subsurface flow*, SIAM J. Sci. Comput., 18 (1997), pp. 1328–1354.
[24] J. Fuhrmann, *On numerical solution methods for nonlinear parabolic problems*, in Modeling and Computation in Environmental Sciences, First GAMM-Seminar at ICA Stuttgart, R. Helmig, W. Jäger, W. Kinzelbach, P. Knabner, and G. Wittum, eds., Vieweg, Braunschweig, 1997, pp. 170–180.
[25] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, 1984.
[26] C. Gräser, U. Sack, and O. Sander, *Truncated nonsmooth Newton multigrid methods for convex minimization problems*, in Domain Decomposition Methods in Science and Engineering XVIII, Lecture Notes in Comput. Sci. Eng. 70, Springer, Berlin, 2009.
[27] C. Kees, M. Farthing, S. Howington, E. Jenkins, and C. Kelley, *Nonlinear multilevel iterative methods for multiscale models of air/water flow in porous media*, in Proceedings of Computational Methods in Water Resources XVI, P. Binning, P. Engesgaard, H. Dahle, G. Pinder, and W. Gray, eds., Copenhagen, Denmark, 2006.
[28] N. Kikuchi and J. T. Oden, *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods*, SIAM Stud. Appl. Math. 8, SIAM, Philadelphia, 1988.
[29] R. Kornhuber, *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems*, B. G. Teubner, Stuttgart, Germany, 1997.
[30] R. Kornhuber, *On constrained Newton linearization and multigrid for variational inequalities*, Numer. Math., 91 (2002), pp. 699–721.
[31] M. Marcus and V. Mizel, *Every superposition operator mapping one Sobolev space into another is continuous*, J. Funct. Anal., 33 (1979), pp. 217–229.
[32] W. Rawls, L. Ahuja, D. Brakensiek, and A. Shirmohammadi, *Infiltration and Soil Water Movement*, in Handbook of Hydrology, D. Maidment, ed., McGraw–Hill, New York, 1993.
[33] L. A. Richards, *Capillary conduction of liquids through porous mediums*, Phys., 1 (1931), pp. 318–333.
[34] E. Schneid, P. Knabner, and F. Radu, *A priori error estimates for a mixed finite element discretization of the Richards' equation*, Numer. Math., 98 (2004), pp. 353–370.
[35] B. Schweizer, *Regularization of outflow problems in unsaturated porous media with dry regions*, J. Differential Equations, 237 (2007), pp. 278–306.
[36] M. van Dijke and S. van der Zee, *Analysis of oil lens removal by extraction through a seepage face*, Comput. Geosci., 2 (1998), pp. 47–72.
[37] M. van Genuchten, *A closed–form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil Sci. Soc. Am. J., 44 (1980), pp. 892–898.
[38] C. Wagner, G. Wittum, R. Fritsche, and H.-P. Haar, *Diffusions–Reaktionsprobleme in ungesättigten porösen Medien*, in Mathematik: Schlüsseltechnologie für die Zukunft., K.-H. H. et al., eds., Springer, Berln, 1997, pp. 243–253.
[39] H. Zheng, D. F. Liu, C. F. Lee, and L. G. Tham, *A new formulation of Signorini's type for seepage problems with free surfaces*, Internat. J. Numer. Methods Engrg., 64 (2005), pp. 1–16.