

Efficient water table evolution discretization using domain transformation

W. M. Boon¹  · N. Balbarini² · P. J. Binning² · J. M. Nordbotten^{1,3}

Received: 16 February 2016 / Accepted: 20 September 2016 / Published online: 2 November 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Domain transformation methods are useful techniques for solving problems on non-stationary domains. In this work, we consider the evolution of the water table in an unconfined aquifer. This nonlinear, time-dependent problem is greatly simplified by using a mapping from the physical domain to a reference domain and is then further reduced to a single, (nonlinear) partial differential equation. We show well-posedness of the approach and propose a stable and convergent discretization scheme. Numerical results are presented supporting the theory.

Keywords Domain transformation · Water table evolution · Finite element method

This work was supported in part by Norwegian Research Council grants 233736 and 228832 and the research project GEOCON, Advancing GEOlogical, geophysical and CONtaminant monitoring technologies for contaminated site investigation (contract 1305-00004B). The funding for GEOCON is provided by Innovation Fund Denmark.

✉ W. M. Boon
wietse.boon@uib.no

¹ Department of Mathematics, University of Bergen, Postboks 7803, Bergen 5020, Norway

² Department of Environmental Engineering, Technical University of Denmark, Lyngby, Denmark

³ Department of Civil and Environmental Engineering, Princeton University, E-208 E-Quad, Princeton, NJ, 08544, USA

1 Introduction

The temporal change of the water table in an unconfined aquifer is challenging to model efficiently while maintaining a desired accuracy. Affected by natural processes including precipitation and subsurface flows combined with artificial factors such as wells and dams, this non-stationary problem emerges in a variety of engineering applications.

From a modeling perspective, the water table evolution problem is challenging because the computational domain will typically depend on the solution to the problem. Specifically, the hydraulic head is the unknown variable, yet it defines the elevation of the water table and thereby the upper boundary of the saturated region. As a result, the domain of computation and the variables of interest cannot be considered independently. Several methods have been developed to solve this problem, which we review concisely in the following summary.

A first approach to describe the non-stationary domain is to use a moving mesh. This way, the computational domain is able to follow the dynamic geometry of the saturated region [6]. However, this approach can be computationally demanding for three-dimensional problems and furthermore, large deformations of the domain may cause instabilities [3]. The primary reason for this is that significant stretching of the mesh may lead to undesirable aspect ratios in the elements and cause oscillations in the solution. For the problem considered here, it is imperative that the solution method is able to handle such large deformations and gradients, for example, those occurring in the vicinity of wells.

Alternatively, it is possible to expand the domain of computation towards the surface and treat this problem as saturated-unsaturated flow governed by Richards' equation.

The solution is then obtained using, for example, finite difference [7, 8] or finite element methods [9]. A drawback with this approach is the nonlinearity of Richards' equation which results in a substantial computational cost to obtain a solution.

A third approach is to assume vertical equilibrium and integrate all equations in the z -direction. The problem is then simplified by reduction of dimensionality which allows for the derivation of analytical solutions [2]. However, this approach is only valid if vertical flows (and head gradients) are negligible. Thus, for the sake of generality, we will not make such an assumption in this work.

Finally, the use of coordinate transformations eliminates the challenges of modeling deformable domains by mapping the problem to a stationary reference domain. This has been applied successfully for saturated-unsaturated flow [1, 10]. In those works, the highly nonlinear Richards' equation is considered leading to a computationally demanding method.

Herein, we propose a discretization method by applying a mapping from the saturated region to a reference domain, therewith considering only saturated flow. We then further reduce the original system of equations to a single, nonlinear equation with a single variable, namely the hydraulic head. These two manipulations simplify the problem to such an extent that the stability and convergence of the proposed discretization follow naturally.

The structure of this paper is as follows. Section 2 introduces the governing equations and presents an explicit choice for the mapping to a reference domain. Next, the problem is reduced to a single equation in Section 3 as part of the derivation of the variational form. Sections 4 and 5 are devoted to the discretizations in time and space, respectively, and introduce the Picard scheme used to handle the included nonlinearity. Numerical results confirming the expected stability and convergence of the scheme are presented in Section 6.

2 Model description

In this section, we introduce the water table evolution problem from a modeling perspective. Here, we state the necessary definitions and elaborate on the governing equations included in the model. First, let $\Omega(t)$ denote the d -dimensional spatial domain representing the saturated region with d equal to 2 or 3. In particular, the time-dependent top boundary, $\Gamma(t)$ is of interest since it represents the water table.

The governing equations for the problem are the mass balance equation, Darcy's law and the water table boundary condition. For the mass conservation equation, we assume incompressibility of both the fluid and the surrounding

matrix. This allows us to reformulate mass conservation as volume conservation given by the equation

$$S_s \frac{\partial h}{\partial t} + \nabla \cdot \mathbf{u} = 0, \quad (1)$$

in which S_s represents the specific storativity [$\frac{1}{L}$], h is the hydraulic head [L], and \mathbf{u} represents the volumetric flux [$\frac{L}{T}$]. Secondly, we assume that the flux is governed by Darcy's law, i.e., it satisfies

$$\mathbf{u} = -\mathbf{K}\nabla h. \quad (2)$$

Here, \mathbf{K} represents the conductivity tensor [$\frac{L}{T}$] which is assumed to be symmetric and positive definite. Finally, a key component of this problem is that the top boundary represents the water table and thus, it is time dependent. To incorporate this in the model, we parametrize the boundary $\Gamma(t)$ by $x_3 = \zeta(\mathbf{x}_{||}, t)$. Here, and later, we use the parallel stroke to indicate the horizontal spatial components of a vector, e.g., $\mathbf{x}_{||} = (x_1, x_2)$ in the 3D case and $\mathbf{x}_{||} = x_1$ in the 2D case. In the following, we will continue with dimensionality $d = 3$. The description in two dimensions is analogous.

