submitted to Computing and Visualization in Science
(special volume dedicated to EMG 2008)

Substructuring of a Signorini-type problem and Robin's method for the Richards equation in heterogeneous soil

Heiko Berninger · Oliver Sander

Submitted: 16 May 2009

Abstract We prove a substructuring result for a variational inequality concerning — but not restricted to — the Richards equation in homogeneous soil and including boundary conditions of Signorini's type. This generalizes existing results for the linear case and leads to interface conditions known from linear variational equalities: continuity of Dirichlet and flux values in a weak sense. In case of the Richards equation these are the continuity of the physical pressure and of the water flux, which is hydrologically reasonable. Therefore, we also apply these interface conditions in the heterogeneous case of piecewise constant soil parameters, which we address by the Robin method. We prove that, for a certain time discretization, the homogeneous problems in the subdomains including Robin and Signorini-type boundary conditions can be solved by convex minimization. As a consequence we are able to apply monotone multigrid in the discrete setting as an efficient and robust solver for the local problems. Numerical results demonstrate the applicability of our approach.

Keywords Domain decomposition methods · saturated-unsaturated porous media flow · convex minimization · monotone multigrid

Mathematics Subject Classification (2000) 65N12 · 65N30 · 65N55

This work was supported by the BMBF–Förderprogramm "Mathematik für Innovationen in Industrie und Dienstleistungen".

Department of Mathematics and Computer Science
Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
E-mail: berninger@math.fu-berlin.de

1 Introduction

Substructuring of global problems into smaller problems, that are defined on subdomains and are coupled properly, is a well-known domain decomposition technique particularly suited for heterogeneous problems [26]. A basic result for the homogeneous case concerning the situation of non-overlapping subdomains states that a linear variational equality on a domain is equivalent to corresponding linear variational equalities on the subdomains together with the continuity of Dirichlet and flux conditions across the interfaces in a weak sense, see [26, p. 7]. Now, these coupling conditions often turn out to be physically reasonable in very general settings. This justifies the use of the multi-domain formulation in heterogeneous cases, too, thus giving a definition of a global problem in these cases which might not have been available otherwise. Finally, if the treatment of the subproblems in the homogeneous case is clear, the heterogeneous problem can then be addressed by non-overlapping domain decomposition methods.

The purpose of this paper is, first, to generalize the equivalence result from homogeneous linear variational equalities to homogeneous nonlinear variational inequalities. Secondly, we choose the treatment of the Richards equation in heterogeneous soil as an application of this result in order to demonstrate that the general strategy just mentioned can also be successfully pursued in a heterogeneous case. In the presentation of our theory, both tasks will be carried out together, i.e. on the one hand, the theory is specified to the Richards equation, and on the other hand, its more general applicability shall become obvious.

The Richards equation is a well-known model for the description of saturated-unsaturated fluid flow in a porous medium [28]. It reads

$$n\theta(p)_t + \operatorname{div} \mathbf{v}(p) = 0 \quad (1)$$

with the water flux

$$\mathbf{v}(p) = -K_h kr(\theta(p))(\nabla p - z) \quad (2)$$

in case of a homogeneous soil. The unknown water or capillary pressure p is a function on $\Omega \times (0,T)$ for a time $T > 0$ and a domain $\Omega \subset \mathbb{R}^d$ $(d = 2,3)$ inhibited by the porous medium. The porosity and the hydraulic conductivity of the soil $n : \Omega \to (0,1)$ and $K_h : \Omega \to \mathbb{R}^+$, respectively, are space-dependent functions. 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 remains constant $\theta(p) = \theta_M$ if p is sufficiently large while the relative permeability $kr : [\theta_m, \theta_M] \to [0,1]$ is an increasing function of θ with $kr(\theta_M) = 1$ and $kr(\theta) \to 0$ for $\theta \to \theta_m$. The homogeneous character of (1), (2) is given by the fact that neither $\theta(\cdot)$ nor $kr(\cdot)$ depends explicitly on $x \in \Omega$, i.e. these parameter functions are fixed on Ω and thus describe the relationships in a single soil-type only. For concrete forms of these functions consult Brooks and Corey [12] or van Genuchten [30].

The reason why the Richards equation fits well into the substructuring idea indicated above is the following. With an otherwise implicit time discretization of equation (1), by which the gravitational part from (2) in z-direction is treated explicitly, and after a Kirchhoff transformation in homogeneous soil, one obtains spatial minimization problems [10]. These can be dealt with by convex analysis and their discrete versions can be solved efficiently and robustly by monotone multigrid methods. As a by-product of this approach, seepage faces around lakes, which lead to Signorini-type problems and variational inequalities, can be quite easlily treated, too. They just lead to additional constraints in the convex minimization problem.

Since Kirchhoff transformation fails to supply minimization problems in heterogeneous soil, this approach cannot be pursued for heterogeneous problems. However, one can think of heterogeneous global problems that can be substructured into local problems on subdomains with homogeneous soil [7]. Hydrologically, this is a quite reasonable situation.

As already mentioned above, the domain decomposition problem motivated by the substructuring result in the homogeneous case can be regarded as a natural definition for the heterogeneous problem. Furthermore, it enables us to set up iteration techniques like Dirichlet–Neumann or Robin methods which are already known from the linear case. A nonlinear Dirichlet–Neumann method for the stationary Richards equation without gravity in heterogeneous soil has first been considered in [9]. In the numerical example contained in this paper we demonstrate that Robin's method can be successfully applied to the time dependent Richards equation in a heterogeneous setting. The Robin condition induces an additional convex contribution to the local minimization problems which can be treated by monotone multigrid, too. Since the Robin method concerns the spatial problems only, we ignore the gravitational part in this example, which we would treat explicitly otherwise (see [10] for details). We point out that by our numerical solution procedure we solve the fully nonlinear problem without any linearization (compare e.g. [2, 15, 16]).

The paper is structured as follows. In Section 2 we prove the equivalence of a global homogeneous variational inequality with local variational inequalities on two subdomains together with suitable interface conditions in a weak form. We also derive some basic restrictions to the result concerning the position of the Signorini-type boundary. Furthermore, we define the heterogeneous problem in terms of the multi-domain formulation motivated by the equivalence result.

Starting with this domain decomposition problem in Section 3, we formulate nonlinear Dirichlet–Neumann and Robin algorithms for the Richards equation in heterogeneous soil. Furthermore, we demonstrate that the local Robin problems can be treated by convex analysis after a reformulation by Kirchhoff transformation. Therefore, with a suitable finite element discretization, monotone multigrid can be applied as a solver to the discrete problems.

In Section 4 we present a numerical example in two space dimensions in which we apply the Robin method to the Richards equation with Signorini-type boundary conditions in heterogeneous soil. We observe good multigrid performance and reasonable convergence rates of the Robin iteration.

2 Substructuring of a Signorini-type problem for the Richards equation in heterogeneous soil

The purpose of this section is to obtain a weak formulation of a Signorini-type problem for the Richards equation in heterogeneous soil. This is achieved via substructuring of a corresponding problem in homogeneous soil, which leads to an equivalence between the global problem and local problems that are coupled by suitable interface conditions. The latter set of problems is then taken as a definition of a Signorini-type problem for the Richards equation in heterogeneous soil.

We start with the global homogeneous problem and some necessary notation in Subsection 2.1. Then, Subsection 2.2 contains the theoretical results for the substructuring of this problem (foremost in Theorem 1). Finally, in Subsection 2.3 we interpret this domain decomposition for the Richards equation and note Definition 1, which gives sense to a Signorini-type problem for the Richards equation in a heterogeneous setting.

2.1 Global problem and notation

The setting that we want to consider is given by a decomposition of a bounded open Lipschitz domain $\Omega \subset \mathbb{R}^d$ into two non-overlapping open and nonempty subdomains Ω_1 and Ω_2 (i.e. $\overline{\Omega}_1 \cup \overline{\Omega}_2 = \overline{\Omega}$ and $\Omega_1 \cap \Omega_2 = \emptyset$) with the interface

$$\Gamma := \overline{\Omega}_1 \cap \overline{\Omega}_2 .$$

As in [26, p. 6] we assume that Γ is a $(d-1)$-dimensional Lipschitz manifold so that common results on trace spaces (see [26, p. 339/340] and [11, pp. 1.56–1.65]) are applicable. In addition, both Ω_1 and Ω_2 are assumed to have Lipschitz boundaries. Figure 1 displays such a situation for $d = 2$, which in case of the Richards equation can be interpreted as two horizontal layers of different soil types.

Fig. 1 2D-domain Ω decomposed into two subdomains

As already stated in the introduction, many of our considerations here are motivated by and generalizations of the corresponding theory presented in Quarteroni and Valli [26] for the linear case. However, since we consider weak formulations of Signorini-type problems, that involve nonlinearities and convex sets in Sobolev spaces rather than the full spaces, our notation needs to be different. We use the notation in [26] wherever it is appropriate, for example, we often abbreviate the "restriction" of a $p \in H^1(\Omega)$ to $\Sigma \subset \partial\Omega$ by $p_{|\Sigma}$ where more precisely the trace $tr_\Sigma\, p$ with the trace operator $tr_\Sigma : H^1(\Omega) \to H^{1/2}(\Sigma)$ is meant. Moreover, for different domains and $i \in \mathbb{N}$ we denote H^1-norms by

$$\|v\|_{1,\Omega} := \|v\|_{H^1(\Omega)} \quad \forall v \in H^1(\Omega)$$

and

$$\|v_i\|_{1,\Omega_i} := \|v_i\|_{H^1(\Omega_i)} \quad \forall v_i \in H^1(\Omega_i)\,.$$

Finally, for Lipschitz submanifolds $\gamma \subset \partial\Omega$ we also need the space

$$H^1_\gamma(\Omega) := \{v \in H^1(\Omega) : v_{|\gamma} = 0\}$$

incorporating homogeneous Dirichlet boundary conditions (analogously for Ω_i, $i = 1, 2$).

We start with the homogeneous case, i.e. we assume constant soil parameters on the whole domain Ω which lead to globally space-independent nonlinearities $\theta(\cdot)$ and $kr(\cdot)$. For simplicity, we set $n = K_h = 1$ from now on. Moreover, we consider a time discretization of (1) which is implicit except for the gravitational part of (2) in z-direction which we treat explicitly. Concretely, with a time step $\tau > 0$, a known physical pressure $\tilde{p} \in H^1(\Omega)$ from the previous time step and the definition $e_z := \nabla z$ we obtain

$$\frac{\theta(p) - \theta(\tilde{p})}{\tau} + \operatorname{div} \tilde{\mathbf{v}} = 0 \tag{3}$$

as the semi-discrete version of (1) in which the time-discretized water flux $\tilde{\mathbf{v}}$ is given by

$$\tilde{\mathbf{v}}(p) = -kr(\theta(p))\nabla p + kr(\theta(\tilde{p}))e_z\,. \tag{4}$$

Now, given a decomposition of $\partial\Omega$ into non-overlapping Lipschitz manifolds γ_D, γ_N and γ_S and suitable $p_D \in H^{1/2}(\gamma_D)$ as well as $f_N \in L^2(\gamma_N)$ we impose the boundary conditions

$$p = p_D \quad \text{on } \gamma_D \tag{5}$$

$$\tilde{\mathbf{v}} \cdot \mathbf{n} = f_N \quad \text{on } \gamma_N \tag{6}$$

$$p \le 0\,, \quad \tilde{\mathbf{v}} \cdot \mathbf{n} \ge 0\,, \quad p \cdot (\tilde{\mathbf{v}} \cdot \mathbf{n}) = 0 \quad \text{on } \gamma_S \tag{7}$$

for the spatial problem (3), (4). Here, $\mathbf{n}$ is the outward normal of Ω on its boundary. The conditions on $\tilde{\mathbf{v}}$ are completely analogous to the ones imposed on $\mathbf{v}$ in the continuous case [10].

Condition (7) is an outflow condition which occurs e.g. at seepage faces around lakes where water might flow out of the soil ($\tilde{\mathbf{v}} \cdot \mathbf{n} > 0$, in which case we have $p = 0$) but not into it, and where the water is at most at atmospheric pressure $p = 0$. Mathematically, this condition is the same as the Signorini boundary conditions for obstacle problems (compare e.g. [19]). Therefore, we call it a Signorini-type boundary condition here, and we refer to problems containing them as problems of Signorini's type. Observe that if $\overline{\gamma}_D \cup \overline{\gamma}_S \neq \emptyset$ this condition restricts the admissible set of Dirichlet values p_D.

Now, with the convex set

$$\mathcal{K}_0 := \{v \in H^1(\Omega) : v_{|\gamma_D} = p_D \,\wedge\, v_{|\gamma_S} \le 0\}$$

we consider the variational inequality

$$
\begin{aligned}
&p \in \mathcal{K}_0 : \\
&\int_\Omega \theta(p)\,(v-p)\,dx + \tau \int_\Omega kr(\theta(p))\nabla p\,\nabla(v-p)\,dx \geq \\
&\int_\Omega \theta(\tilde{p})\,(v-p)\,dx + \tau \int_\Omega kr(\theta(\tilde{p}))e_z\nabla(v-p)\,dx \\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall v \in \mathcal{K}_0\,. \quad (8)
\end{aligned}
$$

Setting $f_N(t) = 0$ for simplicity it is not hard to see that (8) is a weak formulation for the time-discretized Richards equation (3), (4) with boundary conditions (5)–(7). For further justification and discussion we refer the reader to [7, Prop. 1.5.3].

