Name?

G. B. Diaz Cortes[1], C. Vuik[1] and J. D. Jansen[2]

[1]Department of Applied Mathematics, TU Delft
[2]Department of Geoscience & Engineering, TU Delft

March 2016

Abstract

We study fast and robust iterative solvers for large systems of linear equations resulting from reservoir simulation trough porous media. We propose the use of pre-conditioning and deflation techniques based on information obtained from the system to reduce the time spent in the solution of the linear system.

We consider compressible and incompressible single-phase flow in a layered model with large variations in the permeability coefficients and the SPE10 benchmark model (the Markovinovic problem).

Introduction.

Often, most computational time in the simulation of multi-phase flow through porous media is taken up by solution of the pressure equation. This involves primarily solving large systems of linear equations as part of the iterative solution of the time and space discretized governing nonlinear partial differential equations. The time spent in solving the linear systems depends on the size of the problem and the variations of permeability within the medium. Solution of problems with extreme contrast in the grid block permeability values may lead to very large computing times.

A potential approach to reduce the computing time for large-scale problems with the aid of Proper Orthogonal Decomposition (POD) is investigated in [1] and [2]. Astrid et al. [1] propose the use of a POD-based preconditioner for the acceleration of the solution to the pressure equation. The preconditioner is constructed from a set of snapshots, obtained from solutions to the pressure equation in previous time steps. Once the snapshots are computed, the POD method is used to obtain a set of basis vectors that capture the most relevant features of the system, which can be used to improve the simulation time of the subsequent time steps. A similar approach is used in [2], where a set of POD-based basis vectors is obtained from the initial time steps. However, in this case, the acceleration is

1

achieved by only improving the initial guess.

Problems with a high contrast between the permeability coefficients are sometimes approached through the use of deflation techniques; see, e.g., [3]. The use of deflation techniques involves the search of good deflation vectors, which are usually problem-dependent. In [3], subdomain based deflation vectors are used for layered problems with a large contrast between the permeability coefficients. However, these deflation vectors cannot be used if the distribution of the permeability coefficients is not structured as is the case in, e.g., the well-known SPE 10 benchmark problem [4].

Following the ideas of [1] and [2], we propose the use of POD of many snapshots to capture the system's behavior. The basis obtained with POD is studied as an alternative choice of deflation vectors to accelerate the convergence of the pressure solution in a porous medium with high-contrast variations in the permeability coefficients. We consider incompressible and compressible single-phase flow through a porous medium with large variations in the permeability coefficients for layered academic problems and the SPE 10 benchmark (Markovinovic problem?).

In Section 1 we give some theory about the linear solver used for this work and we introduce preconditioning and deflation techniques. We also demonstrate two lemmas that will help us in the choice of good deflation vectors for the incompressible case, necessary for the deflation techniques.

Section 2 is devoted to the description of the models used to simulate flow throug porous media. In Section 3 we present numerical experiments. We describe the problem that is studied, the solver that is used and the preconditioning and deflation techniques used for the speedup of the solver. The results are also presented in this section.

Finally, we end with the conclusions.

1 Iterative solution methods

The solution of partial differential equations PDE's can be performed with numerical methods. Some of these methods, as the finite differences method, transform our equations into a linear system of the form:

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1,$$

$$\cdots$$

$$a_{n1}x_1 + \cdots + a_{nn}x_n = b_n,$$

which is written in matrix form as:

$$\mathbf{Ax} = \mathbf{b}. \tag{1}$$

System (1) can be solved with direct or iterative methods. Direct methods achieve a final solution, while the iterative ones are stopped if the error is less than a given value (tolerance).

Some of the iterative methods are: Jacobi, Gauss Seidel, and if the matrix is Symetric Positive Definite SPD[1] we can use the Conjugate Gradient (CG) method. This section is devoted to iterative methods, in particular CG, that is the method used in this work.

In this section, we also describe the Preconditioning and Deflation techniques for the acceleration of the CG method.

1.1 Krylov subspace Methods

If we have two subspaces $\mathcal{K}_k$, $\mathcal{L}_k$ of $\mathbb{R}^n$ and we want to solve the Equation (1), with $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$ we can use a projection method onto $\mathcal{K}_k$. This method allows us to find an approximate solution $\mathbf{x}^k$ from an arbitrary initial guess solution $\mathbf{x}^0$. This approximate solution lies in the Krylov subspace of dimension k of the matrix $\mathbf{A}$ and residual $\mathbf{r}^0$,

$$\mathbf{x}^k \in \mathbf{x}^0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}^0),$$

with $\mathcal{K}_k(\mathbf{A}, \mathbf{r}^0)$ defined as:

$$\mathcal{K}_k(\mathbf{A}, \mathbf{r}^0) = span\{\mathbf{r}^0, \mathbf{Ar}^0, \ldots, \mathbf{A}^{k-1}\mathbf{r}^0\}.$$

Where the residual $\mathbf{r}^k = \mathbf{b} - \mathbf{Ax}^k$ is orthogonal to the subspace $\mathcal{L}_k$, with

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{B}^{-1}r^k, \qquad \mathbf{r}^k = \mathbf{b} - \mathbf{Ax}^k.$$

The subspace $\mathcal{L}_k$ is chosen depending on the Krylov subspace method that is used.

1.2 Conjugate Gradient Method

The Conjugate Gradient (CG) method is a Krylov subspace method for SPD matrices, such that

$$||\mathbf{x} - \mathbf{x}^k||_{\mathbf{A}}{}^2 \tag{2}$$

---

[1] $\mathbf{A}^T = \mathbf{A}$, and $(\mathbf{Ax}, \mathbf{x}) > 0,$ $\quad \forall \mathbf{x} \in \mathbb{R}^n,$ $\quad \mathbf{x} \neq 0.$
[2] $||\mathbf{x}||_{\mathbf{A}} = \sqrt{(\mathbf{x}, \mathbf{x})_{\mathbf{A}}} = \sqrt{\mathbf{x}^T \mathbf{Ax}}.$

is minimal, with $\mathbf{x}$ the solution of the system and $\mathbf{x}^k$ the approximate solution after k iterations and the error of this iteration is bounded by:

$$||\mathbf{x} - \mathbf{x}^k||_{\mathbf{A}} \leq 2||\mathbf{x} - \mathbf{x}^0||_{\mathbf{A}} \left(\frac{\sqrt{\kappa_2(\mathbf{A})} - 1}{\sqrt{\kappa_2(\mathbf{A})} + 1} \right)^k . ^3 \tag{3}$$

Algorithm 1 Conjugate Gradient (CG) method, solving $\mathbf{Ax} = \mathbf{b}$.

Give an initial guess $\mathbf{x}^0$.
Compute $\mathbf{r}^0 = \mathbf{b} - \mathbf{Ax}^0$ and set $\mathbf{p}^0 = \mathbf{r}^0$.
   for $k = 0, ...,$ until convergence
      $\alpha^k = \frac{(\mathbf{r}^k, \mathbf{r}^k)}{(\mathbf{Ap}^k, \mathbf{p}^k)}$
      $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{p}^k$
      $\mathbf{r}^{k+1} = \mathbf{r}^k - \alpha^k \mathbf{Ap}^k$
      $\beta^k = \frac{(\mathbf{r}^{k+1}, \mathbf{r}^{k+1})}{(\mathbf{r}^k, \mathbf{r}^k)}$
      $\mathbf{p}^{k+1} = \mathbf{r}^{k+1} + \beta^k \mathbf{p}^k$
   end

1.3 Preconditioning

If we want to accelerate the convergence of an iterative method, we can transform the system into another one containing a better spectrum, i.e, a smaller condition number. This can be done by multiplying the original system (1) by a matrix $\mathbf{M}^{-1}$.

$$\mathbf{M}^{-1}\mathbf{Ax} = \mathbf{M}^{-1}\mathbf{b}. \tag{4}$$

The new system has the same solution but provides a substantial improvement on the spectrum. For this preconditioned system, the convergence is given by:

$$||\mathbf{x} - \mathbf{x}^{k+1}||_{\mathbf{A}} \leq 2||\mathbf{x} - \mathbf{x}^0||_{\mathbf{A}} \left(\frac{\sqrt{\kappa(\mathbf{M}^{-1}\mathbf{A})} - 1}{\sqrt{\kappa(\mathbf{M}^{-1}\mathbf{A})} + 1} \right)^{k+1} . \tag{5}$$

If the system matrix is SPD, $\mathbf{M}$ is chosen as an SPD matrix such that $\kappa(\mathbf{M}^{-1}\mathbf{A}) \leq \kappa(\mathbf{A})$, and $\mathbf{M}^{-1}b$ is easy to compute.

---

[3]The condition number $\kappa_2(\mathbf{A})$ is defined as $\kappa_2(\mathbf{A}) = \frac{\sqrt{\lambda_{max}(\mathbf{A}^T\mathbf{A})}}{\sqrt{\lambda_{min}(\mathbf{A}^T\mathbf{A})}}$. If $\mathbf{A}$ is SPD, $\kappa_2(\mathbf{A}) = \frac{\lambda_{max}(\mathbf{A})}{\lambda_{min}(\mathbf{A})}$.

