Journal of Computational Physics **172,** 426–450 (2001)
doi:10.1006/jcph.2001.6795, available online at http://www.idealibrary.com on IDEAL®

The Construction of Projection Vectors for a Deflated ICCG Method Applied to Problems with Extreme Contrasts in the Coefficients

C. Vuik,* A. Segal,* J. A. Meijerink,† and G. T. Wijma*

**Delft University of Technology, Faculty of Information Technology and Systems, Department of Applied Mathematical Analysis, Mekelweg 4, 2628 CD Delft, The Netherlands; and †Shell International Exploration and Production, P.O. Box 60, 2280 AB Rijswijk, The Netherlands*
E-mail: c.vuik@math.tudelft.nl

Received January 4, 2000; revised April 12, 2001

To predict the presence of oil and natural gas in a reservoir, it is important to know the fluid pressure in the rock formations. A mathematical model for the prediction of the fluid pressure history is given by a time-dependent diffusion equation. Application of the finite-element method leads to systems of linear equations. A complication is that the underground consists of layers with very large contrasts in permeability. This implies that the symmetric and positive definite coefficient matrix has a very large condition number. Bad convergence behavior of the ICCG method has been observed, and a classical termination criterion is not valid in this problem. In [19] we have shown that the number of small eigenvalues of the diagonally scaled matrix is equal to the number of high-permeability domains, which are not connected to a Dirichlet boundary. In this paper the proof is extended to an Incomplete Cholesky decomposition. To annihilate the bad effect of these small eigenvalues on the convergence, the Deflated ICCG method is used. In [19] we have shown how to construct a deflation subspace for the case of a set of more or less parallel layers. That subspace proved to be a good approximation of the span of the "small" eigenvectors. As a result of this, the convergence of DICCG is independent of the contrasts in the permeabilities. In this paper it is shown how to construct deflation vectors even in the case of very irregular shaped layers, and layers with so-called inclusions. A theoretical investigation and numerical experiments show that the DICCG method is not sensitive to small perturbations of the deflation vectors. The efficiency of the DICCG method is illustrated by numerical experiments. © 2001 Academic Press

Key Words: deflation; IC preconditioned Conjugate Gradients; Poisson equation; porous media; discontinuous coefficients across layers; sensitivity analysis.

1. INTRODUCTION

Knowledge of the fluid pressure history in the subsurface is important for an oil company to predict the presence of oil and natural gas in reservoirs and a key factor in safety and environmental aspects of drilling a well. A mathematical model for the prediction of fluid pressures in a geological time scale is based on conservation of mass and Darcy's law [2, 7]. The resulting time-dependent three-dimensional nonlinear diffusion equation is linearized and integrated in time by the backwards Euler method. For the space discretization, the finite element method is applied. As a consequence in each time step, a large, sparse linear system of equations has to be solved.

We simplify the problem considerably, taking into account that its characteristic properties are kept. Assume that we have to solve the stationary linear diffusion equation,

$$-\text{div}(\sigma \nabla p) = 0 \quad \text{on } \Omega, \tag{1}$$

with boundary conditions

$$p = f \quad \text{on } \partial\Omega^D \text{ (Dirichlet)} \quad \text{and} \quad \frac{\partial p}{\partial n} = g \quad \text{on } \partial\Omega^N \text{ (Neumann)},$$

where $\partial\Omega = \partial\Omega^D \cup \partial\Omega^N$. In our problems, $\partial\Omega^D$ is the top boundary of the domain, unless stated otherwise. The fluid pressure and permeability are denoted by p and σ, respectively. The domain Ω consists of a number of subdomains in which σ is constant. Two values for σ are considered: $\sigma^h = 1$ for high-permeability subdomains and $\sigma^l = \varepsilon$ for low-permeability subdomains (e.g., the permeabilities ratio for shale and sandstone: ε is of the order 10^{-7}; see [19]). The subdomains are denoted by the disjoint sets Ω_i, $i \in \{1, \ldots, k\}$, which are such that: $\cup_{i=1}^{k} \overline{\Omega}_i = \overline{\Omega}$ and when $\overline{\Omega}_i \cap \overline{\Omega}_j \neq \emptyset$, then $\sigma_i \neq \sigma_j$. Note that in real-life applications, the permeability σ is slightly varying in the subdomains. It is straightforward to adapt our method to construct the projection vectors for such a problem.

After a finite element discretization of (1), the linear system

$$Ax = b, \tag{2}$$

with $A \in \mathbb{R}^{n \times n}$ has to be solved. In practical applications we are faced with large regions in a three-dimensional space and as a consequence a large number of finite elements is necessary. The matrix itself is sparse, but because of fill-in a direct method requires too much memory to fit in core. Therefore, only iterative methods are acceptable candidates for the solution of the linear systems of equations. Since the coefficient matrix of this system is symmetric and positive definite, an Incomplete Cholesky Conjugate Gradient method (ICCG) [13] seems to be a suitable iterative method. At this moment, we have not extended our method to nonsymmetrical problems. However, we expect that application to nonsymmetric methods such as GMRES and BI-CGSTAB does not cause extra difficulties. Unfortunately, the Earth's crust consists of layers with large contrasts in permeability. Hence, a large difference of the extreme eigenvalues is common in the system of equations to be solved. This leads to slow convergence of ICCG, and conventional termination criteria [10] are no longer reliable (see [19] for details).

In [9, 19] it has been proven that the number of small eigenvalues of the diagonally scaled matrix is equal to the number (k^s) of high-permeability domains, which boundaries do not contain a part of the Dirichlet boundary. A comparable spectrum has been observed for the

IC preconditioned matrix. In this paper, we shall prove this observation. The bad effect of these eigenvalues on the convergence of ICCG can be annihilated by using the corresponding "small" eigenvectors as projection vectors in Deflated ICCG [19]. In [19] we have shown how to construct approximate eigenvectors in case we have a set of nice, more or less parallel layers. In this paper we shall introduce a construction that is not limited to this kind of layer, but can be used for fairly general types of layers. Even inclusions of sandstone in shale can be tackled by this new construction method, something that was not possible with our previous method. Our new method involves the solution of a set of well-conditioned subproblems. It appears that we may solve these subproblems with a very limited accuracy without a deterioration of the good convergence properties of the method.

For a literature survey of iterative methods and applications where deflation is used, we refer to [19]. The DICCG method has already been successfully used for complicated magnetic field simulations [6]. A related method is recently presented in [15]. For multigrid type methods for elliptic problems with highly discontinuous coefficients, we refer to [1, 20]. Finally, in [4, 12] a preconditioner is analyzed for problems with large jumps in the permeabilities arising from reservoir simulation. The preconditioner is the inverse of M, where M is the matrix corresponding to a finite-element discretization of (1) with $\sigma \equiv 1$. This preconditioner is only applicable when a fast solution method is available to solve x from $Mx = b$. Another application, where large differences in the coefficients occur, is the fictitious domain method applied to metal casting [14].

To define the Deflated ICCG method, we need a set of projection vectors $v_1, \ldots, v_m$ that form an independent set. The projection on the space A-perpendicular to span $\{v_1, \ldots, v_m\}$ is defined as

$$P = I - VE^{-1}(AV)^T \quad \text{with } E = (AV)^T V \quad \text{and} \quad V = [v_1 \ldots v_m].$$

The solution vector x can be split into two parts $x = (I - P)x + Px$. The first part can be calculated as follows: $(I - P)x = VE^{-1}V^T\ Ax = VE^{-1}V^T b$. For the second part, we project the solution x_j obtained from DICCG to Px_j.

The Deflated ICCG algorithm reads (see Reference [19]):

```
DICCG
$j = 0,\ \hat{r}_0 = P^T r_0,\ p_1 = z_1 = L^{-T}L^{-1}\hat{r}_0;$
while $\|\hat{r}_j\|_2 >$ accuracy do
    $j = j + 1; \alpha_j = \frac{(\hat{r}_{j-1}, z_{j-1})}{(p_j, P^T\ Ap_j)};$
    $x_j = x_{j-1} + \alpha_j p_j;$
    $\hat{r}_j = \hat{r}_{j-1} - \alpha_j P^T\ Ap_j;$
    $z_j = L^{-T}L^{-1}\hat{r}_j; \beta_j = \frac{(\hat{r}_j, z_j)}{(\hat{r}_{j-1}, z_{j-1})};$
    $p_{j+1} = z_j + \beta_j p_j;$
end while
```

Summary of the paper. In Section 2 it is proven that not only the diagonally scaled matrix has k^s small eigenvalues, but that this is also true for the IC preconditioned matrix. Thereafter, a new and efficient construction of the projection vectors in general configurations is given in Section 3. Furthermore, it is proven that the span of these vectors is a good approximation of the "small" eigenspace. A corollary of this is that the convergence of DICCG does not depend on the permeabilities ratio ε. The sensitivity of the method with respect to perturbations of the projection vectors is investigated in Section 4. Finally, in Section 5 some numerical experiments are given.

2. ANALYSIS OF THE IC PRECONDITIONED ITERATION MATRIX

First we repeat the definition of the Incomplete Cholesky (IC) preconditioner. After that it is shown that the preconditioned matrix is scaling invariant. Thereafter, the results of [19] with respect to a simple diagonal preconditioner are generalized to the more complicated case of the IC preconditioner. In [19] it has already been observed numerically that this result is also valid for the IC preconditioner, but no proof has been given. Finally, a remark is given concerning an estimate of the condition of the matrix.

The IC preconditioner is defined as follows [13]:

DEFINITION 2.1. Determine the lower triangular matrix L with the following properties:

- $l_{ij} = 0$, when $a_{ij} = 0$,
- $(LL^T)_{ij} = a_{ij}$, when $a_{ij} \neq 0$, and $l_{ii} > 0$.

Let us define $\hat{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ where D is an arbitrary diagonal matrix with positive elements on the diagonal. The IC factor of $\hat{A}$ is denoted by $\hat{L}$. In the next theorem, we prove that the preconditioned matrix is scaling invariant.