For notational reasons and in order to make clear that only the general form of this variational inequality is of importance we introduce the following abbreviations. We begin by setting

$$
\tilde{a}(p, v-p) := \int_\Omega \theta(p)\,(v-p)\,dx + \tau \int_\Omega kr(\theta(p))\nabla p\,\nabla(v-p)\,dx
$$

and

$$
\ell(v-p) := \int_\Omega \theta(\tilde{p})\,(v-p)\,dx + \tau \int_\Omega kr(\theta(\tilde{p}))e_z\nabla(v-p)\,dx
$$

for all $v \in \mathcal{K}_0$ with the solution $p \in \mathcal{K}_0$ in (8) if it exists. However, rather than this special definition, it turns out that the following properties of $\tilde{a}(\cdot,\cdot)$ and $\ell(\cdot)$ are essential. We assume that $\tilde{a}(\cdot,\cdot)$ is a form on $(H^1(\Omega))^2$ which may be nonlinear in the first but has to be linear in the second entry, while $\ell(\cdot)$ is a linear form on $H^1(\Omega)$. Analogously, we introduce the convex sets

$$
\mathcal{K}_i := \{v \in H^1(\Omega_i) : v_{|\gamma_{D_i}} = p_{D|\gamma_{D_i}} \ \wedge\ v_{|\gamma_{S_i}} \leq 0\}
$$

as well as the forms

$$
\tilde{a}_i(p_i, v_i - p_i) := \int_{\Omega_i} \theta(p_i)\,(v_i - p_i)\,dx + \tau \int_{\Omega_i} kr(\theta(p_i))\nabla p_i\,\nabla(v_i - p_i)\,dx
$$

and

$$
\ell_i(v_i - p_i) := \int_{\Omega_i} \theta(\tilde{p}_i)\,(v_i - p_i)\,dx + \tau \int_\Omega kr(\theta(\tilde{p}_i))e_z\nabla(v_i - p_i)\,dx
$$

for all $v_i \in \mathcal{K}_i$ with $p_i \in \mathcal{K}_i$ and given $\tilde{p}_i := \tilde{p}_{|\Omega_i}$ for $i = 1, 2$, which correspond to the subdomains Ω_1 and Ω_2. Here we have used the definitions $\gamma_{D_i} := \partial\Omega_i \cap \gamma_D$ and $\gamma_{S_i} := \partial\Omega_i \cap \gamma_S$.

We also need to introduce convex sets with prescribed Dirichlet values on the interface which we define as

$$
\mathcal{K}_i^{p_j} := \{v \in \mathcal{K}_i : v_{|\Gamma} = p_{j|\Gamma}\}
$$

for $p_j \in \mathcal{K}_j$ and $i, j \in \{1, 2\}$. In addition, we introduce the convex set of traces

$$
\Lambda_0 := \{\eta \in H^{1/2}(\Gamma) : \eta = v_{|\Gamma} \text{ for a } v \in \mathcal{K}_0\}
$$

and its translated copy

$$
\begin{aligned}
\tilde{\Lambda} :&= \Lambda_0 - p_{|\Gamma} \\
&= \{\eta \in H^{1/2}(\Gamma) : \eta = v_{|\Gamma} \text{ for a } v \in \mathcal{K}_0 - p\}\,.
\end{aligned}
$$

with a $p \in \mathcal{K}_0$ (which will later be the assumed solution of (8)). We refer to Lemma 1 to make sure that traces of $H^1(\Omega)$-functions in the interior of Ω are well defined.

With respect to the setting for the Poisson problem considered in [26, p. 6] we note that our convex sets degenerate and fit into that setting if we only have homogeneous Dirichlet values imposed on $\partial\Omega$. More concretely, for $i = 1, 2$ we obtain the spaces

$$
\begin{aligned}
&\mathcal{K}_0 = H_0^1(\Omega) \\
&\mathcal{K}_i = \{v \in H^1(\Omega_i) : v_{i|\partial\Omega\cap\partial\Omega_i} = 0\} \\
&\mathcal{K}_i^{p_i} - p_i = H_0^1(\Omega_i) \\
&\Lambda_0 = \tilde{\Lambda} = \{\eta \in H^{1/2}(\Gamma) : \eta = v_{|\Gamma} \text{ for a } v \in H_0^1(\Omega)\}
\end{aligned}
$$

in this case. We recall that the latter trace space is $H^{1/2}(\Gamma)$ for $\Gamma \cap \partial\Omega = \emptyset$ and $H_{00}^{1/2}(\Gamma)$ if $\Gamma \cap \partial\Omega \neq \emptyset$ which is the case we mostly consider here (compare [11, pp. 1.60] and [26, pp. 6/7]). Obviously, the structure of $\tilde{\Lambda}$ is more delicate in our general case since here we can only guarantee that $\tilde{\Lambda}$ is a convex subset of $H^{1/2}(\Gamma)$. For our equivalence result (Theorem 1), however, we need the vector space structure of $\tilde{\Lambda}$ which we cannot expect if $\Gamma \cap \overline{\gamma}_S \neq \emptyset$.

2.2 Substructuring equivalence result in a homogeneous setting

We start with a result which guarantees that $\tilde{\Lambda}$ is a vector space.

Proposition 1 *We assume $\Gamma \cap \overline{\gamma}_S = \emptyset$. Then $\tilde{\Lambda}$ is a subspace of $H^{1/2}(\Gamma)$ with the property*

$$
\tilde{\Lambda} = \{\eta \in H^{1/2}(\Gamma) : \eta = v_{|\Gamma} \text{ for a } v \in H^1_{\gamma_D \cup \gamma_S}(\Omega)\}\,, \qquad (9)
$$

i.e. containing $H_{00}^{1/2}(\Gamma)$. If, in addition, $\Gamma\cap\overline{\gamma}_N=\emptyset$ and $\Gamma\cap\partial\Omega\neq\emptyset$, or $\Gamma\cap\overline{\gamma}_D=\emptyset$, then we have $\tilde{\Lambda}=H_{00}^{1/2}(\Gamma)$, or $\tilde{\Lambda}=H^{1/2}(\Gamma)$, respectively. In the general case $\tilde{\Lambda}$ is a Hilbert space with the quotient norm

$$\|\eta\|_{\tilde{\Lambda}}=\inf\left\{\|v\|_{1,\Omega}:v\in H^1_{\gamma_D\cup\gamma_S}(\Omega)\wedge\eta=v_{|\Gamma}\right\}. \quad (10)$$

Moreover, with the subspace

$$\tilde{H}^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i):=\{v\in H^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i):v_{i|\Gamma}\in\tilde{\Lambda}\}$$

of the Hilbert space $H^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i)$ the trace operator

$$tr_\Gamma:H^1_{\gamma_D\cup\gamma_S}(\Omega)\to\tilde{\Lambda} \quad (11)$$

induces continuous linear trace operators

$$tr_{\Gamma,i}:\tilde{H}^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i)\to\tilde{\Lambda},\quad i=1,2,$$

for which, in addition, continuous linear extension operators

$$R_i:\tilde{\Lambda}\to H^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i),\quad i=1,2, \quad (12)$$

with $tr_{\Gamma,i}R_i\eta=\eta$ for all $\eta\in\tilde{\Lambda}$ exist.

Proof In order to see "$\supset$" in (9) observe that $v+p\in\mathcal{K}_0$ for $p\in\mathcal{K}_0$ and any $v\in H^1_{\gamma_D\cup\gamma_S}(\Omega)$. Conversely, since we have $\mathrm{dist}(\Gamma,\gamma_S)>0$ or $\gamma_S=\emptyset$ there are open neighbourhoods O_Γ and O_{γ_S} of Γ and γ_S, respectively, with $O_\Gamma\cap O_{\gamma_S}=\emptyset$ and an open ball $B\subset\mathbb{R}^d$ with $\overline{O}_\Gamma\cup\overline{O}_{\gamma_S}\cup\overline{\Omega}\subset B$. It is well known that there is a $\varphi\in C_0^\infty(B)$ with a range in $[0,1]$ satisfying $\varphi_{|O_{\gamma_S}}=0$ and $\varphi_{|O_\Gamma}=1$, consult e.g. [20, p. 277]. Let $\eta\in\tilde{\Lambda}$ and $v\in\mathcal{K}_0$ with $v_{|\Gamma}=\eta$. Then one can check with the Leibniz rule (compare e.g. [1, p. 21]) and $\varphi_{|\Omega}\in W^{1,\infty}(\Omega)$ that $\varphi v\in H^1_{\gamma_D\cup\gamma_S}(\Omega)$ holds. Moreover, we have $(\varphi v)_{|\Gamma}=v_{|\Gamma}=\eta$. In particular, (9) entails $\tilde{\Lambda}\supset H_{00}^{1/2}(\Gamma)$. Note that the arguments can also be applied to the case $\gamma_S=\emptyset$.

If, in addition, $\Gamma\cap\overline{\gamma}_N=\emptyset$, then we can replace $H^1_{\gamma_D\cup\gamma_S}(\Omega)$ by $H_0^1(\Omega)$ in (9) and obtain $\tilde{\Lambda}=H_{00}^{1/2}(\Gamma)$ due to $\Gamma\cap\partial\Omega\neq\emptyset$. This can be seen in the same manner as above: If, now, O_Γ and O_{γ_N} are chosen analogously as O_Γ and O_{γ_S} before, and a function $\varphi\in C_0^\infty(B)$ with $\varphi_{|O_{\gamma_N}}=0$ and $\varphi_{|O_\Gamma}=1$ is at hand, then for any $v\in H^1_{\gamma_D\cup\gamma_S}(\Omega)$ the function $\varphi v\in H_0^1(\Omega)$ satisfies $(\varphi v)_{|\Gamma}=v_{|\Gamma}$.

If, instead, we have $\Gamma\cap\overline{\gamma}_D=\emptyset$, then we can replace $H^1_{\gamma_D\cup\gamma_S}(\Omega)$ by $H^1(\Omega)$ in (9). Now we choose O_Γ and $O_{\gamma_D\cup\gamma_S}$ analogously as O_Γ and O_{γ_S} above and a $\varphi\in C_0^\infty(B)$ with $\varphi_{|O_{\gamma_D\cup\gamma_S}}=0$ and $\varphi_{|O_\Gamma}=1$. As a consequence, for any $v\in H^1(\Omega)$ we have $\varphi v\in H^1_{\gamma_D\cup\gamma_S}(\Omega)$ and $(\varphi v)_{|\Gamma}=v_{|\Gamma}$.

With regard to the general case it is easily checked that the quotient norm in (10) is indeed a norm (compare [31, p. 34]). With this norm, tr_Γ in (11) is a quotient map and therefore $\tilde{\Lambda}$ is isometrically isomorphic to the quotient $H^1_{\gamma_D\cup\gamma_S}(\Omega)/\ker(tr_\Gamma)$, see [31, pp. 54, 56]. Since $H^1_{\gamma_D\cup\gamma_S}(\Omega)$ is a Hilbert space we have the canonical representation $H^1_{\gamma_D\cup\gamma_S}(\Omega)=\ker(tr_\Gamma)\oplus\ker(tr_\Gamma)^\perp$ in which $\ker(tr_\Gamma)^\perp$ is the orthogonal complement of the (closed) kernel $\ker(tr_\Gamma)$, see [31, p. 221]. Therefore, we can conclude the isometric isomorphisms

$$\ker(tr_\Gamma)^\perp\cong H^1_{\gamma_D\cup\gamma_S}(\Omega)/\ker(tr_\Gamma)\cong\tilde{\Lambda}$$

in which tr_Γ induces the isomorphism $\ker(tr_\Gamma)^\perp\cong\tilde{\Lambda}$. In particular, $\tilde{\Lambda}$ is a Hilbert space. The inverse

$$R:\tilde{\Lambda}\to\ker(tr_\Gamma)^\perp\subset H^1_{\gamma_D\cup\gamma_S}(\Omega)$$

of tr_Γ restricted to $\ker(tr_\Gamma)^\perp$ is a continuous linear map with the property $tr_\Gamma R\eta=\eta$ for all $\eta\in\tilde{\Lambda}$. The definition $R_i\eta:=(R\eta)_{|\Omega_i}$ for all $\eta\in\tilde{\Lambda}$ and $i=1,2$ provides continuous linear operators (12) with the properties

$$\|R_i\eta\|_{1,\Omega_i}\leq\|R\eta\|_{1,\Omega}=\|\eta\|_{\tilde{\Lambda}}\quad\forall\eta\in\tilde{\Lambda}$$

and (with a glance at Lemma 1) $tr_{\Gamma,i}R_i\eta=tr_\Gamma R\eta=\eta$ as required. □

Observe that for the existence of the extension operators R_i, $i=1,2$, we needed to use the Hilbert space structure of $H^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i)$, in particular the existence of an orthogonal complement of $\ker(tr_\Gamma)$. In contrast, a closed subspace in a general Sobolev space $W^{1,p}_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega)$ for $p\geq 1$ and $p\neq 2$ does not necessarily have a complemented subspace in $W^{1,p}_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega)$, see [31, pp. 162, 248]. In such a case one would be forced to define extension operators as in (12) or, equivalently, projections from $W^{1,p}_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega)$ on $\ker(tr_\Gamma)$ more explicitly. Concerning this question, which becomes relevant for corresponding generalizations of our main result in Theorem 1, we refer to Lions and Magenes [23, pp. 19–23, 38–43].

Remark 1 We suppose that (9) is still true if at least $\Gamma\cap\overline{\gamma}_S\subset\Gamma\cap\overline{\gamma}_D$ holds. Furthermore, although one might be used to definitions of trace spaces as collections of traces from functions in another Sobolev space (compare e.g. [11, p. 1.60]), we recall that the space $H_{00}^{1/2}(\Gamma)$ is also intrinsically definable (see [26, p. 7]) and thus only dependent on Γ. Considering this, it becomes plausible that $\tilde{\Lambda}=H_{00}^{1/2}(\Gamma)$ holds whenever $\Gamma\cap\overline{\gamma}_D=\Gamma\cap\partial\Omega$ is satisfied (clearly, at least if we have $\Gamma\cup\overline{\gamma}_D\subset\partial\tilde{\Omega}$ with some Lipschitz domain $\tilde{\Omega}$). In case of $(\Gamma\cap\overline{\gamma}_S)\backslash\overline{\gamma}_D\neq\emptyset$, however, it seems that $\tilde{\Lambda}$ is no longer a vector space. Then there are no extension operators as in (12), either, see Proposition 2.