1.4 Conjugate Gradient Method

The Conjugate Gradient (CG) method is a Krylov subspace method for Symmetric Positive Definite SPD matrices, such that

$$||\mathbf{x} - \mathbf{x}^k||_{\mathbf{A}}, \, ^4 \tag{6}$$

is minimal, with $\mathbf{x}$ the solution of the system and $\mathbf{x}^k$ the $k-th$ iteration.
Given an initial guess $\mathbf{x}^0$ the next approximations can be computed following the search directions $\mathbf{p}^i$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{p}^k.$$

For the CG method, the search directions $\mathbf{p}^k$ are orthogonal with respect to the $\mathbf{A}$ inner product, i.e.

$$(\mathbf{A}\mathbf{p}^k, \mathbf{p}^j) = 0, \qquad k \neq j,$$

and the residuals form an orthogonal set, i.e.

$$(\mathbf{r}^k, \mathbf{r}^j) = 0, \qquad k \neq j.$$

The constant α^k that satisfies (6) is given by:

$$\alpha^k = \frac{(\mathbf{r}^k, \mathbf{r}^k)}{(\mathbf{A}\mathbf{p}^k, \mathbf{p}^k)},$$

and the new search directions can be computed via the residuals,

$$\mathbf{p}^{k+1} = \mathbf{r}^{k+1} + \beta_k \mathbf{p}^k,$$

where

$$\beta_k = \frac{(\mathbf{r}^{k+1}, \mathbf{r}^{k+1})}{(\mathbf{r}^k, \mathbf{r}^k)}.$$

After $k + 1$ iterations of the CG method, the error of the iteration will be bounded by:

$$||\mathbf{x} - \mathbf{x}^{k+1}||_{\mathbf{A}} \leq 2||\mathbf{x} - \mathbf{x}^0||_{\mathbf{A}} \left(\frac{\sqrt{\kappa_2(\mathbf{A})} - 1}{\sqrt{\kappa_2(\mathbf{A})} + 1} \right)^{k+1} \, ^5. \tag{7}$$

1.5 Deflation

Deflation is used to annihilate the effect of extreme eigenvalues on the convergence of an iterative method ([3]). Given an SPD matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the deflation matrix $\mathbf{P}$ is defined as follows ([5]):

$$\mathbf{P} = \mathbf{I} - \mathbf{A}\mathbf{Q}, \qquad \mathbf{P} \in \mathbb{R}^{n \times n}, \qquad \mathbf{Q} \in \mathbb{R}^{n \times n},$$

---

[4] $||\mathbf{x}||_{\mathbf{A}} = \sqrt{(\mathbf{x}, \mathbf{x})_{\mathbf{A}}} = \sqrt{\mathbf{x}^T \mathbf{A} \mathbf{x}}$.

[5] The condition number $\kappa_2(\mathbf{A})$ is defined as $\kappa_2(\mathbf{A}) = \frac{\sqrt{\lambda_{max}(\mathbf{A}^T \mathbf{A})}}{\sqrt{\lambda_{min}(\mathbf{A}^T \mathbf{A})}}$. If $\mathbf{A}$ is SPD, $\kappa_2(\mathbf{A}) = \frac{\lambda_{max}(\mathbf{A})}{\lambda_{min}(\mathbf{A})}$.

where
$$\mathbf{Q} = \mathbf{Z}\mathbf{E}^{-1}\mathbf{Z}^T, \qquad \mathbf{Z} \in \mathbb{R}^{n \times m}, \qquad \mathbf{E} \in \mathbb{R}^{m \times m},$$

with
$$\mathbf{E} = \mathbf{Z}^T \mathbf{A} \mathbf{Z}.$$

The matrix $\mathbf{E}$ is known as the *Galerkin* or *coarse* matrix that has to be invertible. If $\mathbf{A}$ is SPD and $\mathbf{Z}$ is full rank then $\mathbf{E}$ is invertible. The full rank matrix $\mathbf{Z}$ is called the *deflation−subspace* matrix, and it's $l < n$ columns are the *deflation* vectors or *projection* vectors.

Some properties of the previous matrices are ([5, pag. 27]):

a) $\mathbf{P}^2 = \mathbf{P}$.

b) $\mathbf{A}\mathbf{P}^T = \mathbf{P}\mathbf{A}$.

c) $(\mathbf{I} - \mathbf{P}^T)\mathbf{x} = \mathbf{Q}\mathbf{b}$.

d) $\mathbf{P}\mathbf{A}\mathbf{Z} = \mathbf{0}^{n \times m}$.

e) $\mathbf{P}\mathbf{A}$ is $SPSD^6$.

We can split the vector $\mathbf{x}$ as:

$$\mathbf{x} = \mathbf{I}\mathbf{x} - \mathbf{P}^T\mathbf{x} + \mathbf{P}^T\mathbf{x} = (\mathbf{I} - \mathbf{P}^T)\mathbf{x} + \mathbf{P}^T\mathbf{x}. \tag{8}$$

Multiplying the expression above by $\mathbf{A}$, using the properties above, we have:

$$\begin{aligned}
\mathbf{A}\mathbf{x} &= \mathbf{A}(\mathbf{I} - \mathbf{P}^T)\mathbf{b} + \mathbf{A}\mathbf{P}^T\mathbf{x}, & Property: \\
\mathbf{A}\mathbf{x} &= \mathbf{A}\mathbf{Q}\mathbf{b} + \mathbf{A}\mathbf{P}^T\mathbf{x}, & c) \\
\mathbf{b} &= \mathbf{A}\mathbf{Q}\mathbf{b} + \mathbf{P}\mathbf{A}\mathbf{x}, & b),
\end{aligned}$$

multiplying by $\mathbf{P}$ and using the properties $\mathbf{P}\mathbf{A}\mathbf{Q} = \mathbf{0}^{n \times n}$ and $\mathbf{P}^2 = \mathbf{P}$, properties $d)$ and $e)$, we have:

$$\begin{aligned}
\mathbf{P}\mathbf{A}\mathbf{Q}\mathbf{b} + \mathbf{P}^2\mathbf{A}\mathbf{x} &= \mathbf{P}\mathbf{b}, \\
\mathbf{P}\mathbf{A}\mathbf{x} &= \mathbf{P}\mathbf{b},
\end{aligned} \tag{9}$$

where $\mathbf{P}\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{b}$ is the deflated system. Since $\mathbf{P}\mathbf{A}$ is singular, the solution $\mathbf{x}$ can contain components of the null space of $\mathbf{P}\mathbf{A}$. A solution to this system, called deflated solution, is denoted by $\hat{\mathbf{x}}$. The deflated system for $\hat{\mathbf{x}}$ is:

$$\mathbf{P}\mathbf{A}\hat{\mathbf{x}} = \mathbf{P}\mathbf{b}. \tag{10}$$

---

[6]Symmetric Positive Semi-Definite, $(\mathbf{A}\mathbf{x}, \mathbf{x}) \geq 0$, for all $\mathbf{x}$.

As mentioned above, the solution to Equation (9) can contain components of $\mathcal{N}(\mathbf{PA})$. Therefore, a solution to Equation (10), $\hat{\mathbf{x}}$ can be decomposed as:

$$\hat{\mathbf{x}} = \mathbf{x} + \mathbf{y}, \tag{11}$$

with $\mathbf{y} \in \mathcal{R}(\mathbf{Z}) \subset \mathcal{N}(\mathbf{PA})$, and $\mathbf{x}$ the solution to Equation (1).
Note: If $\mathbf{y} \in \mathcal{R}(\mathbf{Z})$, then

$$\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{z}_i,$$

$$\mathbf{PA}y = \mathbf{PA}(\mathbf{z}_1 \alpha_1 + \ldots + \mathbf{z}_m \alpha_m) = \mathbf{PAZ}\alpha, \tag{12}$$

from 1.5 h) $\mathbf{PAZ} = \mathbf{0}^{n \times l}$, then

$$\mathbf{PAy} = \mathbf{0}. \tag{13}$$

Therefore $\mathcal{R}(\mathbf{Z}) \subset \mathcal{N}(\mathbf{PA})$.
Multiplying Equation (11) by $\mathbf{P}^T$ we obtain:

$$\mathbf{P}^T \hat{\mathbf{x}} = \mathbf{P}^T \mathbf{x} + \mathbf{P}^T \mathbf{y},$$

combining Equation (13) with 1.5 f), we have:

$$\mathbf{PA}y = \mathbf{AP}^T \mathbf{y} = \mathbf{0}.$$

Therefore

$$\mathbf{P}^T \hat{\mathbf{x}} = \mathbf{P}^T \mathbf{x}. \tag{14}$$

Substitution to Equation (14) and 1.5 g) in Equation (8) leads to:

$$\mathbf{x} = \mathbf{Qb} + \mathbf{P}^T \hat{\mathbf{x}}, \tag{15}$$

which gives us a relation between $\hat{\mathbf{x}}$ and $\mathbf{x}$.

1.6   Deflated CG Method

To obtain the solution to the linear system (1), we solve the deflated system:

$$\mathbf{PA}\hat{\mathbf{x}} = \mathbf{Pb}. \tag{16}$$

with the CG method, for a deflated solution $\hat{\mathbf{x}}$. Therefore, the solution $\mathbf{x}$ to the original system is obtained from (15):

$$\mathbf{x} = \mathbf{Qb} + \mathbf{P}^T \hat{\mathbf{x}}.$$

Deflated PCG Method

The deflated linear system can also be preconditioned by an SPD matrix $\mathbf{M}$.
The deflated preconditioned system to solve with CG is [5, pag. 30]:

$$\tilde{\mathbf{P}}\tilde{\mathbf{A}}\hat{\tilde{\mathbf{x}}} = \tilde{\mathbf{P}}\tilde{\mathbf{b}},$$

where:

$$\tilde{\mathbf{A}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{M}^{-\frac{1}{2}}, \qquad \hat{\tilde{\mathbf{x}}} = \mathbf{M}^{\frac{1}{2}}\hat{\mathbf{x}}, \qquad \tilde{\mathbf{b}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{b}$$

This method is called the Deflated Preconditioned Conjugate Gradient $DPCG$ method, and the error is bounded by:

$$||\mathbf{x} - \mathbf{x}^{i+1}||_{\mathbf{A}} \leq 2||\mathbf{x} - \mathbf{x}^0||_{\mathbf{A}} \left(\frac{\sqrt{\kappa_{eff}(\mathbf{M}^{-1}\mathbf{PA})} - 1}{\sqrt{\kappa_{eff}(\mathbf{M}^{-1}\mathbf{PA})} + 1} \right)^{i+1},$$

were $\mathbf{C}_{eff} = \frac{\lambda_{max}(M^{-1}PA)}{\lambda_{min}(M^{-1}PA)}$ is the effective condition number and $\lambda_{min}(M^{-1}PA)$ is the smallest non-zero eigenvalue of $M^{-1}PA$.

