The Localized Reduced Basis Multiscale method for two-phase flows in porous media

S. Kaulmann∗†, B. Flemisch‡, B. Haasdonk†, K.-A. Lie§, and M. Ohlberger∗

In this work, we propose a novel model order reduction approach for two-phase flow in porous media by introducing a formulation in which the mobility, which realizes the coupling between phase saturations and phase pressures, is regarded as a parameter to the pressure equation. Using this formulation, we introduce the Localized Reduced Basis Multiscale method to obtain a low-dimensional surrogate of the high-dimensional pressure equation. By applying ideas from model order reduction for parametrized partial differential equations, we are able to split the computational effort for solving the pressure equation into a costly offline step that is performed only once and an inexpensive online step that is carried out in every time step of the two-phase flow simulation, which is thereby largely accelerated. Usage of elements from numerical multiscale methods allows us to displace the computational intensity between the offline and online step to reach an ideal runtime at acceptable error increase for the two-phase flow simulation.

1 Introduction

Real-world applications of two-phase flow in a porous medium are often characterized by large disparities in physical scales in the sense that local variations in the permeability and porosity of the porous medium may strongly affect flow patterns on scales that are orders of magnitude larger. To accurately resolve the effects of local variations, one easily ends up with equation systems (for pressures and phase saturations) that may contain millions of unknowns. Although (non)linear systems of this size obviously can

∗Institute for Computational and Applied Mathematics, Einsteinstr. 62, D-48149 Münster, Germany, {sven.kaulmann,mario.ohlberger}@uni-muenster.de
†Institute for Applied Analysis and Numerical Simulation, Pfaffenwaldring 57, D-70569 Stuttgart, Germany, {sven.kaulmann,haasdonk}@mathematik.uni-stuttgart.de
‡Institute for Modelling Hydraulic and Environmental Systems, Pfaffenwaldring 61, D-70569 Stuttgart, Germany, bernd@iws.uni-stuttgart.de
§SINTEF ICT, PO Box 124, Blindern, N-0314 Oslo, Norway, knut-andreas.lie@sintef.no
be solved within a reasonable time-frame using high-performance computing, there are also many engineering workflows in which high-performance computing is infeasible or impossible. In this work, we will therefore discuss methods for accelerating the computation. To this end, we will focus on sequential simulation methods in which the flow and transport are solved in separate steps, and in which updating the flow pattern is the most computationally expensive part.

The classical approach to reducing the size of the discrete simulation model is to upscale the problem. That is, use a numerical method for calculating effective parameters and functions on a coarser scale that represent the local flow effect of the unresolved scale in an averaged sense. Standard upscaling methods solve representative fine-scale (flow or transport) problems to calculate averaged quantities (such as permeabilities, transmissibilities, etc.). There are both local and global upscaling methods. Local methods choose sub-domains of a size much smaller than the global scale (e.g., the size of one coarse grid block) and calculate effective parameters locally for each of these sub-domains. Global methods solve representative fine-scale problems on the global scale. Local methods can be further extended to (adaptive) local-global methods. In this case, a downscaling step is added to the method to approximate local fine-scale boundary conditions from the global coarse-scale solution. Excellent overviews of upscaling method can be found in, e.g., [1, 2].

A related, but more recent approach is to use so-called multiscale methods that try to incorporate fine-scale information into the set of coarse-scale equations in a way that is consistent with the local property of the differential operator [3, 4]. There are many different multiscale methods: Hughes et al. [5] introduced the Variational Multiscale method as a general framework for constructing hierarchical approximation spaces for Dirichlet-type problems. The Multiscale Finite-Element method [6] uses solutions to local fine-scale problems to incorporate fine-scale details into coarse-scale basis functions; a recent development in this direction is the Generalized Multiscale Finite-Element method [7]. The Mixed Multiscale Finite-Element method [8, 9] is based on the same principles but in addition allows for mass-conserving reconstruction of the fine-scale velocities. Related methods include the Numerical Subgrid Upscaling method [10] and the Multiscale Mortar method [11]. In the Multiscale Finite-Volume method [12], the underlying idea is to construct coarse-scale transmissibilities that account for fine-scale effects and lead to a multi-point approximation for the Finite-Volume discretization on the coarse scale. Recently, a more robust Two-Point Finite-Volume method was proposed [13]. In [14], an adaptive version of the Heterogeneous Multiscale method [15], which aims at capturing the macroscopic scale of the problem by estimating necessary information from a microscopic model, was applied to immiscible two-phase flows in porous media. Finally, Henning et al. [16] recently applied the Partition of Unity method in the multiscale context as a means for reliable numerical homogenization for elliptic equations with rough coefficients.

Reduced Basis (RB) methods represent an approach for model order reduction for parametrized partial differential equations, which has gained large popularity in the last decade. In numerous publications, these methods have proven to be versatile tools for reducing the computational effort for stationary problems [17, 18, 19, 20, 21] as well as for

The key idea of RB methods is to introduce two computational phases: an offline and an online phase. During the computationally demanding offline phase, a set of solutions to the parametrized equation at hand, so-called snapshots, are computed for different parameters using a Greedy-type algorithm [29]. From these snapshots, a reduced basis is computed and the offline-phase is concluded by projecting the high-dimensional discretization onto the reduced basis, giving rise to a reduced-dimensional equation.

During the online phase, a solution for a given parameter is computed by solving the reduced equation. Using the common assumption of parameter-separability or affine parameter dependence [20], a complete decoupling of the offline and online phases becomes straightforward. Hence, the computational effort during the online phase does not depend on the dimension of the underlying fine discretization and decisive speedups over the latter are possible at marginal loss of accuracy. If the assumption of parameter-separability cannot be fulfilled for the equation at hand, techniques like the Empirical Interpolation method [30, 31] can be used.

Herein, we will study the Localized Reduced Basis Multiscale method (LRBMS), which was originally introduced for general parametrized elliptic problems in [32, 33], and which is not limited to the current two-phase flow setting. The method allows parametrizations like boundary, initial or source data, but those will be neglected in this work in order to concentrate on the new parametrization introduced below. Application of the LRBMS method yields a potentially costly offline phase, in which localized reduced-dimensional bases for the pressure are computed. These bases then operate as low-dimensional surrogates of the high-dimensional Discontinuous Galerkin discretization during the actual two-phase flow simulation (here referred to as the online phase). The online phase is carried out using a sequential splitting scheme to decouple the high-dimensional saturation equation and reduced-dimensional pressure equation.

A key point in this contribution is the use of the so-called time-of-flight $\tau$ to parametrize our two-phase model. The time-of-flight is the time it takes for an inert particle to travel along a pathline from the closest point on the inflow boundary and to a given point inside the domain, or alternatively, the time it takes for a particle to travel from a given interior point and to the closest point on the outflow boundary. In the absence of gravity, the three-dimensional Eulerian equations that describe how fluid phases are transported, can be transformed into a family of one-dimensional equations in Lagrangian coordinates (i.e., along streamlines). Here, $\tau$ plays the role as the spatial coordinate for each one-dimensional transport equation. The time-of-flight therefore carries valuable information about the characteristics of the flow, and our idea is to use $\tau^s(x)$, defined for a given spatial point $x$ and a given saturation snapshot $s$, as a means to efficiently compute a limited number of mobility profiles $\Lambda^1, \ldots, \Lambda^M$, which are assumed to approximate the actual mobility $\lambda(s)$ for a given saturation $s$ via $\lambda(s) \approx \sum \mu_i \Lambda^i$ for a given parameter $\mu \in \mathbb{R}^M$. The coupling between the saturation and pressure equation via the mobility is now replaced by a coupling via the parameter $\mu$ and we apply model

\text{3}
order reduction techniques for parametrized problems to the pressure equation, now being parametrized by $\mu$.

The approximation quality of the reduced system is typically controlled using \textit{a-posteriori} error estimators. In a recent work, Schindler et al. \cite{34} introduced a localized \textit{a-posteriori} error estimate perfectly fit for the localized method introduced in Section 4. Another suitable estimate was recently introduced in \cite{35}.

The rest of the paper is structured as follows: After introducing the details of the two-phase flow setting in the next section, we will describe its high-dimensional discretization in Section 3. Section 4 is dedicated to the derivation of the Localized Reduced Basis Multiscale method and Section 5 gives details on the coupling between the reduced pressure equation and the high-dimensional saturation equation. Finally, in Section 6, we demonstrate the applicability of the method in numerical experiments.