Although the following property of Sobolev functions is elementary, we note it here since it is crucial for any substructuring in $H^1(\Omega)$.

Lemma 1 *If $p \in H^1(\Omega)$, then we have $p_i := p_{|\Omega_i} \in H^1(\Omega_i)$ for $i = 1, 2$ and $p_{1|\Gamma} = p_{2|\Gamma}$. Conversely, if $p_i \in H^1(\Omega_i)$ for $i = 1, 2$ and $p_{1|\Gamma} = p_{2|\Gamma}$ holds, then*

$$p := \begin{cases} p_1 & \text{on } \Omega_1 \\ p_2 & \text{on } \Omega_2 \end{cases} \tag{13}$$

is contained in $H^1(\Omega)$.

Proof The first assertion is easy to see by considering a sequence of functions $(\varphi_n)_{n\in\mathbb{N}} \subset C^\infty(\overline{\Omega})$ converging to p in $H^1(\Omega)$ and observing that their restrictions to Ω_i or to Γ converge to p_i or to $p_{i|\Gamma}$, respectively, for $i = 1, 2$ in the corresponding norms. Conversely, to see that p as defined by (13) is weakly differentiable we apply partial integration (see e.g. [26, p. 340]) in $H^1(\Omega_i)$ to the weak derivatives of p_i for $i = 1, 2$ tested with test functions in $C_0^\infty(\overline{\Omega})$ and observe that the contributions on Γ cancel each other due to $p_{1|\Gamma} = p_{2|\Gamma}$. □

We can now prove the main result of this section which is a generalization of Lemma 1.2.1 in Quarteroni and Valli [26] to problems of Signorini's type for nonlinear equations. Recall that extension operators are defined as right inverses to corresponding trace maps.

Theorem 1 *Let $\Gamma \cap \overline{\gamma}_S = \emptyset$. Then the variational problem (8) which in short reads*

$$p \in \mathcal{K}_0 : \quad \tilde{a}(p, v - p) - \ell(v - p) \geq 0 \quad \forall v \in \mathcal{K}_0 \tag{14}$$

can be equivalently reformulated as: Find $p_1 \in \mathcal{K}_1$ and $p_2 \in \mathcal{K}_2$ such that

$$\tilde{a}_1(p_1, v_1 - p_1) - \ell_1(v_1 - p_1) \geq 0 \quad \forall v_1 \in \mathcal{K}_1^{p_1} \tag{15}$$

$$p_1 = p_2 \quad \text{on } \Gamma \tag{16}$$

$$\tilde{a}_2(p_2, v_2 - p_2) - \ell_2(v_2 - p_2) \geq 0 \quad \forall v_2 \in \mathcal{K}_2^{p_2} \tag{17}$$

$$\tilde{a}_2(p_2, R_2\mu) = \ell_2(R_2\mu) + \ell_1(R_1\mu) - \tilde{a}_1(p_1, R_1\mu) \quad \forall \mu \in \tilde{\Lambda} \tag{18}$$

where R_i denotes any possible extension operator from $\tilde{\Lambda}$ to $H^1_{\gamma_{D_i}\cup\gamma_{S_i}}(\Omega_i)$ for $i = 1, 2$.

Note that R_i, $i = 1, 2$, exist and can be chosen as the continuous linear extension operators given by Proposition 1.

Proof First let p be a solution of (14). Then we have $p_i := p_{|\Omega_i} \in \mathcal{K}_i$ for $i = 1, 2$ and (16) due to Lemma 1. Let $v_1 \in \mathcal{K}_1^{p_1}$. Since (16) holds, the function

$$v := \begin{cases} v_1 & \text{on } \Omega_1 \\ p_2 & \text{on } \Omega_2 \end{cases}$$

is contained in $\mathcal{K}_0$ (by Lemma 1), and (15) follows from (14). Analogously, we obtain (17). Now, for each $\mu \in \tilde{\Lambda}$ the function $R\mu$ defined by

$$R\mu := \begin{cases} R_1\mu & \text{on } \Omega_1 \\ R_2\mu & \text{on } \Omega_2 \end{cases} \tag{19}$$

belongs to $H^1_{\gamma_D\cup\gamma_S}(\Omega)$ due to Lemma 1 and we have $\pm R\mu + p \in \mathcal{K}_0$. The variational inequality (14) applied to both $v = R\mu + p \in \mathcal{K}_0$ and $v = -R\mu + p \in \mathcal{K}_0$ provides (18).

Conversely, let $p_i \in \mathcal{K}_i^{p_i}$, $i = 1, 2$, be solutions of (15)–(18). Setting

$$p := \begin{cases} p_1 & \text{on } \Omega_1 \\ p_2 & \text{on } \Omega_2 \end{cases}$$

we obtain $p \in \mathcal{K}_0$ due to (16), the definitions of $\mathcal{K}_i^{p_i}$ and Lemma 1. Choosing a $v \in \mathcal{K}_0$ we set $\mu := v_{|\Gamma}$ and $\lambda := p_{|\Gamma}$ and obtain $\mu - \lambda \in \tilde{\Lambda}$ by definition of $\tilde{\Lambda}$. Defining $R(\mu - \lambda)$ according to (19) we see that

$$v_i := v_{|\Omega_i} - R_i(\mu - \lambda) \in \mathcal{K}_i^{p_i}, \quad i = 1, 2,$$

holds. Now, (15), (17) and (18) entail

$$\begin{aligned} \tilde{a}(p, v - p) - \ell(v - p) &= \sum_{i=1}^{2} \tilde{a}_i(p_i, v_{|\Omega_i} - p_i) - \ell_i(v_{|\Omega_i} - p_i) \\ &= \sum_{i=1}^{2} \Big(\tilde{a}_i(p_i, v_i - p_i) - \ell_i(v_i - p_i) \\ &\qquad + \tilde{a}_i(p_i, R_i(\mu - \lambda)) - \ell_i(R_i(\mu - \lambda)) \Big) \geq 0 \end{aligned}$$

as required. □

Remark 2 We point out that it seems unrealistic to generalize Theorem 1 in a satisfying way to situations in which Γ and $\overline{\gamma}_S$ have a nonempty intersection (and thus $\tilde{\Lambda}$ is in general no longer a vector space). Observe that for the second part of the proof we need extension operators

$$R_i : \tilde{\Lambda} \to \mathcal{K}_i - \mathcal{K}_i^{p_i} \subset H^1_{\gamma_{D_i}}(\Omega_i), \quad i = 1, 2, \tag{20}$$

not necessarily linear or continuous, with the property

$$(v - R(v_{|\Gamma} - p_{|\Gamma}))_{|\gamma_S} \leq 0 \quad \forall v \in \mathcal{K}_0 \tag{21}$$

(with R as in (19)) and for which (18) is satisfied with "$\geq$" instead of "$=$". Indeed, with this modified condition (18) we obtain equivalence — if such extension operators exist. However, we have the following proposition, in which the second assertion presumably also holds in all cases where the intersection of Γ and $\overline{\gamma}_S$ leads to a $\tilde{\Lambda}$ without a vector space structure.

Proposition 2 *Let $p \in \mathcal{K}_0$ and $R : \tilde{\Lambda} \to H^1_{\gamma_D}(\Omega)$ be some (not necessarily linear or continuous) extension operator, i.e. $tr_\Gamma R\mu = \mu$ for all $\mu \in \tilde{\Lambda}$. In addition, assume with (19) that R satisfies (20) and (21). Then we have*

$$R : \tilde{\Lambda} \to \{v \in H^1_{\gamma_D}(\Omega) : v_{|\gamma_S} \geq 0\} . \quad (22)$$

In particular, if $(\Gamma \cap \overline{\gamma}_S)\backslash\overline{\gamma}_D \neq \emptyset$ holds, then such a map R does not exist.

Proof Without loss of generality we start by assuming $p = 0$. Then (21) reads

$$v_{|\gamma_S} \leq (R(v_{|\Gamma}))_{|\gamma_S} \quad \forall v \in \mathcal{K}_0 . \quad (23)$$

The following construction can be carried out as in the proof of Proposition 1. Assuming that γ_S is open in the relative topology of $\partial\Omega$, we consider neighbourhoods $O_n \subset \mathbb{R}^d$ of $\overline{\gamma}_S\backslash\gamma_S$ and neighbourhoods U_n of Γ for $n \in \mathbb{N}$ with Lebesgue measure $|O_n|$, $|U_n| \to 0$ for $n \to \infty$. Now, for any $\eta \in \tilde{\Lambda}$ it is possible to construct a sequence of functions $(v_n)_{n\in\mathbb{N}} \subset \mathcal{K}_0$ that satisfy $v_{n|\Gamma} = \eta$ and $v_{n|\gamma_S\backslash(O_n\cup U_n)} = 0$ for all $n \in \mathbb{N}$. Now, it follows from (23) that $(R\eta)_{|\gamma_S} \geq 0$ holds for all $\eta \in \tilde{\Lambda}$. This proves the first assertion (22) of Proposition 2, which can analogously be obtained for arbitrary $p \in \mathcal{K}_0$.

Consequently, $\tilde{\Lambda}$ is contained in

$$tr_\Gamma\Big(\{v \in H^1_{\gamma_D}(\Omega) : v_{|\gamma_S} \geq 0\}\Big) .$$

However, by definition, $\tilde{\Lambda}$ is the set

$$tr_\Gamma\Big(\{v \in H^1_{\gamma_D}(\Omega) : (v - p)_{|\gamma_S} \leq 0\}\Big) .$$

This leads to a contradiction if $\gamma := (\Gamma \cap \overline{\gamma}_S)\backslash\overline{\gamma}_D \neq \emptyset$. Indeed, we can pick a point $x \in \gamma$, a neighbourhood U_x of this point and a function $\eta \in \tilde{\Lambda}$ satisfying $\eta \leq -1$ in $U_x \cap \Gamma$ where $\tilde{\eta} := (R\eta)_{|\gamma_S} \geq 0$ holds although we have $\eta = v_{|\Gamma}$ and $\tilde{\eta} = v_{|\gamma_S}$ for a $v \in H^1_{\gamma_D}(\Omega)$. As a consequence $tr_{\Gamma\cup\overline{\gamma}_S} v \in H^{1/2}(\Gamma \cup \overline{\gamma}_S)$ has a discontinuity of first order in x. This, however, cannot be the case for $H^{1/2}$-functions, for which the Sobolev–Slobodeckij–norm has to be finite (see e.g. [1, p. 214]) □

Remark 3 Although our considerations in Remark 2 lead to the non-existence result in Proposition 2 they still contain a positive message for the discrete setting. If we discretize the problems (14) and (15)–(18) appropriately with finite elements (see e.g. [7, Sec. 2.5]), then Theorem 1 and Remark 2 can be established accordingly in the discrete setting. Now, however, the properties (20) and (21) are satisfied if we consider R to be the trivial extension, setting $R\mu$ as 0 on the nodes in $\Omega\backslash\Gamma$ while respecting the Dirichlet values. In this case we also obtain an equivalence between the discretized version of (14) and the discretized versions of (15)–(18), where "=" is replaced by "≥" in the discretization of (18).

2.3 Domain decomposition for the Richards equation in heterogeneous soil

As far as the interpretation of the interface conditions is concerned, it is clear that (16) indicates the continuity of the pressure p across the interface Γ. Furthermore, with the help of Green's formula (see [11, p. 1.65]) one can easily verify that (18) is the weak formulation for the continuity of the implicit-explicitly time-discretized flux

$$-\Big(kr(\theta(p_1))\nabla p_1 - kr(\theta(\tilde{p}_1))e_z\Big) \cdot \mathbf{n} = -\Big(kr(\theta(p_2))\nabla p_2 - kr(\theta(\tilde{p}_2))e_z\Big) \cdot \mathbf{n} \quad \text{on } \Gamma \quad (24)$$

corresponding to the implicit-explicitly time-discretized Richards equation (3), (4). Here, we adapt to the usual convention that $\mathbf{n}$ is the outward normal of the subdomain Ω_1 (see Figure 1 and compare [26, pp. 1/2]).

We remark that one can also establish Theorem 1 for the variational form of the Richards equation (1), (2) before time discretization. Then, we have $p_i = \tilde{p}_i$ for $i = 1, 2$ in (24), and the strong form of (18) is just the continuity of the water flux (2) at the time t.

Now we turn to the case of heterogeneous soil, i.e. to the case of possibly different parameter functions $\theta_1(\cdot)$ and $kr_1(\cdot)$ in Ω_1, and $\theta_2(\cdot)$ and $kr_2(\cdot)$ in Ω_2. In analogy to the common term *jumping coefficients* we refer to this situation as *jumping nonlinearities* (see [7]). We assume that $\tilde{a}_i(\cdot,\cdot)$ and $\ell_i(\cdot)$, $i = 1, 2$, are defined according to Subsection 2.1. With these ingredients one can also define $\tilde{a}(\cdot,\cdot)$ and $\ell(\cdot)$ on Ω and give sense to the corresponding variational inequality (14).