1.7   Choices of Deflation Vectors

The deflation method is used to treat the most unfavorable eigenvalues of $\mathbf{A}$. If the matrix $\mathbf{Z}$ contains eigenvectors corresponding to the unfavorable eigenvalues, the convergence of the iterative method is achieved faster. However, to obtain and to apply the eigenvectors is costly in most of the cases. Therefore, a good choice of the matrix $\mathbf{Z}$ that does not contain the eigenvectors is essential for the acceleration of the convergence.
A good choice of the deflation vectors is usually problem-dependent. Available information on the system is, in general, used to obtain these vectors. Most of the techniques used to choose deflation vectors are based on approximating eigenvectors, recycling ([6]), subdomain deflation vectors ([7]) or multigrid and multilevel based deflation techniques ([8, 9]). A summary of these techniques is given below.

Recycling Deflation. A set of vectors previously used is reused to build the deflation-subspace matrix ([6]). The vectors could be, for example, $q-1$ solution vectors of the linear system with different right-hand sides or of different time steps. The matrix $\mathbf{Z}$ containing these solutions is:

$$\mathbf{Z} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(q-1)}].$$

Subdomain Deflation. The domain is divided into several subdomains, each corresponding to one or more deflation vectors. For each subdomain, there is a deflation vector that contains ones for cells in the subdomain and zeros for cells outside ([7]).

Multi Grid and Multilevel Deflation. For the multigrid and multilevel methods, the prolongation and restriction matrices are used to pass from one level or grid to another. These matrices can be used as the deflation-subspace matrices $\mathbf{Z}$ ([8]).

1.8 Analysis of the snapshots

As mentioned in the previous section, it is important to choose 'good' deflation vectors if we want to speedup an iterative method.

We can use solutions of a system slightly different from the original (snapshots) as deflation vectors. For this, we need to choose the way of selecting these snapshots wisely. The idea behind this selection is to obtain a small number of snapshots and at the same time obtain the largest amount of information from the system.

In this section two lemmas are proved. The lemmas could help us to select the systems that we are going to solve for the snapshots.

Lemma 1. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a non-singular matrix, such that

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{17}$$

and $\mathbf{x}_i, \mathbf{b}_i \in \mathbb{R}^n$, $i = 1, ..., m$, where the vectors $\mathbf{b}_i$ are linearly independent ($l.i.$) such that:

$$\mathbf{A}\mathbf{x}_i = \mathbf{b}_i, \tag{18}$$

The following equivalence holds

$$\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i \qquad \Leftrightarrow \qquad \mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{b}_i. \tag{19}$$

Proof $\Rightarrow$

$$\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i \Rightarrow \mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{b}_i. \tag{20}$$

Substituting $\mathbf{x}$ from (20) into $\mathbf{A}\mathbf{x} = \mathbf{b}$ leads to:

$$\mathbf{A}\mathbf{x} = \sum_{i=1}^{m} \mathbf{A} c_i \mathbf{x}_i = \mathbf{A}(c_1 \mathbf{x}_1 + ... + c_m \mathbf{x}_m).$$

Using the linearity of $\mathbf{A}$ the equation above can be rewritten as:

$$\mathbf{A} c_1 \mathbf{x}_1 + ... + \mathbf{A} c_m \mathbf{x}_m = c_1 \mathbf{b}_1 + ... + c_m \mathbf{b}_m = \mathbf{B}\mathbf{c}. \tag{21}$$

where $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{c} \in \mathbb{R}^m$, and the columns of $\mathbf{B}$ are the vectors $\mathbf{b}_i$.
From (17) and (21) we get:

$$\mathbf{A}\mathbf{x} = \mathbf{b} = c_1 \mathbf{b}_1 + ... + c_m \mathbf{b}_m = \sum_{i=1}^{m} c_i \mathbf{b}_i.$$

Proof $\Leftarrow$

$$\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i \Leftarrow \mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{b}_i. \tag{22}$$

9

Substituting $\mathbf{b}$ from (22) into $\mathbf{Ax} = \mathbf{b}$ leads to:

$$\mathbf{Ax} = \sum_{i=1}^{m} c_i \mathbf{b}_i. \tag{23}$$

Since $\mathbf{A}$ is non-singular, multiplying (18) and (22) by $\mathbf{A}^{-1}$ we obtain:

$$\mathbf{x}_i = \mathbf{A}^{-1} \mathbf{b}_i, \tag{24}$$

$$\mathbf{x} = \mathbf{A}^{-1} \sum_{i=1}^{m} c_i \mathbf{b}_i = \sum_{i=1}^{m} c_i \mathbf{A}^{-1} \mathbf{b}_i = \sum_{i=1}^{m} c_i \mathbf{x}_i. \tag{25}$$

$$Q.E.D.$$

Lemma 2. If the the deflation matrix $\mathbf{Z}$ is constructed with a set of m vectors

$$\mathbf{Z} = \begin{bmatrix} \mathbf{x}_1 & \dots & \dots & \mathbf{x}_m \end{bmatrix}, \tag{26}$$

such that $\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i$, with $\mathbf{x}_i$ $l.i.$, then the solution of system $\mathbf{Ax} = \mathbf{b}$ is achieved within one iteration of DCG.
Proof.
The relation between $\hat{\mathbf{x}}$ and $\mathbf{x}$ is given in Equation (15):

$$\mathbf{x} = \mathbf{Qb} + \mathbf{P}^T \hat{\mathbf{x}}.$$

For the first term $\mathbf{Qb}$, taking $\mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{b}_i$ we have:

$$\begin{aligned}
\mathbf{Qb} &= \mathbf{Z}\mathbf{E}^{-1}\mathbf{Z}^T \left(\sum_{i=1}^{m} c_i \mathbf{b}_i \right) \\
&= \mathbf{Z}(\mathbf{Z}^T\mathbf{A}\mathbf{Z})^{-1}\mathbf{Z}^T \left(\sum_{i=1}^{m} c_i \mathbf{A}\mathbf{x}_i \right) \qquad \text{using Lemma 1} \\
&= \mathbf{Z}(\mathbf{Z}^T\mathbf{A}\mathbf{Z})^{-1}\mathbf{Z}^T \left(\mathbf{A}\mathbf{x}_1 c_1 + \dots + \mathbf{A}\mathbf{x}_m c_m \right) \\
&= \mathbf{Z}(\mathbf{Z}^T\mathbf{A}\mathbf{Z})^{-1}\mathbf{Z}^T (\mathbf{A}\mathbf{Z}\mathbf{c}) \\
&= \mathbf{Z}(\mathbf{Z}^T\mathbf{A}\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{A}\mathbf{Z})\mathbf{c} \\
&= \mathbf{Z}\mathbf{c} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + c_3 \mathbf{x}_3 + c_4 \mathbf{x}_4 + c_5 \mathbf{x}_5 \\
&= \sum_{i=1}^{m} c_i \mathbf{x}_i = \mathbf{x},
\end{aligned}$$

which is the solution to the original system.

For the second term of Equation (15), $\mathbf{P}^T\hat{\mathbf{x}}$, we compute $\hat{\mathbf{x}}$ from Equation (16):

$$\mathbf{P}\mathbf{A}\hat{\mathbf{x}} = \mathbf{P}\mathbf{b}$$
$$\mathbf{A}\mathbf{P}^T\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A}\mathbf{Q})\mathbf{b} \qquad \text{using 1.5 f) and definition of } \mathbf{P},$$
$$\mathbf{A}\mathbf{P}^T\hat{\mathbf{x}} = \mathbf{b} - \mathbf{A}\mathbf{Q}\mathbf{b}$$
$$\mathbf{A}\mathbf{P}^T\hat{\mathbf{x}} = \mathbf{b} - \mathbf{A}\mathbf{x} = 0 \qquad \text{taking } \mathbf{Q}\mathbf{b} = \mathbf{x} \text{ from above,}$$
$$\mathbf{P}^T\hat{\mathbf{x}} = 0 \qquad \text{as } \mathbf{A} \text{ is invertible.}$$

Then we have achieve the solution $\mathbf{x}$ in one step of DICCG.

1.8.1   Accuracy of the snapshots.

If we use an iterative method to obtain an approximate solution $\mathbf{x}^k$ for the system $\mathbf{A}\mathbf{x} = \mathbf{b}$, we cannot compute the relative error of the approximation with respect to the true solution because we do not know the true solution,

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2}.$$

Instead, we compute the relative residual,

$$\frac{||\mathbf{r}^k||_2}{||\mathbf{b}||_2} \leq \epsilon,$$

and we set a stopping criterium ϵ or tolerance, that is related to the relative error as follows (see Appendix B),

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2} \leq \kappa_2(\mathbf{A})\epsilon = \varepsilon,$$

where $\kappa_2(\mathbf{A})$ is the condition number of the matrix $\mathbf{A}$

Diverse tolerance values can be used in the experiments for the snapshots as well as for the solution of the original system.

If the maximum residual for the snapshots is $\epsilon = 10^{-\eta}$ then the error of the snapshots is given by

$$\frac{||\mathbf{x}_i - \mathbf{x}_i^k||_2}{||\mathbf{x}_i||_2} \leq \kappa_2(\mathbf{A}) \times 10^{-\eta} = \varepsilon.$$

If we use m snapshots obtained with an iterative method to compute the solution of $\mathbf{x}$, after one iteration of DCG we obtain

$$\hat{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i^{k(i)}.$$

The error of this solution is given by:

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2} = \frac{||\sum_{i=1}^{m} c_i(\mathbf{x}_i - \mathbf{x}_i^k)||_2}{||\sum_{i=1}^{m} c_i \mathbf{x}_i||_2} \leq \frac{\sum_{i=1}^{m} |c_i| \times \kappa_2(\mathbf{A}) \times 10^{-\eta}}{||\sum_{i=1}^{m} c_i \mathbf{x}_i||_2}.$$

11

Which means that the approximation has an error of $\mathcal{O}(\kappa_2(\mathbf{A}) \times 10^{-\eta})$.

From Lemma 2 we know that if we use the snapshots $\mathbf{x}_i$ as deflation vectors, for the deflation method

$$\mathbf{Qb} = \mathbf{x}.$$

If the approximation has an error $\mathcal{O}(\kappa_2(\mathbf{A}) \times 10^{-\eta})$, then the solution achieved with deflation will have the same error,

$$\mathbf{Qb} = \mathbf{x}^k + \mathcal{O}(\kappa_2(\mathbf{A}) \times 10^{-\eta}).$$

Therefore, is important to take into account the condition number of the matrix related to the system and the accuracy of the deflation vectors.