THEOREM 2.1. *The matrices $L^{-1}AL^{-T}$ and $\hat{L}^{-1}\hat{A}\hat{L}^{-T}$ are identical.*

Proof. The nonzero pattern of L and $\hat{L}$ are the same. Suppose $\hat{L} = D^{-\frac{1}{2}}L$ and check the second statement of Definition 2.1:

$$(\hat{L}\hat{L}^T)_{ij} = \left(D^{-\frac{1}{2}} L L^T D^{-\frac{1}{2}}\right)_{ij} = \frac{1}{\sqrt{d_{ii}}}(LL^T)_{ij}\frac{1}{\sqrt{d_{jj}}} = \frac{1}{\sqrt{d_{ii}}} a_{ij} \frac{1}{\sqrt{d_{jj}}} = \left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right)_{ij}.$$

These identities imply that $\hat{L}$ is equal to $D^{-\frac{1}{2}}L$. From this the theorem follows since

$$\hat{L}^{-1}\hat{A}\hat{L}^{-T} = L^{-1}D^{\frac{1}{2}}\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)D^{\frac{1}{2}}L^{-T} = L^{-1}\ AL^{-T}. \qquad \blacksquare$$

As a consequence of this theorem, ICCG has the same convergence behavior for the original system $Ax = b$ and the diagonally scaled system $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}y = D^{-\frac{1}{2}}b$ with $x = D^{-\frac{1}{2}}y$. Therefore, we consider the diagonally scaled matrix $\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ with $D = diag(A)$ in the remainder of this section. In [19] it is shown that $\hat{A}$ has k^s small eigenvalues. Below we generalize this result to the Incomplete Cholesky preconditioned matrix.

We consider the following characteristic configuration: Ω is a rectangular domain, which consists of $2k^s + 1$ plain layers of equal thickness with a high-permeability layer at the top and alternating low- and high-permeability layers further down (see Fig. 1). In order to simplify the proofs that will be given later on, the unknowns are renumbered in the following way: first, all high-permeability unknowns are numbered per layer from top to bottom and next all low-permeability unknowns. Unknowns on the interface of a low- and high-permeability layer will be considered as high-permeability unknowns. Let Δ_h^N, Δ_h^{DT}, and Δ_h^D be the finite element matrices of the Laplacian on a single layer with respectively homogeneous Neumann boundary conditions on all boundaries, Dirichlet boundary conditions on the top boundary, and Dirichlet boundary conditions on top and bottom boundaries. In Δ_h^{DT} and Δ_h^D, a homogeneous Neumann boundary condition is posed on all other boundaries.

THEOREM 2.2. *For the problem given above, it appears that if ϵ is small enough, the* IC *preconditioned matrix $\hat{L}^{-1}\hat{A}\hat{L}^{-T}$ has only k^s eigenvalues of $O(\epsilon)$, where ϵ is the permeabilities ratio.*

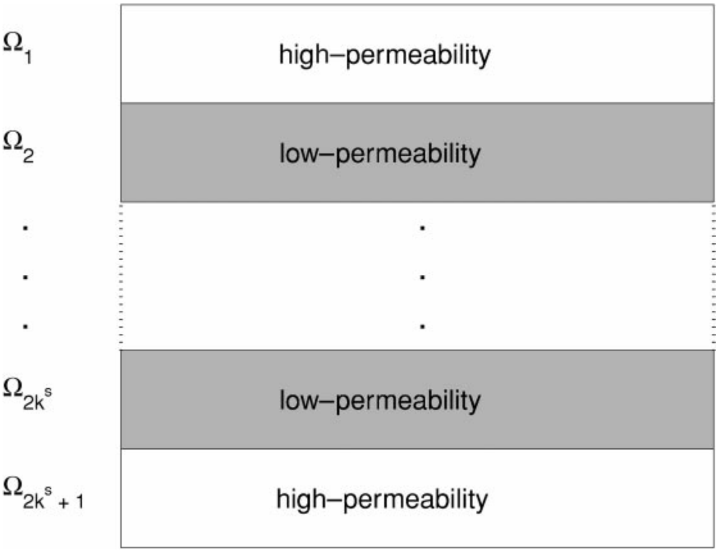

FIG. 1. A problem with $2k^s + 1$ plain layers of equal thickness.

Proof. Using the finite element discretization as given in [19], the matrix $\hat{A}$ can be split into an ϵ dependent and an ϵ independent part,

$$\hat{A} = \hat{\Delta} + \mathcal{E}, \tag{3}$$

where $\hat{\Delta}$ is the block-diagonal matrix with as first block $\hat{\Delta}_h^{DT}$, the diagonally scaled Δ_h^{DT}, and then further down alternating $\hat{\Delta}_h^{D}$, the scaled Δ_h^{D}, and $\hat{\Delta}_h^{N}$, the scaled Δ_h^{N}. From [19] it follows that $\|\mathcal{E}\|_2 = O(\sqrt{\epsilon})$.

Consider the Incomplete Cholesky decomposition $\tilde{L}\tilde{L}^T$ of the matrix $\hat{\Delta}$. Because of its block diagonal structure the Cholesky factor is also block diagonal. The existence of the nonsingular IC decomposition of $\hat{\Delta}^{DT}$ and $\hat{\Delta}^{D}$ follows from [13]. Under mild conditions (which are fullfilled in our problem) Kaasschieter ([11], Theorem 3.2) has proven that an IC decomposition exists for the singular matrix $\hat{\Delta}^N$ and that the resulting Cholesky factor is nonsingular. This implies that $\tilde{L}$ exists and $\lambda_{\min}(\tilde{L}\tilde{L}^T) > 0$.

In order to prove $\hat{L}\hat{L}^T = \tilde{L}\tilde{L}^T + \hat{\mathcal{E}}$, with $\|\hat{\mathcal{E}}\|_2 = O(\sqrt{\epsilon})$, we first show that

$$\hat{L} = \tilde{L} + O(\sqrt{\epsilon}). \tag{4}$$

The elements of $\hat{L}$ are determined by the following formulas:

$$\hat{l}_{ij} = 0, \quad \text{when } \hat{a}_{ij} = 0, \quad \text{else } \hat{l}_{ij} = \left(\hat{a}_{ij} - \sum_{k=1}^{j-1} \hat{l}_{ik}\hat{l}_{jk}\right) \Big/ \hat{l}_{jj} \quad \text{for } j < i, \tag{5}$$

$$\hat{l}_{ii} = \sqrt{\hat{a}_{ii} - \sum_{k=1}^{i-1} \hat{l}_{ik}^2}. \tag{6}$$

Equation (4) is proven by induction. The Induction Hypothesis is

$$\hat{l}_{ij} = \tilde{l}_{ij} + O(\sqrt{\epsilon}), \quad 1 \le j \le i. \tag{7}$$

The start of the induction is possible because the first block of $\hat{A}$ and $\hat{\Delta}$ are identical. For $\hat{l}_{i+1,j},\ 1 \le j \le i$ we distinguish two cases:

(i) $\hat{a}_{i+1,j} = O(1)$

Using (5), (3), and (7) we obtain $\hat{l}_{i+1,j} = \tilde{l}_{i+1,j} + O(\sqrt{\epsilon})$.

(ii) $\hat{a}_{i+1,j} = O(\sqrt{\epsilon})$

In this case $\tilde{l}_{i+1,j} = 0$ and one has to show that $\hat{l}_{i+1,j} = O(\sqrt{\epsilon})$. Because of our ordering of the unknowns $\hat{a}_{i+1,k} = O(\sqrt{\epsilon})$ for $k < j$ and thus $\hat{l}_{i+1,k} = O(\sqrt{\epsilon})$ for $k < j$. Furthermore, $\hat{l}_{jj} = \tilde{l}_{jj} + O(\sqrt{\epsilon})$, together with $\tilde{l}_{jj} > 0$, implies that $\hat{l}_{jj} > 0$ for ϵ small enough. These estimates combined with (5) show that $\hat{l}_{i+1,j} = O(\sqrt{\epsilon})$.

For $\hat{l}_{i+1,i+1}$, we obtain from (6), (3), and (7) that

$$\hat{l}_{i+1,i+1} = \sqrt{\hat{\Delta}_{i+1,i+1} - \sum_{k=1}^{i} \tilde{l}^2_{i+1,k}} + O(\sqrt{\epsilon}).$$

Since $\tilde{l}_{i+1,i+1} \neq 0$, it follows that $\hat{l}_{i+1,i+1} = \tilde{l}_{i+1,i+1} + O(\sqrt{\epsilon})$.

From Eq. (4) it easily follows that

$$\hat{L}\hat{L}^T = \tilde{L}\tilde{L}^T + \hat{\mathcal{E}}, \quad \text{with } \|\hat{\mathcal{E}}\|_2 = O(\sqrt{\epsilon}).$$

The minimax characterization ([8], Theorem 8.1.5) can be used to derive the inequalities

$$\lambda_{\min}(\hat{L}\hat{L}^T) \ge \lambda_{\min}(\tilde{L}\tilde{L}^T) - O(\sqrt{\epsilon}) \quad \text{and} \quad \lambda_{\max}(\hat{L}\hat{L}^T) \le \lambda_{\max}(\tilde{L}\tilde{L}^T) + O(\sqrt{\epsilon}).$$

For ϵ small enough we have the bound $\lambda_{\min}(\hat{L}\hat{L}^T) \ge \frac{1}{2}\lambda_{\min}(\tilde{L}\tilde{L}^T) > 0$, which is independent of ϵ. The Courant–Fisher Minimax Theorem ([8], Theorem 8.1.2) can be used to show that

$$\frac{1}{\lambda_{\max}(\hat{L}\hat{L}^T)}\lambda_k(\hat{A}) \le \lambda_k(\hat{L}^{-1}\hat{A}\hat{L}^{-T}) \le \frac{1}{\lambda_{\min}(\hat{L}\hat{L}^T)}\lambda_k(\hat{A}).$$

So the number and size of small eigenvalues of $\hat{A}$ and $\hat{L}^{-1}\hat{A}\hat{L}^{-T}$ are the same. By using Theorem 3.1 from [19], the theorem is proven. ■

This section is concluded with a remark on estimating the condition of $\hat{A}$. When a linear system is solved by an iterative method, it is important to have an estimate of the condition number of the matrix. This estimate can be combined with the machine precision to derive a lower bound for the norm of the residual. It makes no sense to iterate further when the norm of the residual is below this bound. Since the classical method to estimate the smallest eigenvalue fails [10, 19] we have no cheap test to detect that $\hat{A}$ is ill-conditioned because of the large jumps in the permeability. Another method to estimate the condition of a matrix is to estimate the condition of its Cholesky factor [3, 5]. This method is also not usefull in this problem because it follows from the proof of Theorem 2.2 that the incomplete Cholesky factor $\hat{L}$ is well conditioned.