The newly introduced variable ζ which represents the elevation of the water table is then subject to the evolution equation [13]

$$\phi \frac{\partial \zeta}{\partial t} = I(t) + u_3 - \mathbf{u}_{||} \cdot \nabla_{||} \zeta, \quad (3)$$

in which ϕ represents the porosity, $I(t)$ is a function representing the infiltration, and u_3 is the third component of the flux. $I(t)$ is positive for a downward flux. To close the model, we impose boundary conditions on the boundary $\partial\Omega$. For unconfined aquifers, the elevation of the water table is equal to the hydraulic head at the water table [13]. Thus, our first boundary condition becomes

$$h = \zeta \quad \text{on } \Gamma(t). \quad (4)$$

The remaining boundaries are split up into two parts, namely Γ_D and Γ_N . On Γ_D , we assume that the hydraulic head is given by a known function f . On the other hand, we assume no flow is possible on Γ_N . These two conditions are:

$$h = f, \quad \text{on } \Gamma_D(t), \quad (5a)$$

$$\mathbf{n} \cdot \mathbf{u} = 0, \quad \text{on } \Gamma_N(t). \quad (5b)$$

Here, \mathbf{n} is the outward pointing, unit normal vector on the boundary $\partial\Omega$.

2.1 Transformation to reference space

The problem introduced up to this point may be difficult to solve since the spatial domain is not fixed in time. Therefore, we transform the problem to a stationary domain $\hat{\Omega} = \Omega_{||} \times [0, 1]$, i.e., a domain with unit height. In the subsequent derivation of the transformed governing equations,

the hat notation will consistently refer to variables in the transformed domain.

Let $\Psi(t) : \hat{\Omega} \rightarrow \Omega$ denote any invertible transformation and let $\Phi(t)$ denote its inverse. For simplicity, we will require that $\Psi(t)$ leaves horizontal components unchanged. In particular, this means that $\Psi([\hat{\mathbf{x}}_{||}, 1], t) = [\mathbf{x}_{||}, \zeta(\mathbf{x}_{||}, t)]$. The specific choice of $\Psi(t)$ is discussed in Section 2.2, and an illustration of these definitions is shown in Fig. 1. We continue with the transformation on the variables included in the model to the reference domain. The hydraulic head in reference space \hat{h} will be defined such that $h(\mathbf{x}, t) = \hat{h}(\Phi(\mathbf{x}, t), t)$. For the flux vector $\hat{\mathbf{u}}$, we will use the Piola transform in order to preserve the divergence structure of the original problem [4]. In particular, we define $\hat{\mathbf{u}}$ such that

$$\mathbf{u} = \frac{\hat{\nabla}\Psi}{J} \hat{\mathbf{u}}, \tag{6}$$

with $J = \det(\hat{\nabla}\Psi)$. The transformed version of Eq. 1 then becomes

$$S_s \frac{\partial \hat{h}}{\partial t} + \hat{\nabla} \cdot \hat{\mathbf{u}} = 0. \tag{7}$$

Secondly, we reformulate Darcy’s law given by Eq. 2 for the reference domain. Substitution of $h(\mathbf{x}, t) = \hat{h}(\Phi(\mathbf{x}, t), t)$ leads us to

$$\mathbf{u}(\mathbf{x}, t) = -\mathbf{K} \nabla h(\Phi(\mathbf{x}, t), t). \tag{8}$$

Next, we use Eq. 6 and the chain rule to obtain the following formulation in transform space

$$\frac{\hat{\nabla}\Psi}{J} \hat{\mathbf{u}} = -\mathbf{K}(\nabla\Phi)^T \hat{\nabla}\hat{h}. \tag{9}$$

It follows that we may write Darcy’s law in transform space as

$$\hat{\mathbf{u}} = -\hat{\mathbf{K}} \hat{\nabla}\hat{h}, \tag{10}$$

in which the transformed conductivity $\hat{\mathbf{K}}$ is given by

$$\hat{\mathbf{K}} = J(\hat{\nabla}\Psi)^{-1} \mathbf{K}(\hat{\nabla}\Psi)^{-T}. \tag{11}$$

Here, we have made the final substitution $\nabla\Phi = (\hat{\nabla}\Psi)^{-1}$.

Following the same procedure, we aim to find the equation for the water table evolution in reference space, i.e., the equivalent to Eq. 3. First, we note that since Ψ leaves

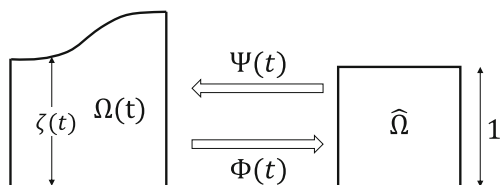


Fig. 1 The transformations between the time-dependent, physical domain $\Omega(t)$ and the stationary, reference domain $\hat{\Omega}$

horizontal components unchanged, we can express Ψ in terms of only the third component Ψ_3

$$\Psi(\hat{\mathbf{x}}) = [\hat{\mathbf{x}}_{||}, \Psi_3(\hat{x}_3, \zeta)]. \tag{12}$$

This allows us to derive the Jacobian matrix

$$\hat{\nabla}\Psi = \begin{bmatrix} I_2 & 0 \\ \nabla_{||}^T \zeta & \frac{\partial}{\partial \hat{x}_3} \Psi_3 \end{bmatrix} \text{ at } \hat{x}_3 = 1. \tag{13}$$

We now continue by rewriting Eq. 3 in the reference space. First, we note that

$$u_3 = \mathbf{e}_3 \cdot \frac{\hat{\nabla}\Psi}{J} \hat{\mathbf{u}} = \frac{[\nabla_{||}^T \zeta, \frac{\partial}{\partial \hat{x}_3} \Psi_3]}{J} \hat{\mathbf{u}}. \tag{14}$$