1.8.2 Boundary conditions.

From Lemma 2, we know that if we use as deflation vectors a set of m snapshots

$$\mathbf{Z} = [\mathbf{x}_1 \quad ... \quad \mathbf{x}_m],$$

such that $\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{x}_i$, where $\mathbf{x}$ is the solution of the system $\mathbf{Ax} = \mathbf{b}$, the solution of the latter system is achieved within one iteration of DICCG.

In our application, only a small number (m) of elements of the right hand side vector ($\mathbf{b}$) are changed for various situations. This implies that every $\mathbf{b}$ can be written as $\mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{b}_i$. Using Lemma 1, this implies that $\mathbf{x}$ is such that $\mathbf{x} \in span\{\mathbf{x}_1, ..., \mathbf{x}_m\}$, which is called the solution span. Therefore, it is necessary to find the solution span of the system, such that the sum of the elements in the solution span and the sum of right hand sides give as result the original system. In this section we explore the subsystems that should be chosen, depending on the boundary conditions of the original system.

Neumann Boundary conditions

When we have Neumann boundary conditions everywhere, the resulting matrix $\mathbf{A}$ is singular, and $\mathbf{A}[1 \quad 1 \quad ... \quad 1 \quad 1]^T = \mathbf{0}$, $Ker(\mathbf{A}) = span([1 \quad 1 \quad ... \quad 1 \quad 1]^T)$. Note that $\mathbf{Ax} = \mathbf{b}$ has only a solution if $\mathbf{b} \in span\{\mathbf{a}_1, ..., \mathbf{a}_n\}$ (with $\mathbf{a}_i$ the $i-th$ column of $\mathbf{A}$), which is equivalent to $\mathbf{b} \perp Ker(\mathbf{A})$ [10]. This implies that if we have m sources with value s_i for the vector $\mathbf{b}_i$, we need that

$$\sum_{j=1}^{m} s_i^j = 0.$$

Then, for each nonzero right hand side we need to have at least two sources. Therefore, we can have at most $m - 1$ linearly independent right hand sides $\mathbf{b}_i$ containing two sources. This means that the solution space has dimension $m - 1$ and it can be spanned by $span\{\mathbf{x}_1, ..., \mathbf{x}_{m-1}\}$. Each of these subsystems will have the same no-flux conditions (Neumann) in all the boundaries. As the original system is a linear combination of the subsystems (Lemma 1), the deflation vectors can be chosen as the solutions corresponding to the subsystems. Therefore, the deflation matrix will be given by:

$$\mathbf{Z} = [\mathbf{x}_1 \quad ... \quad \mathbf{x}_{m-1}],$$

and if the accuracy of the snapshots used as deflation vectors is good enough (see Section 1.8.1), the solution is expected to be achieved within one iteration.

Dirichlet Boundary conditions

In this case, the right hand side of the system can contain the values of the boundary $\mathbf{b}_b$ and the sources of the system $\mathbf{s}_i$. If we have m sources, as in the previous case, the right hand side will be given by:

$$\mathbf{b} = \sum_{i=1}^{m} c_i \mathbf{s}_i + \mathbf{b}_b.$$

The subsystems will be $m+1$, where one of them corresponds to the boundary conditions $\mathbf{A}\mathbf{x}_b = \mathbf{b}_b$, and the other m will correspond to the sources $\mathbf{A}\mathbf{x}_i = \mathbf{s}_i$. Therefore, one of the snapshots will be the system with no sources and the Dirichlet boundary conditions of the original system. The other m snapshots will correspond to the m sources with homogeneous Dirichlet boundary conditions. Then, the solution space will be given by $span\{\mathbf{x}_1, ..., \mathbf{x}_m, \mathbf{x}_b\}$. If we use the solution of the m snapshots as deflation vectors, with the correct accuracy, we will obtain the solution within one iteration.

1.9 Proper Orthogonal Decomposition (POD)

The Proper Orthogonal Decomposition (POD) method is a Model Order Reduction (MOR) method, where a high-order model is projected onto the space spanned by a set of orthonormal basis vectors. The high dimensional variable $\mathbf{x} \in \mathbb{R}^n$ is approximated by a linear combination of $l << n$ orthonormal basis vectors [1]:

$$\mathbf{x} \approx \sum_{i=1}^{l} z_i \phi_i, \tag{27}$$

where $\phi_i \in \mathbb{R}^n$ are the basis vectors and z_i are their corresponding coefficients. In matrix notation, equation (27) is rewritten as :

$$\mathbf{x} \approx \Phi\mathbf{z},$$

where $\Phi = [\phi_1 \ \phi_2 \ .. \ \phi_l]$, $\Phi \in \mathbb{R}^{n \times l}$ is the matrix containing the basis vectors, and $\mathbf{z} \in \mathbb{R}^l$ is the vector containing the coefficients of the basis vectors.

The basis vectors are computed from a set of 'snapshots' $\{\mathbf{x_i}\}_{i \in \mathbb{N}}$, obtained by simulation or experiments [2]. In POD, the basis vectors $\{\phi_j\}_{j=1}^{l}$, are l eigenvectors corresponding to the largest eigenvalues $\{\lambda_j\}_{j=1}^{l}$ of the data snapshot correlation matrix $\mathbf{R}$.

$$\mathbf{R} := \frac{1}{m}\mathbf{X}\mathbf{X}^T \equiv \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i\mathbf{x}_i^T, \qquad \mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_m], \tag{28}$$

where $\mathbf{X} \in \mathbb{R}^{n \times m}$ is an SPSD matrix containing the previously obtained snapshots. The l eigenvectors should contain almost all the variability of the snapshots. Usually, they are chosen as the eigenvectors of the maximal number (l) of eigenvalues satisfying [2]:

$$\frac{\sum_{j=1}^{l} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \leq \alpha, \qquad 0 < \alpha \leq 1, \tag{29}$$

with α close to 1. The eigenvalues are ordered from large to small with λ_1 the largest eigenvalue of $\mathbf{R}$. It is not necessary to compute the eigenvalues from $\mathbf{XX}^T$, instead, it is possible to compute the eigenvalues of the much smaller matrix $\mathbf{X}^T\mathbf{X}$ (see Appendix C).

2 Flow through porous media

To describe single-phase flow through a porous medium, we use the continuity equations:

$$\frac{\partial(\rho\phi)}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = q, \qquad \mathbf{v} = -\frac{\mathbf{K}}{\mu}(\nabla\mathbf{p} - \rho g \nabla z), \tag{30}$$

or

$$\frac{\partial(\rho\phi)}{\partial t} - \nabla \cdot \left(\frac{\rho\mathbf{K}}{\mu}(\nabla\mathbf{p} - \rho g \nabla z) \right) = q. \tag{31}$$

Where the primary unknown is the pressure $\mathbf{p}$. The fluid $\rho = \rho(\mathbf{p})$ and rock $\phi = \phi(\mathbf{p})$ compressibilities can be pressure dependent. Rock compressibility is defined by:

$$c_r = \frac{1}{\phi}\frac{d\phi}{dp} = \frac{ln(\phi)}{dp},$$

If the rock compressibility is constant, the previous equation can be integrated as:

$$\phi(p) = \phi_0 e^{c_r(p-p_0)}.$$

Fluid compressibility is defined as:

$$c_f = \frac{1}{\rho}\frac{d\rho}{dp} = \frac{ln(\rho)}{dp}. \tag{32}$$

If the fluid compressibility is constant, the previous equation can be integrated as:

$$\rho(\mathbf{p}) = \rho_0 e^{c_f(\mathbf{p}-\mathbf{p}_0)}. \tag{33}$$

Using implicit discretization, Equation (30) becomes:

$$\frac{(\phi\rho)^{n+1} - (\phi\rho)^n}{\Delta t^n} + \nabla \cdot (\rho\mathbf{v})^{n+1} = \mathbf{q}^n, \qquad \mathbf{v}^{n+1} = -\frac{\mathbf{K}}{\mu^{n+1}}[\nabla(\mathbf{p}^{n+1}) - g\rho^{n+1}\nabla(\mathbf{z})]. \tag{34}$$

If ϕ and ρ depend nonlinearly on $\mathbf{p}$ we have a nonlinear system of equations to be solved for each time step.

Assuming no gravity terms, Equation (34) can be rewritten as:

$$\frac{\phi(\mathbf{p}^{n+1})\rho(\mathbf{p}^{n+1}) - \phi(\mathbf{p}^n)\rho(\mathbf{p}^n)}{\Delta t^n} + \nabla \cdot (\rho \mathbf{v})^{n+1} = \mathbf{q}^n, \qquad \mathbf{v}^{n+1} = -\frac{\mathbf{K}}{\mu^{n+1}}\nabla(\mathbf{p}^{n+1}). \quad (35)$$

Or:

$$\frac{\phi(\mathbf{p}^{n+1})\rho(\mathbf{p}^{n+1}) - \phi(\mathbf{p}^n)\rho(\mathbf{p}^n)}{\Delta t^n} - \nabla \cdot (\rho(\mathbf{p}^{n+1})\frac{\mathbf{K}}{\mu^{n+1}}\nabla(\mathbf{p}^{n+1})) - \mathbf{q}^n = 0. \quad (36)$$

The latter system can be written in short vector form as:

$$\mathbf{F}(\mathbf{p}^{n+1}; \mathbf{p}^n) = 0, \quad (37)$$

with $\mathbf{p}^n$ the vector of unknown state variables at time step n.

This non-linear system can be solved by Newton Rhapson (NR) method, the $(i+1)$-th iteration approximation is obtained from:

$$\frac{\partial \mathbf{F}(\mathbf{x}^i)}{\partial \mathbf{x}^i}\delta \mathbf{x}^i = -\mathbf{F}(\mathbf{x}^i), \qquad \mathbf{x}^{i+1} = \mathbf{x}^i + \delta \mathbf{x}^{i+1},$$

where $\mathbf{J}(\mathbf{x}^i) = \frac{\partial \mathbf{F}(\mathbf{x}^i)}{\partial \mathbf{x}^i}$ is the Jacobian matrix, and $\delta \mathbf{x}^{i+1}$ is the NR update at iteration step $i+1$.