However, our solution theory for the Richards equation in homogeneous soil (compare [7, Ch. 4] or [10]) cannot be applied in this heterogeneous setting since the Kirchhoff transformation does not lead to a minimization problem if applied globally on Ω. Therefore, we turn to the corresponding substructuring problem (15)–(18) for which arising local Dirichlet, Neumann and Robin problems turn out to be uniquely solvable after suitable (domain-dependent) Kirchhoff transformations in the subdomains. The interface conditions discussed above also seem hydrologically justified. Continuity of the (time-discretized) water flux represents the mass conservation whereas continuity of the physical pressure (16) is assumed in several standard models [17]. (See [18] for cases with discontinuity which could in principle also be considered here.) Therefore, we note

Definition 1 Let $\Gamma \cap \overline{\gamma}_S = \emptyset$ and $\theta_i(\cdot)$, $kr_i(\cdot)$, $i = 1, 2$, be given (possibly different) parameter functions on Ω_i. Let $\tilde{a}_i(\cdot, \cdot)$, $\ell_i(\cdot)$, $i = 1, 2$, be defined with these functions according to Subsection 2.1. We call a function p defined a.e. on Ω with $p_i := p_{|\Omega_i}$, $i = 1, 2$, a *weak solution of the corresponding Signorini-type problem for the Richards equation in heterogeneous soil* on Ω if we have $p_1 \in \mathcal{K}_1$ and $p_2 \in \mathcal{K}_2$ such that

$$\tilde{a}_1(p_1, v_1 - p_1) - \ell_1(v_1 - p_1) \geq 0 \quad \forall v_1 \in \mathcal{K}_1^{p_1} \tag{25}$$

$$p_1 = p_2 \quad \text{on } \Gamma \tag{26}$$

$$\tilde{a}_2(p_2, v_2 - p_2) - \ell_2(v_2 - p_2) \geq 0 \quad \forall v_2 \in \mathcal{K}_2^{p_2} \tag{27}$$

$$\tilde{a}_2(p_2, R_2\mu) = \ell_2(R_2\mu) + \ell_1(R_1\mu) - \tilde{a}_1(p_1, R_1\mu) \quad \forall \mu \in \tilde{\Lambda} \tag{28}$$

where R_i denotes any possible extension operator from $\tilde{\Lambda}$ to $H^1_{\gamma_{D_i} \cup \gamma_{S_i}}(\Omega_i)$ for $i = 1, 2$.

As usual, for more than two subdomains, one considers convex problems for each subdomain and imposes continuity of the pressure and the (time-discretized) water flux on each interface.

3 Solution of the Richards equation in heterogeneous soil

This section is devoted to the question of how the domain decomposition problem for the Richards equation in heterogeneous soil as given in Definition 1 can be solved. To this end, we first introduce nonlinear Dirichlet–Neumann and Robin schemes in Subsection 3.1. These address the treatment of the heterogeneity. Then, in Subsection 3.2, we make clear that the local Robin problems (33) and (34) for the homogeneous Richards equation are associated to convex minimization problems which are uniquely solvable under reasonable assumptions on the parameter functions. Finally, in Subsection 3.3, we introduce the finite element discretization of these problems and make clear that the monotone multigrid method can be applied to the discrete problems.

3.1 Dirichlet–Neumann and Robin schemes for the domain decomposition problem

As in the linear case in [26, p. 7], the subproblems (25) and (27) are underdetermined because of lacking boundary values for p_i, $i = 1, 2$, on Γ. If these problems involve nonempty Signorini-type boundaries, then even the convex sets of test functions $\mathcal{K}_i^{p_i}$ are unknown a priori. This is not the case if $\gamma_{S_i} = \emptyset$ since then $\mathcal{K}_i^{p_i} - p_i = H^1_0(\Omega_i)$ is a vector space. Nevertheless, in order to tackle the domain decomposition problem (25)–(28) analytically or numerically, it is essential to formulate uniquely solvable local subproblems.

As in the linear case this can be done by establishing iterative schemes with the help of the Dirichlet interface condition (26) and the Neumann interface condition (28). Given an iterate $p_2^k \in \mathcal{K}_2$ for a $k \geq 0$, one can compute an iterate $p_1^{k+1} \in \mathcal{K}_1$ by solving the variational inequality

$$p_1^{k+1} \in \mathcal{K}_1^{p_2^k}: \ \tilde{a}_1(p_1^{k+1}, v_1 - p_1^{k+1}) - \ell_1(v_1 - p_1^{k+1}) \geq 0 \quad \forall v_1 \in \mathcal{K}_1^{p_2^k} \tag{29}$$

imposing the Dirichlet condition $p_1^{k+1} = p_2^k$ on Γ induced by (26). Unique solvability of this problem is extensively discussed in [7, Th. 1.5.18, 2.3.16]. On the other hand, given $p_1^{k+1} \in \mathcal{K}_1$, one can obtain $\tilde{p}_2^{k+1} \in \mathcal{K}_2$ by solving the variational inequality

$$\tilde{p}_2^{k+1} \in \mathcal{K}_2: \ \tilde{a}_2(\tilde{p}_2^{k+1}, \tilde{v}_2 - \tilde{p}_2^{k+1}) - \ell_2(\tilde{v}_2 - \tilde{p}_2^{k+1}) - \Big(\ell_1(R_1(\tilde{v}_2 - \tilde{p}_2^{k+1})_{|\Gamma}) - \tilde{a}_1(p_1^{k+1}, R_1(\tilde{v}_2 - \tilde{p}_2^{k+1})_{|\Gamma})\Big) \geq 0 \quad \forall \tilde{v}_2 \in \mathcal{K}_2 \tag{30}$$

into which the weak form (28) of the Neumann condition (24) (with p_1 and p_2 replaced by p_1^{k+1} and $\tilde{p}_2^{k+1}$, respectively, and kr, θ indexed accordingly) has been inserted. To see this we first assume that (30) is uniquely solvable (see again [7, Th. 1.5.18, 2.3.16] for suitable conditions). Now, for any $\tilde{v}_2 \in \mathcal{K}_2$ we consider the trace function

$$\mu := (\tilde{v}_2 - \tilde{p}_2^{k+1})_{|\Gamma} \in \Lambda_0 - \tilde{p}_{2|\Gamma}^{k+1}$$

and

$$v_2 := \tilde{v}_2 - R_2\mu \in \mathcal{K}_2^{\tilde{p}_2^{k+1}}.$$

Then, with these functions, adding (27) and (28) leads to (30).

In order to obtain a reasonable iterative procedure by (29) and (30) for $k \geq 0$, we need an initial guess p_2^0 and additional damping, i.e. for a $\vartheta \in (0, 1)$ we replace the intermediate iterate $\tilde{p}_2^{k+1}$ by

$$p_2^{k+1} = \vartheta\, \tilde{p}_2^{k+1} + (1 - \vartheta)\, p_2^k \tag{31}$$

before carrying out (29) for the next iterate. The damping is necessary even in the linear case in order to get convergence (see [26, p. 12]). Now, with the initial iterate p_2^0 the scheme (29)–(31), $k \geq 0$, provides a damped *nonlinear Dirichlet–Neumann method* in a weak formulation applied to the Signorini-type problem (25)–(28)

for the time-discretized Richards equation. We have indicated a convergence proof for this method applied to a stationary case in 1D and given a numerical example for the applicability of the method in 2D in [9].

It is also possible to combine the two interface conditions (26) and (28) in order to obtain convex subproblems with Robin boundary conditions on Γ. In our setting, given non-negative γ_1 and γ_2 with $\gamma_1+\gamma_2>0$, this approach seeks to realize the interface conditions

$$\tau\Big(kr_1(\theta_1(p_1))\nabla p_1 - kr_1(\theta_1(\tilde p_1))e_z\Big)\cdot\mathbf{n} + (-1)^{i+1}\gamma_i\, p_1 = \tau\Big(kr_2(\theta_2(p_2))\nabla p_2 - kr_2(\theta_2(\tilde p_2))e_z\Big)\cdot\mathbf{n} + (-1)^{i+1}\gamma_i\, p_2 \tag{32}$$

on Γ, $i=1,2$, for the fixed point

$$p=\begin{cases} p_1 & \text{on } \Omega_1\\ p_2 & \text{on } \Omega_2\end{cases}$$

of a corresponding iterative procedure (compare [26, p. 16]). In the following we present such a *nonlinear Robin method* for the Signorini-type problem (25)–(28) in a weak formulation. For this purpose we abbreviate the L^2-scalar product on Γ as

$$(\eta,\mu)_\Gamma := \int_\Gamma \eta\mu\, d\sigma \quad \forall \eta,\mu\in L^2(\Gamma)$$

and also write

$$(v,w)_\Gamma := (tr_\Gamma v, tr_\Gamma w)_\Gamma \quad \forall v,w\in H^1(\Omega_i)\,,\quad i=1,2\,.$$

Now, given an initial iterate p_2^0 on Ω_2 the Robin method reads as follows. For $k\ge 0$ solve successively the subproblems

$$\begin{aligned} &p_1^{k+1}\in\mathcal{K}_1:\\ &\tilde a_1(p_1^{k+1},\tilde v_1-p_1^{k+1})-\ell_1(\tilde v_1-p_1^{k+1})+\gamma_1(p_1^{k+1},\tilde v_1-p_1^{k+1})_\Gamma\\ &\quad+\tilde a_2(p_2^k,R_2(\tilde v_1-p_1^{k+1})_{|\Gamma})-\ell_2(R_2(\tilde v_1-p_1^{k+1})_{|\Gamma})\\ &\qquad\qquad-\gamma_1(p_2^k,v_1-p_1^{k+1})_\Gamma\ge 0\quad \forall \tilde v_1\in\mathcal{K}_1\,, \end{aligned}\tag{33}$$

$$\begin{aligned} &p_2^{k+1}\in\mathcal{K}_2:\\ &\tilde a_2(p_2^{k+1},\tilde v_2-p_2^{k+1})-\ell_2(\tilde v_2-p_2^{k+1})+\gamma_2(p_2^{k+1},\tilde v_2-p_2^{k+1})_\Gamma\\ &\quad+\tilde a_1(p_1^{k+1},R_1(\tilde v_2-p_2^{k+1})_{|\Gamma})-\ell_1(R_1(\tilde v_2-p_2^{k+1})_{|\Gamma})\\ &\qquad\qquad-\gamma_2(p_1^{k+1},\tilde v_2-p_2^{k+1})_\Gamma\ge 0\quad \forall \tilde v_2\in\mathcal{K}_2\,. \end{aligned}\tag{34}$$

Just as the Dirichlet and Neumann problems (29) and (30) these Robin subproblems for the Richards equation in homogeneous soil are also related to convex minimization problems which are uniquely solvable under natural conditions. In the next subsection we will address this topic in more detail and provide a numerical example which involves the Robin method for the Richards equation in heterogeneous soil. Analytic convergence results will be presented in a forthcoming paper [8].

3.2 Homogeneous problem of Signorini's type with Robin boundary conditions

In Section 2.1 we have considered boundary value problems for the implicit-explicitly time-discretized Richards equation (8) in which Dirichlet, Neumann and Signorini-type boundary conditions (5)–(7) can be contained. In (33) and (34) these problems have been generalized by incorporating Robin boundary conditions, which (we drop the indices k and $k+1$ from now on) assign a certain value to

$$\tau\Big(kr_i(\theta_i(p_i))\nabla p_i\Big)\cdot\mathbf{n} + (-1)^{i+1}\gamma_i\, p_i \quad \text{on } \Gamma\,,\quad i=1,2\,.$$

With the discrete water fluxes

$$\tilde{\mathbf{v}}_i := -kr_i(\theta_i(p_i))\nabla p_i + kr_i(\theta_i(\tilde p_i))e_z$$

on Ω_i, $i=1,2$, this is equivalent to assigning a value to the linear combinations

$$-\tau\tilde{\mathbf{v}}_i\cdot\mathbf{n} + (-1)^{i+1}\gamma_i\, p_i \quad \text{on } \Gamma\,,\quad i=1,2\,.$$

Kirchhoff's transformation for (8)

In [7, Ch. 1, 2] and [10] problems of the kind (8) have been analysed extensively. As indicated in the introduction, the basic approach is as follows. First, (8) is transformed into generalized variables u on Ω by Kirchhoff's transformation

$$\kappa: p\mapsto u := \int_0^p kr(\theta(q))\,dq\,. \tag{35}$$

Concretely, we denote the transformed saturation by

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

set the transformed Dirichlet value as $u_D := \kappa(p_D)$ on $\gamma_D\neq\emptyset$ as well as $\tilde u := \kappa(\tilde p)$ and define the convex set

$$\tilde{\mathcal{K}} := \{v\in H^1(\Omega) : tr_{\gamma_D} v = u_D \,\wedge\, tr_{\gamma_S} v\le 0\}\,. \tag{36}$$

Now, by the chain rule, the transformed variational inequality (8) is no longer quasilinear but semilinear and reads

$$u\in\tilde{\mathcal{K}}: \int_\Omega M(u)\,(v-u)\,dx + \tau\int_\Omega \nabla u\nabla(v-u)\,dx \ge \int_\Omega M(\tilde u)\,(v-u)\,dx + \tau\int_\Omega kr(M(\tilde u))e_z\nabla(v-u)\,dx \quad \forall v\in\tilde{\mathcal{K}}\,. \tag{37}$$

In the following we assume θ, kr to be Borelfunctions with kr satisfying

$$c\le kr(\cdot)\le 1 \tag{38}$$

for some $c > 0$. Then the Kirchhoff transformation $\kappa : \mathbb{R} \to \mathbb{R}$ is monotonically increasing, Lipschitz continuous, has a Lipschitz continuous inverse, i.e. κ and κ^{-1} induce continuous superposition operators acting on $H^1(\Omega)$, consult [24, 25]. Due to (35) and (38) we also have

$$\tilde{\mathcal{K}} - u \subset \mathcal{K}_0 - p \tag{39}$$

and

$$tr_{\gamma_S(t)}\, p(t) \le 0 \implies tr_{\gamma_S(t)}\, u(t) \le 0 \,.$$

Note that we have equality in (39) if $\gamma_S = \emptyset$ holds. Altogether, with these ingredients one can prove the following result, see [7, Thm. 1.5.18].