Secondly, the last term of Eq. 3 is rewritten using Eqs. 6 and 13 as

$$\mathbf{u}_{||} \cdot \nabla_{||} \zeta = \left(\mathbf{e}_{||}^T \frac{\hat{\nabla}\Psi}{J} \hat{\mathbf{u}} \right) \cdot \nabla_{||} \zeta = \left(\frac{[I_2, 0]}{J} \hat{\mathbf{u}} \right) \cdot \nabla_{||} \zeta. \tag{15}$$

With $J = \frac{\partial}{\partial \hat{x}_3} \Psi_3$, the water table evolution (3) is reformulated as

$$\phi \frac{\partial \zeta}{\partial t} = I(t) + \hat{u}_3. \tag{16}$$

The three equations (7), (10), and (16) form our system in reference space, which we repeat here for convenience:

$$S_s \frac{\partial \hat{h}}{\partial t} + \hat{\nabla} \cdot \hat{\mathbf{u}} = 0, \tag{17a}$$

$$\hat{\mathbf{u}} = -\hat{\mathbf{K}} \hat{\nabla}\hat{h}, \quad \text{in } \hat{\Omega}, \tag{17b}$$

$$\phi \frac{\partial \zeta}{\partial t} = I(t) + \hat{u}_3, \quad \text{on } \hat{\Gamma}. \tag{17c}$$

The associated boundary conditions are obtained from (4) and (2):

$$\hat{h} = \zeta, \quad \text{on } \hat{\Gamma}, \tag{18a}$$

$$\hat{h} = f, \quad \text{on } \hat{\Gamma}_D, \tag{18b}$$

$$\mathbf{n} \cdot \hat{\mathbf{u}} = 0, \quad \text{on } \hat{\Gamma}_N. \tag{18c}$$

2.2 A choice of transformation

The transformation $\Psi(t)$ can in principle be chosen as any monotone function. However, although certain complex choices may hold desirable properties, we aim for a computationally efficient discretization scheme. Therefore, we choose $\Psi(t)$ to be a linear transformation in the vertical coordinate, such that:

$$\mathbf{x} = \Psi(\hat{\mathbf{x}}, t) = [\hat{\mathbf{x}}_{||}, \hat{x}_3 \zeta(\hat{\mathbf{x}}_{||}, t)]. \tag{19}$$

The inverse transformation Φ is then given by

$$\hat{\mathbf{x}} = \Phi(\mathbf{x}, t) = \begin{bmatrix} \mathbf{x}_{||}, \\ \frac{x_3}{\zeta(\mathbf{x}_{||}, t)} \end{bmatrix}. \tag{20}$$

Furthermore, the Jacobian matrices can be stated explicitly as follows.

$$\hat{\nabla}\Psi = \begin{bmatrix} I_2 & 0 \\ \hat{x}_3 \nabla_{\parallel}^T \zeta & \zeta \end{bmatrix} \text{ and } \nabla\Phi = \begin{bmatrix} I_2 & 0 \\ -x_3 \zeta^{-2} \nabla_{\parallel}^T \zeta & \zeta^{-1} \end{bmatrix}. \tag{21}$$

A more complex choice, adapted for layered systems, will be considered in the numerical examples.

3 Time-continuous formulation

In this section, we establish some basic properties for the time-continuous problem. We start by stating the variational form and rewrite the system (17a–17c) to a single equation with a single unknown variable. In turn, this gives us a noteworthy advantage in the consequent analysis of the well-posedness of the problem.

3.1 Variational formulation

The first step is to note that according to Eq. 18a, the variable ζ is the trace of \hat{h} on $\hat{\Gamma}$. Thus, the definition of ζ as an independent variable is obsolete and it becomes convenient to discard the notation of ζ and use the variable \hat{h} , instead. However, it is important to note that the transformation Ψ and therewith the conductivity tensor $\hat{\mathbf{K}}$ only depend on the trace of \hat{h} on $\hat{\Gamma}$ and not the whole function \hat{h} . To emphasize this, we will occasionally revert to the notation

$$\hat{\mathbf{K}}(\zeta) = \hat{\mathbf{K}}(\hat{h}|_{\hat{\Gamma}}). \tag{22}$$

In the following steps, we will continue to simplify the system (17a–17c). With the aim of solving this problem with the use of the finite element method, we derive the variational formulation. For this, we employ the conventional inner product notation

$$(h, g)_{\hat{\Omega}} = \int_{\hat{\Omega}} h(\hat{x})g(\hat{x})d\hat{x}, \tag{23}$$

and omit hats on variables and gradients for brevity. Let us multiply Eq. 7 with a test function $g \in H^1$ and integrate over the domain $\hat{\Omega}$. We pay special attention to the second term in Eq. 7, in which we first substitute Darcy’s law (10). With the application of partial integration and the associated boundary conditions, this term becomes

$$\begin{aligned} & -(\nabla \cdot (\hat{\mathbf{K}}(h)\nabla h), g)_{\hat{\Omega}} \\ &= (\hat{\mathbf{K}}(h)\nabla h, \nabla g)_{\hat{\Omega}} - (\mathbf{n} \cdot (\hat{\mathbf{K}}(h)\nabla h), g)_{\hat{\Gamma}} \\ &= (\hat{\mathbf{K}}(h)\nabla h, \nabla g)_{\hat{\Omega}} + (u_3, g)_{\hat{\Gamma}} \\ &= (\hat{\mathbf{K}}(h)\nabla h, \nabla g)_{\hat{\Omega}} + \phi \left(\frac{\partial h}{\partial t}, g \right)_{\hat{\Gamma}} - (I(t), g)_{\hat{\Gamma}}. \end{aligned} \tag{24}$$

Here, we have used Eq. 16 in order to eliminate u_3 . We are now ready to write the weak form of Eq. 7 as

$$S_s \left(\frac{\partial h}{\partial t}, g \right)_{\hat{\Omega}} + (\hat{\mathbf{K}}(h)\nabla h, \nabla g)_{\hat{\Omega}} + \phi \left(\frac{\partial h}{\partial t}, g \right)_{\hat{\Gamma}} = (I(t), g)_{\hat{\Gamma}}. \tag{25}$$

Thus, we have reduced the system of equations derived in the previous section to a single equation. The advantage of this becomes evident in the next sections.