2 Model Problem

In this section, we introduce the mathematical model of two-phase flow that will be used throughout the rest of this work. Our model problem is the flow of two phases in a spatio-temporal domain $\Omega \times [0, T] \subset \mathbb{R}^d \times \mathbb{R}^+$ where $d \in \{2, 3\}$ denotes the space dimension. We use a global pressure, total velocity formulation for two incompressible, immiscible fluids that includes gravity but no capillary effects. This yields the equation for the unknown pressure $p$

$$-\nabla \cdot \left( \lambda(s) \kappa \nabla p - \kappa \left[ \lambda_w(s) \varrho_w + \lambda_o(s) \varrho_o \right] G \right) = q_1 \quad \text{in } \Omega \times [0, T], \quad (1)$$

where $\varrho_w$ and $\varrho_o$ denote the densities for the wetting and non-wetting phases, respectively, and $G$ is the gravitational force vector $G = (0, 0, -g)^T$, with $g$ being the gravitational acceleration. Furthermore, $\kappa$ denotes the total permeability and the functions $\lambda_w, \lambda_o, \lambda : [0, 1] \rightarrow \mathbb{R}^+$ denote the wetting, non-wetting and total mobility, given by

$$\lambda_w(s) = \frac{k_{rw}(s)}{\eta_w}, \quad \lambda_o(s) = \frac{k_{ro}(s)}{\eta_o}, \quad \lambda(s) = \lambda_w(s) + \lambda_o(s), \quad (2)$$

where $\eta_w$ and $\eta_o$ are the viscosities of the wetting and non-wetting phase, respectively and the relative permeabilities $k_{rw}$ and $k_{ro}$ are given functions of the saturation $s$ of the wetting phase.

Using Darcy’s law, the total velocity $u$ is given by

$$u = -\lambda(s) \kappa \nabla p + \kappa \left[ \lambda_w(s) \varrho_w + \lambda_o(s) \varrho_o \right] G \quad \text{in } \Omega \times [0, T], \quad (3)$$

and enters the transport equation for the saturation $s$,

$$\phi \partial_t s + \nabla \cdot \left( f_w(s) \left[ u + \kappa \lambda_o(s) (\varrho_w - \varrho_o) G \right] \right) = q_2 \quad \text{in } \Omega \times [0, T]. \quad (4)$$

Here, $\phi$ denotes the porosity of the porous medium and the fractional flow of water $f_w$ is given by

$$f_w(s) = \frac{\lambda_w(s)}{\lambda_w(s) + \lambda_o(s)}. \quad (5)$$
The above three equations for the pressure (1), velocity (3) and saturation (4) are equipped with the boundary conditions

\begin{align*}
    s &= s_D \quad \text{in } \partial \Omega_{s,d} \times [0,T], \\
    -\lambda(s)\kappa \nabla p \cdot n &= v_N \quad \text{in } \partial \Omega_{p,n} \times [0,T], \\
    p &= p_D \quad \text{in } \partial \Omega_{p,d} \times [0,T],
\end{align*}

on the inflow boundary for the saturation \( \partial \Omega_{s,d} \) and on the Dirichlet and Neumann boundaries for the pressure \( \partial \Omega_{p,d}, \partial \Omega_{p,n} \). Further, we impose initial conditions

\begin{align*}
    s(\cdot,0) &= s_0 \quad \text{in } \Omega.
\end{align*}

### 3 High-Dimensional Discretization

This section is dedicated to the exposition of the so-called high-dimensional discretization of problem (1)–(6). The term high-dimensional is used here to indicate the antonym of the reduced (low-dimensional) discretization, cf. Section 1.

Different schemes are frequently used to discretize Equations (1)-(6). A tendency towards finite-volume-type schemes can be observed in the literature because they are motivated by local conservation properties. As it will prove advantageous in the analysis of our reduced method, we will use a Discontinuous Galerkin (DG) discretization with arbitrary local polynomial degree. In particular, we have found the Symmetric Weighted Interior-Penalty (SWIP) DG method [36], including the total velocity reconstruction presented therein, to be very robust and it is therefore our method of choice in the following.

#### 3.1 Discretization

As a first step towards a discrete version of Equations (1)–(6), we introduce an admissible tessellation \( T_h \) of the computational domain \( \Omega \). We assume \( T_h \) to be a set of non-overlapping elements \( e \) with width \( h_e = \text{diam}(e) \) and define the grid size \( h = \max_{e \in T_h} h_e \). The intersections of co-dimension one of an element \( e \) are denoted by \( f \) and their width by \( h_f = \text{diam}(f) \). With each intersection \( f \) we associate a unique normal vector \( n_f \), for which the subscript \( f \) will be dropped whenever no ambiguity arises. We let \( \Gamma_{p,d} \) and \( \Gamma_{p,n} \) denote all boundary intersections of \( T_h \) where Dirichlet and Neumann conditions for the pressure shall be implied, respectively, and \( \Gamma_{s,d} \) denote all boundary intersections on which we impose a fixed saturation. Finally, by \( \Gamma_i \) we denote the inner intersections and by \( \Gamma_a = \Gamma_{p,n} \cup \Gamma_{p,d} \cup \Gamma_i \) the set of all intersections.

Additionally we introduce a temporal discretization \( t^0, \ldots, t^{NT} \in [0,T] \) and space-time-discrete approximations \( s_h^0 \approx s(\cdot, t^0), p_h^0 \approx p(\cdot, t^0) \) of the saturation and the pressure, both stemming from the space of piecewise polynomials of maximum local order \( k \)

\begin{equation}
    V_h = \left\{ v \in L^2(\Omega) \left| v_e \in \mathbb{P}^k(e) \quad \forall e \in T_h \right. \right\}.
\end{equation}
Functions in $V_h$ are two-valued on intersections $f = e_1 \cap e_2$ (with $n_f$ pointing from $e_1$ to $e_2$). We define jump $\left[ \cdot \right]_f$ and (weighted) mean $\left\{ \cdot \right\}_f$ values for $w \in V_h$, i.e.,

$$\left[ w \right]_f = w|_{e_1} - w|_{e_2}, \quad \left\{ w \right\}_f = \tau_{e_1,f} w|_{e_1} + \tau_{e_2,f} w|_{e_2}, \quad \tau_{e_\ell,f} = \frac{a_{\ell,f}}{a_{1,f} + a_{2,f}}; \quad a_{\ell,f} = ||(\kappa)|_{e_\ell}||_{L^{\infty}(f)}, \ \ell = 1, 2.$$  

For the sake of readability, the subscript $f$ in $\left[ \cdot \right]_f$ and $\left\{ \cdot \right\}_f$ will be dropped in the following whenever no ambiguity arises.

### 3.2 Pressure Equation

Following the ideas in [36], we introduce discrete formulations of Equations (1)–(6). For a given general mobility function $\gamma : \Omega \to \mathbb{R}^+$ the bilinear form $b_h(\cdot, \cdot ; \gamma) : V_h \times V_h \to \mathbb{R}$ is defined as

$$b_h(v, w; \gamma) = \sum_{e \in T_h} \int_e \gamma \kappa \nabla v \nabla w \, dx + \sum_{f \in \Gamma_i \cup \Gamma_{p,d}} \frac{\sigma_f}{h_f} \int_f \left[ v \right] \left[ w \right] \, dS \quad \text{\quad (9)}$$

$$+ \sum_{f \in \Gamma_i \cup \Gamma_p} \int_f \left( \{ \gamma \kappa \nabla v \cdot n_f \} \left[ w \right] + \{ \gamma \kappa \nabla w \cdot n_f \} \left[ v \right] \right) \, dS.$$  

Here, the penalty parameter $\sigma_f$ is given as

$$\sigma_f = c_f \frac{2a_{1,f} a_{2,f}}{a_{1,f} + a_{2,f}}, \quad \text{\quad (10)}$$

where $c_f > 0$ is a constant that has to be chosen larger than a minimal threshold depending on the regularity of the mesh.

**Remark 1.** Note that in [36], $a_{\ell,f} = ||(\kappa)|_{e_\ell}||_{L^{\infty}(f)}$ was used. As the penalty parameter $\sigma_f$ would not allow an affine decomposition, we refrain from using the mobility in the penalties and instead replace the original constant $c_f$ by $c_f/\max\{\eta_w, \eta_o\}$. Usage of the original weighted averages and penalties would be possible using the Empirical Interpolation technique [30, 31].