3. THE CONSTRUCTION OF THE PROJECTION VECTORS

In [19] the “small” eigenvectors have been computed and analyzed for a simple test problem. This analysis suggests that the eigenvectors are constant in high-permeability layers and vary linearly in low-permeability layers with appropriate boundary conditions. The computation of eigenvectors is expensive, and therefore an efficient way to construct the projection vectors for layered problems (see Fig. 1) is given in [19]. In this construction, the vertical cross sections of the projection vectors v_i are made such that:

- the value of v_i is one in the $i + 1$th high-permeability layer including interface points and zero in the other high-permeability layers,
- the value of v_i in low-permeability layers is a linear interpolation of its values on the interfaces.

Below we give a generalization of the construction of the projection vectors for general geometries. Furthermore, a proof is given that the space spanned by these projection vectors is a good approximation of the "small" eigenspace of the (diagonal or IC) preconditioned matrix.

ASSUMPTION 3.1. We assume that the number of small eigenvalues (k^s) of the diagonally scaled (or IC preconditioned) matrix is equal to the number of high-permeability subdomains not connected to $\partial\Omega^D$.

This assumption holds for a layered structure (see [19] and Theorem 2.2). To construct the projection vectors the subdomains are ordered as follows:

DEFINITION 3.1. The high-permeability subdomains are numbered first: $\Omega_i,\ i \in \{1, \ldots, k^h\}$. Furthermore, the first k^s high-permeability subdomains are such that $\overline{\Omega}_i \cap \partial\Omega^D = \emptyset,\ i \in \{1, \ldots, k^s\}$.

In other words the first k^s subdomains are highly permeable and their boundaries do not contain a part of the Dirichlet boundary. This ordering is a generalization of the ordering used in Section 2.

DEFINITION 3.2. The projection vectors v_i for $i \in \{1, \ldots, k^s\}$ are defined as

- $v_i = 1$ on $\overline{\Omega}_i$ and $v_i = 0$ on $\overline{\Omega}_j,\ j \neq i,\ j \in \{1, \ldots, k^h\}$,
- v_i satisfies the finite element discretization of the equation

$$-\text{div}(\sigma_j \nabla v_i) = 0 \quad \text{on } \Omega_j, \qquad j \in \{k^h + 1, \ldots, k\}, \tag{8}$$

where Dirichlet boundary conditions are used at the interfaces ($\partial\Omega_j \cap \Omega$) and homogeneous Dirichlet and Neumann boundary conditions are used at outer boundaries ($v_i = 0$ on $\overline{\Omega}_j \cap \partial\Omega^D$ and $\frac{\partial v_i}{\partial n} = 0$ on $\overline{\Omega}_j \cap \partial\Omega^N$).

Note that in this definition the linearly varying part is replaced by the solution of the relatively small systems of Eq. (8). Each of these systems itself has a good condition. Later on we shall solve these systems (8) with a limited accuracy, so that only a few ICCG steps per subsystem are necessary. This construction is definitely much more general than the previous one, since it does not require parallel layers anymore. It even allows the possibility of inclusions of sandstone in shale layers, something that was not possible in the old method.

Another important feature of this new approach is that the permeability does not have to be constant per layer. It may be a function of space, and we may even replace the permeability by a full Cartesian permeability tensor that is a function of space. Of course, there is no reason why the permeability in the sandstone layer may not be variable as well.

Note that the projection vectors v_i are independent of ε. In the high-permeability domains, the vectors v_i can also be interpreted as a solution of (1) with Dirichlet boundary conditions equal to 1 for $\partial\Omega_i$ and equal to 0 for $\partial\Omega_j$, $j \neq i$, $j \in \{1, \ldots, k^h\}$.

We first show that $\|D^{-1}Av_i\|_\infty = O(\varepsilon)$, where D = diag(A). Thereafter, we prove that the vectors v_i are independent, so span $\{v_1, \ldots, v_{k^s}\}$ approximates the "small" eigenspace of the diagonally scaled matrix.

ASSUMPTION 3.2. We assume that the finite-element discretization is consistent, which means that the discretization error is zero for a constant function. The subdomains Ω_i are approximated by polygons and each element is contained in only one polygon. Finally, we assume that the off-diagonal elements of A are nonpositive.

This last assumption means that the finite-element grid must satisfy certain requirements with respect to the angles of the elements. This is a sufficient condition for our proofs; however, the method has a wider range of applicability. Let the vector $\mathbf{x}_m$ contain the space coordinates of grid point m. The jth component of vector v_i is denoted by $(v_i)_j$. A consequence of Assumption 3.2 (consistency) is

$$\sum_{j=1}^{n} a_{mj} = 0 \quad \text{for } \mathbf{x}_m \in \bar{\Omega}\backslash\partial\Omega^D. \tag{9}$$

For every projection vector v_i we define an index set $\mathbb{I}_i \subset \{1, \ldots, n\}$, which contains the indices of all points on the interfaces of the low-permeability domains which are neighbors of Ω_i.

THEOREM 3.1. *When Assumption* 3.2 *is fulfilled the vectors v_i as defined in Definition* 3.2 *are such that*

$$\|D^{-1}Av_i\|_\infty = O(\varepsilon), \quad i \in \{1, \ldots, k^s\}.$$

Proof. We first prove

$$(Av_i)_m = 0 \quad \text{for } m \in \{1, \ldots, n\}\backslash\mathbb{I}_i. \tag{10}$$

For $\mathbf{x}_m \in \Omega_j \cup (\bar{\Omega}_j \cap \partial\Omega)$, $j \in \{1, \ldots, k^s\}$ component $(v_i)_m$ is equal to the components of v_i in the neighboring points. So (10) follows from Eq. (9). From Definition 3.2 we see that (10) holds for $\mathbf{x}_m \in \Omega_j \cup (\bar{\Omega}_j \cap \partial\Omega)$, $j \in \{k^s + 1, \ldots, k^h\}$. On a low-permeability subdomain ($\mathbf{x}_m \in \Omega_j \cup (\bar{\Omega}_j \cap \partial\Omega)$, $j \in \{k^h + 1, \ldots, k\}$), Eq. (10) also follows from Definition 3.2. Finally, when $\mathbf{x}_m$ is on an interface where v_i has zero components on the low-permeability subdomain then all components of v_i are zero in the high-permeability subdomain and thus Eq. (10) holds. This implies $(D^{-1}Av_i)_m = 0$ for $m \in \{1, \ldots, n\}\backslash\mathbb{I}_i$.

For an $m \in \mathbb{I}_i$ we split its neighbors into two sets $\mathbb{J}_m^h$ and $\mathbb{J}_m^l$. The sets $\mathbb{J}_m^h$, $\mathbb{J}_m^l$ contain the indices of the neighboring points in the high-, low-permeability subdomain, respectively. The set $\mathbb{J}_m^h$ also contains the indices of the neighboring points at the interface. This implies

$$\begin{aligned} a_{mj} &= O(1), \quad \text{for } j \in \mathbb{J}_m^h, \\ a_{mj} &= O(\varepsilon), \quad \text{for } j \in \mathbb{J}_m^l. \end{aligned} \tag{11}$$

Assumption 3.2 ($a_{mj} \leq 0,\ j \neq m$), together with (9) and (11), implies

$$D_{mm} = a_{mm} = -\sum_{j \in \mathbb{J}_m^h \cup \mathbb{J}_m^l} a_{mj} = O(1). \tag{12}$$

Equation (9) combined with $(v_i)_j = (v_i)_m$ for $j \in \mathbb{J}_m^h$ yields

$$(Av_i)_m = \sum_{j \in \mathbb{J}_m^l} a_{mj}((v_i)_j - (v_i)_m). \tag{13}$$

From the maximum principle we know that $(v_i)_j \in [0, 1]$ so $(Av_i)_m = O(\varepsilon)$. Together with $D_{mm} = O(1)$, the result is proven. ■

COROLLARY 3.1. *For* $m \in \mathbb{I}_i$, *an* $O(1)$ *perturbation of* $(v_i)_j,\ j \in \mathbb{J}_m^h$ *leads to* $\|D^{-1}Av_i\|_\infty = O(1)$, *whereas an* $O(1)$ *perturbation of* $(v_i)_j,\ j \in \mathbb{J}_m^l$ *leads to* $\|D^{-1}Av_i\|_\infty = O(\varepsilon)$.

This means that a large perturbation of v_i in the low-permeability layer leads to a small perturbation in the matrix vector product $D^{-1}Av_i$.

DEFINITION 3.3. The normalized eigenvectors of $D^{-1}A$ are denoted by u_i,

$$D^{-1}Au_i = \lambda_i u_i, \quad i = 1, \ldots, n,$$

where the eigenvalues are ordered: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Two matrices are defined: $V_{k^s} = [v_1 \cdots v_{k^s}]$ and $U_{k^s} = [u_1 \cdots u_{k^s}]$.

In the following theorem, we show that the space spanned by $\{v_1, \ldots, v_{k^s}\}$ is “nearly” a subspace of the “small” eigenspace span$\{u_1, \ldots, u_{k^s}\}$. With “nearly” a subspace we mean that the norm of the component of a vector $v \in \{v_1, \ldots, v_{k^s}\}$ outside the subspace span$\{u_1, \ldots, u_{k^s}\}$ is small.

THEOREM 3.2. *When Assumptions* 3.1 *and* 3.2 *are fulfilled, the expression holds,*

$$V_{k^s} = U_{k^s} Z + E, \tag{14}$$

where $\|E\|_2 = O(\sqrt{\varepsilon})$.