Well model

In reservoirs, wells are typically drilled to extract or inject fluids. Fluids are injected into a well at constant surface rate or constant bottom-hole pressure, and are produced at constant bottom-hole pressure or a constant surface rate.

A widely used model is Peacemans's model, that takes into account the above-mentioned pressure drop. This model is a linear relationship between the bottom-hole pressure and the surface flow rate in a well:

$$q_0 = J(p_R - p_{bhp}),$$

where J is the productivity or injectivity index, $\mathbf{p}_R$ is the reservoir pressure in the cell where the well is located, and $\mathbf{p}_{bhp}$ is the pressure inside the well. We use this well model in this work.

3 Numerical experiments.

We study the solution of systems of linear equations resulting from the discretization of elliptic and parabolic partial differential equations for the description of single-phase flow through a porous media. The solution of the system is performed with the Deflated Conjugate Gradient method preconditioned with Incomplete Cholesky (DICCG). We propose the use of snapshots and the snapshots-based basis functions of POD as deflation vectors

for the above-mentioned method.

In the present section, we give a general overview of the experiments that we perform, but the specifications are presented below for each problem separately. In the first part we solve the elliptic problem (incompressible flow) and the second is devoted to the parabolic problem (compressible flow). For the elliptic problem, a good choice of snapshots, and therefore of deflation vectors, depends on the boundary conditions of the problem. Hence, we study two cases with different boundary conditions. For the first set of elliptic problems, Dirichlet boundary conditions are used for an academic layered model with various contrasts in permeability between the layers. In the second set of elliptic problems, we used Neumann boundary conditions (no-flux) for the previous academic layered problem and for the SPE 10 benchmark problem. We investigate the behavior of the ICCG and DICCG methods with various contrasts between the permeability layers for both cases.

We study the influence of the size of the problem in the performance of the ICCG and DICCG methods changing the grid size of the SPE 10 benchmark, we study with diverse grid sizes of the 2nd layer, and the complete benchmark (85 layers).

For the compressible problem, we solve the problem with Neumann boundary conditions in all the boundaries. We study a layered permeability problem.

The model

The experiments simulate flow through a porous medium with a constant porosity field of 0.2. We model incompressible and compressible single-phase flow. For the incompressible single-phase model the following properties of the fluid are used:

- $\mu = 1cp$

- $\rho = 1014kg/m^3$

In the compressible case, the compressibility of the fluid is:

- $c = 1 \times 10^{-3}$

In these experiments, a Cartesian grid with different grid sizes is used. Wells or sources are added to the system. The matrices corresponding to the linear systems **A** and right-hand sides **b** are obtained with MRST [11].

Snapshots

As mentioned above, for the DICCG method we need a set of deflation vectors. In a first series of experiments (incompressible model), the deflation vectors are solutions of the system with various wells configurations and boundary conditions. These solutions, called snapshots, are obtained with ICCG, the tolerance of the linear solvers is given for each problem. The configuration used to obtain each snapshot depends on the problem that we are solving (see section 1.8). For the compressible problem, the snapshots are the solutions at the first time steps, first with the same well configuration, and then with different wells configurations. Solutions of the same problem with zero compressibility are also used as snapshots. For each case, the configuration of the snapshots, as well as the configuration

of the solved system are presented.

The solver

The solution of the system is approximated with ICCG and DICCG.

For the DICCG method, we need a set of deflation vectors. In a first set of experiments, we use a linear independent set of solutions as deflation vectors. Then, we use as deflation vectors a linearly dependent set of solutions, and finally the deflation vectors are a linearly independent basis of the latter dependent set obtained with POD. As tolerance or stopping criterium we use the relative residual, defined as the 2-norm of the residual of the kth iteration divided by the 2-norm of the right-hand side of the preconditioned system:

$$\frac{||\mathbf{M}^{-1}r^k||_2}{||\mathbf{M}^{-1}b||_2} \leq \epsilon.$$

The stopping criterium is varied for each problem.

Incompressible Problem

Case 1, Dirichlet and Neumann boundary conditions.

In the configuration of *Case 1*, four wells are positioned in a square at distances equal to one-third of the reservoir length and width. Two wells have a bottom hole pressure (bhp) of 5 bars and two have a bhp of -5 bar. No-flux conditions are used at the right and left boundaries and a pressure drop in the vertical direction. The pressure at the lower boundary $(y = 1)$ is 0 bars, and at the upper boundary $(y = ny)$ is 3 bars. The first four snapshots $(z_1 - z_4)$ are obtained setting only one well pressure different from zero, taking no-flux conditions at the right and left boundaries and homogeneous Dirichlet conditions at the other boundaries. A fifth snapshot is obtained setting all the wells pressures to zero and setting the pressure drop in the vertical direction of the original system. A summary is presented below.

Configuration 1:

System configuration

W1 = W2 = -5 bars.

W3 = W4 = +5 bars.

Boundary conditions :

$P(y = 1) = 0$ bars,

$P(y = ny) = 3$ bars,

$\frac{\partial P(x=1)}{\partial n} = \frac{\partial P(x=nx)}{\partial n} = 0.$

Snapshots

$\mathbf{z}_1$: W1 = -5 bars, W2 = W3 = W4 = 0,

$\mathbf{z}_2$: W2 = -5 bars, W1 = W3 = W4 = 0.

$\mathbf{z}_3$: W3 = +5 bars, W1 = W2 = W4 = 0.

$\mathbf{z}_4$: W4 = +5 bars, W1 = W2 = W3 = 0.

$P(y = 1) = 0$ bars,

$P(y = ny) = 0$ bars,

$\frac{\partial P(x=1)}{\partial n} = \frac{\partial P(x=nx)}{\partial n} = 0.$

$\mathbf{z}_5$: W1 = W2 = W3 = W4 = 0.

$P(y = 1) = 0$ bars,

$P(y = ny) = 3$ bars,

$\frac{\partial P(x=1)}{\partial n} = \frac{\partial P(x=nx)}{\partial n} = 0.$

As mentioned above, we studied flow through a porous medium with *heterogeneous permeability* layers. A grid of $nx = ny = 64$ elements is studied. We use 8 layers of the same size, 4 layers with one value of permeability σ_1, followed by a layer with a different permeability value σ_2. Figure 1 shows these layers. The permeability of one set of layers is set to $\sigma_1 = 1mD$, the permeability of the other set σ_2 is changed. Therefore, the contrast in permeability between the layers $(\frac{\sigma_2}{\sigma_1} = \frac{\sigma_2}{1mD})$, depends on the value of σ_2.

Figure 1: Heterogeneous permeability, 4 wells.
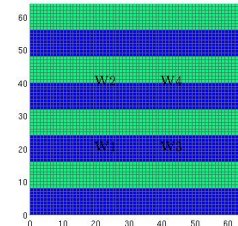
We investigate the dependence on the contrast in permeability value between the layers

for the ICCG and DICCG methods. The permeability σ_2 varies from $\sigma_2 = 10^{-1}mD$ to $\sigma_2 = 10^{-3}mD$. The tolerance is set as 10^{-11} for the snapshots as well as for the original problem.

κ_2 (mD)	10^{-1}	10^{-2}	10^{-3}
ICCG	75	103	110
DICCG	1	1	1

Table 1: Table with the number of iterations for different contrasts in the permeability of the layers for the ICCG and DICCG methods.

Table **??** shows the number of iterations required to achieve convergence for ICCG and DICCG, for various permeability contrasts between the layers.
The plot of the residual and the solution to the problem are presented in Figures 2 and 3 for a value of permeability $\sigma_2 = 10^{-2}$.
In Table **??** we observe that the number of iterations increases when the contrast between the permeability layers increases for ICCG. For DICCG, we observe that we only need one iteration despite the change in permeability contrast between the layers.
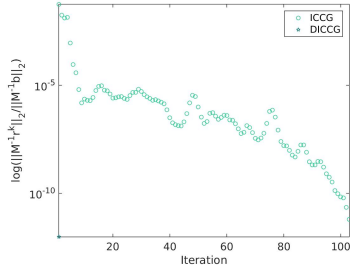
Figure 2: Convergence for the heterogeneous problem, 64 x 64 grid cells, $\sigma_2 = 10^{-2}mD$.
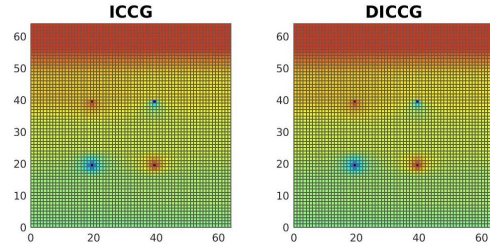
Figure 3: Solution of the heterogeneous problem, 64 x 64 grid cells, $\sigma_2 = 10^{-2}mD$.

Case 2, Neumann boundary conditions only.
In this case, four wells are positioned in the corners and have a bhp of -1 bar. One well is positioned in the center of the domain and has a bhp of +4 bars (see Figure 4). We set Neumann boundary conditions in all boundaries. For this case, we use a set of four linearly independent snapshots as deflation vectors. We also use a linearly dependent set of 15 snapshots and the basis of POD (linearly independent set) obtained from the 15 snapshots. We set the same boundary conditions as in the original problem for all the snapshots. The four linearly independent snapshots $(z_1 - z_4)$ are obtained giving a value of zero to one well and non zero values to the other wells. The set of 15 snapshots are all the possible combinations of wells that satisfy that the flow in equals the flow out the reservoir. A summary of the configurations is presented below.

Configuration 2 :

System configuration

W1 = W2 = W3 = W4 = -1 bar.

W5 = +4 bars.

Boundary conditions :

$$\frac{\partial P(y=1)}{\partial n} = \frac{\partial P(y=ny)}{\partial n} = \frac{\partial P(x=1)}{\partial n} = \frac{\partial P(x=nx)}{\partial n} = 0.$$

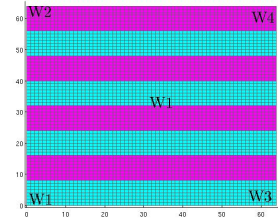*Snapshots* (4 linearly independent)

$\mathbf{z}_1$: W2 = W3 = W4 = -1 bars, W1 = 0 bars, W5 = +3 bars.