The right-hand side for the discrete formulation of the pressure equation is for $s \in V_h, \ \gamma_o, \gamma_w : \Omega \to \mathbb{R}^+$, $\gamma = \gamma_o + \gamma_w$ given via the linear form $l_h(\cdot; \gamma_o, \gamma_w) : V_h \to \mathbb{R},$

$$l_h(w; \gamma_o, \gamma_w) = \sum_{e \in T_h} \int_e q_1 w + (\gamma_o \varrho_o + \gamma_w \varrho_w) \kappa \cdot \nabla w \, dx$$

$$- \sum_{f \in \Gamma_i \cup \Gamma_{p,d}} \int_f \left( (\gamma_o \varrho_o + \gamma_w \varrho_w) \kappa \cdot n_f \right) \left[ w \right] \, dS \quad \text{\quad (11)}$$

$$+ \sum_{f \in \Gamma_p} \int_f \left( \frac{\sigma_f}{h_f} w - \gamma \kappa \nabla w \cdot n_f \right) p_D \, dS + \sum_{f \in \Gamma_{p,n}} \int_f v_N w \, dS.$$  

At this point we are ready to introduce the so-called high-dimensional discretization of (1).
Definition 1. For given functions $\gamma_w, \gamma_o : \Omega \to \mathbb{R}^+$ and $\gamma = \gamma_w + \gamma_o$, we call $p_h \in \mathcal{V}_h$ the high-dimensional pressure solution if it satisfies
\[
b_h(p_h, w; \gamma) = l_h(w; \gamma_o, \gamma_w) \quad \forall w \in \mathcal{V}_h.
\]

3.3 Total Velocity Reconstruction

From a given pressure $p$, which may be either $p_h$ or a reduced pressure later on, we compute the total velocity $u_h$ using the conservative reconstruction from [36], which yields continuity of the normal components. For $k = 1$, i.e., piecewise linear pressure, the velocity is computed in the lowest order Raviart-Thomas space
\[
RT = \left\{ u \in H(\text{div}) \left| u|_T \in [P_0(T)]^d + xP_0(T) \forall T \in \mathcal{T}_h \right. \right\},
\]
where $H(\text{div}) = \{ v \in L^2(\Omega) \left| \nabla \cdot v \in L^2(\Omega) \right. \}$ and $x = (x_1, \ldots, x_d)^T$. Additionally, we introduce
\[
[v]_f^* = \begin{cases} [v]_f, & f \in \Gamma_i, \\ v|_\Omega - v_D, & f \in \Gamma_{v,d}, \\ 0, & f \in \Gamma_{p,n}, \end{cases}
\]
for $v \in \mathcal{V}_h$ with Dirichlet boundary data $v|_{\Gamma_{v,d}} = v_D$.

Definition 2. For a given pressure $p \in \mathcal{V}_h$ and saturation $s \in \mathcal{V}_h$, let the total velocity $u_h \in RT$ be the unique solution to
\[
\int_f (u_h \cdot n) dS = \int_f -n \cdot \{ \lambda(s)\kappa \nabla p - \kappa [\lambda_w(s)\rho_w + \lambda_o(s)\rho_o] G \} + \frac{\sigma_f}{h_f} [p]^*_f dS \quad \forall f \in \Gamma_a.
\]

3.4 Saturation Equation

The saturation equation discretized by a DG scheme in space using an upwind flux in the space of piecewise polynomial functions $\mathcal{V}_h$, and for the temporal discretization, we use an explicit Euler scheme.

Definition 3. For a given saturation $s^n_h \in \mathcal{V}_h$ and velocity $u \in [\mathcal{V}_h]^d$, the saturation $s^{n+1}_h \in \mathcal{V}_h$ is given as the solution to
\[
\sum_{e \in \mathcal{T}_h} \int_e \frac{\phi}{\Delta t} s^{n+1}_h v_h d x = \sum_{e \in \mathcal{T}_h} \int_e \frac{\phi}{\Delta t} s^n_h v_h d x + \sum_{e \in \mathcal{T}_h} q_2 v_h d x \\
+ \sum_{e \in \mathcal{T}_h} \int_e [u + \lambda_o(s^n_h)(\rho_w - \rho_o)\kappa \cdot G] f_w(s^n_h) \cdot \nabla v_h d x \\
- \sum_{f \in \Gamma_a} \int_f n_f \cdot \{ u + \lambda_o(s^n_h)(\rho_w - \rho_o)\kappa \cdot G \} f_w(\chi(s^n_h)) [v_h] f dS \\
- \sum_{f \in \Gamma_i \cup \Gamma_{s,d}} \frac{\sigma_f}{h_f} \int_f [s^n_h]^*_f [v_h] f dS,
\]
for all \( v_h \in \mathcal{V}_h \). Here the upwind function \( \chi \) is given as

\[
\chi(s^n_h) = \begin{cases} 
  s^n_{h}^\uparrow, & f \in \Gamma_i, \\
  s_D, & f \in \Gamma_{s,d},
\end{cases}
\]

(15)

where \( s^n_{h}^\uparrow \) denotes the upwind value of \( s^n_h \). If \( f \in \Gamma_i \), there exist \( e_1, e_2 \in \mathcal{T}_h \) such that \( f = \partial e_1 \cap \partial e_2 \). Then \( s^n_{h}^\uparrow = s^n_h|_{e_1} \) if \( u_h \cdot n_f \geq 0 \) and \( s^n_{h}^\uparrow = s^n_h|_{e_2} \) otherwise. Remember that we assume \( n_f \) to point from \( e_1 \) to \( e_2 \). The penalty \( \sigma_f \) in Equation (14) is the same as for the pressure equation, see (10).

### 3.5 Slope Limiter

Discontinuous Galerkin methods require some kind of stabilization to avoid over- and undershoots if they are to be used in a two-phase flow context; see [37, 38], for example. Both artificial diffusion and slope limiters have been used as a means to this end. We will employ a slope limiter introduced by Dedner et al. [38] that is applicable to different element types and in two and three spatial dimensions. We will shortly outline the most important features of the mass-conservative limiter that we use in our experiments for the case of a piecewise linear saturation, where we use the “DG scheme” (in contrast to “DG + R”, see [38, Section 6.1]). For more details on the method and a discussion of the general case of arbitrary polynomial order on \( s^n_h \) we refer to [38].

The first step is to compute a so-called shock-detector \( \mathcal{S} : \mathcal{T}_h \to \mathbb{R}^+ \), which for a given saturation \( s_h \in \mathcal{V}_h \) and velocity \( u_h \in [\mathcal{V}_h]^d \) reads:

\[
\mathcal{S}(e) = \sum_{\mathcal{I}_e^\uparrow} \left( 0.08 \cdot d \cdot \sqrt{h_e|e|} \right)^{-1} \int_f \| s_h \| \, dS,
\]

(16)

where \( \mathcal{I}_e^\uparrow = \{ f \in \Gamma_u \mid f \subset \partial e, u_h \cdot n_f < 0 \} \) denotes the upstream interfaces of \( e \). Using \( \mathcal{S} \) makes it possible to apply the limiter (i.e., reduce gradients of the saturation) only on cells in which (strong) discontinuities are present and therefore unwanted numerical oscillations may occur. In other regions, the saturation is left unlimited. Cells that will be flagged in the above sense are then all cells \( e \in \mathcal{T}_h \) with \( \mathcal{S}(e) > 1 \) or \( s_h(x) \notin [0,1] \) for some \( x \in e \). Let \( e^* \in \mathcal{T}_h \) be such a cell and \( \{ e_i \mid i = 1, \ldots, N_n \} \) its direct neighbors. We then compute

\[
\begin{align*}
g_i &= \nabla s_h \cdot (b_{e_i} - b_{e^*}) \quad \text{and} \quad d_i = \frac{1}{|e_i|} \int_{e_i} s_h \, dx - \frac{1}{|e^*|} \int_{e^*} s_h \, dx, \\

\text{where} \ b_e \text{ denotes the barycenter of the cell} \ e. \ \text{From} \ g_i \ \text{and} \ d_i \ \text{we compute the gradient scales} \ m_i \ \text{by}
\end{align*}
\]

\[
m_i = \begin{cases} 
  0, & \text{if} \ g_i d_i < 0, |g_i| > 10^{-8} \ \text{and} \ |d_i| > 10^{-8}, \\
  d_i/g_i, & \text{if} \ g_i d_i > 0, |g_i| > |d_i|, |g_i| > 10^{-8} \ \text{and} \ |d_i| > 10^{-8}, \\
  1, & \text{otherwise.}
\end{cases}
\]
Finally, the stabilized saturation $\tilde{s}_h$ is computed as
\[
\int_e \tilde{s}_h \varphi \, dx = \int_e s_h \varphi \, dx + \int_e \min(m_i) \nabla s_h \cdot (x - b_e) \varphi \, dx \quad \forall \varphi \in \mathcal{P}_1(e) \tag{17}
\]
in all flagged cells $e \in \mathcal{T}_h$. In all other cells, we set $\tilde{s}_h = s_h$.