3.2 Well-posedness

Next, we analyze the well-posedness of the derived problem with the use of the following lemma.

Lemma 1 *Let the initial and boundary conditions satisfy $h_0 \geq \delta > 0$ and $f \geq 0$. Then the solution to Eq. 25 exists and is unique.*

Proof The proof is done by showing that the conductivity tensor $\hat{\mathbf{K}}(\zeta)$ in Eq. 25 is symmetric and positive definite under these assumptions. Then, the problem is identified as a diffusion-type problem which is known to be well-posed.

Let us start with the assumption that the physical conductivity tensor \mathbf{K} is symmetric and positive definite. Equation 11 states that $\hat{\mathbf{K}}(\zeta)$ is equal to \mathbf{K} which is left- and right- multiplied by transpose matrices. Thus, if ζ is bounded away from zero, it follows that $\hat{\mathbf{K}}(\zeta)$ is symmetric and positive definite as well.

In order to ensure that ζ is non-zero, it is essential that the solution to Eq. 25 is also non-zero for all time. This property can be shown by using the maximum (or rather, minimum) principle related to problem (25).

From this principle, we deduce that the solution h attains its minimum on the boundary. On this boundary, we distinguish three different cases. First, if the minimum is attained on $\hat{\Gamma}_D$, then it follows $h = f$. Since $f > 0$ by assumption, it follows that h is strictly positive as well.

Secondly, the minimum could be attained in a point \hat{x}_0 on $\hat{\Gamma}$. We then introduce Hopf’s Lemma which states that $\frac{\partial h}{\partial \mathbf{n}}(\hat{x}_0) < 0$, yet this contradicts the homogeneous Neumann boundary condition (2) prescribed there.

Finally, if the minimum is attained on $\hat{x}_0 \in \hat{\Gamma}$, we recall that the boundary condition there is given by

$$I(t) - \mathbf{n} \cdot \hat{\mathbf{K}}(\zeta)\nabla h = \phi \frac{\partial h}{\partial t}. \tag{26}$$

Again, using Hopf’s lemma, it follows that $\frac{\partial h}{\partial \mathbf{n}}(\hat{x}_0) < 0$. The prescribed boundary condition then implies that $\frac{\partial h}{\partial t} > 0$ since the infiltration is a positive function. A positive temporal derivative in combination with the assumed $h_0 \geq \delta > 0$ means that the function h remains strictly positive in time.

Thus, we conclude that the solution h is bounded away from zero. This implies positive definiteness of the conductivity tensor $\hat{\mathbf{K}}(h)$ and therewith the well-posedness of the problem. \square

4 Time-discrete formulation

In this section, we continue by discretizing problem (25) in time. We show that using a Backward Euler approach is sufficient to obtain a stable discretization.

Let us introduce Δt as the time step and h_n as the solution at time t_n . Using a Backward Euler discretization scheme, we solve the following problem for h_{n+1}

$$\begin{aligned} & S_s \Delta t^{-1} (h_{n+1}, g)_{\hat{\Omega}} + \left(\hat{\mathbf{K}}(h_{n+1}) \nabla h_{n+1}, \nabla g \right)_{\hat{\Omega}} \\ & + \phi \Delta t^{-1} (h_{n+1}, g)_{\hat{\Gamma}} \\ = & S_s \Delta t^{-1} (h_n, g)_{\hat{\Omega}} + \phi \Delta t^{-1} (h_n, g)_{\hat{\Gamma}} + (I(t), g)_{\hat{\Gamma}}. \end{aligned} \tag{27}$$

Note that the nonlinearity in the diffusive term requires special attention. In order to solve the problem, we propose using a Picard iterative scheme. In this context, let the superscript k denote the iteration number. The next iterate can then be found by solving the following problem:

$$\begin{aligned} & S_s \Delta t^{-1} (h_{n+1}^{k+1}, g)_{\hat{\Omega}} + \left(\hat{\mathbf{K}}(h_{n+1}^k) \nabla h_{n+1}^{k+1}, \nabla g \right)_{\hat{\Omega}} \\ & + \phi \Delta t^{-1} (h_{n+1}^{k+1}, g)_{\hat{\Gamma}} \\ = & S_s \Delta t^{-1} (h_n, g)_{\hat{\Omega}} + \phi \Delta t^{-1} (h_n, g)_{\hat{\Gamma}} + (I(t), g)_{\hat{\Gamma}}. \end{aligned} \tag{28}$$

Each iterate remains subject to the Dirichlet boundary condition

$$h_{n+1}^{k+1} = f, \text{ on } \hat{\Gamma}_D. \tag{29}$$

A natural choice as an initial guess is $h_{n+1}^0 = h_n$, i.e., the solution at the previous time step.

We are interested in convergence of this scheme to a certain limit solution as this limit solution will solve the nonlinear problem. Therefore, a formal proof concerning the existence and uniqueness of such a limit solution is given by the following lemma.

Lemma 2 *Let the same assumptions from Lemma 1 hold and let the time step Δt be sufficiently small. Then the time-discrete problem (27) has a unique solution.*

Proof For this proof, the essential tool is the set $H_\delta^1(\hat{\Gamma})$ which contains (weakly) differentiable functions bounded away from zero. More precisely, for a given $\delta > 0$, this set is defined as follows

$$H_\delta^1(\hat{\Gamma}) = \{h \in H^1(\hat{\Gamma}) \mid h \geq \delta, \text{ a.e.}\}. \tag{30}$$

If $w \in H^1(\hat{\Omega})$ solves (28) for a given $\zeta \in H_\delta^1(\hat{\Gamma})$, then we can construct the operator $A : H_\delta^1(\hat{\Gamma}) \rightarrow H_\delta^1(\hat{\Gamma})$ given by

$$A[\zeta] = w|_{\hat{\Gamma}}. \tag{31}$$

With these definitions, the lemma states that if $f \in H_\delta^0(\hat{\Gamma}_D)$ and $h_n \in H_\delta^1(\hat{\Omega})$ for some $\delta > 0$, then the system (27) has a unique solution. The proof is done through an application of Banach’s Fixed Point theorem on the introduced, nonlinear mapping A related to system (28).