$\mathbf{z}_2$: W1 = W3 = W4 = -1 bars, W2 = 0 bars, W5 = +3 bars.

$\mathbf{z}_3$: W1 = W2 = W4 = -1 bars, W3 = 0 bars, W5 = +3 bars.

$\mathbf{z}_4$: W1 = W2 = W3 = -1 bars, W4 = 0 bars, W5 = +3 bars.

Dependent snapshots

$\mathbf{z}_5$: W1 = W2 = W3 = W4 = -1 bars, W5 = +4 bars.

$\mathbf{z}_6$: W1 = W2 = 0 bars, W3 = W4 = -1 bars, W5 = +2 bars.

$\mathbf{z}_7$: W2 = W3 = 0 bars, W1 = W4 = -1 bars, W5 = +2 bars.

$\mathbf{z}_8$: W3 = W4 = 0 bars, W1 = W2 = -1 bars, W5 = +2 bars.

$\mathbf{z}_9$: W1 = W3 = 0 bars, W2 = W4 = -1 bars, W5 = +2 bars.

$\mathbf{z}_{10}$: W2 = W4 = 0 bars, W1 = W3 = -1 bars, W5 = +2 bars.

$\mathbf{z}_{11}$: W1 = W4 = 0 bars, W2 = W3 = -1 bars, W5 = +2 bars.

$\mathbf{z}_{12}$: W1 = -1 bars, W2 = W3 = W4 = 0 bars, W5 = +1 bars.

$\mathbf{z}_{13}$: W2 = -1 bars, W1 = W3 = W4 = 0 bars, W5 = +1 bars.

$\mathbf{z}_{14}$: W3 = -1 bars, W1 = W3 = W4 = 0 bars, W5 = +1 bars.

$\mathbf{z}_{15}$: W4 = -1 bars, W1 = W2 = W3 = 0 bars, W5 = +1 bars.

Heterogeneous permeability layers

As in the previous case, single-phase flow through a porous medium with heterogeneous permeability layers is studied. A grid of $nx = ny = 64$ elements is investigated. The deflation vectors used in this case are the 4 snapshots ($\mathbf{z}_1$-$\mathbf{z}_4$), a set of 15 linearly dependent vectors and 4 basis vectors obtained for the POD method from the latter set.

The snapshots and the solutions are obtained with a tolerance of 10^{-11}.

Figure 4: Heterogeneous permeability, 5 wells.

Table **??** shows the number of iterations required to reach convergence for ICCG method and the deflation method with four linearly independent snapshots as deflation vectors DICCG$_4$, 15 linearly dependent snapshots DICCG$_{15}$ and the basis
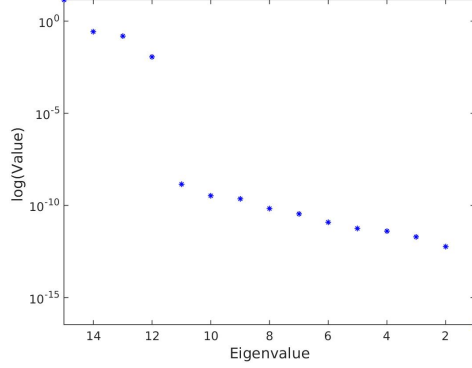
Figure 5: Eigenvalues of the snapshot correlation matrix $\mathbf{R} = \mathbf{XX}^T$, 15 snapshots used.

vectors of POD, DICCG$_{POD}$[7]. For the selection of the deflation vectors of DICCG$_{POD}$ we plot the eigenvalues of the snapshot correlation matrix $\mathbf{R} = \mathbf{X}^T\mathbf{X}$ (see section 1.9) in Figure 5. We observe that there are 4 eigenvalues larger than the rest of the eigenvalues which are the responsible of hamper the convergence of the method, therefore we use the eigenvectors corresponding to these eigenvalues as deflation vectors.

The plot of the residual and the solution of the problem are presented in Figure 6 and 7 for the ICCG and DICCG methods for the case of $\sigma_2 = 10^{-1}$.

σ_2 (mD)	10^{-1}	10^{-2}	10^{-3}
ICCG	90	115	131
DICCG$_4$	1	1	1
DICCG$_{15}$	200*	200*	200*
DICCG$_{POD}$	1	1	1

Table 2: Table with the number of iterations for different contrast in the permeability of the layers for the ICCG, DICCG$_4$, DICCG$_{15}$, and DICCG$_{POD}$ methods, tolerance of solvers and snapshots 10^{-11}.

In Table **??**, for the ICCG method, we observe that the number of iterations increases as the contrast in the permeability increases. For the DICCG method with 4 linearly independent deflation vectors and 4 basis vectors of POD, convergence is reached within one iteration. However, for the case of 15 linearly dependent vectors, the solution is not reached within the 200 iterations allowed for this problem.

---

[7]The * means that the solution is not reached.
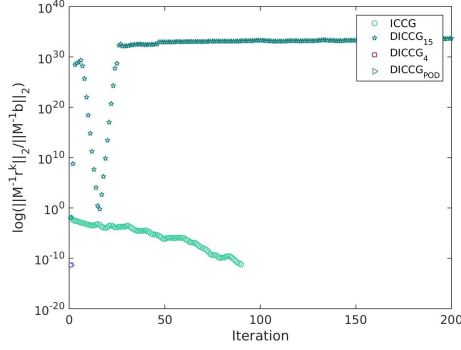
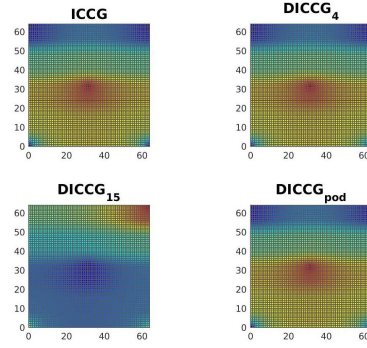Figure 6: Convergence for the heterogeneous problem, 64 x 64 grid cells, $\sigma_2 = 10^{-1}$.

Figure 7: Solution of the heterogeneous problem, 64 x 64 grid cells, $\sigma_2 = 10^{-1}$.

SPE 10 model

This model has large variations in the permeability coefficients, $\mathcal{O}(10^7)$. It has 5 sources or wells, four producers in the corners (negative) and one injector in the center (positive). The model contains 60 x 220 x 85 cells, we study the dependence of the ICCG and the DICCG method on the size of the problem. One layer is studied with various grid sizes 16 x 56, 30 x 110, 46 x 166 and 60 x 220, and the complete model containing 85 layers. Permeability is upscaled averaging the permeability in each grid using the harmonic-arithmetic average algorithm from MRST. The permeability of the coarser grid (16 x 56 cells) is shown in Figure 8 and the complete model in Figure 9. The permeability contrast for the diverse grid size problems is shown in Table 3. From this table, we observe that the contrast in the permeability for different grid sizes varies slightly, but that the order of magnitude remains the same for all the cases.

The snapshots are obtained solving the system with different well configurations (*Configuration 2*). As before, we simulate single-phase incompressible flow.
The system and snapshots are solved with an accuracy of 10^{-7}. In the first experiment with the deflation method, the four linearly independent snapshots are used as deflation vectors (DICCG). Then, 15 linearly dependent vectors and finally 4 vectors of the POD basis are used as deflation vectors (DICCG$_{POD}$).

Grid size	16x56x1	30x110x1	46x166x1	60x220x1	60x220x85
Contrast ($\times 10^7$)	1.04	2.52	2.6	2.8	3

Table 3: Table with the contrast of permeabilities for different grid sizes.

The number of iterations required to achieve convergence with the ICCG and DICCG methods for various grid sizes is presented in Table 4.
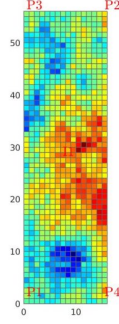
22

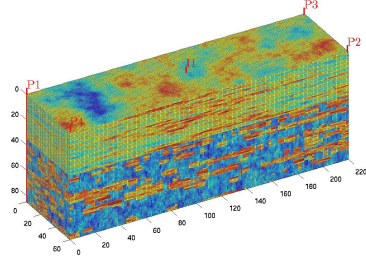Figure 8: SPE 10 benchmark, 2nd layer 16 x 56 grid cells, permeability field.

Figure 9: SPE 10 benchmark, permeability field.

The convergence and the solution to the ICCG and DICCG methods are presented in Figure 10 and Figure 11 for the complete problem. In Table 3 we observe that for the ICCG method the required iterations to reach convergence increases as the size of the grid increases. Meanwhile, for the deflated methods only one iteration is required and it does not change with the size of the grid. We can also note that for the deflated method with 15 linearly dependent snapshots as deflation vectors (DICCG$_{15}$), after the first iteration, the relative residual is close to two orders of magnitude larger than for the other deflation methods (DICCG$_4$, DICCG$_{POD}$). Therefore, more iterations are required to reach the desired tolerance.

Method	16x56x1	30x110x1	46x166x1	60x220x1	60x220x85
ICCG	34	74	128	162	357
DICCG$_{15}$	1	1	1	1	3
DICCG$_4$	1	1	1	1	1
DICCG$_{POD}$	1	1	1	1	1

Table 4: Table with the number of iterations for ICCG and DICCG methods, various tolerance for the snapshots, various grid sizes.
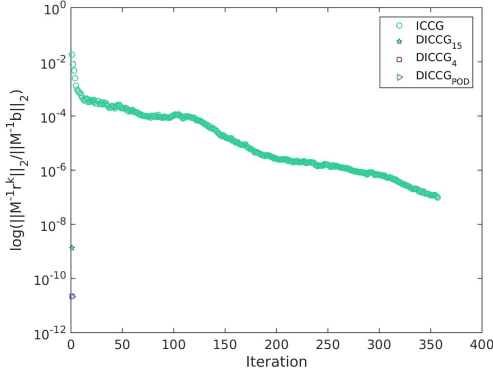
Figure 10: Convergence for the SPE 10 problem, 60 x 220 x 85 grid cells, accuracy of the snapshots 10^{-7}.
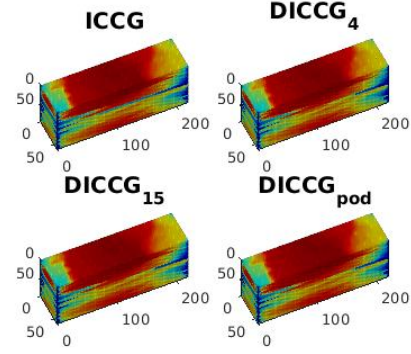

Figure 11: Solution of the SPE 10 benchmark, 16 x 56 grid cells, 2nd layer, accuracy of the snapshots 10^{-7}.