**Algorithm 1** (High-Dimensional Two-Phase Flow Scheme). The high-dimensional simulation scheme for the two-phase flow is as follows:

1. Project the initial data $s_0$ onto the high-dimensional discrete space: Let $s_0^h \in V_h$ be given by
\[
\int_\Omega s_0 \psi \, dx = \int_\Omega s_0^h \psi \, dx \quad \forall \psi \in V_h.
\]

2. For all $n \in \{0, \ldots, N_T - 1\}$,
   a) use the pressure scheme (Definition 1) for $\gamma_o = \lambda_o(s^n_h)$, $\gamma_w = \lambda_w(s^n_h)$ to compute $p_{h}^{n+1}$;
   b) use the velocity scheme (Definition 2) for $p = p_{h}^{n+1}$ to compute the total velocity $u_{h}^{n+1}$;
   c) use the saturation scheme (Definition 3) for $u = u_{h}^{n+1}$ to compute the saturation $s_{h}^{n+1}$;
   d) use the limiter (Equation (17)) for $s_h, u_h = u_{h}^{n+1}$ to compute a stabilized version $\tilde{s}_{h}^{n+1}$ of the saturation and set $s_{h}^{n+1} = \tilde{s}_{h}^{n+1}$.

### 3.6 Time-of-Flight Equation

The so-called time-of-flight $\tau(x)$ is defined as the time it takes for a passive particle to reach a given point $x \in \Omega$, starting from the closest point on the inflow boundary. Here, $\tau(x)$ can be defined by integrating $\int \phi/|u| \, dx$ along streamlines, or by solving $u \cdot \tau = \phi$.

The time-of-flight inherits important information about the flow pattern and will be used here to approximate the spatio-temporal behavior of the saturation without computing a whole temporal evolution the transport equation.

Consistent with our discretization of the saturation and pressure equation we use the DG-discretization introduced in [39]: Find $\tau_{h}^s \in V_h$ such that
\[
-\int_e (\tau_{h}^s u_h) \cdot \nabla \psi_h \, dx + \int_{\partial e} (\tau_{h}^s u_h \cdot n) \psi_h \, dS = \int_e \phi \psi_h \, dx, \quad \forall \psi_h \in V_h \tag{18}
\]
for all $e \in V_h$. Here, $\tau_{h}^s \uparrow$ again denotes the upwind value of the two-valued function $\tau_{h}^s$, see (15), and $\phi$ denotes the porosity of the porous medium.

In the absence of gravitational forces, the discretized time-of-flight equation can be permuted to a lower block-triangular form—if the computational mesh is reordered according to the direction of the flow—and hence solved very efficiently in a per-element fashion by a simple backsubstitution method; see [39] for details.
4 The Localized Reduced Basis Multiscale Method

Reduced Basis (RB) methods as introduced in Section 1 offer good complexity reduction for a lot of different kinds of equations. Nevertheless, for real-world multiscale applications, the number of unknowns in the high-dimensional discretization can be prohibitively large such that the Greedy-type algorithm [29] becomes unfeasible because the number of snapshots computed during the basis generation cannot be controlled.

A range of different extensions to the original RB method that can be seen as a remedy for this problem were developed in recent years. The Reduced Basis Element method [40, 41, 42] introduces a number of representative domains and computes reduced bases for each one of them. The domain under consideration is then built up from the representative domains (possibly using deformation mappings) and computations are carried out in a space built up from the precomputed spaces. The Reduced Basis Hybrid method [43] and the Reduced Basis Domain Decomposition Finite Element method [44] make use of this concept. In [45], Maier and co-workers introduce a RB approach for heterogeneous domain decomposition that could be used in the current setting to reduce the number of global snapshots, too. Another approach for model order reduction for our application is presented in [46], in which Discrete Empirical Interpolation and Proper Orthogonal Decomposition are used in conjunction to build reduced-order models for simulation of viscous fingering in porous media.

In our recent articles [33, 32], we introduced the so-called Localized Reduced Basis Multiscale method (LRBMS). It connects ideas from numerical multiscale methods with the RB approach and aims at reducing the offline time of the RB method while ensuring at least identical approximation quality during the online time, possibly at a slightly increased cost in terms of runtime. (Similarly, in [47], the proper orthogonal decomposition method is combined with a multiscale mixed finite element framework.) The main idea of this approach is to introduce two grids: A fine triangulation for the high-dimensional discretization and a coarse one that is used to build up localized RB spaces. The rest of this section is dedicated to the exposition of the details of this method.

4.1 Parameterization

At this point we provide details on the parametrization indicated in Section 1. As stated earlier, the coupling between the saturation and pressure equation will be realized via a parametrization of the pressure equation: We introduce wetting and non-wetting phase mobilities $\Lambda_w, \Lambda_o : \Omega \to \mathbb{R}^+$ that depend on the saturation $s$ via

$$
\Lambda_w(s)(x) = \sum_{q=1}^{M} \theta_q(s) \Lambda^q_w(x), \quad \Lambda_o(s)(x) = \sum_{q=1}^{M} \theta_q(s) \Lambda^q_o(x), \quad x \in \Omega,
$$

(19)

where $\Lambda^q_w, \Lambda^q_o \in \mathcal{V}_h$, $q = 1, \ldots, M$ denote saturation-independent mobility profiles that were precomputed during the offline phase and the coefficients $\theta_q$ are computed during
the online phase via a least-squares fitting: Let $\Theta(s) \in \mathbb{R}^M$ be given by

$$
\Theta(s) = \arg \min_{\vartheta \in \mathbb{R}^M} \|\lambda(s) - \sum_{q=1}^M \vartheta_q \Lambda^q\|_{L^2},
$$

then $\theta_q(s) = (\Theta(s))_q$. Here we used $\Lambda^q = \Lambda^q_o + \Lambda^q_w$. Details on the computation of the mobility profiles $\Lambda^o, \Lambda^w$ are given below.

### 4.2 Discretization

As mentioned before, we introduce a second, coarser triangulation of the domain $\Omega$: Let $\mathcal{T}_H$ be an admissible triangulation with grid size $H = \max_{E \in \mathcal{T}_H} H_E = \max_{E \in \mathcal{T}_H} \text{diam}(E)$, for $H \gg h$. Cells in $\mathcal{T}_H$ will be denoted by $E$, intersections of two elements by $F$, and the set of all coarse intersections by $\Gamma_A$. Furthermore, we assume the two triangulations to be matching in the sense that for each $F \in \Gamma_A$ there exist $m \in \mathbb{N}$ and $f_1, \ldots, f_m \in \Gamma_a$ such that $F = \bigcup_{i=1}^m f_i$.

Based on the coarse triangulation $\mathcal{T}_H$, we introduce a localized reduced-dimensional function space $\mathcal{W}_N$. For every coarse grid cell $E \in \mathcal{T}_H$ we assume a set of linearly independent functions $\varphi^E_1, \ldots, \varphi^E_N \in \mathcal{V}_h$ with $\text{supp}(\varphi^E_i) \subset E$ to be given so that they form the local reduced basis: $\Phi^E = \{\varphi^E_1, \ldots, \varphi^E_N\}$. From those local bases we build a global reduced basis $\Phi_N$ of size $N = \sum_{E \in \mathcal{T}_H} N^E$ on the whole domain $\Omega$ by setting

$$
\Phi_N = \bigcup_{E \in \mathcal{T}_H} \Phi^E.
$$

We call the space $\mathcal{W}_N = \text{span}(\Phi_N)$, spanned by the global reduced basis, reduced broken space and point out that $\mathcal{W}_N \subset \mathcal{V}_h$ because of our DG ansatz.

Using the reduced basis $\Phi_N$, we compute the coarse-scale pressure approximation $p_H \in \mathcal{W}_N$.

**Definition 4.** For a given reduced broken space $\mathcal{W}_N$ and mobilities $\gamma_o, \gamma_w : \Omega \to \mathbb{R}^+$ where $\gamma = \gamma_o + \gamma_w$, the reduced solution $p_H \in \mathcal{W}_N$ to Equation (1) is given by

$$
b_h(p_H, w; \gamma) = l_h(w; \gamma_o, \gamma_w), \quad \forall w \in \mathcal{W}_N.
$$

### 4.3 Basis Construction

One of the key ingredients of the method proposed in this paper is a basis construction algorithm that is a modification of the Greedy-type algorithm used in conventional RB methods as mentioned in Section 1.