Let us consider two functions $\zeta_1, \zeta_2 \in H_\delta^1(\hat{\Gamma})$ and denote the corresponding solutions to Eq. 28 by w_1 and w_2 , respectively. Subtraction of these two equations leads us to

$$\begin{aligned} & (\hat{\mathbf{K}}(\zeta_1) \nabla w_1 - \hat{\mathbf{K}}(\zeta_2) \nabla w_2, \nabla g)_{\hat{\Omega}} \\ & + \Delta t^{-1} (w_1 - w_2, g)_{\hat{\Gamma}} = 0. \end{aligned} \tag{32}$$

We then set $g = w_1 - w_2$ to obtain

$$\begin{aligned} & S_s \Delta t^{-1} \|w_1 - w_2\|_{0, \hat{\Omega}}^2 + \phi \Delta t^{-1} \|w_1 - w_2\|_{0, \hat{\Gamma}}^2 \\ = & - \left(\hat{\mathbf{K}}(\zeta_1) \nabla w_1 - \hat{\mathbf{K}}(\zeta_2) \nabla w_2, \nabla (w_1 - w_2) \right)_{\hat{\Omega}} \\ = & - \frac{1}{2} \left((\hat{\mathbf{K}}(\zeta_1) - \hat{\mathbf{K}}(\zeta_2)) \nabla (w_1 + w_2), \nabla (w_1 - w_2) \right)_{\hat{\Omega}} \\ & - \frac{1}{2} \left((\hat{\mathbf{K}}(\zeta_1) + \hat{\mathbf{K}}(\zeta_2)) \nabla (w_1 - w_2), \nabla (w_1 - w_2) \right)_{\hat{\Omega}}. \end{aligned} \tag{33}$$

Next, we observe that $(\hat{\mathbf{K}}(\zeta_1) + \hat{\mathbf{K}}(\zeta_2))$ is positive definite. By rearranging the terms and applying the Cauchy-Schwarz-Young inequality, we derive

$$\begin{aligned} & S_s \Delta t^{-1} \|w_1 - w_2\|_{0, \hat{\Omega}}^2 + \phi \Delta t^{-1} \|w_1 - w_2\|_{0, \hat{\Gamma}}^2 \\ & + C_1 \|\nabla (w_1 - w_2)\|_{0, \hat{\Omega}}^2 \\ \leq & - \frac{1}{2} \left((\hat{\mathbf{K}}(\zeta_1) - \hat{\mathbf{K}}(\zeta_2)) \nabla (w_1 + w_2), \nabla (w_1 - w_2) \right)_{\hat{\Omega}} \\ \leq & \frac{1}{2} \|\hat{\mathbf{K}}(\zeta_1) - \hat{\mathbf{K}}(\zeta_2)\|_{0, \hat{\Omega}} \|\nabla (w_1 - w_2)\|_{0, \hat{\Omega}} \|\nabla (w_1 + w_2)\|_{0, \hat{\Omega}} \\ \leq & \frac{1}{2} \left(\frac{1}{2\epsilon} \|\hat{\mathbf{K}}(\zeta_1) - \hat{\mathbf{K}}(\zeta_2)\|_{0, \hat{\Omega}}^2 \right. \\ & \left. + \epsilon \|\nabla (w_1 - w_2)\|_{0, \hat{\Omega}}^2 \|\nabla (w_1 + w_2)\|_{0, \hat{\Omega}}^2 \right). \end{aligned} \tag{34}$$

The next step is to choose the constant ϵ . In order to do this, we form a bound for $\|\nabla (w_1 + w_2)\|_{0, \hat{\Omega}}^2$. Let us reconsider Eq. 28 and set $g = w_1 - h_n$. This choice gives us

$$\begin{aligned} & S_s \Delta t^{-1} \|w_1 - h_n\|_{0, \hat{\Omega}}^2 + \|\hat{\mathbf{K}}(\zeta_1)^{1/2} \nabla w_1\|_{0, \hat{\Omega}}^2 \\ & + \phi \Delta t^{-1} \|w_1 - h_n\|_{0, \hat{\Gamma}}^2 \\ = & (I(t), w_1 - h_n)_{\hat{\Gamma}} + \left(\hat{\mathbf{K}}(\zeta_1) \nabla w_1, \nabla h_n \right)_{\hat{\Omega}} \\ \leq & \frac{1}{2} \phi^{-1} \Delta t \|I(t)\|_{0, \hat{\Gamma}}^2 + \frac{1}{2} \phi \Delta t^{-1} \|w_1 - h_n\|_{0, \hat{\Gamma}}^2 \\ & + \frac{1}{2} \|\hat{\mathbf{K}}(\zeta_1)^{1/2} \nabla w_1\|_{0, \hat{\Omega}}^2 + \frac{1}{2} \|\hat{\mathbf{K}}(\zeta_1)^{1/2} \nabla h_n\|_{0, \hat{\Omega}}^2. \end{aligned} \tag{35}$$

Using the positive definiteness of $\hat{\mathbf{K}}$ once more, we obtain

$$\begin{aligned} \|\nabla w_1\|_{0,\hat{\Omega}}^2 &\leq C_2^{-1} \|\hat{\mathbf{K}}(\zeta_1)^{1/2} \nabla w_1\|_{0,\hat{\Omega}}^2 \\ &\leq C_2^{-1} \left(\phi^{-1} \Delta t \|I(t)\|_{0,\hat{\Gamma}}^2 + \|\hat{\mathbf{K}}(\zeta_1)^{1/2} \nabla h_n\|_{0,\hat{\Omega}}^2 \right). \end{aligned} \quad (36)$$