Proof. The vector v_1 can be written as a linear combination of the eigenvectors

$$v_1 = \sum_{j=1}^{n} \alpha_j u_j. \tag{15}$$

Theorem 3.1 implies $\|D^{-1}Av_1\|_2^2 \leq O(\varepsilon^2)$. Substitution of (15) gives

$$\|D^{-1}Av_1\|_2^2 = \left\| \sum_{j=1}^{n} \lambda_j \alpha_j u_j \right\|_2^2.$$

The eigenvectors $D^{\frac{1}{2}}u_j$ of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ are orthogonal. This property is used in the derivation

$$\begin{aligned}\|D^{-1}Av_1\|_2^2 &\geq \lambda_{\min}(D^{-1})\left\|\sum_{j=1}^{n}\lambda_j\alpha_j D^{\frac{1}{2}}u_j\right\|_2^2\\ &= \lambda_{\max}(D)\sum_{j=1}^{n}\lambda_j^2\alpha_j^2\left\|D^{\frac{1}{2}}u_j\right\|_2^2 \geq \lambda_{\max}(D)\lambda_{\min}(D)\sum_{j=1}^{n}\lambda_j^2\alpha_j^2.\end{aligned}$$

Since $\lambda_{\max}(D) = O(1)$ and $\lambda_{\min}(D) = O(\varepsilon)$, we obtain

$$\sum_{j=1}^{k^s}\lambda_j^2\alpha_j^2 + \sum_{j=k^s+1}^{n}\lambda_j^2\alpha_j^2 \leq O(\varepsilon).$$

Rearranging the terms shows that

$$\lambda_{k^s+1}^2\sum_{j=k^s+1}^{n}\alpha_j^2 \leq \sum_{j=k^s+1}^{n}\lambda_j^2\alpha_j^2 \leq O(\varepsilon),$$

because $\sum_{j=1}^{k^s}\alpha_j^2$ is bounded. Since $\lambda_{k^s+1} = O(1)$, it follows that $\sum_{j=k^s+1}^{n}\alpha_j^2 = O(\varepsilon)$. This can be shown for every v_i, $i \in \{1, \ldots, k^s\}$ so the theorem is proven. ■

Note that the projection vectors $v_1, \ldots, v_{k^s}$ are linearly independent, because for $\mathbf{x}_m \in \Omega_i$, $i \in \{1, \ldots, k^s\}$, $(v_i)_m = 1$, and $(v_j)_m = 0$ for $j \neq i$. As a consequence of this the matrix $V_{k^s}^T V_{k^s}$ is nonsingular. This can be used to show that the “small” eigenspace span$\{u_1, \ldots, u_{k^s}\}$ is “nearly” a subspace of the space spanned by $\{v_1, \ldots, v_{k^s}\}$.

THEOREM 3.3. *When Assumptions* 3.1 *and* 3.2 *are fulfilled, the expression holds,*

$$U_{k^s} = V_{k^s}Z^{-1} + \hat{E}, \tag{16}$$

where $\|\hat{E}\|_2 = O(\sqrt{\varepsilon})$.

Proof. From Theorem 3.2 it follows that

$$V_{k^s}^T V_{k^s} = Z^T U_{k^s}^T U_{k^s} Z + \tilde{E} = Z^T Z + \tilde{E},$$

where $\|\tilde{E}\|_2 = O(\sqrt{\varepsilon})$. From the minimax characterization ([8], Theorem 8.1.5) we obtain the bound

$$\lambda_{\min}(Z^T Z) \geq \lambda_{\min}\left(V_{k^s}^T V_{k^s}\right) - O(\sqrt{\varepsilon}).$$

Since $V_{k^s}^T V_{k^s}$ is nonsingular, $\lambda_{\min}(Z^T Z) > 0$ for ε small enough, and Z is nonsingular. Postmultiplying (14) by Z^{-1} gives

$$U_{k^s} = V_{k^s}Z^{-1} - EZ^{-1}. \tag{17}$$

The theorem now follows from (16) and (17) because

$$\|\hat{E}\|_2 = \|EZ^{-1}\|_2 \leq (\lambda_{\min}(Z^T Z))^{-\frac{1}{2}}\|E\|_2 = O(\sqrt{\varepsilon}). \qquad ■$$

This theorem motivates the use of the vectors $\{v_1, \ldots, v_{k^s}\}$ in the Deflated CG method applied to the diagonally scaled matrix to annihilate the effect of the small eigenvalues.

In Section 2, it has been shown that the IC preconditioned matrix has k^s eigenvalues of $O(\varepsilon)$. Below we prove that the vectors $\{v_1, \ldots, v_{k^s}\}$ can also be used as projection vectors in the DICCG method.

THEOREM 3.4. *When Assumption* 3.2 *is fulfilled, the vectors v_i as defined in Definition* 3.2 *are such that*

$$\|L^{-T}L^{-1}Av_i\|_2 = O(\varepsilon), \quad i \in \{1, \ldots, k^s\}.$$

Proof. The proof of this theorem is based on the results presented in Theorem 3.1. To use these results we note

$$\|L^{-T}L^{-1}Av_i\|_2 = \|L^{-T}L^{-1}DD^{-1}Av_i\|_2 \leq \lambda_{\max}(L^{-T}L^{-1}D)\|D^{-1}Av_i\|_2.$$

Since $L^{-T}L^{-1}D$ and $\hat{L}^{-T}\hat{L}^{-1}$ are similar, their spectra are identical. From the proof of Theorem 2.2 we have that $\lambda_{\max}(\hat{L}^{-T}\hat{L}^{-1}) = 1/\lambda_{\min}(\hat{L}\hat{L}^T)$ is bounded. This combined with Theorem 3.1 leads to the inequality:

$$\|L^{-T}L^{-1}Av_i\|_2 \leq \lambda_{\max}(\hat{L}^{-T}\hat{L}^{-1})\sqrt{n}\|D^{-1}Av_i\|_\infty = O(\varepsilon). \qquad \blacksquare$$

Analogous to Theorem 3.3 one can prove that the “small” eigenspace of $L^{-T}L^{-1}A$ is “nearly” a subspace of the span$\{v_1, \ldots, v_{k^s}\}$. This suggests that the convergence of DICCG is independent of the ratio of the high and low permeability. This is confirmed by numerical experiments in Section 5.1.

We conclude this section with some remarks about an efficient implementation of the DICCG method. It follows from Definition 3.2, that each projection vector v_i is sparse because it is zero everywhere except on $\overline{\Omega}_i$ and its neighboring subdomains. The application of the projection P to a vector consists of inner products and vector updates with v_i and Av_i. The matrix vector product Av_i is less sparse than v_i. The fill-in occurs at the grid points connected to the domain where v_i is nonzero. The number of grid points where Av_i is nonzero is defined as N_i. It is easy to see that the application of P costs approximately $3\sum_{i=1}^{k^s} N_i$ floating point operations. To store v_i and Av_i, $2\sum_{i=1}^{k^s} N_i$ memory positions are necessary. In many applications, $\sum_{i=1}^{k^s} N_i$ is less than $2n$. However, one can have problems where $N_i \approx n$, which makes DICCG unattractive. An example of this: assume that the domain Ω consists of a low-permeability subdomain which contains k^s subdomains with a high-permeability. Inspection of the computed projection vectors shows that large parts of them are close to zero. Ignoring the small components in the projection vectors, makes DICCG feasible again. To that end we introduce a tolerance δ and ignore all elements of the projection vectors less than δ. In Section 5.3 such a problem is investigated by numerical experiments.

Another important point is the cost to compute the projection vectors on the low-permeability subdomains. The construction implies that a number of diffusion problems have to be solved. The amount of work to solve these subproblems is small with respect to the total amount of work. The reasons for this are

- the size of the subproblems is small,
- the submatrices are obtained by copying the relevant part of the original matrix,

- the preconditioned submatrices are well conditioned,
- it is sufficient to have an approximation of the subsolution with a low accuracy.

The final reason is investigated in more detail in the following section.

4. SENSITIVITY INVESTIGATION OF DICCG

In Section 3 a method is given to construct projection vectors, which span is close to the "small" eigenspace. An important question is how sensitive is DICCG to perturbations in these approximations. This is a very important issue since our construction includes the solution of a subsystem per shale layer, and we would like to do that as cheap as possible. In Section 4.1, we analyze the dependence of DICCG to perturbations of the projection vectors in a simple case. Some numerical experiments to validate our analysis are given in Section 4.2. Finally in Section 4.3 we investigate numerically the occurrence of small eigenvalues when a high-permeability domain is only weakly connected to a Dirichlet boundary.

4.1. *Analysis of a Perturbed Projection Vector*

We consider a problem where the matrix $A_{IC} = L^{-1}AL^{-T}$ has one small eigenvalue λ_1. The normalized eigenvectors of A_{IC} are denoted by w_i. For simplicity we assume that $\lambda_2 = 1$, and that the perturbed projection vector is given by

$$v_1 = w_1 + \alpha w_2. \tag{18}$$

Since A_{IC} is symmetric, the eigenvectors w_i are orthonormal. The perturbed projection operator is

$$P = I - v_1 E^{-1}(A_{IC}v_1)^T,$$

where $E = (A_{IC}v_1)^T v_1 = \lambda_1 + \alpha^2$. Consider the eigenvectors of $P^T A_{IC}$. From the definition of P, it follows that $P^T A_{IC} v_1 = 0$. Furthermore, (18) implies that $P^T A_{IC} w_i = \lambda_i w_i$, for $i = 3, \ldots, n$. Finally, the vector $\alpha w_1 - w_2$ is also an eigenvector of $P^T A_{IC}$ and its eigenvalue is $\lambda_{per} = \frac{\lambda_1(1+\alpha^2)}{\lambda_1+\alpha^2}$. For α small $\lambda_{per} \approx 1 = \lambda_2$. In Table I we give λ_{per} for $\lambda_1 = 10^{-9}$ and some values of α. This analysis teaches us the following: when the projection vectors are perturbed the smallest eigenvalue remains exactly zero, however the smallest but one eigenvalue can change considerably. So if the perturbation of a projection vector is too large, deflation with this perturbed vector does not help.