**Algorithm 2 (LRBMS Basis Construction).** Based on the coarse triangulation $\mathcal{T}_H$ and an error measure $\Delta_N$, we compute the global reduced basis $\Phi_N$ using the following algorithm:

1. Initialization
a) Given the initial saturation \( s_0 \), compute the initial pressure \( p_0^h \in V_h \) and velocity \( u_0^h \in \mathbb{R} \).

b) From \( u_0^h \) compute the time-of-flight \( \tau_{h}^{s_0} \).

c) From the time-of-flight and \( M \in \mathbb{N}, M \geq 1 \), compute mobility profiles \( \Lambda_1^\alpha, \ldots, \Lambda_M^\alpha \in V_h \) and \( \Lambda_1^w, \ldots, \Lambda_M^w \in V_h \) via

\[
\Lambda_\alpha^1(x) = \lambda_\alpha(0), \quad \Lambda_\alpha^M(x) = \lambda_\alpha(1)
\]

\[
\Lambda_q^\alpha(x) = \begin{cases} 
\lambda_\alpha(0), & \text{if } \tau_{h}^{s_0}(x) > (q-1) \cdot \frac{T}{M-2}, \\
\lambda_\alpha(1), & \text{otherwise}, 
\end{cases} \quad \forall q \in 2, \ldots, M-1
\]

where \( \lambda_\alpha \) denotes the linear mobility (2) of phase \( \alpha \). Set \( \Lambda_q^\alpha = \Lambda_q^w + \Lambda_q^o \) for \( q = 1, \ldots, M \).

d) Choose a desired maximum basis size \( N_{\text{max}} \in \mathbb{N} \), an approximation tolerance \( \epsilon_{\text{tol}} \), a parameter set \( P = \{ \mu \in [\mathbb{R}^+]^M | \sum_\mu \mu_i \leq 1 \} \) and a discrete subset \( P_{\text{tr}} \subset P \), the so-called training set. Furthermore let \( \Phi_E = \emptyset \) for all \( E \in \mathcal{T}_H \), \( \Xi = \bigcup_{E \in \mathcal{T}_H} \Phi_E \), \( W = \text{span}(\Xi) \) and \( \Psi = \emptyset \).

2. Basis Extension

a) For each \( \mu \in P_{\text{tr}} \) compute \( p_h(\gamma) \in W \) as the solution to the coarse-scale pressure equation (22), that is

\[
b_h(p_h(\gamma), w; \gamma) = l_h(w; \gamma_o, \gamma_w) \quad \forall w \in W
\]

with \( \gamma_\alpha = \sum_{q=1}^M \mu_q \Lambda_q^\alpha \) and \( \gamma = \gamma_w + \gamma_o \).

b) Evaluate the error estimator for each parameter in the training set: \( \epsilon_\mu := \Delta_N(\mu, W) \) and find the parameter worst approximated in the current basis: \( \mu_{\text{max}} = \arg \max_{\mu \in P_{\text{tr}}} \epsilon_\mu \).

c) If \( \epsilon_{\mu_{\text{max}}} > \epsilon_{\text{tol}} \) compute \( p_h(\gamma_{\text{max}}) \) as solution to the high-dimensional equation (see Definition 1) for the mobilities \( \gamma_{w_{\text{max}}} = \sum_{\mu_{\text{max}}} \mu_{\text{max}}^q \Lambda_q^w \) and \( \gamma_{o_{\text{max}}} = \sum_{\mu_{\text{max}}} \mu_{\text{max}}^q \Lambda_q^o \). Otherwise go to step 3a.

d) Extend the set of snapshots: \( \Psi = \Psi \cup \{ p_h(\gamma_{\text{max}}) \} \).

e) For each \( E \in \mathcal{T}_H \):

i. Use the Gram-Schmidt algorithm to orthonormalize the restriction of the pressure snapshot \( p_h(\gamma_{\text{max}}) \mid_E \) with respect to the current local basis \( \Phi^E \):

\[
\tilde{p} = \text{ortho}(p_h(\gamma_{\text{max}}) \mid_E ; \Phi^E).
\]

ii. Extend the local reduced basis: \( \Phi^E = \Phi^E \cup \{ \tilde{p} \} \).

f) Set \( \Xi = \bigcup_{E \in \mathcal{T}_H} \Phi^E \) and \( W = \text{span}(\Xi) \).
3. Data Compression (optional)
   a) Set $\tilde{\Phi}^E = \Psi|_E$, that is: Define local bases $\tilde{\Phi}^E$ per coarse element $E \in T_H$ as restrictions of the snapshots $\Psi$ to the coarse elements.
   b) Apply the principal component analysis with tolerance $\epsilon_{PCA}$ on each element:

   $$\Phi^E = \text{PCA}(\tilde{\Phi}^E, \epsilon_{PCA}).$$

   The details of this step can be found in the literature, see [48], for example.

4. Finalization
   a) Define the global reduced basis $\Phi_N$ for $N = \sum_{E \in T_H} |\Phi|^E$ as

   $$\Phi_N = \bigcup_{E \in T_H} \Phi^E.$$

   b) Define the global reduced basis space $W_N$ as

   $$W_N = \text{span}(\Phi_N).$$

In summary, the basis construction performs the following: In an “initialization” step, we compute pressure and velocity from the initial saturation data. From the velocity, we compute approximate mobility profiles using the time-of-flight. This will incorporate important features of the problem, like low-permeability-lenses, for example, into the mobility profiles and therefore also into the reduced basis for the pressure. After fixing some input data like the desired basis size, we proceed to the basis extension step.

In the basis extension step, we add localized orthonormalizations of high-dimensional snapshots to an initially empty local bases until either the maximum total basis size is reached or a prescribed error tolerance, measured by an \textit{a-posteriori} error estimator, is fulfilled in the current overall basis, which is the joint of all local bases. Additionally, we save the original snapshots.

In an optional “compression” step we define local per-coarse-element bases by restricting the global untouched snapshots from the last step to each coarse element. Next, we apply a data compression algorithm, the principal component analysis, to each local basis to reduce the basis size in regions where redundant information may be present, like in regions far from sinks and sources, for example. At this point, it is important that we did not use the orthonormalized local bases from the extension step for the local compressions. Redundancies were canceled out in those bases and therefore the data compression would not give meaningful results.

We conclude the algorithm with the “finalization” step by defining one global reduced basis as the joint of all local bases and the reduced broken space as its linear span.

The main ideas of this novel method are the restriction of the basis to elements of a coarser grid and subsequent per-element data compression. By the restriction to a
coarse grid we reduce the number of snapshots needed to fulfill a desired error tolerance on a prescribed training set of parameters during the offline phase. This is easy to see for the limit case in which \( \mathcal{T}_H = \mathcal{T}_h \), as in this case the reduced space coincides with the high-dimensional discrete function space after a finite number of basis extensions. For \(|\mathcal{T}_H| \leq |\mathcal{T}_h|\), this effect was demonstrated in [33] and can be seen again in the numerical experiments in Section 6.

The downside of this approach is that the size of the global reduced basis increases with the number of coarse elements. This is where the idea of per-element data compression comes into play and allows us to keep the total basis size \( N \), which is the main factor in online computation complexity, in an agreeable range by reducing the local basis size in regions where little or no variation is inherent in the local bases, as may be the case in the absence of sinks or sources in the neighboring elements. The effectiveness of the PCA-step will be backed up by the experiments in Section 6.

### 4.4 Offline-Online Decomposition

In this section, we will demonstrate how the computations for the LRBMS method can be split into computationally demanding parts that will be executed during the offline phase and computationally inexpensive parts to be performed during the online phase. This is possible because of the affine splitting (19).

Given a reduced basis \( \Phi = \{\varphi_1, \ldots, \varphi_N\} \) of size \( N \in \mathbb{N} \), we compute all parameter-independent parts of Equation (22), that is: all parts that depend on the location in space but not on \( \Theta(s) \). We can identify the terms \( b_q \in \mathbb{R}^{N \times N} \), \( q = 1, \ldots, M \),

\[
(b_q)_{ij} = \sum_{e \in \mathcal{T}_h} \int_e \Lambda^q \kappa \nabla \varphi_i \nabla \varphi_j \, dx + \sum_{f \in \Gamma \cup \Gamma_{p,d}} \int_f \left( \{ \Lambda^q \kappa \nabla \varphi_i \cdot n_f \} \nabla \varphi_j \right) \, dS,
\]

and the terms \( d_q, \xi \in \mathbb{R}^N \) for the right hand side:

\[
(d_q)_i = \sum_{e \in \mathcal{T}_h} \int_e \left( \Lambda^q \kappa \varphi_o + \Lambda^q \kappa \varphi_w \right) G \cdot \nabla \varphi_i \, dx - \sum_{f \in \Gamma \cup \Gamma_{p,d}} \int_f \left( \{ \Lambda^q \kappa \varphi_i \cdot n_f \} \right) \, dS - \sum_{f \in \Gamma_{p,n}} \int_f \left( \Lambda^q \kappa \nabla \varphi_i \cdot n_f \right) p_D \, dS,
\]

\[
(\xi)_i = \sum_{e \in \mathcal{T}_h} \int_e q_i \varphi_i \, dx + \sum_{f \in \Gamma_{p,d}} \int_f \frac{\sigma_f}{h_f} \varphi_o p_D \, dS + \sum_{f \in \Gamma_{p,n}} \int_f v_N \varphi_i \, dS.
\]

During the online phase, for a given saturation \( s \in \mathcal{V}_h \), the solution \( p_H \in \mathbb{R}^N \) of the
discrete equivalent of the reduced equation (22) is then given by

\[
\left( c + \sum_{q=1}^{M} \theta_q(s)b_q \right) p_H = \varepsilon + \sum_{q=1}^{M} \theta_q(s)d_q.
\]  

(23)

While the computation of the quantities \(b_q, c, d_q\) and \(\varepsilon\) has a complexity polynomial in \(1/h\), computing the sums in the reduced system (23) has a complexity polynomial in \(N\). The only critical part in (23) is the computation of the coefficients \(\theta_i(s), i = 1, \ldots, M\) that depends on the grid size \(h\) because of the least-squares approximation (20). Nevertheless, as the complexity of the least-squares fit is only \(O(1/h)\), we still expect largely accelerated computations compared to a high-dimensional pressure solve, which has a complexity of \(O(1/h^2)\) or even \(O(1/h^3)\).