A bound on $\|\nabla w_2\|_{0,\hat{\Omega}}^2$ is formed analogously. This allows us to construct the desired bound on $\|\nabla(w_1 + w_2)\|_{0,\hat{\Omega}}^2$ which takes the following form:

$$\begin{aligned} \|\nabla(w_1 + w_2)\|_{0,\hat{\Omega}}^2 &\leq 2 \left(\|\nabla w_1\|_{0,\hat{\Omega}}^2 + \|\nabla w_2\|_{0,\hat{\Omega}}^2 \right) \\ &\leq C_3 \Delta t + C_4. \end{aligned} \quad (37)$$

Note that if $C_3 \Delta t + C_4 = 0$, then $\nabla(w_1 + w_2) = 0$ as well. Consequently, we obtain $w_1 = w_2$ from inequality (34) and conclude that A is a contraction.

Otherwise, we set $\epsilon = C_1(C_3 \Delta t + C_4)^{-1}$ and derive

$$\begin{aligned} \Delta t^{-1} \|w_1 - w_2\|_{0,\Gamma}^2 &\leq \frac{C_3 \Delta t + C_4}{4C_1 \phi} \|\hat{\mathbf{K}}(\zeta_1) - \hat{\mathbf{K}}(\zeta_2)\|_{0,\hat{\Omega}}^2 \\ &\leq (C_5 \Delta t + C_6) \|\zeta_1 - \zeta_2\|_{0,\hat{\Gamma}}^2. \end{aligned} \quad (38)$$

Here, we used the property that $\hat{\mathbf{K}}(\zeta)$ is Lipschitz continuous when ζ is bounded away from zero. (Note, however, that the values of C_5 and C_6 depend on $\delta > 0$.) After multiplication with the time step Δt , we obtain

$$\|w_1 - w_2\|_{0,\Gamma}^2 \leq (C_5 \Delta t^2 + C_6 \Delta t) \|\zeta_1 - \zeta_2\|_{0,\hat{\Gamma}}^2. \quad (39)$$

Substitution of $A[\zeta] = w|_{\hat{\Gamma}}$ and taking the square root leads us to the main result

$$\|A[\zeta_1] - A[\zeta_2]\|_{0,\Gamma} \leq (C_5 \Delta t^2 + C_6 \Delta t)^{1/2} \|\zeta_1 - \zeta_2\|_{0,\hat{\Gamma}}. \quad (40)$$

Thus, A is a contraction if we choose the time step Δt sufficiently small. Since we are considering the continuous problem, this constraint is independent of grid size. Now, by Banach’s Fixed Point Theorem, it follows that A has a unique, fixed point $\zeta \in H_\delta^1(\hat{\Gamma})$ satisfying

$$A[\zeta] = \zeta. \quad (41)$$

The unique solution h can then be obtained by solving problem (28) with ζ as the trace value. \square

5 Finite element formulation

The analytical results shown up to this point all concern the continuous case. The next step is to discretize the problem spatially using the finite element method leading to a discrete problem. First, the domain $\hat{\Omega}$ is partitioned using a Delaunay triangulation.

The Galerkin method is then applied to approximate the solution by a combination of N known basis functions $\varphi_j(\hat{\mathbf{x}})$

with $j = 1, \dots, N$. In particular, we assume that the solution can be written as follows:

$$h(\hat{\mathbf{x}}) = \sum_j^N h_j \varphi_j(\hat{\mathbf{x}}), \quad (42)$$

with h_j constants representing the values of h at node $\hat{\mathbf{x}}_j$. These values are collected in a vector denoted by \mathbf{h} . Finally, the test functions g will be chosen as φ_i and the system (28) is rewritten to the following matrix-vector multiplication.

$$A(\mathbf{h}^k) \mathbf{h}^{k+1} = \mathbf{b}. \quad (43)$$

The elements in the matrix A and vector \mathbf{b} can be computed as follows

$$\begin{aligned} A_{i,j}(\mathbf{h}^k) &= S_s \Delta t^{-1} (\varphi_i, \varphi_j)_{\hat{\Omega}} + \left(\hat{\mathbf{K}}(\mathbf{h}^k) \nabla \varphi_i, \nabla \varphi_j \right)_{\hat{\Omega}} \\ &\quad + \phi \Delta t^{-1} (\varphi_i, \varphi_j)_{\hat{\Gamma}}, \\ b_i &= S_s \Delta t^{-1} (h_n, \varphi_i)_{\hat{\Omega}} + \phi \Delta t^{-1} (h_n, \varphi_i)_{\hat{\Gamma}} + (I(t), \varphi_i)_{\hat{\Gamma}}. \end{aligned} \quad (44)$$

Here, in an abuse of notation, $\hat{\mathbf{K}}$ acts upon the function h^k given by the vector \mathbf{h}^k .

During the transition to a discrete problem, it is not trivial that the resulting, discrete problem is also well-posed. We therefore present the following, separate lemma concerning the discrete problem.

Lemma 3 *Let the the assumptions from Lemma 1 hold and let the spatial discretization be linear Lagrange elements on a Delaunay triangulation. If, furthermore, a sufficiently small time step is chosen, then the solution to the fully discrete problem (43) exists and is unique.*

Proof The choice of linear Lagrange elements on a Delaunay triangulation ensures that the matrix A is an M-matrix [11, 14]. In turn, the discrete maximum principle holds [5, 14]. Thus, following the same arguments from the proofs of Lemmas 1 and 2, we obtain the result. \square

Finally, we note that Eq. 43 is a diffusion-type problem and Lagrangian finite elements are a classic choice to obtain a stable discretization. Together with the well-posedness from the Lemma 3, quadratic convergence of the hydraulic head and linear converge of its gradient are to be expected.