Compressible Problem

A grid of 35×35 cells is studied with Neumann boundary conditions only. The domain consists of layers with two different permeability values (see Figure 12). The first layer has a permeability of $\sigma_1 = 30mD$, and the permeability of the second layer is $\sigma_2 = 3mD$. Therefore the permeability contrast between the layers is 10^{-1}. The initial pressure of the reservoir is set as 200 bars. Four wells are positioned in the corners of the domain with a bhp of 100 bars, and a well is placed in the center of the domain with a bhp of 600 bars. For the first set of problems, we use the updated solution of the first 10 time steps with the same configuration as the problem as deflation vectors. We solve the rest of the time

Figure 12: Heterogeneous permeability, 5 wells, compressible problem.

steps with DICCG_{10} method with the 10 snapshots as deflation vectors and with 5 basis POD vectors as deflation vectors, DICCG_{POD}. For the second set of experiments, we use the same configuration as in the original problem but we vary the configuration of the wells. One corner well has the same pressure as the reservoir (200 bar), the other corner wells have a pressure of 100 bars and the central well has a pressure of 500 bars. A summary is presented below.
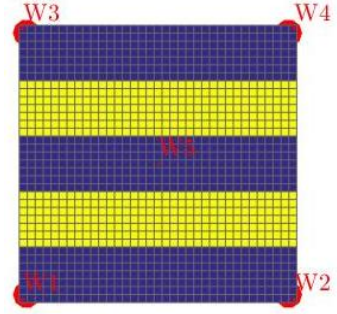
System configuration

Initial pressure 200 bar.

W1 = W2 = W3 = W4 = 100 bar.

W5 = 600 bars.

Snapshots (second set of experiments)

$\mathbf{z}_1$: W2 = W3 = W4 = 100 bars, W1 = 200 bars, W5 = 500 bars.

$\mathbf{z}_2$: W1 = W3 = W4 = 100 bars, W2 = 200 bars, W5 = 500 bars.

$\mathbf{z}_3$: W1 = W2 = W4 = 100 bars, W3 = 200 bars, W5 = 500 bars.

$\mathbf{z}_4$: W1 = W2 = W3 = 100 bars, W4 = 200 bars, W5 = 500 bars.

Boundary conditions :

$$\frac{\partial P(y=1)}{\partial n} = \frac{\partial P(y=ny)}{\partial n} = \frac{\partial P(x=1)}{\partial n} = \frac{\partial P(x=nx)}{\partial n} = 0.$$

The simulation was performed during 152 days with 52 time steps and a time step of 3 days. The tolerance of the NR method and the linear solvers is 10^{-5}.

Case 1

In Figure 13, the solution obtained with the ICCG method is presented, the solution is the same for all methods. The upper left figure represents the pressure field at the final time step. The upper right figure represents the pressure across the diagonal joining the (0,0) and (35,35) grid cells for all the timesteps. We observe the initial pressure (200 bars) in this diagonal and the evolution of the pressure field through time. In the lower figure we observe the surface volume rate for the five wells during the simulation.
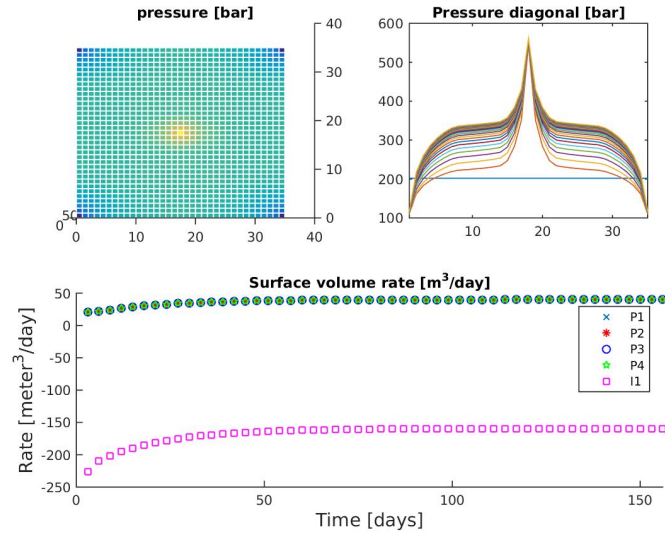

Figure 13: Solution of the compressible problem solved with the ICCG method.

25

The number of iterations necessary to reach convergence with the linear solvers is presented for the first four NR iterations in Figure 15 for the ICCG method, Figure 17 for the deflated method DICCG$_{10}$ using 10 snapshots as deflation vectors and Figure 19 DICCG$_{POD}$ using the first 5 basis vectors of POD as deflation vectors. The eigenvalues of the matrices are presented in Figure 14 for the original system matrix $\mathbf{J}$ for the first timestep, Figure 16 for the preconditioned system, Figure 18 the deflated system $DICCG_{10}$ and Figure 20 the deflated system combined with POD $DICCG_{POD}$ using the eigenvectors corresponding to the first five largest eigenvalues. The eigenvalues of the snapshot correlation matrix $\mathbf{R} = \frac{1}{m}\mathbf{ZZ}^{T}$ are presented in Figure ??.
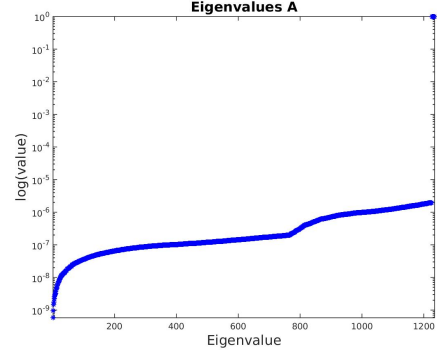
Figure 14: Eigenvalues of the original matrix $\mathbf{J}$, time step 1.
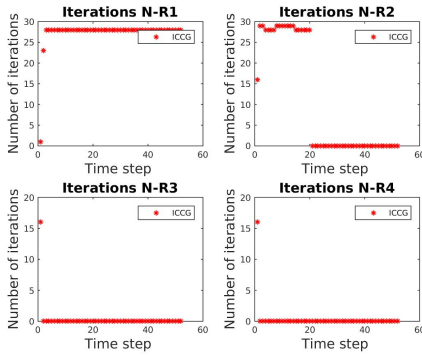

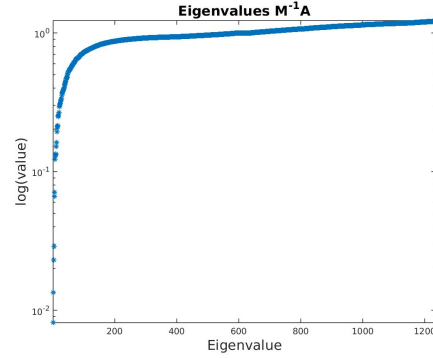Figure 15: Number of iterations of the ICCG method for the first four NR iterations.

Figure 16: Eigenvalues of the preconditioned matrix, time step 1.
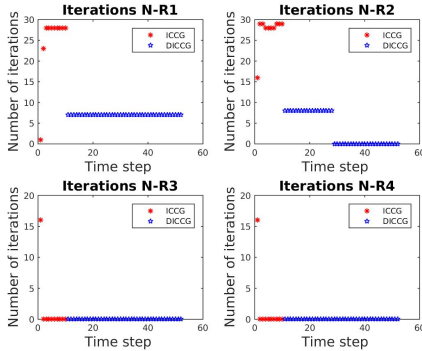

Figure 17: Number of iterations of the DICCG$_{10}$ method for the first four NR iterations.
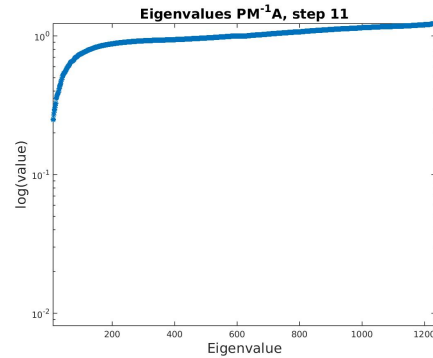

Figure 18: Eigenvalues of the deflated system DICCG$_{10}$ with 10 deflation vectors.
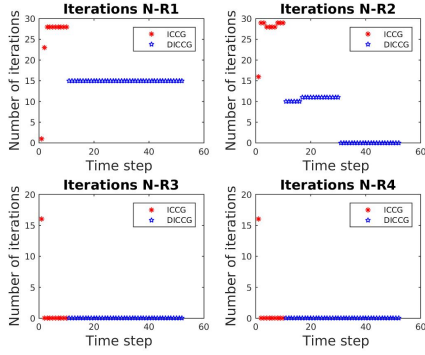
26

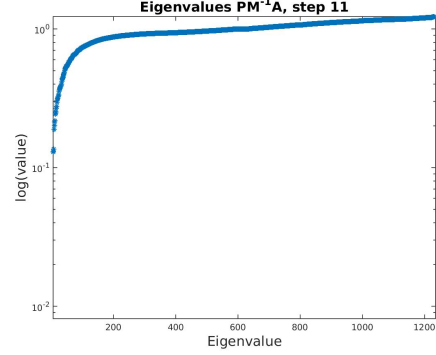Figure 19: Number of iterations of the DICCG$_{POD}$ method for the first four NR iterations, eigenvectors [6-10].

Figure 20: Eigenvalues of the deflated system DICCG$_{POD}$ with 10 deflation vectors, eigenvectors [6-10].