Proposition 3 *With the definitions and assumptions above, every solution p of (8) gives a solution $u = \kappa(p)$ of (37). Furthermore, we have equivalence of (8) and (37) in case of $\gamma_S = \emptyset$.*

Convex minimization for (8)

Now, we assume additionally that $M : \mathbb{R} \to \mathbb{R}$ is monotonically increasing, continuous and bounded (which by (38) is equivalent to imposing the same conditions on $\theta(\cdot)$). Then (37) turns out to be equivalent to a convex minimization problem which is uniquely solvable, see [7, Thm. 2.3.16]. More concretely, we define the coercive and continuous bilinear form

$$a(u,v) := \tau \int_\Omega \nabla u \nabla v \, dx \quad \forall u,v \in \tilde{\mathcal{K}}$$

and the continuous linear form

$$\ell(v) := \int_\Omega M(\tilde{u})\, v\, dx + \tau \int_\Omega kr(M(\tilde{u})) e_z \nabla v \, dx$$

for $v \in H^1(\Omega)$ as well as the convex functionals

$$\mathcal{J}(v) := \frac{1}{2} a(v,v) - \ell(v) \quad \forall v \in H^1(\Omega) \tag{40}$$

and

$$\phi(v) := \int_\Omega \Phi(v(x))\, dx \quad \forall v \in \tilde{\mathcal{K}}\,, \tag{41}$$

the latter with a (convex) primitive $\Phi : \mathbb{R} \to \mathbb{R}$ of M. Then we have

Proposition 4 *With the preceeding definitions and assumptions, the variational inequality (37) is equivalent to the uniquely solvable convex minimization problem*

$$u \in \tilde{\mathcal{K}} : \quad \mathcal{J}(u) + \phi(u) \le \mathcal{J}(v) + \phi(v) \quad \forall v \in \tilde{\mathcal{K}}\,. \tag{42}$$

Kirchhoff's transformation for (33) and (34)

Now we turn to the Signorini-type problems (33) and (34) with additional Robin boundary conditions. Since these problems are symmetric we concentrate on (33) and skip the indices k and $k+1$ for convenience. From now on we must assume that the extension operator

$$R_2 : \tilde{\Lambda} \to H^1_{\gamma_{D_2} \cup \gamma_{S_2}}(\Omega_2) \quad \text{is linear and continuous}\,.$$

Then, by linearity and continuity of the trace operator $tr_\Gamma : H^1(\Omega_1) \to H^{1/2}(\Gamma)$, the contributions coming from all summands in (33) except for

$$\tilde{a}_1(p_1, \tilde{v}_1 - p_1) \quad \text{and} \quad \gamma_1(p_1, \tilde{v}_1 - p_1)_\Gamma$$

provide a continuous linear form on $H^1(\Omega_1)$ applied to $\tilde{v}_1 - p_1$.

On the other hand, by definition of $\tilde{a}_1(\cdot,\cdot)$ in Section 2.1 the contribution of this term in (33) has exactly the same structure as the right hand side in (37) if a local Kirchhoff transformation κ_1 (coming from $kr_1(\theta_1(\cdot))$ in (35)) is applied to this subproblem on Ω_1.

Altogether, apart from the summand

$$\gamma_1(p_1, \tilde{v}_1 - p_1)_\Gamma \quad \forall \tilde{v}_1 \in \mathcal{K}_1$$

the variational inequality is related to a convex minimization problem like (42). In an untransformed version such a summand provides an additional coercive and continuous bilinear form in the problem, thus guaranteeing its unique solvability. Here, we have to deal with the transformed counterpart

$$\gamma_1(\kappa_1^{-1}(u_1), w_1)_\Gamma \quad \forall w_1 \in \tilde{\mathcal{K}}_1 - u_1 \subset \mathcal{K}_1 - p_1$$

if we define $\tilde{\mathcal{K}}_1$ just as $\tilde{\mathcal{K}}$ in (36) and analogously assume (38) which entails (39) with the subscript 1. Again, with arguments as in the proof of [7, Thm. 1.5.18], we obtain

Proposition 5 *With the preceeding assumptions, if the variational inequality (33) has a solution p_1, then $u_1 = \kappa_1(p_1)$ is a solution of its transformed version*

$$u_1 \in \tilde{\mathcal{K}}_1 : \int_{\Omega_1} M_1(u_1)(v_1 - u_1)\, dx + \gamma_1(\kappa_1^{-1} u_1, v_1 - u_1)_\Gamma$$
$$+ a_1(u_1, v_1 - u_1) - \ell_1(v_1 - u_1) \ge 0 \quad \forall v_1 \in \tilde{\mathcal{K}}_1\,. \tag{43}$$

In addition, this latter problem is equivalent to (33) if $\gamma_{S_1} = \emptyset$.

Convex minimization for (33) and (34)

As above, the variational inequality (43) can be treated by convex analysis.

Theorem 2 *Let $M_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2$, be monotonically increasing, continuous and bounded. Furthermore, let kr_i, $i = 1, 2$, be Borelfunctions with $kr_i \geq c > 0$ for some $c > 0$. Then the Robin subproblem (43) corresponding to (33) for $i = 1$ as well as its counterpart for $i = 2$ are equivalent to a uniquely solvable convex minimization problems.*

Proof As before we restrict ourselves to $i = 1$. Due to the preliminary arguments we only need to deal with the term $\gamma_1(\kappa_1^{-1}u_1, v_1 - u_1)_\Gamma$. In view of (41) we first consider a primitive Ψ_1 of κ_1^{-1}. Since by the assumptions on kr_1 the real function κ_1^{-1} is monotonically increasing and Lipschitz continuous, Ψ_1 is convex and differentiable with

$$|\Psi_1(z) - \Psi_1(z')| \leq a(|z| + |z'|) + b \quad \forall z, z' \in \mathbb{R} \tag{44}$$

for some positive a, b. Consequently, the functional ψ_1 defined on $\tilde{\mathcal{K}}_1$ by

$$\psi_1 : v_1 \mapsto \int_\Gamma \Psi_1(tr_\Gamma v_1(s))\, d\sigma(s) \quad \forall v_1 \in \tilde{\mathcal{K}}_1 \tag{45}$$

assumes finite values (even for all functions in $L^2(\Gamma)$), is convex and continuous and has an affine minorant

$$\psi_1(v_1) \geq \alpha - \beta\, \|v_1\|_{1,\Omega_1} \quad \forall v_1 \in \tilde{\mathcal{K}}_1 \tag{46}$$

with some real α, β (consult [22, Prop. 1.1, 1.5] and use the continuity of $tr_\Gamma : \tilde{\mathcal{K}}_1 \to L^2(\Gamma)$). Furthermore, the differentiability of Ψ_1 entails the existence of the directional derivative $\partial_{v-w}\psi_1(w)$ for any $v, w \in \tilde{\mathcal{K}}_1$ with

$$\begin{aligned}\partial_{v-w}\psi_1(w) &= \int_\Gamma \kappa_1^{-1}(tr_\Gamma w) \cdot tr_\Gamma(v - w)\, d\sigma \\ &= (\kappa_1^{-1}w, v - w)_\Gamma\,,\end{aligned} \tag{47}$$

see e.g. [7, Prop. 2.3.9, 1.5.16].

Therefore, if we choose a convex primitive Φ_1 of M_1 as above, which leads to a convex and differentiable functional ϕ_1 on $\tilde{\mathcal{K}}_1$, one can verify that (43) is the condition

$$\partial_{v_1 - u_1} F(u_1) \geq 0 \quad \forall v_1 \in \tilde{\mathcal{K}} \tag{48}$$

on the directional derivatives of the convex functional

$$F_1 : v_1 \mapsto \phi_1(v_1) + \gamma_1\psi_1(v_1) + \frac{1}{2}a_1(v_1, v_1) - \ell_1(v_1) \quad \forall v_1 \in \tilde{\mathcal{K}}_1 \tag{49}$$

in the point $u_1 \in \tilde{\mathcal{K}}_1$ and all directions $v_1 - u_1$ with $v_1 \in \tilde{\mathcal{K}}_1$. This, however, is equivalent to u_1 solving the convex minimization problem

$$u_1 \in \tilde{\mathcal{K}}_1 : \quad F_1(u_1) \leq F_1(v_1) \quad \forall v_1 \in \tilde{\mathcal{K}}_1\,, \tag{50}$$

compare e.g. [14, Prop. 2.1].

Now, a classical result (see e.g. [14, Prop. 1.2]) shows that (50) is uniquely solvable. This is because the functional $F_1 : \tilde{\mathcal{K}}_1 \to \mathbb{R}$ is strictly convex, coercive and continuous. Apart from the contribution of the functional ψ_1 this has been discussed above. However, we already know that ψ_1 is convex and continuous. Since, in addition, it has an affine minorant (46) just as ϕ_1, the coercivity of F_1 is guaranteed by the coercivity of the bilinear form $a_1(\cdot, \cdot)$, compare [7, Thm. 2.3.16]. □

Summary and generalizations

At the end of this subsection it seems appropriate to point out the main steps one has to take on the way from the variational inequality (33) to the unique solvability of (43) since each of them requires its own special conditions. In particular, Theorem 2 actually embraces two different steps.

First, the variational inequality is Kirchhoff–transformed. If it has a solution p_1 then $u_1 = \kappa_1(p_1)$ is a solution of the transformed variational inequality (43) provided that θ_1, kr_1 are Borelfunctions with kr_1 satisfying $0 \leq kr_1(\cdot) \leq 1$. This guarantees at least that no solution of (33) gets lost by the transformation which is relevant on the level of physics. We note this here since it applies to the parameter functions given by Brooks and Corey [12, 30], even though in this case we only have $\kappa(\mathbb{R}) = (u_c, \infty)$ with a $u_c < 0$ such that M is only defined on an interval (u_c, ∞). We use the Brooks–Corey functions in our numerical example below, see Figures (2), (3).

Conversely, a solution u_1 of (43) provides a solution $p_1 = \kappa_1^{-1}(u_1)$ of (33) if we have additionally $c \leq kr_1(\cdot)$ for a $c > 0$ and $\gamma_S = \emptyset$ which is of course a strong restriction. Note, however, that $\kappa_1^{-1}(u_1)_{|\Gamma} \in L^2(\Gamma)$ in (43) is not guaranteed if κ_1^{-1} has a singularity.

Secondly, the variational inequality (43) is reformulated as a convex minimization problem (50) which requires to identify it as the corresponding condition (48) on the directional derivatives of F_1 in u_1. In order to achieve this, differentiation under the integral is necessary, for which continuity of M_1 and κ_1^{-1} is needed. Here, $\gamma_S \neq \emptyset$ is allowed, $kr_1(\cdot) \leq 1$ is no longer needed, and in case of $\gamma_1 = 0$ the reformulation also works for the Brooks–Corey functions.

Thirdly, we strive for unique solvability of the convex minimization problem (50), for which the continuity condition on M_1 and κ_1^{-1} is no longer necessary. Here, M_1 and κ_1^{-1} have to be monotonically increasing (so that ϕ_1 and ψ_1 are convex) and bounded or else Hölder continuous outside of a bounded Intervall (so that (44) is satisfied). Again, unique solvability is also obtained in case of $\gamma_1 = 0$ for the Brooks–Corey functions.

Finally, we remark that by above arguments one can see how the domain decomposition problem (25)–(28) is related to its Kirchhoff–transformed counterpart and formulate conditions for their equivalence, even if (25) and (27) also contain Robin boundary conditions on subsets of $\partial\Omega_i$, $i = 1, 2$ (for which the Theorem 1 also holds). See [6] for a more detailed discussion on this topic also including a proof for the commutativity

$$tr_\Gamma \kappa_i^{-1} u_i = \kappa_i^{-1} tr_\Gamma u_i \quad \forall u_i \in H^1(\Omega_i)\,, \ \ i = 1, 2\,, \qquad (51)$$

which gives the equivalence

$$tr_\Gamma p_1 = tr_\Gamma p_2 \quad \Longleftrightarrow \quad \kappa_1^{-1} tr_\Gamma u_1 = \kappa_2^{-1} tr_\Gamma u_2\,.$$

We highlight property (51) because, although we have used it almost unnoticeable in (47), it is not straightforward. For example, the definition

$$\psi_1 : v_1 \mapsto \int_\Gamma tr_\Gamma \Psi_1(v_1(s))\, d\sigma(s) \quad \forall v_1 \in \tilde{\mathcal{K}}_1$$

instead of (45) would be meaningless since in general $\Psi_1(v_1) \in H^1(\Omega_1)$ does not hold for all $v_1 \in H^1(\Omega_1)$. Indeed, this can only be guaranteed for Lipschitz continuous Ψ_1 (or locally Lipschitz continuous Ψ_1 in 1D), see [24]. In this case, however, the commutativity (51) holds.