6 Numerical results

In this section, we introduce a model problem in order to demonstrate the theoretical results concerning stability and convergence. To capture the importance of vertical flows, we base our example on the water table evolution in the vicinity of an extraction well. A broader comparison to

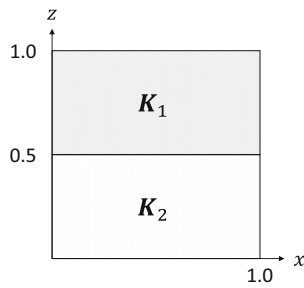


Fig. 2 The computational domain is split into two layers with (possibly different) permeabilities

existing numerical techniques and application to field data is forthcoming.

As an initial condition, we will assume that the hydraulic head is equal to one uniformly, i.e.,

$$h_0 \equiv 1. \tag{46}$$

The extraction well is incorporated as a Neumann boundary condition and starts to operate at time $t = 0$. We will assume no flow conditions on the bottom boundary and a constant unit head on the boundaries far from the well, i.e. $f = 1$. Furthermore, the top boundary is governed by the water table evolution as prescribed by Eq. 16.

The (dimensionless) parameters are as follows, $\mathbf{K} = \mathbf{I}$ where \mathbf{I} is the $d \times d$ identity tensor, $S_s = 0.1$, $\phi = 0.1$, and we introduce a small, constant infiltration $I(t) = 10^{-3}$. As a result of the previous analysis, the magnitude of the extraction rate may not be too large in order to ensure that the hydraulic head remains strictly positive. Thus, we set this rate to -0.3 in the 2D case and to -1.0 for the 3D equivalent.

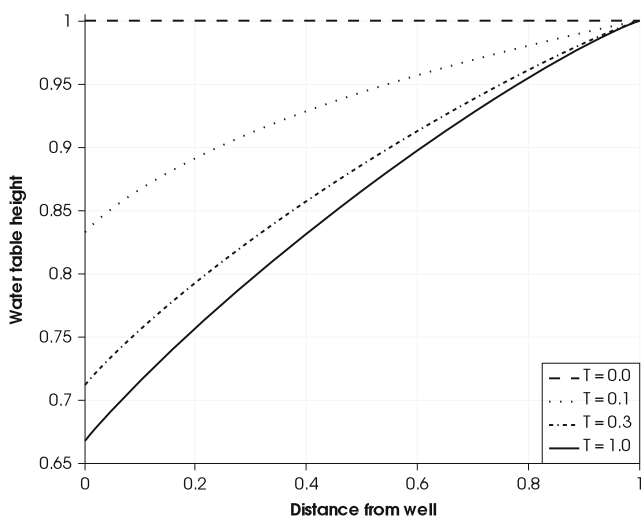


Fig. 3 Water table evolution for the two-dimensional, homogeneous case at different times. The solution is monotone, approaching an equilibrium state at $t = 1.0$

Table 1 Convergence results for the 2D, homogeneous test case showing the expected quadratic convergence of the hydraulic head and linear converge of its gradient

h_{mesh}	h		∇h	
	error	rate	error	rate
1/2	7.02E-03		1.04E-01	
1/4	1.89E-03	1.89	5.57E-02	0.90
1/8	4.86E-04	1.96	2.90E-02	0.94
1/16	1.21E-04	2.00	1.48E-02	0.97
1/32	2.91E-05	2.06	7.38E-03	1.01
1/64	6.13E-06	2.25	3.35E-03	1.14

Although these values may not coincide with physical values, this choice suffices to demonstrate the performance of the scheme. All computations are implemented within the open-source environment FEniCS [12]. In the following subsections, the model problem is considered on a two-dimensional domain followed by a 3D equivalent.

6.1 2D test case

Let us introduce the two-dimensional, reference domain $\hat{\Omega}$, which we define as the unit square. Heterogeneity is incorporated by defining two permeability tensors \mathbf{K}_1 and \mathbf{K}_2 (Fig. 2). We emphasize that radial symmetry is not assumed here and therefore the left boundary can be viewed as an infinite strip sink rather than an extraction well.

The problem is solved on a uniform, triangular mesh using a fixed time step $\Delta t = 0.1$. Adaptive time stepping or local grid refinement may hold certain advantages, but this exceeds the focus of this work.

Let us first consider the homogeneous case, i.e., $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{I}$. As depicted in Fig. 3, the solution obtained is stable in space and time, showing no sign of oscillations. Furthermore, the water table evolves rapidly to a steady state, which is to be expected.

Table 2 The scheme remains convergent when heterogeneities are included with a slight deterioration in the rates

h_{mesh}	h		∇h	
	error	rate	error	rate
1/2	7.70E-03		3.44E-01	
1/4	4.88E-03	0.66	2.48E-01	0.47
1/8	3.24E-03	0.59	1.66E-01	0.58
1/16	5.34E-04	2.60	7.96E-02	1.06
1/32	1.67E-04	1.68	5.20E-02	0.62
1/64	6.86E-05	1.28	3.25E-02	0.68

Table 3 The convergence rates can be recovered in the heterogeneous case by only applying the coordinate transformation in the upper part of the domain

h_{mesh}	h		∇h	
	error	rate	error	rate
1/2	3.28E-03		8.30E-02	
1/4	8.84E-04	1.89	4.45E-02	0.90
1/8	2.31E-04	1.94	2.32E-02	0.94
1/16	5.90E-05	1.97	1.19E-02	0.97
1/32	1.49E-05	1.99	5.93E-03	1.00
1/64	3.60E-06	2.05	2.70E-03	1.14

Next, we consider the convergence of the discretization. Due to the nonlinearity of the problem, an analytical solution is not at our disposal. Instead, the relative error is computed with respect to a solution obtained on a fine mesh at time $t = 1.0$. The results in Table 1 show that the scheme is convergent with rates as expected from linear finite elements, namely second-order convergence of the L^2 -error in h and first-order convergence of its gradient.