We observe that the number of iterations for the first and second NR iterations is lower for the deflated methods compared with the ICCG method. However, we observe that the time step when convergence is achieved for the NR cycle is larger for these methods with respect to the ICCG method. We also observe that for the first NR iteration, the decrease is larger for the deflated method with 10 snapshots as deflation vectors.

Case 2

In the second case, we compute the first 4 time steps varying the bhp of the wells, as described above. Figure 21 presents the solution obtained with ICCG method is presented. The upper left figure represents the pressure field at the final time step. The upper right figure represents the pressure across the diagonal joining the (0,0) and (35,35) grid cells. We can observe the initial pressure (200 bars) in this diagonal and the evolution of the pressure field through the time. In the lower figure, we observe the surface volume rate for the five wells during the simulation.
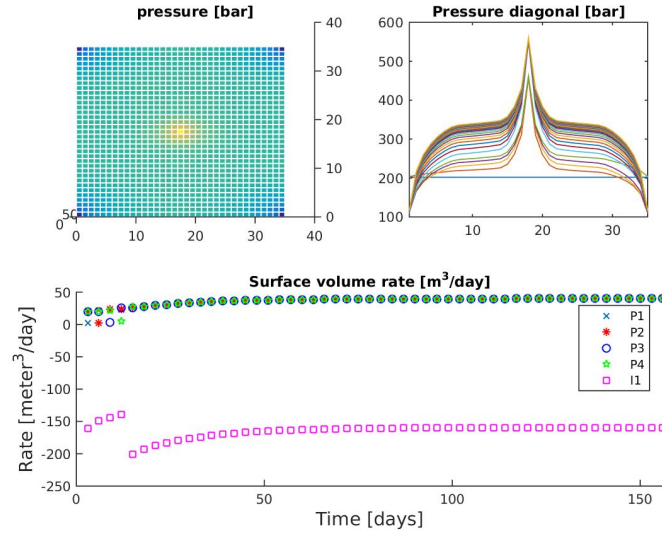

Figure 21: Solution of the compressible problem solved with the ICCG method, varying the bhp of the first 4 time steps.

The number of iterations necessary to achieve convergence with the linear solvers for this second problem is presented for the first four NR iterations in Figure 22 for the ICCG method, Figure 23 for the deflated method $DICCG_{10}$ using 10 snapshots as deflation vectors and Figure 24 $DICCG_{POD}$ using 5 basis vectors of POD as deflation vectors.

As in the previous case, we observe that the number of iterations for the first and second NR iterations is lower for the deflated methods compared with the ICCG method. However, we observe that the time step when convergence is achieved for the NR cycle is larger for these methods with respect to the ICCG method. We also observe that for the first NR iteration, the reduction is larger for the deflated method with 10 snapshots as deflation vectors.
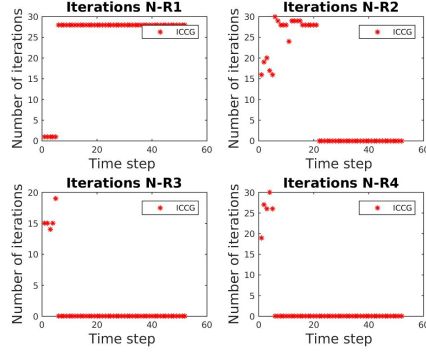
Figure 22: Number of iterations of the ICCG method for the first four NR iterations.
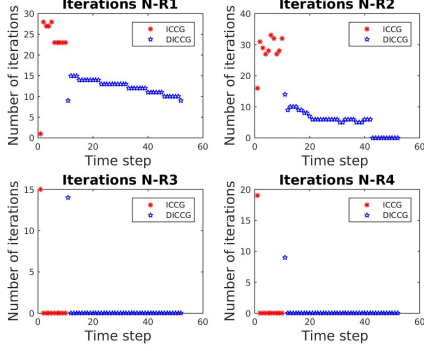

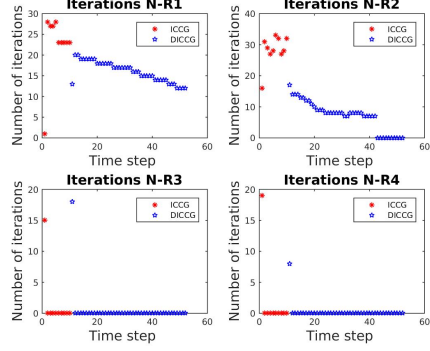Figure 23: Number of iterations of the DICCG$_{10}$ method for the first four NR iterations.

Figure 24: Number of iterations of the DICCG$_{POD}$ method for the first four NR iterations.

A   List of notation

Symbol	Quantity	Unit
ϕ	Rock porosity	
$\mathbf{K}$	Rock permeability	$Darcy\ (D)$
c_r	Rock compressibility	Pa^{-1}
$\mathbf{v}$	Darcy's velocity	m/d
α	Geometric factor	
ρ	Fluid density	kg/m^3
μ	Fluid viscosity	$Pa \cdot s$
p	Pressure	Pa
g	Gravity	m/s^2
d	Reservoir depth	m
c_l	Liquid compressibility	Pa^{-1}
q	Sources	

Table 5: Notation

B   Stopping criteria

When we use an iterative method, we always want that our approximation is close enough to the exact solution. In other words, we want that the error [12, pag. 42]:

$$||\mathbf{e}^k||_2 = ||\mathbf{x} - \mathbf{x}^k||_2,$$

or the relative error:

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2},$$

is small.

When we want to chose a stopping criteria, we could think that the relative error is a good candidate, but is has the disadvantage that we need to know the exact solution to compute it. What we have instead is the residual

$$\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k,$$

that is actually computed in each iteration of the CG method. There is a relationship between the error and the residual that can help us with the choice of the stopping criteria.

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2} \leq \kappa_2(A)\frac{||\mathbf{r}^k||_2}{||\mathbf{b}||_2}.$$

With this relationship in mind, we can choose the stopping criteria as an ϵ for which

$$\frac{||\mathbf{r}^k||_2}{||\mathbf{b}||_2} \leq \epsilon.$$

But we should keep to have in mind the condition number of the matrix $\mathbf{A}$, because the relative error will be bounded by:

$$\frac{||\mathbf{x} - \mathbf{x}^k||_2}{||\mathbf{x}||_2} \leq \kappa_2(A)\epsilon.$$

C   Singular Value Decomposition for POD

If we perform SVD in $\mathbf{X}$, we have
$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \qquad \mathbf{U} \in \mathbb{R}^{n\times n}, \qquad \mathbf{\Sigma} \in \mathbb{R}^{n\times m}, \qquad \mathbf{V} \in \mathbb{R}^{m\times m}.$
Then we have

$$\begin{aligned}
\mathbf{R} &= \mathbf{X}\mathbf{X}^T \\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T \\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T, \ \mathbf{V}^T\mathbf{V} = \mathbf{I} \\
&= \mathbf{U}\Lambda\mathbf{U}^T, \ \Lambda = \mathbf{\Sigma}\mathbf{\Sigma}^T \in \mathbb{R}^{n\times n}
\end{aligned} \qquad \begin{aligned}
\mathbf{R}^T &= \mathbf{X}^T\mathbf{X} \\
&= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \ \mathbf{U}^T\mathbf{U} = \mathbf{I} \\
&= \mathbf{V}\Lambda^T\mathbf{V}^T, \ \Lambda^T = \mathbf{\Sigma}^T\mathbf{\Sigma} \in \mathbb{R}^{m\times m}.
\end{aligned}$$

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$
$$\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{\Sigma}^{-1}$$
$$\mathbf{U} = \mathbf{X}\mathbf{V}\Lambda^{-\frac{1}{2}}$$

If we compute Λ^T, we can compute U as follows:

$$\mathbf{U} = \mathbf{X}\mathbf{V}(\Lambda^T)^{-\frac{T}{2}} = \mathbf{X}\mathbf{V}(\Lambda^T)^{\frac{1}{2}}$$

References

[1] J. C Vink J.D. Jansen P. Astrid, G. Papaioannou. Pressure Preconditioning Using Proper Orthogonal Decomposition. In *2011 SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA*, January 2011.

[2] R. Markovinović and J. D. Jansen. Accelerating iterative solution methods using reduced-order models as solution predictors. *International journal for numerical methods in engineering*, 68(5):525–541, 2006.

[3] A. Segal C. Vuik and J. A. Meijerink. An Efficient Preconditioned CG Method for the Solution of a Class of Layered Problems with Extreme Contrasts in the Coefficients. *Journal of Computational Physics*, 152:385, 1999.

[4] M.A. Christie and M.J. Blunt. Tenth spe comparative solution project: a comparison of upscaling techniques. *SPE Reservoir Engineering and Evaluation*, 4(4):308–317, 2001.

[5] J. Tang. *Two-Level Preconditioned Conjugate Gradient Methods with Applications to Bubbly Flow Problems*. PhD thesis, Delft University of Technology, 2008.

[6] R. Schuhmann M. Clemens, M. Wilke and T. Weiland. Subspace projection extrapolation scheme for transient field simulations. *IEEE Transactions on Magnetics*, 40(2):934–937, 2004.

[7] L. Yaakoubi C. Vuik, A. Segal and E. Dufour. A comparison of various deflation vectors applied to elliptic problems with discontinuous coefficients. *Applied Numerical Mathematics*, 41(1):219–233, 2002.

[8] C. Vuik J.M. Tang, R. Nabben and Y. Erlangga. Comparison of two-level preconditioners derived from deflation, domain decomposition and multigrid methods. *Journal of scientific computing*, 39(3):340–370, 2009.

[9] W. Gropp B. Smith, P. Bjorstad. *Domain decomposition: parallel multilevel methods for elliptic partial differential equations*. Cambridge University Press New York, 1996.

[10] G. Strang. *Linear Algebra and Its Applications*. Wellesley-Cambridge Press, 2009.

[11] K. A. Lie. *An Introduction to Reservoir Simulation Using MATLAB: User guide for the Matlab Reservoir Simulation Toolbox (MRST)*. SINTEF ICT, 2013.

[12] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics Philadelphia, PA, USA. 2nd edition, 2003.