In order to use DICCG we approximate the eigenvectors u_i of $L^{-T}L^{-1}A$. The eigenvectors u_i and w_i are related: $w_i = L^T u_i = \hat{L}^T D^{\frac{1}{2}} u_i$. From Section 2 we know that the elements of $\hat{L}$ are $O(1)$, whereas the elements of $D^{\frac{1}{2}}$ are $O(1)$ in high-permeability domains and $O(\sqrt{\epsilon})$ in low-permeability domains. So perturbations of u_i in a low-permeability domain lead to small perturbations of w_i and λ_{per}. In the following section, we compare the results obtained from this analysis with numerical experiments.

TABLE I
Value of λ_{per} for Various Values of α

α	0	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
λ_{per}	1	0.0909	10^{-4}	10^{-5}	10^{-7}	$2 \cdot 10^{-9}$

TABLE II
Newly Introduced Small Eigenvalue (λ_{per}) and Number of Iterations (n) Needed before DICCG (or ICCG) Reaches the Required Accuracy (Perturbation α Is Restricted to the Shale Layers)

α	0	10^{-2}	10^{-1}	1	ICCG
λ_{per}	0.164	0.164	0.164	$8.2 \cdot 10^{-3}$	$1.6 \cdot 10^{-9}$
n	14	14	15	24	54

4.2. *Validation of the Perturbation Analysis*

As a test problem we consider the straight layer problem as given in [19]. This problem consists of seven horizontal layers with a sandstone layer ($\sigma = 1$) at the top and alternately shale ($\sigma = 10^{-7}$) and sandstone layers further down.

Two experiments are done. In the first experiment a random vector is added to the projection vector in the shale layers. The amplitude of this vector is $\frac{\alpha}{2}$ and it is zero at the interfaces. The results are given in Table II. The number of iterations increases for increasing α, however even for $\alpha = 1$ DICCG is much faster than ICCG. We observe that the smallest eigenvalue only changes considerably for $\alpha = 1$. For this choice λ_{per} is of the same order as the square root of the smallest eigenvalue of the original matrix. This agrees well with our analysis. Moreover, it appears that the difference between the estimated and exact error is relatively small compared to the case of nodeflation.

In the second example we perturb the nonzero parts of the projection vectors v_i also in the sandstone layer Ω_i. In Table III the smallest nonzero eigenvalue λ_{per} and the number of iterations are given. Note that qualitatively there is a good correspondence between the results given in Table I and Table III.

4.3. *The Influence of the Geometry on Small Eigenvalues*

Initially we have assumed that a high-permeability domain, which is connected to a Dirichlet boundary does not lead to a small eigenvalue of the diagonally scaled matrix (Assumption 3.1). In this section this assumption is investigated in more detail. The reason for this is that in a groundwater flow problem (see Section 5.4), we observe a small eigenvalue, although the high-permeability inclusion is connected with a Dirichlet boundary. In our first experiments we consider configurations as in Fig. 2, with permeability constant $\sigma_h = 1$. In these tests, no small eigenvalues occur, which is expected from [19]. Therefore,

TABLE III
Newly Introduced Small Eigenvalue (λ_{per}) and Number of Iterations (n) Needed before DICCG (or ICCG) Reaches the Required Accuracy (Perturbation α in the Whole Domain)

α	0	10^{-4}	10^{-3}	10^{-2}	10^{-1}	ICCG
λ_{per}	0.164	0.0825	$9 \cdot 10^{-4}$	$9 \cdot 10^{-6}$	$9 \cdot 10^{-8}$	$1.6 \cdot 10^{-9}$
n	14	18	27	38	56	54

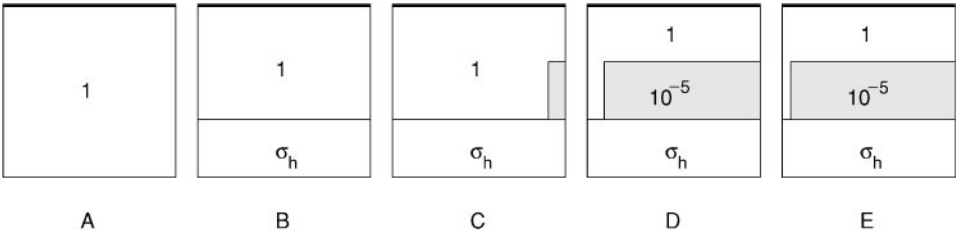

FIG. 2. The test configurations with the permeability constants.

we use $\sigma_h = 100$ in the rest of our experiments (in analogy to the groundwater flow problem). We consider five test problems. The geometry of the problems is presented in Fig. 2. All problems are solved on a grid of constant spacing. The first problem, with a medium permeability, is only used for reference. In the second geometry a high-permeability domain is added. In the remaining problems, the connection width between the medium- and high-permeability domain is decreased because of a low-permeability layer. The smallest eigenvalue of the diagonally scaled matrices are given in Table IV. The results show clearly that adding a high-permeability domain with no Dirichlet boundary conditions decreases the smallest eigenvalue by a factor 55. This is of the same order as the contrast in the permeability constants in both domains. On top of that, the smallest eigenvalue decreases proportionally to a decrease of the connection width. Both effects enhance each other. Since the ratio between the eigenvalues of the various configurations is independent of the grid-size, we expect that the smallest eigenvalue of the continuous problem has the same behavior.

Based on these results, we make the following observations on the eigenvalues of the diagonally scaled matrix when a problem is considered with high (σ_h), medium (σ_m) and low-permeability (σ_l) domains:

- A small eigenvalue occurs when a high-permeability domain is only connected to a Dirichlet boundary via a medium or a low-permeability domain.
- A high-permeability domain which is "weakly connected" (which means that the connection width is small) via a medium-permeability domain to a Dirichlet boundary leads to an eigenvalue of order $\frac{\sigma_l}{\sigma_h}$.

These observations imply that the method to construct the projection vectors (Definition 3.2) should be refined in more general problems. In Section 5.4 a groundwater flow problem is solved, which contains high-, medium-, and low-permeability domains, together with "weakly connected" high-permeability domains.

TABLE IV
The Smallest Eigenvalues for Various Test Problems, Including Variations of the Connection Width

Configuration	A	B	C	D	E
Connection width [%]	—	100	90	10	1
50 × 45 grid	2.7×10^{-4}	4.9×10^{-6}	4.6×10^{-6}	6.9×10^{-7}	—
100 × 90 grid	6.8×10^{-5}	1.2×10^{-6}	1.1×10^{-6}	1.7×10^{-7}	2.3×10^{-8}

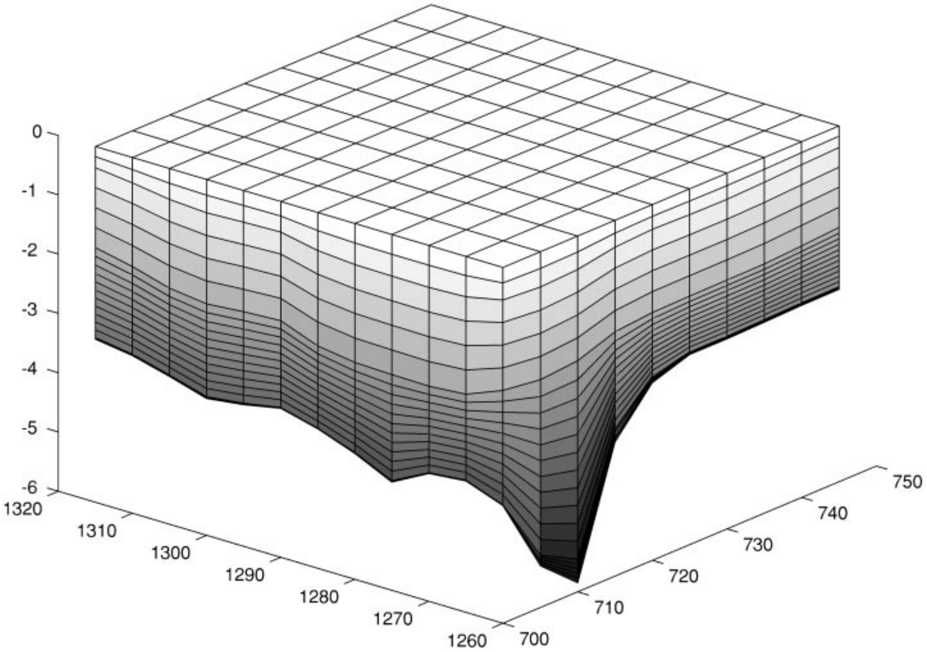

FIG. 3. The geometry of an oil flow problem.

5. NUMERICAL EXPERIMENTS

In this section we consider four test problems. The first problem is a three-dimensional layered problem motivated by transport of oil in a reservoir. Secondly a two-dimensional problem is considered with variable anisotropic permeabilities. Thirdly a test problem is considered with many high-permeability inclusions in a low-permeability layer. Finally the fourth problem is the simulation of a two-dimensional groundwater flow.

5.1. *An Oil Flow Problem*

For the flow simulation of oil and natural gas in a reservoir it is necessary to predict fluid pressures in rock layers. Therefore the diffusion equation (1) has to be solved in large three-dimensional geometries with a layered structure. This problem may also be solved with the method treated in [19], however, we have never applied that method to 3D problems, since the definition of linearly varying vectors is much more complicated than in 2D. With our new method there is no problem with the definition of the approximate eigenvectors all.

As a first problem we consider two sandstone layers ($\sigma = 10^{-4}$ and $\sigma = 10$) separated by shale layers ($\sigma = 10^{-7}$). The layers vary in thickness and orientation (see Figs. 3 and 4). At the top of the first sandstone layer a Dirichlet boundary condition is posed, so there is only one small eigenvalue. The number of iterations and the total CPU time for (D)ICCG are shown in Table V for various grid sizes. The CPU time to construct the projection vector is given too. We see that the construction time is relatively small.