The incorporation of heterogeneity is done by choosing a higher permeability in the upper domain. Specifically, we set $\mathbf{K}_1 = 4\mathbf{K}_2 = 4\mathbf{I}$. As the numerical results show in Table 2, the scheme remains convergent with a moderate decrease in the convergence rates. The deterioration is mainly due to the fact that the heterogeneities are not stationary in the reference domain. Nevertheless, the results show that convergence is maintained despite the unresolved heterogeneities.

If the water table remains within the upper region, a full recovery of the convergence rates from the homogeneous

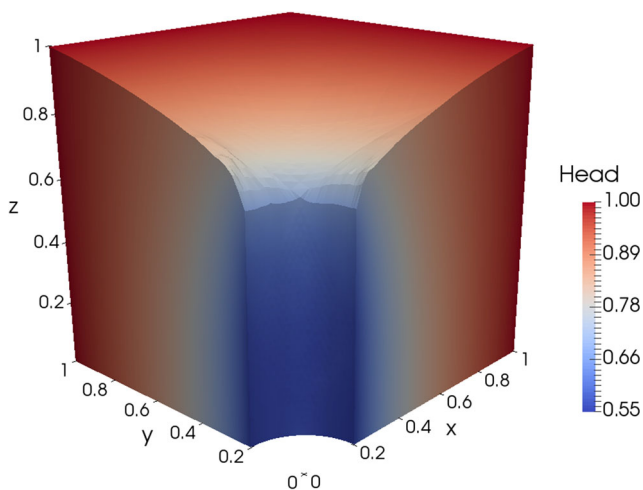


Fig. 4 The hydraulic head in the 3D test case represented in the physical domain at time $t = 1.0$

Table 4 Convergence results for the 3D test case. The observed rates approach the rates expected from the theory

h_{mesh}	h		∇h	
	error	rate	error	rate
1	3.62E-02		6.45E-01	
1/2	1.66E-02	1.13	4.61E-01	0.48
1/4	6.30E-03	1.40	2.87E-01	0.68
1/8	1.95E-03	1.69	1.60E-01	0.84

problem can be achieved by choosing a transformation which allows the heterogeneity to be stationary in the computational domain. For this purpose, the transformation is chosen such that it only applies to the upper layer. The recovered convergence rates are shown in Table 3.

We emphasize that this choice of coordinate transformation gives rise to a natural coupling with the saturated region situated below. This is an advantage compared to approaches with Richards' equation where the coupling with deeper, saturated layers is typically less trivial.

6.2 3D test case

In the 3D case, we consider $\hat{\Omega}$ to be a unit cube. The well, positioned along the z -axis with radius 0.2 is removed from the domain and is incorporated as a Neumann boundary condition.

As shown in Fig. 4, the scheme once again produces a stable solution. Qualitatively, a decrease in the water table can be seen in the vicinity of the well, as expected. The equilibrium state, depicted in Fig. 4, is attained at time $t = 1.0$.

Table 4 shows that the convergence rates approach the asymptotic rates as the nonlinearities are resolved. These numerical results therefore support the theory presented in previous sections.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- An, H., Ichikawa, Y., Tachikawa, Y., Shiiba, M.: Three-dimensional finite difference saturated-unsaturated flow modeling with nonorthogonal grids using a coordinate transformation method. *Water Resour. Res.* **46**(11) (2010)
- Bear, J.: *Dynamics of Fluids in Porous Media*. Dover Civil and Mechanical Engineering Series. Dover (1972)

3. Bresciani, E., Davy, P., Dreuzy, J.R.: A finite volume approach with local adaptation scheme for the simulation of free surface flow in porous media. *Int. J. Numer. Anal. Methods Geomech.* **36**(13), 1574–1591 (2012)
4. Ciarlet, P.G.: *Mathematical Elasticity: Three-Dimensional Elasticity*, vol. 1. Elsevier (1993)
5. Ciarlet, P.G., Raviart, P.A.: Maximum principle and uniform convergence for the finite element method. *Comput. Methods Appl. Mech. Eng.* **2**(1), 17–31 (1973)
6. Darbandi, M., Torabi, S., Saadat, M., Daghighi, Y., Jarrahbashi, D.: A moving-mesh finite-volume method to solve free-surface seepage problem in arbitrary geometries. *Int. J. Numer. Anal. Methods Geomech.* **31**(14), 1609–1629 (2007)
7. Dogan, A., Motz, L.H.: Saturated-unsaturated 3d groundwater model. i: development. *J. Hydrol. Eng.* **10**(6), 492–504 (2005)
8. Freeze, R.A.: Three-dimensional, transient, saturated-unsaturated flow in a groundwater basin. *Water Resour. Res.* **7**(2), 347–366 (1971)
9. Keating, E., Zyvoloski, G.: A stable and efficient numerical algorithm for unconfined aquifer analysis. *Ground Water* **47**(4), 569–579 (2009)
10. Kinouchi, T., Kanda, M., Hino, M.: Numerical simulation of infiltration and solute transport in an s-shaped model basin by a boundary-fitted grid system. *J. Hydrol.* **122**(1), 373–406 (1991)
11. Křížek, M., Qun, L.: On diagonal dominance of stiffness matrices in 3D. *East-West J. Numer. Math.* **3**(1), 59–69 (1995)
12. Logg, A., Mardal, K.A., Wells, G.N., et.al.: *Automated Solution of Differential Equations by the Finite Element Method*. Springer (2012) doi:[10.1007/978-3-642-23099-8](https://doi.org/10.1007/978-3-642-23099-8)
13. Nordbotten, J.M., Celia, M.A.: *Geological Storage of CO₂: Modeling Approaches for Large-Scale Simulation*. Wiley (2011)
14. Xu, J., Zikatanov, L.: A monotone finite element scheme for convection-diffusion equations. *Mathematics of Computation of the American Mathematical Society* **68**(228), 1429–1446 (1999)