Remark 2. As mentioned before, different parametrizations like boundary value parametrizations would be possible for Problem (1-6). As long as these parametrizations are affine-linear, the offline-online decomposition works analogously. For other kinds of parametrizations, see e.g., [20, 32, 25].

5 LRBMS Two-Phase Flow Scheme

We now have all parts together to introduce our overall reduced approximation scheme for two-phase flow in porous media using an IMPES-type coupling.

Algorithm 3 (LRBMS Two-Phase Flow Scheme). Compute the LRBMS pressures \(p_{H1}^1, \ldots, p_{HT}^N\) and saturations \(s_{H1}^1, \ldots, s_{HT}^N\) as follows.

1. Choose a maximum basis size \(N_{\text{max}} \in \mathbb{N}\), the number of mobility profiles \(M \in \mathbb{N}\) and the number of time steps \(N_T \in \mathbb{N}\).

2. Use Algorithm 2 to compute a reduced basis \(\Phi = \{\varphi_1, \ldots, \varphi_N\} \subset \mathcal{V}_h\) of size \(N \in \mathbb{N}\) for mobility profiles \(\Lambda^1_\alpha, \ldots, \Lambda^M_\alpha \in \mathcal{V}_h\).

3. Compute the quantities \(b_q, c, d_q\) and \(\varepsilon\) from Section 4.4.

4. Project the initial data \(s_0\) to the fine grid:

\[
\int_{\Omega} s_0^H v \, dx = \int_{\Omega} s_0 v \, dx, \quad \forall v \in \mathcal{V}_h.
\]

5. For \(n = 0, \ldots, N_T - 1\),

a) compute the reduced-dimensional pressure solution \(p_{H1}^{n+1} \in \mathbb{R}^N\) as solution to Equation (23) for \(s = s_H^n\). This means that we need to compute the least-squares fit (20) to the given saturation \(s_H^n \in \mathcal{V}_h\) and then solve the reduced dimensional system (23).
b) given the reduced-dimensional pressure solution $p_H^{n+1}$, reconstruct a function $p_h^r \in V_h$ by setting

$$p_h^r = \sum_{i=1}^{N} \left( p_H^{n+1} \right)_i \varphi_i.$$  

c) from the reconstructed pressure solution, compute the velocity $u_h^{n+1} \in RT$ using (13) for $s = s_H^r$ and $p = p_h^r$.

d) compute the fine-scale saturation $s_H^{n+1} \in V_h$ using (14).

Notice the difference to the quantities computed in the high-dimensional two-phase flow scheme: While in Algorithm 1 we used the linear mobilities (2) in the DG-bilinear-form and in the right hand side for the pressure, we now use the parametrized mobilities (19). We use the notation $s_H^n$ for saturations computed with the LRBMS scheme—although the saturation itself is not computed in a reduced space—to point out the dependency on the coarse-scale pressure.

The novelties in this scheme are the coupling between pressure and saturation via a parametrized mobility and the replacement of the pressure equation by a reduced-dimensional substitute. While both the least-squares approximation in the basis construction step (2) and the reconstruction step (5b) still need to prove their efficiency, we expect this scheme to allow largely accelerated computations with acceptable additional error. The validity of this assumption will be investigated in the numerical experiments.

We therefore propose Scheme 3 as an alternative to the reduction approaches mentioned in Section 1.

6 Numerical Experiments

In this section, we demonstrate the advantages of our approach introduced in Sections 4 and 5 by means of a 2D-benchmark problem. All implementation was done using the Distributed and Unified Numerics Environment (DUNE), see [49, 50, 51, 52]. For the results shown in this section we will use the true error in an energy norm as error measure:

$$\Delta_N(\mu, W) = \left[ \int_{\Omega} (p_H(\gamma) - p_h(\gamma), p_H(\gamma) - p_h(\gamma)) \, d\Omega \right]^{1/2}, \quad \gamma = \sum_{q=1}^{M} \bar{\mu}_q \Lambda_q$$

where $\bar{\lambda} = \sum_{q=1}^{M} \bar{\mu}_q \Lambda_q$ for a fixed parameter $\bar{\mu}$, $p_H(\gamma)$ denotes the LRBMS-pressure-solution in the space $W$, see Definition 4, and $p_h(\gamma)$ denotes the high-dimensional pressure solution, see Definition 1. For the exposition of well-applicable a-posteriori estimators for our setting we refer to [34, 35].

Our benchmark models the replacement of the non-wetting phase by the wetting phase in $\Omega = [0, 300] \times [0, 60]$ for $T = 3 \cdot 10^5$. The fine mesh $T_h$ consists of $400 \cdot 160 = 64000$ rectangles and we use $N_T = 6000$ time steps for the temporal discretization. The densities are $\varrho_w = 999.749$, $\varrho_o = 890$, the viscosities are $\eta_w = 0.00130581$ and $\eta_o = 0.008$. The permeability and porosity fields are shown in Figures 1 and 2, respectively.
The domain is initially fully saturated with the non-wetting phase \( (s_0 \equiv 0) \), which is then displaced by the wetting phase entering from the left boundary, modelled via the boundary conditions

\[
\begin{align*}
  s &= 1.0 \quad \text{in } \Gamma_{s,d} \times [0, T], & -\lambda(s) \kappa \nabla p \cdot n &= 3 \cdot 10^{-4} \quad \text{in } \Gamma_{p,n}^1 \times [0, T], \\
  p &= 10 \quad \text{in } \Gamma_{p,d} \times [0, T], & -\lambda(s) \kappa \nabla p \cdot n &= 0 \quad \text{in } \Gamma_{p,n}^2 \times [0, T],
\end{align*}
\]

on \( \Gamma_{s,d} = \Gamma_{p,d} = \{0\} \times [0,60] \) and \( \Gamma_{p,n} = \Gamma_{p,n}^1 \cup \Gamma_{p,n}^2 \), where \( \Gamma_{p,n}^1 = \{300\} \times [0,60] \), \( \Gamma_{p,n}^2 = [0,300] \times \{0\} \cup [0,300] \times \{60\} \). In this benchmark no sources are used \( (q_1 \equiv q_2 \equiv 0) \) and we neglect gravity so that \( G = (0,0,0)^T \). The relative permeabilities in Equation (2) are given here via the linear relations \( k_{rw}(s) = s \) and \( k_{ro}(s) = 1 - s \) for simplicity. Different relations like the Brooks-Corey or van Genuchten law would be possible. Figure 3 shows the saturation \( s_h \) computed with the full scheme (Algorithm 1) after approximately 3.5, 10, 15, and 48 hours.

From the time-of-flight, which is depicted in Figure 4, we compute eight wetting and non-wetting mobility profiles using Algorithm 2. The resulting wetting mobility profiles are depicted in Figure 5. Using the profiles we compute the reduced basis \( \Phi \) using coarse meshes with sizes \( |T_H| = 1, 4, 8, 16, 32 \). We use the tolerance \( \epsilon_{tol} = 10^{-4} \), a training set \( \mathcal{P}_{tr} \) consisting of 300 randomly distributed parameters \( \mu \) in \( [0.0001, 1]^8 \) with \( \sum(\mu)_i = 1 \), and the maximum size \( N_{\text{max}} = \infty \) for Algorithm 2. The resulting basis sizes can be seen in Table 1. We observe that with increasing size of the coarse mesh (first column), the number of snapshots computed in Step 2c of Algorithm 2 (fourth column) decreases significantly from 134 to 69. At the same time, the overall basis size (the sum of all local basis sizes, second column) increases. Notice that the basis size is not necessarily equal the product between the number of snapshots and the coarse grid size. This is because we orthonormalize each new snapshot with respect to the existing basis in each extension during the basis generation, reject local extensions with norms below a certain threshold.
Figure 3: Wetting phase saturation $s_h$ computed using the LRBMS method after 3.5, 10, 15 and 48 hours (from left to right and top to bottom) including contour lines

Figure 4: The time-of-flight $\tau_h$ for the 2D benchmark problem for $s = 0$

Figure 5: Wetting mobility profiles computed with Algorithm 2 for the time-of-flight depicted in Figure 4. Not shown are the profiles for $q = 1$ and $q = M$ which have constant values $\Lambda^0_w \equiv 0$ and $\Lambda^M_w \equiv 765.808$, respectively.
Table 1: Basis sizes \(|\Psi|\) resulting from Algorithm 2 for the 2D benchmark problem and a fine triangulation with 64000 elements: Size of the coarse mesh, sum of all local basis sizes before and after application of the PCA, number of snapshots computed during the basis generation.