Next we consider the same geometry, but now the domain consists of nine layers. Five sandstone layers are separated by four shale layers. The matrix of this problem has four small eigenvalues. Only 1.5 *n* memory positions are required to store the four projection vectors. The (D)ICCG results are given in Table VI. There is a large gain in number of

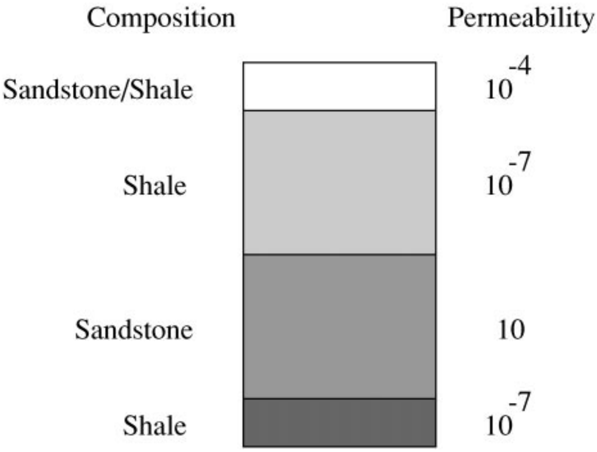

FIG. 4. Permeabilities for each layer.

iterations and CPU time. The convergence history of both methods is shown in Figs. 5 and 6. The norm of the ICCG residual has four bumps (corresponding to the number of small eigenvalues) before the true error decreases. In this example, the second and third bump nearly coincide. The observed properties of (D)ICCG correspond well with the observations made in [19] for simple two-dimensional test problems.

For the 9-layer problem we also investigate the effect of the jump in the permeabilities. We use the grid with 19665 nodal points, take the permeability in the sandstone layers equal to 1, and vary the permeability in the shale layers. The results in Table VII confirm that DICCG is independent of the value of σ_{shale}. In our experiments a relative high accuracy (10^{-5}) is used. In Table VIII the results are given for other accuracies. We conclude that the gain in CPU time is very large for accuracies (10^{-1}, 10^{-2}), which are sufficient in many applications.

5.2. *A Flow Problem with Variable Anisotropic Permeabilities*

In this problem we investigate the applicability of the DICCG method to a problem with variable anisotropic permeabilities. We consider the equation

$$-\frac{\partial}{\partial y}\sigma_{yy}\frac{\partial p}{\partial y} - \frac{\partial}{\partial z}\sigma_{zz}\frac{\partial p}{\partial z} = 0, \quad y \in [0, 10000], \quad z \in [-2000, 0],$$

TABLE V

Number of Iterations and CPU Time for Various Grid Sizes

	ICCG		DICCG		
Nodal points	Iterations	CPU	Iterations	CPU	CPU construction
2760	21	0.57	10	0.36	0.08
19665	38	9.01	20	5.80	0.60
148185	86	163	43	99.6	5.4

TABLE VI
Number of Iterations and CPU Time for the 9-Layer Problem

	ICCG		DICCG		
Nodal points	Iterations	CPU	Iterations	CPU	CPU construction
2760	47	1.19	10	0.37	0.12
19665	83	19.1	20	6.22	1.29
148185	189	350	44	108	12.7

with

$$p(y,z) = 1, \quad z = 0,$$

and homogeneous Neumann boundary conditions on all other boundaries. In a porous media flow, the permeability is a function of the depth z. We use the following function ([16], [18]): $\sigma_{yy} = 30\sigma_{zz}$ and

$$\sigma_{zz}(z) = \frac{k(z)\left(0.5 - 0.1e^{(-60(z-0.3)^2)}\right)^3 e^{\left(14 - \frac{3800}{283+0.03z}\right)}}{0.5 + 0.1e^{(-60(z-0.3)^2)}}.$$

The function k is defined by

$$k(z) = \begin{cases} 10^{-7} & z \in [-1000, 0], \\ 1 & z \in [-2000, -1000]. \end{cases}$$

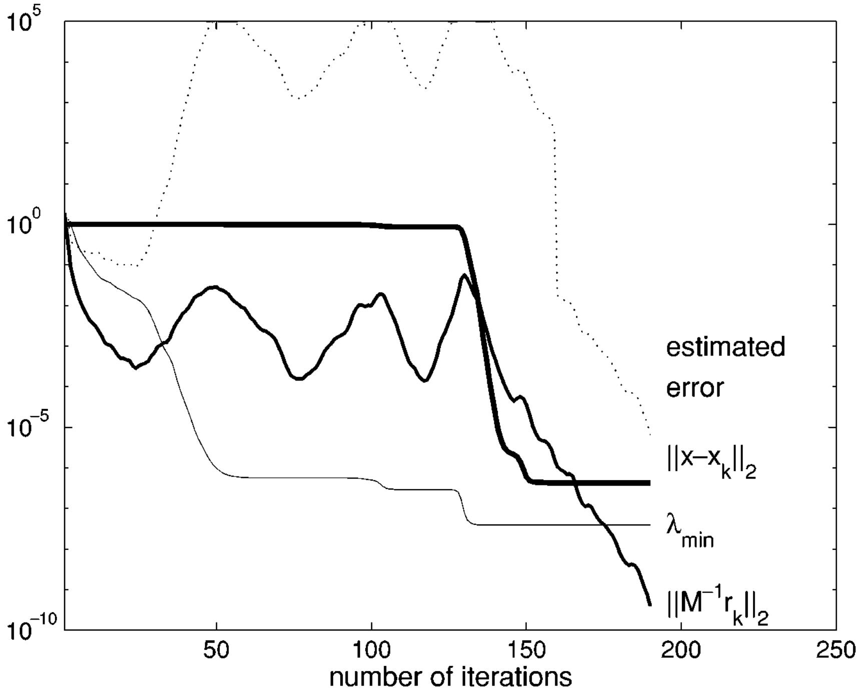

FIG. 5. The convergence behavior for ICCG.

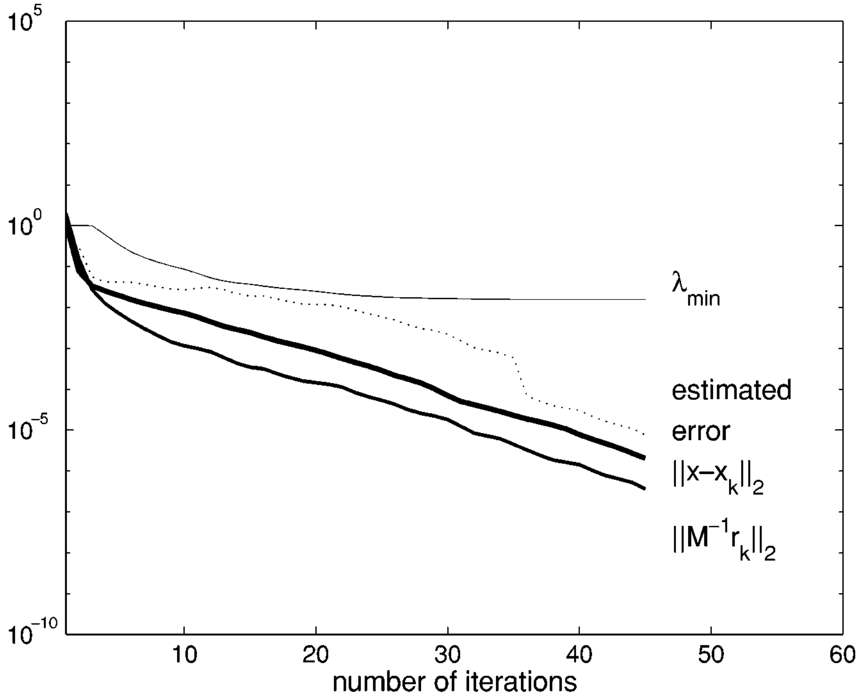

FIG. 6. The convergence behavior for DICCG.

The results for (D)ICCG are given in Figs. 7 and 8. The same behavior is observed as for the previous example. Using ICCG we can see the true error only decreases after 40 iterations and the estimated error is not reliable. In the case of DICCG, the true error decreases from the beginning, and the estimated error is close to the true error. This proves that our new approach can also be applied for anisotropic and variable permeabilities.

5.3. *A Problem with Many High-Permeability Inclusions*

At the end of Section 3 we have noted that there are problems in which the amount of memory to store the projection vectors is proportional to the number of small eigenvalues.

TABLE VII
The Smallest Nonzero Eigenvalue and the Number of Iterations for the 9-Layer Problem

	ICCG		DICCG	
σ_{shale}	$\lambda_{\min}$	Iterations	$\lambda_{\min}$	Iterations
10^{-3}	$1.5 \cdot 10^{-2}$	26	$6.9 \cdot 10^{-2}$	20
10^{-4}	$2.0 \cdot 10^{-3}$	39	$8.7 \cdot 10^{-2}$	19
10^{-5}	$2.2 \cdot 10^{-4}$	59	$7.7 \cdot 10^{-2}$	20
10^{-6}	$2.2 \cdot 10^{-5}$	73	$7.8 \cdot 10^{-2}$	20
10^{-7}	$2.3 \cdot 10^{-6}$	82	$7.7 \cdot 10^{-2}$	20

TABLE VIII
Varying the Accuracy for the 9-Layer Problem

	ICCG		DICCG	
Accuracy	Iterations	CPU	Iterations	CPU
10^{-5}	82	18.9	20	6.3
10^{-4}	80	18.4	16	5.2
10^{-3}	78	18.0	12	4.1
10^{-2}	77	17.8	3	1.5
10^{-1}	75	17.2	2	1.2

To diminish the required amount of memory, we have proposed the use of a tolerance (δ). The components of the projection vectors which are less than δ are set equal to zero. In this subsection, we investigate the influence of δ on the properties of DICCG.

We consider a three-layer problem, where the shale layer contains eight sandstone inclusions (see Fig. 9). There are 12585 nodal points used in the finite element discretization of this problem. For one of the projection vectors, a contour plot is given in Fig. 10. From this figure we see that the value of the projection vector is very small in the shale layer except in the vicinity of one sand inclusion. This observation has motivated us to delete the small components of the projection vectors.