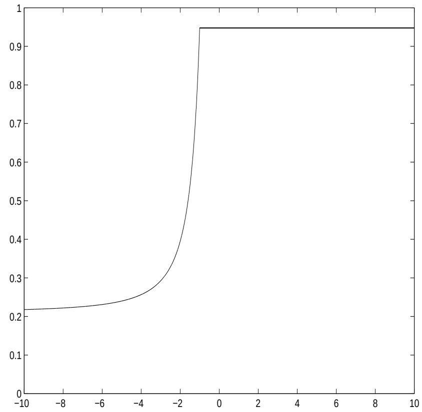

Fig. 2 Brooks–Corey function $p \mapsto \theta(p)$

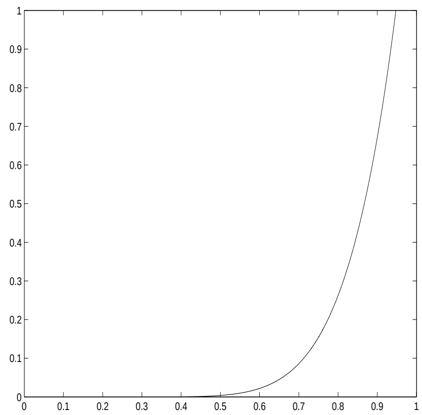

Fig. 3 Brooks–Corey function $\theta \mapsto kr(\theta)$

3.3 Discretization and numerical treatment

As far as the discretization and the numerical treatment of the Robin method is concerned, it is enough to consider this with respect to one subproblem (33) or else (34). However, since our approach aims at exploiting the convex minimization in the subdomains, we choose a certain finite element discretization of the transformed problem (43) instead, that preserves the convex structure. Then the corresponding discrete problem can be solved by monotone multigrid methods. The question what kind of discretized problem with respect to the physical variables we actually solve with this approach is discussed in [10]. Here, one can also find convergence results for the discretizations that can be extended to Robin problems.

For simplicity, if we refer to the variational inequality (43) and related terms from now on, we think of the index 1 skipped everywhere. By setting the Robin parameter $\gamma = 0$ we recover a Signorini-type problem like (37). The discretization and the numerical treatment of such problems is given in [10], a detailed exposition can be found in [7]. In the proof of Theorem 2 we have seen that the term induced by the Robin boundary condition induces a convex functional just as the other nonlinearity $M(\cdot)$ in (43). Therefore, it is natural to treat it analogously as the latter in the discretization.

Brooks–Corey parameter functions

Before we can give the discretization in concrete terms, we need to introduce the parameter functions since they induce a more general setting than discussed above. They have been introduced by Brooks–Corey [12, 30] with results due to Burdine [13]. The functions depend on a bubbling pressure $p_b < 0$ and a pore size distribution factor $\lambda > 0$ as the soil parameters. Concretely, the saturation θ as a function of the pressure p is given by

$$\theta(p) = \begin{cases} \theta_m + (\theta_M - \theta_m)\left(\frac{p}{p_b}\right)^{-\lambda} & \text{for } \ p \le p_b \\ \theta_M & \text{for } \ p \ge p_b \end{cases} \qquad (52)$$

and the relative permeability kr as a function of the saturation θ reads

$$kr(\theta) = \left(\frac{\theta - \theta_m}{\theta_M - \theta_m}\right)^{3+\frac{2}{\lambda}}\,, \quad \theta \in [\theta_m, \theta_M]\,. \qquad (53)$$

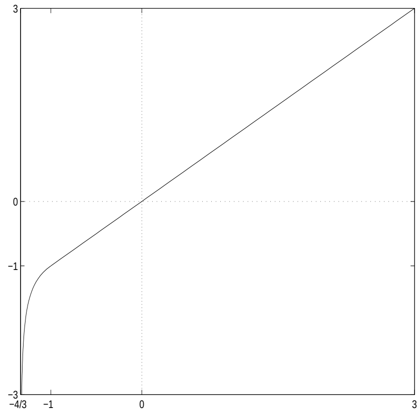

Fig. 4 Inverse Kirchhoff transformation $u \mapsto \kappa^{-1}(u)$

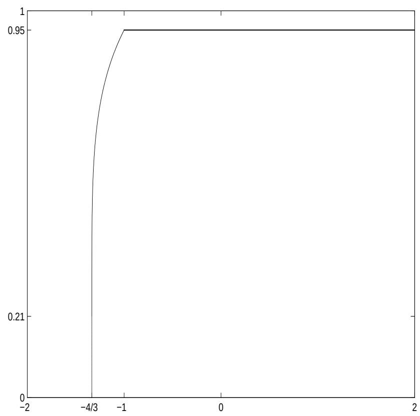

Fig. 5 Generalized saturation $u \mapsto M(u)$

Typical shapes of these nonlinearities are depicted in Figures 2 and 3.

Now, in contrast to the cases considered so far, the image $\kappa(\mathbb{R})$ of the Kirchhoff transformation turns out to be a strict subset (u_c, ∞) of $\mathbb{R}$ and, consequently, κ^{-1} and M are only defined on (u_c, ∞). Obviously, the critical pressure $u_c < 0$ corresponds to $p = -\infty$ and, therefore, $M(u_c) = \theta_m$ is a sensible definition. The situation is more extreme than before since M has unbounded derivatives and κ^{-1} is ill-conditioned around u_c, compare Figures 4 and 5.

With these parameter functions the convex set $\tilde{\mathcal{K}}$ in (36) is replaced by

$$\mathcal{K} := \{v \in H^1(\Omega) : v \geq u_c \wedge tr_{\gamma_D} v = u_D \wedge tr_{\gamma_S} v \leq 0\} .$$

and the first statement in Proposition 3 as well as Proposition 4 remain true. Although the variational inequality (43) for the Robin problem does no longer make sense with this generalized convex set, its corresponding minimization problem (50) does and is still uniquely solvable (since κ^{-1} is improperly integrable on $(u_c, 0)$ and, therefore, Ψ is continuous in u_c). This minimization problem shall be discretized in the following.

Finite element discretization

For simplicity, we assume that $\Omega \subset \mathbb{R}^2$ is a polygonal domain. Let $\mathcal{T}_j$, $j \in \mathbb{N}_0$, be a partition of Ω into triangles $t \in \mathcal{T}_j$ with minimal diameter of order $\mathcal{O}(2^{-j})$. We assume the triangulation $\mathcal{T}_j$ to be conform, i.e. the intersection of two different triangles in $\mathcal{T}_j$ is either empty or consists of a common edge or a common vertex. The set of all vertices of the triangles in $\mathcal{T}_j$ is denoted by $\mathcal{N}_j$ and we set $\mathcal{N}_j^D := \mathcal{N}_j \cap \gamma_D$ as well as $\mathcal{N}_j^S := \mathcal{N}_j \cap \gamma_S$.

Let $\mathcal{S}_j \subset H^1(\Omega)$ be the subspace of all continuous functions in $H^1(\Omega)$ which are linear on each triangle $t \in \mathcal{T}_j$. It is spanned by the nodal basis functions $\lambda_p^{(j)}$, $p \in \mathcal{N}_j$. For the definition of the finite dimensional analogue of $\mathcal{K}$ we assume that the Dirichlet boundary condition u_D is continuous in each node $p \in \mathcal{N}_j^D$ so that we can evaluate $u_D(p)$ in these nodes. Then we define this convex set $\mathcal{K}_j \subset \mathcal{S}_j$ by

$$\mathcal{K}_j := \Big\{ v \in \mathcal{S}_j : v(p) \geq u_c \, \forall p \in \mathcal{N}_j \wedge \\ v(p) = u_D(p) \, \forall p \in \mathcal{N}_j^D \wedge v(p) \leq 0 \, \forall p \in \mathcal{N}_j^S \Big\} .$$

The convex functionals ϕ and ψ in (49) are discretized by a quadrature formula arising from $\mathcal{S}_j$-interpolation of the integrands in (41) and (45), respectively. In this way, we arrive at the discrete convex functionals $\phi_j, \psi_j : \mathcal{K}_j \to \mathbb{R}$ defined by the weighted sums

$$\phi_j(v) := \sum_{p \in \mathcal{N}_j} \Phi(v(p))\, h_p \quad \forall v \in \mathcal{K}_j \tag{54}$$

and

$$\psi_j(v) := \sum_{p \in \mathcal{N}_j \cap \Gamma} \Psi(v(p))\, h_{\Gamma,p} \quad \forall v \in \mathcal{K}_j \tag{55}$$

which contain the weights

$$h_p := \int_\Omega \lambda_p^{(j)}(x)\, dx \quad \text{and} \quad h_{\Gamma,p} := \int_\Gamma \lambda_p^{(j)}\, d\sigma ,$$

respectively. Recall that Γ is the part of the boundary of Ω where Robin boundary conditions are given.

With these definitions our finite element discretization of (49) reads

$$u \in \mathcal{K}_j : \quad \mathcal{J}(u_j) + \phi_j(u_j) + \gamma\, \psi_j(u_j) \leq \\ \mathcal{J}(v) + \phi_j(v) + \gamma\, \psi_j(v) \quad \forall v \in \mathcal{K}_j , \tag{56}$$

where $\mathcal{J}$ is the quadratic functional defined in (40). For this discrete convex minimization problem the existence and uniqueness result from Theorem 2 carries over to the discrete case, even if we consider Brooks–Corey functions (consult e.g. [14, Prop. 1.2]).

Proposition 6 *Let $M : \mathbb{R} \to \mathbb{R}$ or $M : (u_c, \infty) \to \mathbb{R}$, $u_c < 0$, be monotonically increasing, continuous and bounded. Furthermore, let $kr : M(\mathbb{R}) \to \mathbb{R}$ be a bounded Borelfunction. Then the discrete convex minimization problem (56) has a unique solution.*

Finally, we mention that with some regularity assumptions on u_D and $\mathcal{K}_j$ one can prove H^1-convergence of the discrete solutions u_j, $j \geq 0$, to the solution of the continuous problem (49) under natural conditions (compare [7, Sec. 2.5]).

Monotone multigrid

The standard reference for the numerical treatment of discrete minimization problems like (56) by monotone multigrid methods is [21]. Therefore, we can restrict ourselves to a very basic description of such methods and mention the special situation given by the use of the Brooks–Corey functions. In particular, we note that Φ in (54), which is considered in [21], may well depend on $p \in \mathcal{N}_j$.

The smoother used in the monotone multigrid is the nonlinear Gauss–Seidel method. Starting with a given iterate, this method minimizes the convex functional

$$F_j := \mathcal{J}(\cdot) + \phi_j(\cdot) + \gamma\, \psi_j(\cdot)$$

successively in the directions of the nodal basis functions $\lambda_p^{(j)}$ for $p \in \mathcal{N}_j \backslash \mathcal{N}_j^D$. Thanks to a decoupling provided by the definitions (54) and (55) of the discrete convex functionals, this leads to successive one-dimensional problems of finding the zero of the functions

$$M(\cdot)\, h_p + \tau\gamma\, \kappa^{-1}(\cdot)\, h_{\Gamma,p} + g_p(\cdot)$$

where $g_p(\cdot)$ are certain affine functions. Here, $M(\cdot)$ has to be extended to a monotone graph in u_c (and in 0 if $p \in \mathcal{N}_j^S$). The smoother on the fine grid guarantees convergence of the method.

As in the linear case, coarse grid corrections are carried out in order to increase the convergence speed. These are provided by considering constrained quadratic approximations of the functional F_j around smoothed iterates (in nodes where F_j is C^2), which result in quadratic obstacle problems. The latter are solved on the coarse grid within the constraints induced by the critical value u_c, the bubbling pressure p_b (where M and κ^{-1} are non-differentiable and, therefore, F_j is not C^2) and 0 in Signorini-nodes. These constraints, the coarse grid obstacles, can be obtained from the fine grid obstacles by different quasioptimal truncation techniques. Additional damping, locally for each coarse grid direction, ensures that the iterates provided by the coarse grid correction lead to a further decreasing energy F_j.

Under some technical and non-degeneracy conditions one can prove that the coarse grid corrections of the monotone multigrid eventually become Newton multigrid steps applied to smooth problems for which convergence rates can be derived. Concretely, one can prove that these asymptotic convergence rates only degenerate very mildly with j.

4 Numerical example in 2D: Robin's method for the Richards equation without gravity

In this last section, we present a numerical example in 2D which we obtained by a successful application of Robin's method to the Richards equation without gravity in a heterogeneous setting with two different soil types in two subdomains. It is based on the analytical and numerical approaches described in Section 3. We present convergence rates of the monotone multigrid method used for the local problems and convergence rates of the Robin iteration. Since our focus is on the performance of the Robin method, which deals with the heterogeneity of the spatial problems, we ignore the gravitational term for simplicity since its contribution would only be treated explicitly in our approach [10].

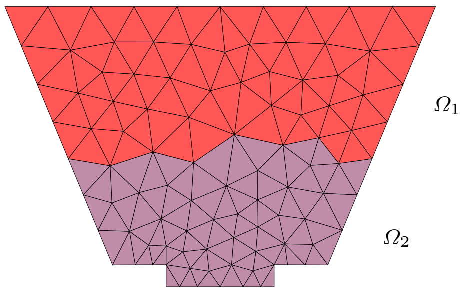

Fig. 6 Coffee filter like domain Ω

Concretely, we consider the domain $\Omega \subset \mathbb{R}^2$ depicted in Figure 6 which resembles a coffee filter. We assume that the top subdomain Ω_1 is filled with sandy loam while the bottom subdomain Ω_2 contains loamy sand. For the parametrization we use the Brooks–Corey functions $p \mapsto \theta(p)$ and $\theta \mapsto kr(\theta)$ according to Burdine given in (52) and (53). The corresponding hydrological data are chosen according to the USDA soil texture triangle [27, Tables 5.3.2 and 5.5.5] and collected in Table 1 with the corresponding indices $i = 1, 2$. The maximal water contents $\theta_{M,i}$ are constant with $\theta_{M,i} = 1$ for $i = 1, 2$ according to [27, Table 5.1.1]. We remark that the variation of residual water contents $\theta_{m,i}$ and the porosity-values n_i does not effect the performance

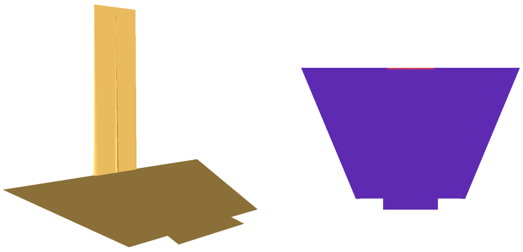

Fig. 7 $t = 0$

of the Robin method as strongly as the other more relevant soil parameters λ_i, $p_{b,i}$ and $K_{h,i}$, $i = 1, 2$, which influence the spatial derivative in (1).