<table>
<thead>
<tr>
<th>Coarse Grid Size</th>
<th>Basis Size w/o PCA</th>
<th>Basis Size w/ PCA</th>
<th>Number of Snapshots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x1</td>
<td>134</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>4x1</td>
<td>353</td>
<td>155</td>
<td>105</td>
</tr>
<tr>
<td>8x1</td>
<td>565</td>
<td>264</td>
<td>84</td>
</tr>
<tr>
<td>8x2</td>
<td>1030</td>
<td>443</td>
<td>78</td>
</tr>
<tr>
<td>16x2</td>
<td>1771</td>
<td>751</td>
<td>69</td>
</tr>
</tbody>
</table>

Figure 6: Foreground: local basis sizes after the basis generation performed for Table 1, last line, including the PCA. Background: permeability field.

to avoid linear dependencies in the resulting basis and hence reduce the number of local basis functions.

In Table 1 we also see the impact of the local data compression using PCA (Step 3b in Algorithm 2): With increasing coarse mesh size, the PCA is able to reduce the local basis sizes significantly on the different coarse elements.

This effect can be seen again in Figure 6 where we plot the local basis sizes after the PCA step for the test run presented in Table 1, last line. We observe that the local basis size varies strongly from left to right, due to the fact that the peak saturation does not reach the right half of the computational domain before the end time \(T\). Also, the basis size differs from bottom to top, especially in regions where the permeability, which is plotted in the background of Figure 6, shows strong variation from bottom to top.

In Tables 2 and 3 we see the resulting discrepancies between the saturation \(s^n_H\) computed by the SWIP-DG method on the fine mesh (Algorithm 1) and the saturation \(s^n_{H'}\) computed with the reduced scheme 3, as well as the discrepancy between the SWIP-DG pressure and the reduced pressure during the two-phase flow simulation. We present both the relative \(L^2\) and \(H^1\) discrepancies for the saturation

\[
\Delta^n_{L^2} = \frac{\|s^n_h - s^n_{H'}\|_{L^2(\Omega)}}{\|s^n_h\|_{L^2(\Omega)}}, \quad \Delta^n_{H^1} = \frac{\|s^n_h - s^n_{H'}\|_{H^1(\Omega)}}{\|s^n_h\|_{H^1(\Omega)}},
\]
and the respective quantities for the pressure. For different coarse grids, the second column in each table displays the number of snapshots computed during the extension step 2c in Algorithm 2 as a measure for the amount of work needed in that step. Then, in the two sections for the $L^2$ and $H^1$ relative discrepancies, we present mean discrepancies over all time steps and the discrepancy at the last time step $t^{NT}$ for basis generations without and with usage of the PCA.

Over the different coarse mesh configurations, we consistently observe a mean $L^2$-discrepancy for the saturation of approximately 5.6% (standard deviation: 1.6%), both with and without the PCA, which is reduced to 4.3% at end time. The respective $H^1$-discrepancies are slightly bigger with a mean of 7.9% (standard deviation: 2.9%) and 5.5% at end time. The mean $L^2$-discrepancy for the pressure is approximately 1.4% (standard deviation: 0.3%) for all coarse grid configurations, and reduces to 1.3% at end time. The respective $H^1$-discrepancies are in the same ranges. In conclusion, we can say that the data compression step using the PCA increases neither the error for the pressure nor the error for the saturation noticeably.

While we can ensure mass conservation on the coarse grid by adding per-coarse-cell unit basis functions to the reduced basis, the method is not mass-conservative on the fine grid. In Figure 7 we plot the relative mass loss as a measure of the lack of mass conservation at end time $t^n = 3 \cdot 10^5$ for the velocity computed with the LRBMS scheme on a $16 \times 2$ coarse grid. We see that the velocity lacks mass conservation mainly on coarse cell intersections and in regions with large gradients in the mobility profiles $\Lambda^i$. Overall, the relative lack of mass conservation is well below $2 \cdot 10^{-5}$. Although we therefore found this to be only a minor problem in the tests—the saturation usually did not grow above a value of 1.05 and hardly ever fell below zero—future work will include research on how to ensure mass-conservation on the fine grid.

As the discrepancies are increased neither by introducing more coarse cells, nor by application of the PCA, we find the Localized Reduced Basis Multiscale method to be well applicable in this context: Introduction of more coarse cells reduces the number of costly snapshots to be computed during the basis construction, while application of the PCA yields smaller, more compact bases, leading to faster computations during the two-phase flow simulation.
<table>
<thead>
<tr>
<th>Coarse Grid Size</th>
<th>Number of Snapshots</th>
<th>Relative L2 Errors</th>
<th>Relative H1 Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No PCA</td>
<td>PCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Error</td>
<td>Error at End Time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1x1</td>
<td>134</td>
<td>5.59 %</td>
<td>4.26 %</td>
</tr>
<tr>
<td>4x1</td>
<td>105</td>
<td>5.59 %</td>
<td>4.25 %</td>
</tr>
<tr>
<td>8x1</td>
<td>84</td>
<td>5.58 %</td>
<td>4.25 %</td>
</tr>
<tr>
<td>8x2</td>
<td>78</td>
<td>5.58 %</td>
<td>4.25 %</td>
</tr>
<tr>
<td>16x2</td>
<td>69</td>
<td>5.59 %</td>
<td>4.25 %</td>
</tr>
</tbody>
</table>

Table 2: Relative $L^2$ and $H^1$ discrepancies between the saturation computed with Algorithm 3 and the saturation computed with the fine-scale algorithm 1 for different coarse mesh configurations: Number of snapshots needed during the basis construction, mean relative $L^2$ discrepancy and relative $L^2$ discrepancy at end time $T$ without using the PCA, respective quantities after usage of the PCA, respective quantities for the $H^1$ norm

<table>
<thead>
<tr>
<th>Coarse Grid Size</th>
<th>Number of Snapshots</th>
<th>Relative L2 Errors</th>
<th>Relative H1 Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No PCA</td>
<td>PCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Error</td>
<td>Error at End Time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1x1</td>
<td>134</td>
<td>1.36 %</td>
<td>1.33 %</td>
</tr>
<tr>
<td>4x1</td>
<td>105</td>
<td>1.36 %</td>
<td>1.33 %</td>
</tr>
<tr>
<td>8x1</td>
<td>84</td>
<td>1.35 %</td>
<td>1.32 %</td>
</tr>
<tr>
<td>8x2</td>
<td>78</td>
<td>1.36 %</td>
<td>1.32 %</td>
</tr>
<tr>
<td>16x2</td>
<td>69</td>
<td>1.36 %</td>
<td>1.32 %</td>
</tr>
</tbody>
</table>

Table 3: Relative $L^2$ and $H^1$ discrepancies between the pressure computed with Algorithm 3 and the pressure computed with the fine-scale algorithm 1 for different coarse mesh configurations: Number of snapshots needed during the basis construction, mean relative $L^2$ discrepancy and relative $L^2$ discrepancy at end time $T$ without using the PCA, respective quantities after usage of the PCA, respective quantities for the $H^1$ norm
Test runs with a different tolerance $\epsilon_{\text{tol}} = 10^{-5}$ for the Greedy basis construction (Algorithm 2) exhibit the same discrepancies in both the $L^2$ and $H^1$ norms. Further, computations with higher numbers of mobility profiles ($M = 10, 12, 20$) give roughly the same discrepancies. This gives rise to the assumption that the error is dominated by different phenomena, which we consider to be twofold: First, the assumption that the time-of-flight is invariant for the whole simulation is not valid in some regions. Therefore, the position of the saturation front and its impact on the mobility cannot be represented correctly. Second, the profile of the mobility along streamlines is not approximated well enough due to the ad-hoc profile generation in Algorithm 2.

To support the first statement, we present in Figure 8 the absolute value of the difference between the SWIP-DG approximation $s^{N_T}_h$ and our approximation of the saturation $s^{N_T}_h$ at end time $t^{N_T}$. We see that huge discrepancies arise in three distinct positions: around the point $x = (30, 35)$, the point $x = (40, 10)$, and the point $x = (80, 1)$. These positions are located directly downwind along the streamlines from points where the time-of-flight changes significantly during the test run, see Figure 9. The error produced in those regions is then transported through the domain along the streamlines, hence the error distribution to be seen in Figure 8 is established.