It is clear that this example cannot be treated with the method introduced in [19]. It is impossible to define a linearly varying vector field in the shale region that also has

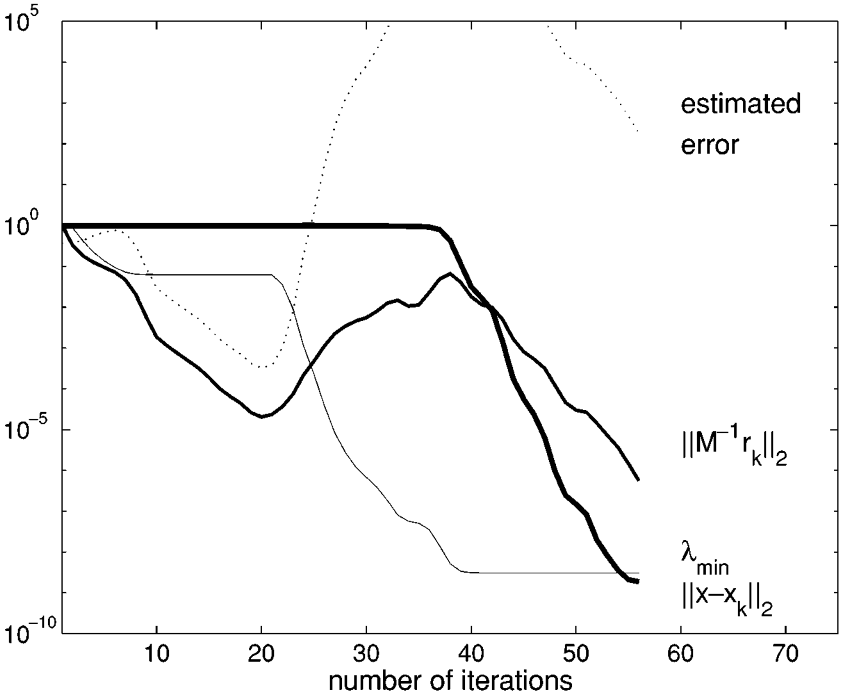

FIG. 7. The convergence behavior for ICCG.

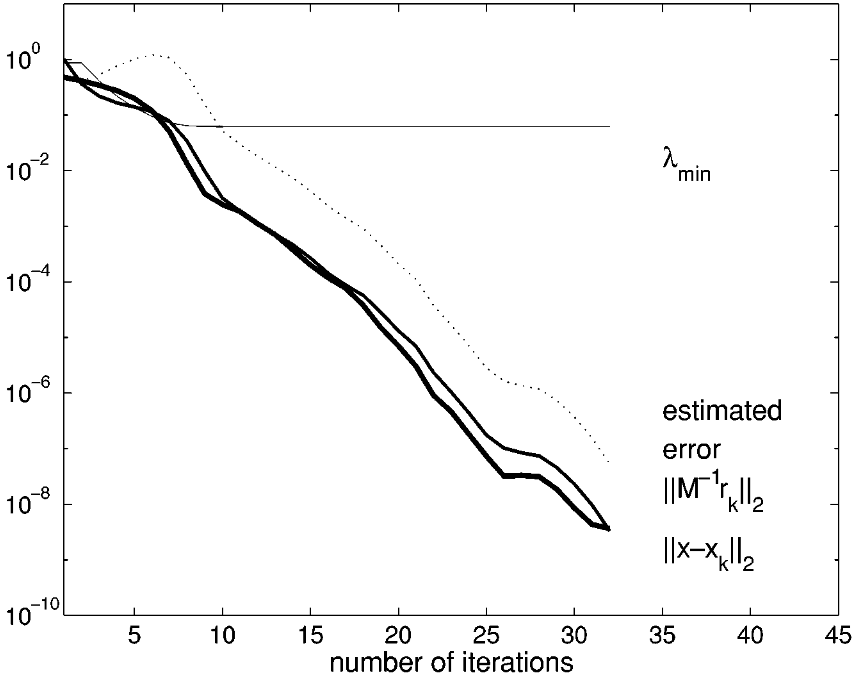

FIG. 8. The convergence behavior for DICCG.

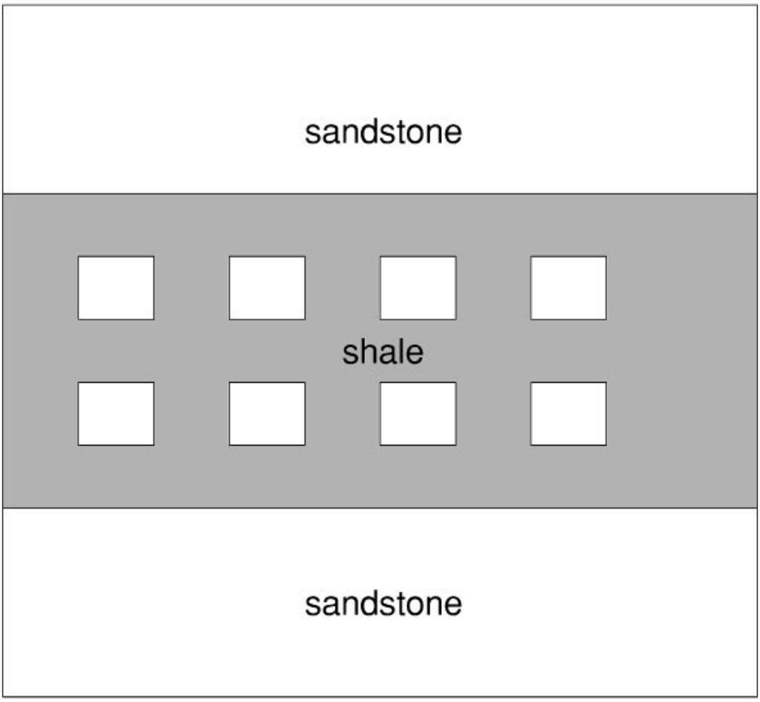

FIG. 9. A problem with eight sandstone inclusions in the shale layer.

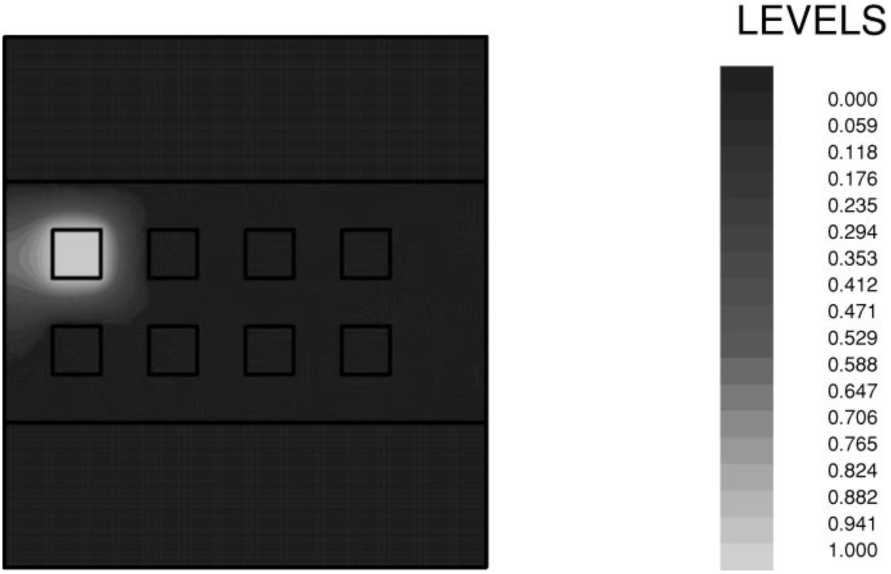

FIG. 10. Contour plot of one of the projection vectors.

prescribed values in the sandstone inclusions. Only our new approach is able to solve this problem.

In Table IX the relevant results are given for ICCG and DICCG for various values of δ. We see again a large decrease in CPU time and number of iterations when deflation (DICCG) is used. However in DICCG, 3.9 n extra memory positions are required to store the projection vectors. Increasing the tolerance leads to the same number of iterations, less CPU time, a small increase in the true error, and a large decrease of the amount of extra memory. So this combination of DICCG with a tolerance leads to an efficient solution method even when the problem has many high-permeability inclusions.

5.4. *A Groundwater Flow Problem*

In Section 4.3 we have seen that a small eigenvalue occurs when a high-permeability layer is “weakly connected” to a Dirichlet boundary. In this section we will examine the validity of this assumption in more detail. In [17], a Poisson-like equation has been solved

TABLE IX
Results of ICCG and DICCG for Various Values of δ

	ICCG	DICCG		
δ		0	10^{-2}	10^{-1}
CPU	44	12	8.3	8.1
CPU construction	0	3.1	2.9	2.9
Iterations	616	76	76	76
$\lambda_{\min}$	4×10^{-9}	8×10^{-3}	8×10^{-3}	8×10^{-3}
True error	7×10^{-7}	1.86×10^{-5}	1.87×10^{-5}	2.28×10^{-5}
Extra memory	0	3.9 n	1.6 n	1.2 n

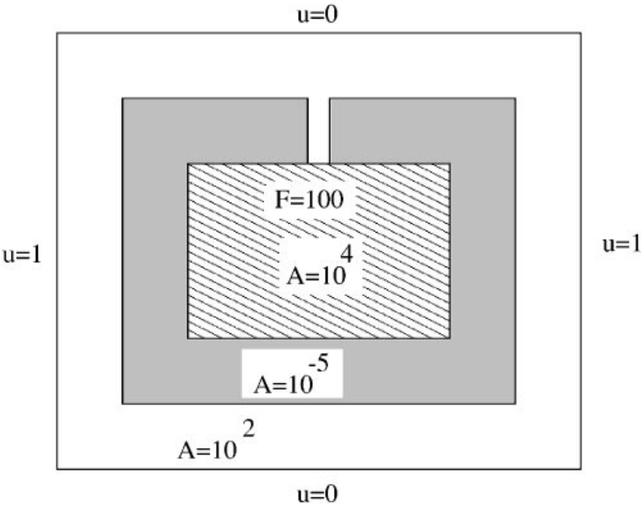

FIG. 11. The coefficients and geometry of the groundwater problem.

on the structure shown in Fig. 11. The solution satisfies the equation

$$-\nabla \cdot (A\nabla u) + B(x, y)u_x = F, \quad \text{where} \quad B(x, y) = 2e^{2(x^2+y^2)}.$$

The coefficient A is defined as shown in Fig. 11. The function F is everywhere zero except in the center section where $F = 100$. We have Dirichlet conditions on the complete outer