Ω_i	n_i	$\theta_{m,i}$	λ_i	$p_{b,i}$	$K_{h,i}$
$i = 1$	0.453	0.091	0.378	−0.147	$6.06 \cdot 10^{-6}$
$i = 2$	0.437	0.080	0.553	−0.087	$1.66 \cdot 10^{-5}$

Table 1 Brooks–Corey soil parameters for sandy loam ($i = 1$) and loamy sand ($i = 2$)

The domain Ω is chosen to be situated in the quadrilateral $[-1, 1] \times [-0.74, 0.56]$ and the top boundary is $[-1, 1] \times \{-0.74\}$, the z-axis is directed downward. We start with a practically dry soil as the initial condition given by $p_0 = -20$ on Ω except for $p_0 = 100$ on the subset $[-0.21, 0.21] \times \{-0.74\}$ of the top boundary, see Figure 7. The latter is treated as a Dirichlet boundary γ_D with constant data $p_D = 100$ for all time steps. The pressure unit is one meter of a water column. Apart from the bottom boundary γ_S situated on $[-0.25, 0.25] \times \{0.56\}$, which is chosen as a Signorini-type boundary where outflow is possible, we assume homogeneous Neumann boundary conditions $\mathbf{v} \cdot \mathbf{n} = 0$ on $\partial\Omega \backslash (\gamma_D \cup \gamma_S)$. This situation results in an evolution process with an increasing saturation due to flow of water into Ω with possible outflow across γ_S until Ω is fully saturated and a stationary solution is obtained.

We treat the problem as described in the previous sections using an implicit time discretization (since there is no gravity) with the constant time step size $\tau = 1\,[s]$ and a space discretization with linear finite elements. The discrete Robin problems for the Richards equation in each time step result in convex minimization problems which are solved by monotone multigrid. We use 4 levels of a grid hierarchy with 112 nodes on the coarse grid in Figure 6 and about 5500 nodes on the finest grid with a mesh size of $h = (10 \cdot 2^4)^{-1} = 1/160$ obtained by uniform refinement. Moreover, a constant Robin parameter $\gamma = \gamma_1 = \gamma_2 = 3 \cdot 10^{-4}$ suggested by numerical experiments is chosen.

Time evolution in physical pressure

In Figures 8–19 one can see the evolution of the physical pressure at equidistant time steps (except for the last one) in heightplots on the left and colourplots on the right. One can clearly detect the wetting front (cf. [5, p. 303]), where a pressure difference of almost $\Delta p = 20$ occurs, moving from the top to the bottom. More concretely, the wetting front marks the free boundary which separates the unsaturated from the fully saturated regime and, thus, around which we encounter the pressure difference between the initial condition p_0 and the bubbling pressure $p_{b,i}$ on Ω_i for $i = 1, 2$. We need 684 time steps until the stationary situation with a fully saturated Ω is reached. At about $t = 133$ the wetting front reaches the interface, and starting with $t = 473$ the top subdomain Ω_1 is fully saturated. The range of p is between -20 and 100 until shortly before the last time step, and in the stationary case it is in the interval $[26.3, 100]$ on Ω_1 and $[0.0, 32.2]$ on Ω_2. One can see in the heightplots that the physical pressure is nonsmooth across the interface, at least in the saturated regime.

Multigrid convergence rates in generalized pressure

The local problems on the subdomains with homogeneous soil are treated by a monotone multigrid solver which we already mentioned in Subsection 3.3. More concretely, we use truncated monotone multigrid with a $V(3, 3)$-cycle, i.e. containing 3 presmoothing and 3 postsmoothing steps. As a stopping criterion for this local solver we require the relative error to satisfy

$$\frac{|u_i^j - u_i^{j-1}|_{1,\Omega_i}}{|u_i^{j-1}|_{1,\Omega_i}} \leq 10^{-12}\,, \quad i = 1, 2\,,$$

for the last multigrid iterate u_i^j with $j \geq 1$ where $|\cdot|_{1,\Omega_i}$ is the energy norm on Ω_i induced by the bilinear form $a_i(\cdot, \cdot)$. The initial condition is given by the solution from the previous time step.

Figures 20 and 21 show averaged multigrid convergence rates $\rho_{m,1}$ and $\rho_{m,2}$ as well as maximal multigrid convergence rates $\rho_{M,1}$ and $\rho_{M,2}$ per time step for Ω_1 and Ω_2, respectively. They are determined in the following way. For each domain decomposition step $l \in \mathbb{N}$ in a fixed Ω_i, $i = 1, 2$, the geometric mean

$$\bar{\rho}_{l,j} = \left(\prod_{k=2}^{j} \rho_k \right)^{1/j}$$

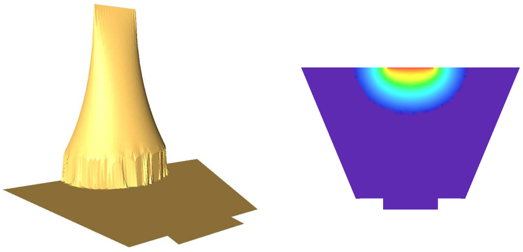

Fig. 8 $t = 60$

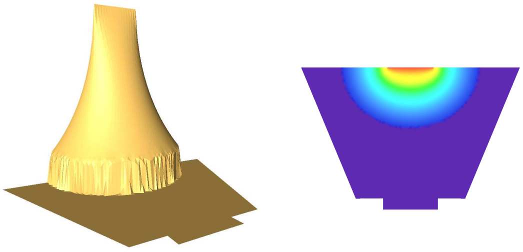

Fig. 9 $t = 120$

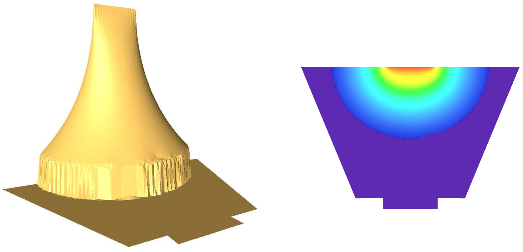

Fig. 10 $t = 180$

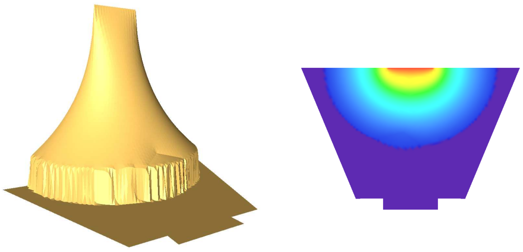

Fig. 11 $t = 240$

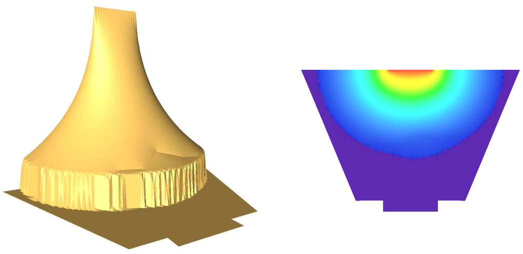

Fig. 12 $t = 300$

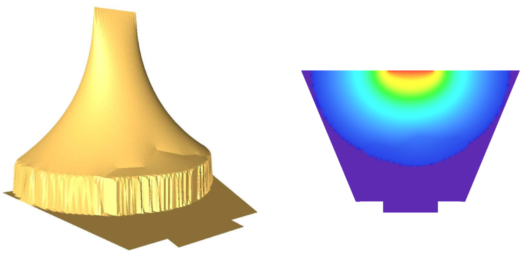

Fig. 13 $t = 360$

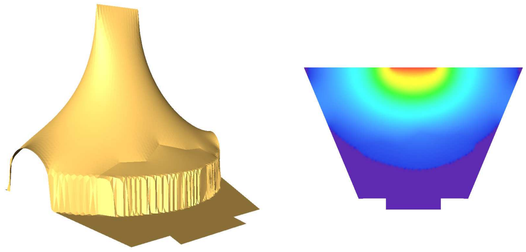

Fig. 14 $t = 420$

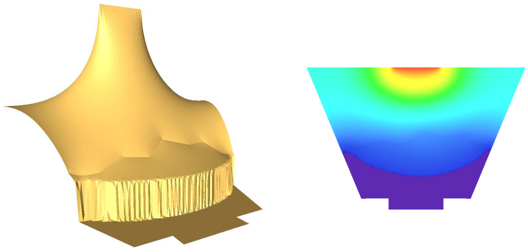

Fig. 15 $t = 480$

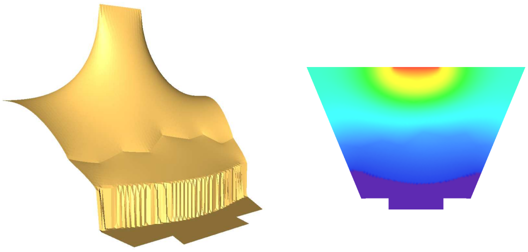

Fig. 16 $t = 540$

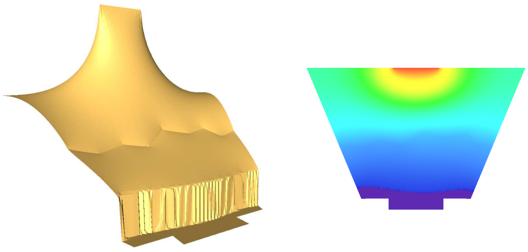

Fig. 17 $t = 600$

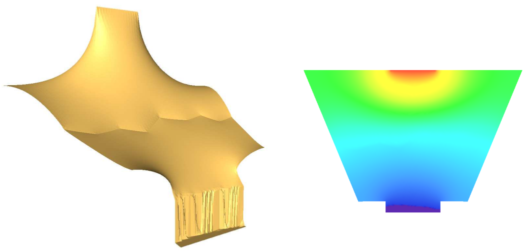

Fig. 18 $t = 660$

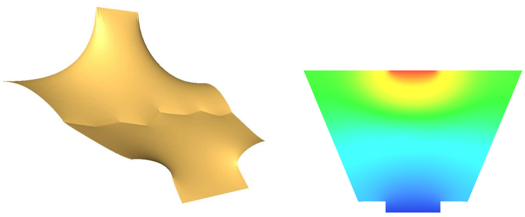

Fig. 19 $t = 684$

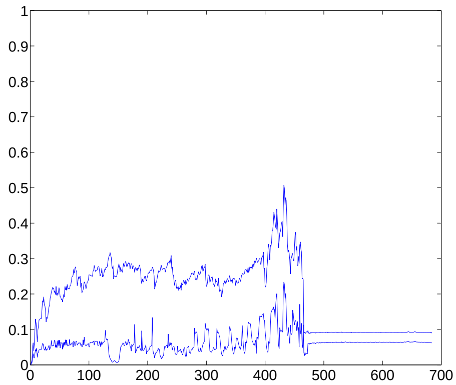

Fig. 20 Multigrid convergence rates per time step in Ω_1 (top: maximal, bottom: averaged)

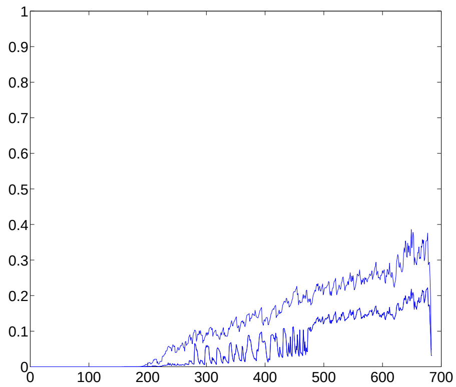

Fig. 21 Multigrid convergence rates per time step in Ω_2 (top: maximal, bottom: averaged)

of the approximated rates

$$\rho_k = \frac{|u_i^k - u_i^{k-1}|_{1,\Omega_i}}{|u_i^{k-1} - u_i^{k-2}|_{1,\Omega_i}}, \quad i = 1, 2,$$

with the multigrid iterates u_i^k is calculated for $k \geq 2$ as long as $\bar{\rho}_{l,j}$ increases. We set $\bar{\rho}_{l,1} = 0$ if one multigrid step is needed only. With the maximum obtained in this way, which we call $\bar{\rho}_l$, we determine $\rho_{m,i}$ and $\rho_{M,i}$ as

$$\rho_{m,i} = \frac{1}{n} \sum_{l=1}^{n} \bar{\rho}_l \quad \text{and} \quad \rho_{M,i} = \max_{1 \leq l \leq n} \bar{\rho}_l, \quad i = 1, 2,$$

where n is the number of Robin steps needed for the corresponding time step.

Since we use the solution from the previous time step as the initial condition for the next time step, we already have a good approximation for the solution at that time step. With this choice we obtain fast multigrid convergence as one can see from Figures 20 and 21. Here, $\rho_{M,i}$ often occurs in the first few Robin steps,

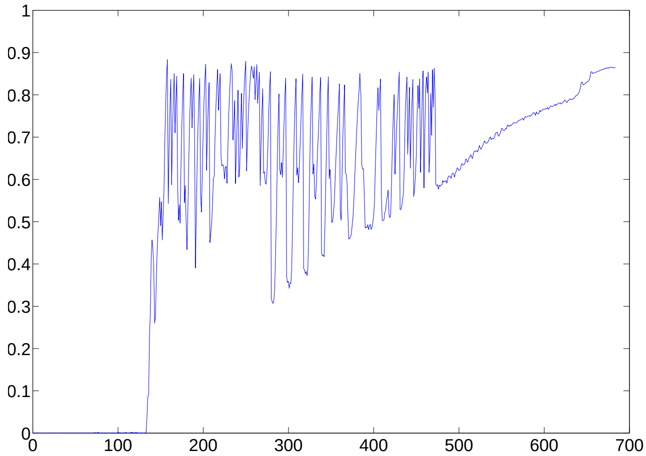

Fig. 22 Convergence rates ρ per time step for Robin's method measured in generalized variables with stopping criterion (57)

and the multigrid convergence rates can improve quite a lot for higher accuracies in the domain decomposition iteration history (where finally often one multigrid step is enough). This explains that the difference between $\rho_{M,i}$ and $\rho_{m,i}$ can be quite considerable.