The second statement can be justified by using $M$ high-dimensional saturation approximations $s^{n_1}_h, \ldots, s^{n_M}_h$ at points $t^1, \ldots, t^M$ in time to form the profiles in Step 1c of Algorithm 2:

$$\Lambda^q_w = \lambda_w(s^q_h), \quad \Lambda^q_o = \lambda_o(s^q_h).$$

In doing so, we ensure that the shape of the mobility profiles along streamlines is correct and the error of the LRBMS two-phase flow scheme should decrease drastically. Indeed, for $M = 8$, this procedure decreases the mean relative $L^2$-discrepancy to 2.3% with a standard deviation of 2.0%. Even more: The $L^2$-discrepancy at $t = T$ is reduced by a factor of ten to 0.45%. Obviously, using multiple high-dimensional saturation profiles is not possible in general as the model order reduction would become superfluous, but two other remedies could be implemented: One could be to recompute the mobility profiles and the reduced basis after a certain number of time steps $t^n$ using the time-of-flight for $s^n_h$. Another approach could be to solve the full high-dimensional two-phase flow problem along some streamlines in one dimension in a preparatory step and use a combination of the resulting mobility profiles as a replacement for those in Step 1c of Algorithm 2.

In Table 4 we present runtimes for the basis construction part of the LRBMS two-phase flow scheme. Again, we give the coarse grid configuration (column one) and the number of snapshots needed (column two). The third column then shows the time needed for the basis construction (Algorithm 2). The timings presented here include the time for the computation of the mobility profiles (about 13 seconds), the so-called “training”-step (computation of all reduced solutions, evaluation of the error measure, selection of the parameter for basis extension), and the extension-step including the computation of a high-dimensional snapshot and application of the Gram-Schmidt procedure. It does not contain the application of the PCA which is consistently below one second for all coarse grid configurations and hence can be considered negligible. The time needed to compute
Figure 8: Absolute value of the difference in the saturation $s_{hT}^{N_T}$ computed with the full high-dimensional scheme and the saturation $s_{H}^{N_T}$ computed with the LRBMS scheme at end time $T$.

Figure 9: Absolute difference between the time-of-flight for the initial saturation $s_0$ and the time-of-flight for the saturation computed with the LRBMS scheme for a coarse grid with $16 \times 2$ cells at time $t = 10000s$.

the reduced basis $\Phi$ is 50 minutes on one coarse cell (which corresponds to a standard Reduced Basis method), goes up to 1 hour 7 minutes for a coarse grid of size $4 \times 1$ and is then reduced to 45 minutes for all other coarse grids. The increase in total runtime from line one to line two can explained by the relatively high number of iterations that is still needed to reach the error tolerance with an increased per-step cost due to the larger reduced bases. This effect would vanish if the snapshot computation (column four) was more expensive, for larger fine grids, for example. As mentioned earlier, we use the true error as error measure. The time for error estimation in column five includes the time needed to reconstruct a high-dimensional snapshot from each reduced solution and compute the difference in the $L^2$-norm. The time for computation of the high-dimensional snapshots itself is not included.

In conclusion we can say that for the test case at hand, introduction of more coarse cells yields slight reductions in terms of runtime. As the computation of the snapshots shows a speedup by a factor of two for the computation on 32 coarse cells compared to the computation on one coarse cell, we can expect the runtime-gain to increase drastically as the size of the fine mesh—and hence the time for the computation of one high-dimensional snapshot—is increased.

Finally, in Table 5 we present runtimes for the two-phase flow simulation (Step 5 in Algorithm 3) for uncompressed bases (that is: bases that were computed without usage of the PCA) and compressed bases on different coarse grids and, for comparison, the same runtimes for a full high-dimensional computation. We see that for the reduced
Table 4: Runtimes for the basis construction (Algorithm 2): Coarse grid size, number of snapshots computed, total runtime, total time for snapshot computation, total time for error estimation.

simulations, more than 50% of the overall runtime of about 2 hours 40 minutes is spend in the application of the ODE solver and slope limiter for the saturation equation. About one hour is spend for the elliptic equation with about 45 minutes for the reconstruction of the flux (see Definition 2). For uncompressed bases, computing all pressure solutions takes five to 15 minutes, depending on the coarse grid and respective basis size, for compressed bases those times drop to two to five minutes.

The time to solution for one reduced pressure computation therefore ranges from 20 milliseconds on the $1 \times 1$, $4 \times 1$ and $8 \times 1$ coarse grids to 50 milliseconds on the $8 \times 2$ and $16 \times 2$ coarse grids using the PCA. Comparing these runtimes to the time needed for a high-dimensional simulation we see the advantage of our method: The high-dimensional test run takes approximately 16 hours with nearly 90% of the time spent in the treatment of the elliptic equation: more than two hours are spent assembling the pressure system, solving it takes approximately 11 hours in total (approximately seven seconds per solve). The speed-up for the solution of the pressure equation is approximately a factor 140. Because nothing was done to speed up the transport solve, the overall speed-up for the two-phase flow simulation (high-dimensional vs. basis construction and reduced simulation) is five. To also speed up the transport solve significantly, one could replace the explicit temporal discretization by a backward Euler scheme and utilize the fact that the resulting nonlinear system share the same unidirectional flow properties as the time-of-flight equation and hence can be computed in a per-element fashion with local control over the nonlinear iterations; see [53] for details. Note that in this case, the slope limiter presented in Section 3 needs to be replaced by a different kind of stabilization as it is not compliant with implicit time stepping schemes.

7 Summary and Outlook

We introduced a localized version of the Reduced Basis method. The idea of our approach is to make use of two computational meshes: A fine mesh is used to compute detailed
Table 5: Runtimes for Step 5 of Algorithm 3 for uncompressed and compressed bases: 
Coarse grid size, total runtime, time spent for treatment of the elliptic equation, thereof time spent in pressure solve and velocity reconstruction and time spent in saturation part. Last line: same numbers for a full high-dimensional simulation

Table: 

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1x1</td>
<td>2h 37m</td>
<td>1h 0m</td>
<td>5m 28s</td>
<td>46m 22s</td>
<td>1h 36m 35s</td>
<td>2h 34m</td>
<td>0h 56m</td>
<td>1m 59s</td>
<td>45m 11s</td>
<td>1h 38m 5s</td>
</tr>
<tr>
<td>4x1</td>
<td>2h 41m</td>
<td>1h 1m</td>
<td>5m 5s</td>
<td>47m 11s</td>
<td>1h 40m 23s</td>
<td>2h 31m</td>
<td>0h 55m</td>
<td>2m 0s</td>
<td>44m 21s</td>
<td>1h 36m 24s</td>
</tr>
<tr>
<td>8x1</td>
<td>2h 38m</td>
<td>1h 0m</td>
<td>5m 25s</td>
<td>46m 31s</td>
<td>1h 36m 57s</td>
<td>2h 41m</td>
<td>0h 55m</td>
<td>3m 0s</td>
<td>43m 53s</td>
<td>1h 45m 49s</td>
</tr>
<tr>
<td>8x2</td>
<td>2h 40m</td>
<td>1h 4m</td>
<td>10m 40s</td>
<td>44m 45s</td>
<td>1h 35m 59s</td>
<td>2h 37m</td>
<td>0h 57m</td>
<td>4m 42s</td>
<td>44m 26s</td>
<td>1h 39m 11s</td>
</tr>
<tr>
<td>16x2</td>
<td>2h 45m</td>
<td>1h 9m</td>
<td>15m 3s</td>
<td>45m 7s</td>
<td>1h 35m 59s</td>
<td>2h 37m</td>
<td>0h 57m</td>
<td>4m 42s</td>
<td>44m 26s</td>
<td>1h 39m 11s</td>
</tr>
</tbody>
</table>

High-Dim. | 15h 58m | 14h 23m | 11h 13m | 43m 29s | 1h 34m 50s

The quality of the reduced approximations for saturation and pressure is mainly decided by the quality of the approximation of the original mobility by our parametrized surrogate, therefore future work will include improvements of the mobility approximation.

Mass conservation can only be guaranteed on the coarse mesh of our scheme, in future work we will address the problem of mass conservation on the fine mesh for the reduced simulations. Applying the LRBMS to the saturation equation also would be an extension promising additional speedup.

Acknowledgement

The author S. Kaulmann would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart and within the German Priority Programme 1648, “SPPEXA - Software for Exascale Computing”. Furthermore, financial support by the Baden-Württemberg Stiftung gGmbH is gratefully acknowledged. Parts of the work was carried out while the first author visited SINTEF in Oslo, Norway.
References