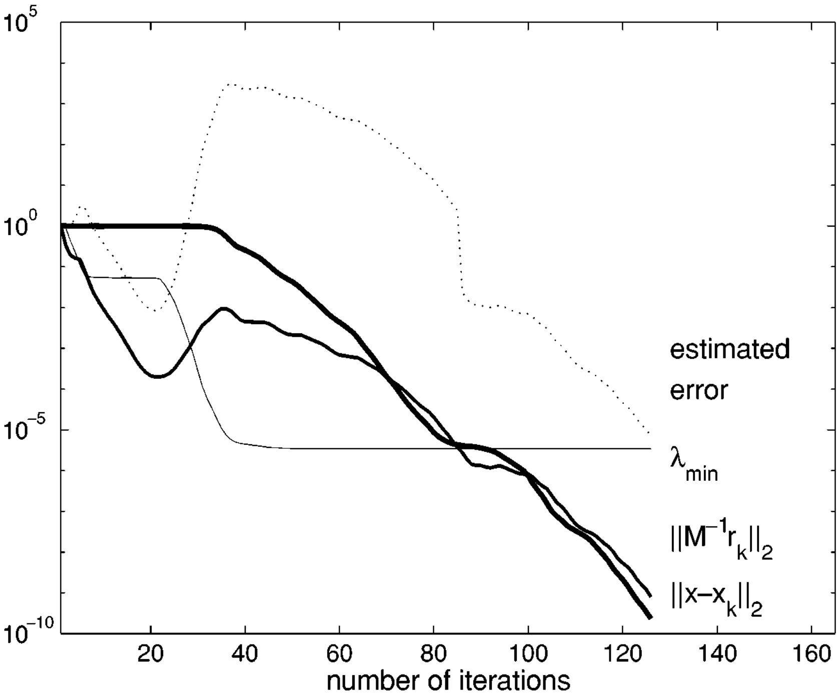

FIG. 12. The convergence behavior for ICCG.

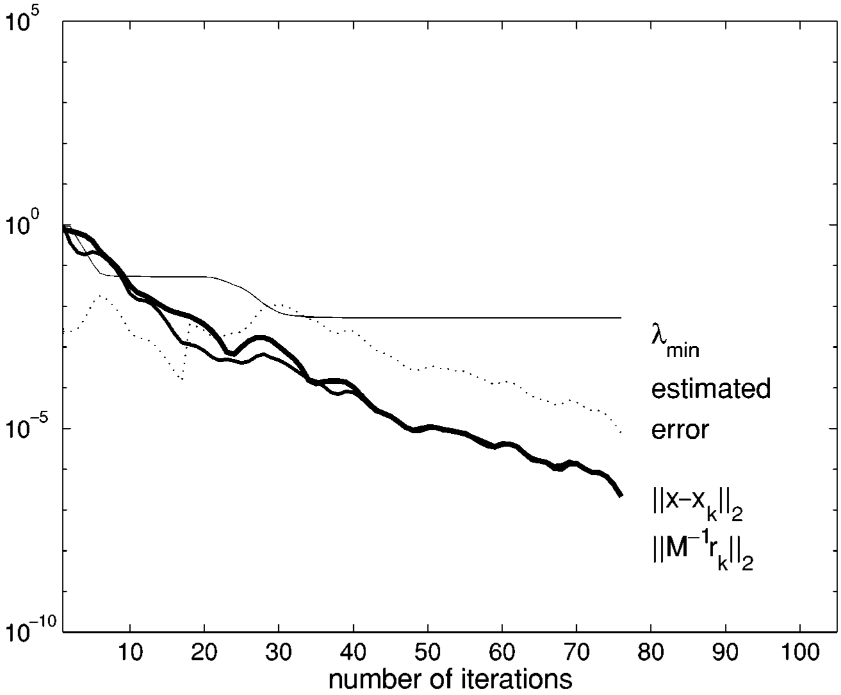

FIG. 13. The convergence behavior for DICCG.

boundary. In [17], BI-CGSTAB and CGS have been used to solve the discretized system. In both cases, an incomplete LU-factorization has been used as preconditioner. If we take a look at the convergence behavior of BI-CGSTAB and CGS as reported in [17], there is a strong resemblance with the convergence behavior of ICCG applied to layered problems.

Since at this moment we are only interested in symmetric problems we will analyze the Poisson equation: $-\nabla \cdot (A\nabla u) = F$. Like before, the smallest eigenvalue of the discretized system has been calculated. Both the presence of the clay section ($A = 10^{-5}$) as well as the jump in permeabilities between the two sand sections have an influence on the smallest eigenvalue. To annihilate the effect of the smallest eigenvalue on the convergence, DICCG has been applied to this problem. The projection vector has been constructed by neglecting the small gap in the low-permeability layer. The convergence behavior for ICCG and DICCG (applied to the original geometry) is plotted in Figs. 12 and 13. The convergence behavior of DICCG is much better than that of ICCG. The number of iterations decreases with a factor of two, and a proper termination criterion can be used.

Note that in the shaded region we use a coefficient $A = 10^4$, while in the outer region $A = 10^2$. The reason for this is that the erratic behavior we observe from the convergence of ICCG is only present if these coefficients differ. If both values of A are equal, no small eigenvalues are present. See also the remark given in Section 4.3.

6. CONCLUSIONS

It has been shown that the Incomplete Cholesky (IC) preconditioned matrix is a scaling invariant. This property is used to show that the number of small eigenvalues of the IC preconditioned matrix is equal to the number of high-permeability domains, which are not connected to a Dirichlet boundary.

A detailed description has been given as to how to construct the projection vectors in a cheap way. A proof is given to show that the span of these vectors approximates the "small" eigenspace of the diagonal or IC preconditioned matrix. This implies that the convergence behavior of DICCG is independent of the size of the jump in the coefficients.

It has been shown that perturbations of the projection vectors in the low-permeability part have only a limited influence on the convergence properties of DICCG. This has important consequences for the efficiency of the method:

- It is sufficient to compute a low-accuracy solution of the subdomain problems, which are used in the construction of the projection vectors.
- Small components of the projection vectors can be neglected to save work and memory requirements.

The use of our projection vectors, in combination with the DICCG method, makes the solver robust for elliptic problems with highly discontinuous coefficients. For this kind of problem, a robust stopping criterion is available, which is not the case for the ICCG method. For high accuracies, the DICCG method converges considerably faster than the ICCG method. However, for practical accuracies, the CPU time decreases with a factor of 10 to 20. This means that in the context of nonlinear problems or time-dependent problems, DICCG is far superior above ICCG. It has been shown that the construction of the projection vectors can be done fully automatically and that the method can be applied to practical problems. We conclude that DICCG, ignoring of the small components in the projection vectors, is a very robust and extremely efficient method to solve problems with extreme contrasts in the coefficients.

ACKNOWLEDGMENTS

We thank the referees for helpful suggestions, which have improved the paper considerably, and Fred Vermolen for posing the problem with variable anisotropic permeabilities.

REFERENCES

1. R. E. Bank and C. Wagner, Multilevel ILU decomposition, *Numer. Math.* **82**, 543 (1999).
2. J. Bear, *Dynamics of Fluids in Porous Media* (Elsevier Amsterdam/New York, 1972).
3. C. H. Bischof, J. G. Lewis, and D. J. Pierce, Incremental condition estimation for sparse matrices, *SIAM J. Matrix Anal. Appl.* **11**, 644 (1990).
4. X. Cai, B. F. Nielsen, and A. Tveito, An analysis of a preconditioner for the discretized pressure equation arising in reservoir simulation, *IMA J. Numer. Anal.* **19**, 291 (1999).
5. A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson, An estimate for the condition number of a matrix, *SIAM J. Numer. Anal.* **16**, 368 (1979).
6. H. De Gersem and K. Hameyer, A deflated iterative solver for magnetostatic finite element models with large differences in permeability, *Eur. Phys. J. Appl. Phys.* **13**, 45 (2000).
7. Melvyn R. Giles, Diagenesis: A Quantitative Perspective; Implications for Basin Modelling and Rock Property Prediction (Kluwer Academic, Dordrecht/Norwell, MA) 1997.
8. G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins Press, Baltimore, 1996).
9. I. G. Graham and M. J. Hagger, Unstructured additive Schwarz-conjugate gradient method for elliptic problems with highly discontinuous coefficients, *SIAM J. Sci. Comp.* **20**, 2041 (1999).

10. E. F. Kaasschieter, A practical termination criterion for the conjugate gradient method, *BIT* **28**, 308 (1988).
11. E. F. Kaasschieter, Preconditioned conjugate gradients for solving singular systems, *J. Comp. Appl. Math.* **24**, 265 (1988).
12. G. M. Kobelkov, On the solution of the boundary value problem for the diffusion equation with highly varying coefficient, *Russian J. Numer. Anal. and Math. Model.* **11**, 487 (1996).
13. J. A. Meijerink and H. A. Van der Vorst, An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix, *Math. Comp.* **31**, 148 (1977).
14. B. F. Nielsen and A. Tveito, On the approximation of the solution of the pressure equation by changing the domain, *SIAM J. Appl. Math.* **57**, 15 (1997).
15. A. Padiy, O. Axelsson, and B. Polman, Generalized augmented matrix preconditioning approach and its application to iterative solution of ill-conditioned algebraic systems, *SIAM J. Matrix Anal. Appl.* **22**, 793 (2000).
16. R. C. Reid, J. M. Prausnitz, and B. E. Poling, *The Properties of Gases and Liquids*, 4th ed. (McGraw-Hill, New York, 1987).
17. H. A. Van der Vorst, Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for solution of non-symmetric linear systems, *SIAM J. Sci. Stat. Comp.* **13**, 631 (1992).
18. M. A. Vegas-Landeau, R. Propp, T. W. Patzek, and P. Colella, Sequential semi-implicit algorithm for computing discontinuous flows in porous media, *Soc. Pet. Eng. J.* **June**, 200 (1998).
19. C. Vuik, A. Segal, and J. A. Meijerink, An efficient preconditioned CG method for the solution of a class of layered problems with extreme contrasts in the coefficients, *J. Comput. Phys.* **152**, 385 (1999).
20. W. L. Wan, Interface preserving coarsening multigrid for elliptic problems with highly discontinuous coefficients, *Num. Lin. Alg. Appl.* **7**, 727 (2000).