Convergence rates of Robin's method in generalized pressure

With initial iterates u_i^0 for $i = 1, 2$, given as the solutions from the previous time step, the Robin iteration is carried out until the relative error satisfies

$$\frac{\left(\sum_{i=1}^{2} a_i(u_i^n - u_i^{n-1}, u_i^n - u_i^{n-1})\right)^{1/2}}{\left(\sum_{i=1}^{2} a_i(u_i^{n-1}, u_i^{n-1})\right)^{1/2}} < 10^{-12} \tag{57}$$

for some $n \geq 0$. Then we calculate ρ as the maximum of the geometric means of the rates

$$\frac{\left(\sum_{i=1}^{2} a_i(u_i^k - u_i^n, u_i^k - u_i^n)\right)^{1/2}}{\left(\sum_{i=1}^{2} a_i(u_i^{k-1} - u_i^n, u_i^{k-1} - u_i^n)\right)^{1/2}} \tag{58}$$

for $1 \leq k \leq \tilde{n}$ over all $\tilde{n} < n$ (note that we get zero for $\tilde{n} = n$).

Figure 22 displays the average convergence rates ρ for the domain decomposition iteration given by the Robin method at each time step. For $t \in [138, 472]$, when the location of the wetting front has a nontrivial intersection with the interface Γ, the convergence rates vary quite a lot between around 0.3 and 0.9. These big variations can also be observed in the Robin iteration history at various time steps. In Figures 23 and 24 we illustrate two examples of such cases for $t = 197$

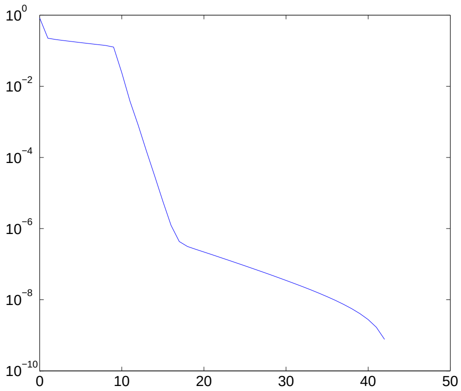

Fig. 23 Error (59) vs. Robin iteration step at time $t = 197$

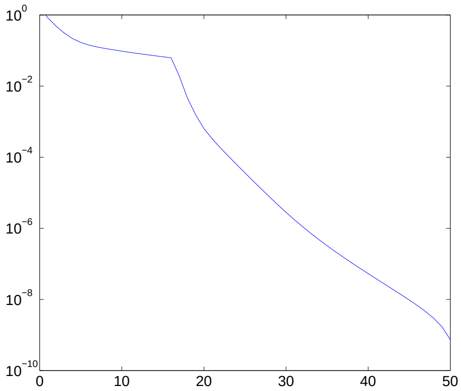

Fig. 24 Error (59) vs. Robin iteration step at time $t = 443$

and $t = 443$ where we have the average convergence rates 0.56 and 0.63, respectively. One can see that different error reduction rates (i.e. convergence rates) are obtained for different accuracies, i.e. absolute errors

$$\left(\sum_{i=1}^{2} a_i(u_i^k - u_i^n, u_i^k - u_i^n)\right)^{1/2} , \quad k = 0, ..., n-1 , \quad (59)$$

(u_i^n being the last Robin iterate) in the iteration history. We assume that these effects occur because the pressure values for nodes directly at the wetting front probably depend quite sensitively on the solution of the previous time step, the precise Robin conditions at the interface and the required accuracy given by the stopping criterion. In addition, our measuring (57), (58) of the convergence rates in the generalized variables u_i seems to be particularly sensitive in this respect, see the paragraph on the convergence rates in the physical pressure below.

In general the first convergence rate in the iteration history is considerably smaller than the following ones for any time step. In particular, for time steps $t < 133$ the error reduction in the first Robin step is such that it already provides an almost vanishing average convergence rate. For $t \geq 473$, when Ω_1 is fully saturated and the wetting front is entirely located in Ω_2, the convergence rates do no longer oscillate, neither with respect to t nor in the iteration history for fixed t. Furthermore, they increase as the wetting front approaches the Signorini-type boundary until the stationary solution is attained at $t = 684$ (for $t > 684$ vanishing convergence rates are observed as expected).

Convergence rates of Robin's method in physical pressure

Quite natually, one is as interested in the convergence rates of the Robin iteration measured in the physical variables p_i as in the generalized pressure variables u_i, $i = 1, 2$. However, one has to be careful here because the inverse transformations κ_i^{-1}, by which u_i is transformed into p_i for $i = 1, 2$, are ill-conditioned for small generalized pressure values now, compare Figures **??** and **??**. As can be seen in these figures, small perturbations in u_i can result in big variations of p_i in the unsaturated regime (we refer to [10] where this phenomenon is investigated numerically in detail).

Therefore, the stopping criterion (57) expressed in p_i may correspond to a much more restrictive stopping criterion in u_i which might require a higher accuracy than provided by the local solvers given by (**??**). In the example above, a certain absolute error (59) in u_i usually corresponds to a much bigger absolute error calculated in p_i, $i = 1, 2$. In fact, they can differ by several orders of magnitude. We even observe numerical instabilities if we choose the same accuracy 10^{-12} in the stopping criterion (57) with u_i replaced by p_i. If, instead, we choose the stopping criterion

$$\frac{\left(\sum_{i=1}^{2} a_i(p_i^n - p_i^{n-1}, p_i^n - p_i^{n-1})\right)^{1/2}}{\left(\sum_{i=1}^{2} a_i(p_i^{n-1}, p_i^{n-1})\right)^{1/2}} < 10^{-9} \qquad (60)$$

rather than (57) and measure the convergence rates as in (58) with u_i replaced by p_i, we obtain a time evolution (with 684 time steps) which practically does not differ from the one above (i.e., the first few digits of the obtained pressure values usually coincide). Interestingly, however, the convergence rates ρ_p per time step measured in the physical pressure p and displayed in Figure 25 do not show as big oscillations as the ones measured in u in Figure 22. Considerable oscillations only occur shortly before time step $t = 473$ when Ω_1 is fully saturated. In addition, the convergence rates measured p are more stable in the iteration history for fixed time steps than the ones measured in u.

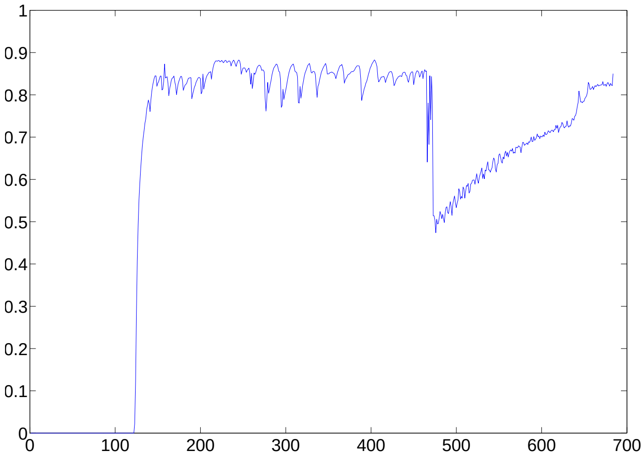

Fig. 25 Convergence rates ρ_p per time step for Robin's method measured in physical variables with stopping criterion (60)

Limitations for the applicability of the method

Unfortunately, the convergence rates of the Robin iteration deteriorate for higher levels. They also deteriorate if we choose sets of parameters corresponding to more extreme pairs of soil-types like very coarse sand and very fine clay. Furthermore, the convergence rates depend on the choice of the time step size for the variation of which we have to alter γ as well. (Observe that the time step size occurs as an additional factor in front of the water flux in the Robin conditions (32).)

In addition, we observe deteriorating convergence rates if the pressure difference Δp of the wetting front is too big. Note, however, that the situation $\Delta p = 20$ in our example above where we have the Dirichlet value $p = 100$, measured in meters of a water column, is already chosen quite extreme.

Therefore, it seems possible that the Robin method can be successfully applied in reasonable hydrological settings which are not too extreme (compare e.g. [7, Sec. 4.3]). Finally, ongoing research indicates that the convergence rates of the Robin method can be considerably improved if different Robin parameters γ_1 and γ_2 corresponding to the subdomains are suitably chosen.

We close this section by noting that the implementation for the numerical example has been performed in the numerics environment DUNE [4] using the grid manager from UG [3]. For the visualization of the results we made use of the toolbox AMIRA [29].

References

1. R.A. Adams. *Sobolev Spaces*. Academic Press, 3rd edition, 1980.
2. P. Bastian, O. Ippisch, F. Rezanezhad, H.J. Vogel, and K. Roth. Numerical simulation and experimental studies of unsaturated water flow in heterogeneous systems. In W. Jäger, R. Rannacher, and J. Warnatz, editors, *Reactive Flows, Diffusion and Transport*, pages 579–598. Springer, 2005.
3. 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(1):27–40, 1997.
4. 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. Preprint 404, MATHEON, 2007, Berlin. Submitted to *Computing*.
5. J. Bear. *Dynamics of Fluids in Porous Media*. Dover Publications, 1988.
6. H. Berninger. Non-overlapping domain decomposition for the Richards equation via superposition operators. In O.B. Widlund and D.E. Keyes, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, LNCSE. Springer, 2009.
7. H. Berninger. *Domain Decomposition Methods for Elliptic Problems with Jumping Nonlinearities and Application to the Richards Equation*. PhD thesis, Freie Universität Berlin, 2007.
8. H. Berninger and M. Discacciati. Convergence of the Robin method for nonlinear transmission problems. In preparation.
9. H. Berninger, R. Kornhuber, and O. Sander. On nonlinear Dirichlet–Neumann algorithms for jumping nonlinearities. In O.B. Widlund and D.E. Keyes, editors, *Domain Decomposition Methods in Science and Engineering XVI*, volume 55 of *LNCSE*, pages 483–490. Springer, 2007.
10. H. Berninger, R. Kornhuber, and O. Sander. A fast and robust solver for the Richards equation in homogeneous soil. *to appear*, 2009.
11. F. Brezzi and G. Gilardi. Functional spaces. In H. Kardestuncer and D.H. Norrie, editors, *Finite Element Handbook*, chapter 2 (part 1), pages 1.29–1.75. Springer, 1987.
12. R.J. Brooks and A.T. Corey. Hydraulic properties of porous media. Technical Report Hydrology Paper No. 3, Colorado State University, Civil Engineering Department, Fort Collins, 1964.

13. N.T. Burdine. Relative permeability calculations from pore-size distribution data. *Petr. Trans., Am. Inst. Mining Metall. Eng.*, (198):71–77, 1953.
14. I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North–Holland, 1976.
15. J. Fuhrmann. *Zur Verwendung von Mehrgitterverfahren bei der numerischen Behandlung elliptischer partieller Differentialgleichungen mit variablen Koeffizienten.* PhD thesis, TU Chemnitz–Zwickau, 1994.
16. J. Fuhrmann and H. Langmach. Stability and existence of solutions of time-implicit finite volume schemes for viscous nonlinear conservation laws. *Appl. Numer. Math.*, 37(1–2):201–230, 2001.
17. R. Helmig, A. Weiss, and B. Wohlmuth. Dynamic capillary effects in heterogeneous porous media. *Comput. Geosci.*, 11:261–274, 2007.
18. R. Helmig, A. Weiss, and B. Wohlmuth. Variational inequalities for modeling flow in heterogeneous porous media with entry pressure. Technical Report 2008/1, Universität Stuttgart, NUPUS, 2008. To appear in *Comput. Geosci.*
19. N. Kikuchi and J.T. Oden. *Contact Problems In Elasticity: A Study Of Variational Inequalities And Finite Element Methods*, volume 8. SIAM Studies in Applied Mathematics, 1988.
20. K. Königsberger. *Analysis 1.* Springer, 1990.
21. R. Kornhuber. On constrained Newton linearization and multigrid for variational inequalities. *Numer. Math.*, 91:699–721, 2002.
22. R. Kornhuber. *Adaptive Monotone Multigrid Methods for Nonlinear Variational Problems.* Teubner, 1997.
23. J.-L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*, volume I. Springer, 1972.
24. M. Marcus and V.J. Mizel. Complete characterization of functions which act, via superposition, on Sobolev spaces. *Trans. Amer. Math. Soc.*, 251: 187–218, 1979.
25. M. Marcus and V.J. Mizel. Every superposition operator mapping one Sobolev space into another is continuous. *J. Funct. Anal.*, 33:217–229, 1979.
26. A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations.* Oxford Science Publications, 1999.
27. W.J. Rawls, L.R. Ahuja, D.L. Brakensiek, and A. Shirmohammadi. Infiltration and soil water movement. In D.R. Maidment, editor, *Handbook of Hydrology*, chapter 5. McGraw–Hill, 1993.
28. L.A. Richards. Capillary conduction of liquids through porous mediums. *Physics*, 1:318–333, 1931.
29. D. Stalling, M. Westerhoff, and H.-C. Hege. Amira: A highly interactive system for visual data analysis. In C. Hansen and C. Johnson, editors, *The Visualization Handbook*, chapter 38, pages 749–767. Elsevier, 2005.
30. M.T. van Genuchten. A closed–form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.*, 44:892–898, 1980.
31. D. Werner. *Funktionalanalysis.* Springer, 5th edition, 2005.