DELFT UNIVERSITY OF TECHNOLOGY

REPORT 03-10

A comparison of Deflation and Coarse Grid Correction applied to porous media flow

R. Nabben, C. Vuik

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2003

A comparison of Deflation and Coarse Grid Correction applied to porous media flow

R. Nabben*   C. Vuik†

June 26, 2003

Abstract

In this paper we compare various preconditioners for the numerical solution of partial differential equations. We compare a coarse grid correction preconditioner used in domain decomposition methods with a so-called deflation preconditioner. We prove that the effective condition number of the deflated preconditioned system is always, i.e. for all deflation vectors and all restrictions and prolongations, below the condition number of the system preconditioned by the coarse grid correction. This implies that the Conjugate Gradient method applied to the deflated preconditioned system converges always faster than the Conjugate Gradient method applied to the system preconditioned by the coarse grid correction. Numerical results for porous media flows emphasize the theoretical results.

Keywords. deflation, coarse grid correction, preconditioners, Conjugate Gradients, porous media flow, scalable parallel preconditioner

AMS subject classifications. 65F10, 65F50, 65N22

1 Introduction

It is well known that the convergence rate of the Conjugate Gradient method is bounded as a function of the condition number of the system matrix to which it is applied. Let $A \in \mathbb{R}^{n \times n}$ be symmetric positive definite. We assume that the vector $b \in \mathbb{R}^n$ represents a discrete function on a grid Ω and that we are searching for the vector $x \in \mathbb{R}^n$ on Ω which solves the linear system

$$Ax = b.$$

Such systems are encountered, for example, when a finite volume/difference/element method is used to discretize an elliptic partial differential equation defined on the continuous analog of Ω.

Let us denote the ith eigenvalue in nondecreasing order by $\lambda_i(A)$ or simply by λ_i when it is clear to which matrix we are referring. After k iterations of the Conjugate Gradient method, the error is bounded by (cf. [8], Thm. 10.2.6):

$$\|x - x_k\|_A \le 2\,\|x - x_0\|_A \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \tag{1}$$

where $\kappa = \kappa(A) = \lambda_n / \lambda_1$ is the spectral condition number of A and the A-norm of x is given by $\|x\|_A = (x^T A x)^{1/2}$. The convergence may be significantly faster if the eigenvalues of A are clustered [23].

*Fakultät für Mathematik Universität Bielefeld Postfach 100131, 33501 Bielefeld, Germany, phone: 49 521 106 4785, fax: 49 521 106 6498 (nabben@mathematik.uni-bielefeld.de)

†Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Applied Mathematical Analysis, P.O. Box 5031, 2600 GA Delft, The Netherlands, phone: 31 15 2785530, fax: 31 15 2787209 (c.vuik@math.tudelft.nl)

If the condition number of A is large it is advisable to solve, instead, a preconditioned system $M^{-1}Ax = M^{-1}b$, where the symmetric positive definite preconditioner M is chosen such that $M^{-1}A$ has a more clustered spectrum or a smaller condition number than that of A. Furthermore, M must be cheap to solve relative to the improvement it provides in convergence rate. A final desirable property in a preconditioner is that it should parallelize well, especially on distributed memory computers. Probably the most effective preconditioning strategy in common use is to take $M = LL^T$ to be an incomplete Cholesky (IC) factorization of A [14]. We denoted the Preconditioned Conjugate Gradient method as PCG.

With respect to the known preconditioners at least two problems remain:

- if there are large jumps in the coefficients of the discretized PDE the convergence of PCG becomes very slow,
- if a block preconditioner is used in a domain decomposition algorithm the condition number of the preconditioned matrix deteriorates if the number of blocks increases.

Both problems can be solved by a deflation technique or a suitable coarse grid correction. In this section we describe both methods, which are compared in the next sections. To describe the deflation method we define the projection P_D by

$$P_D = I - AZ(Z^TAZ)^{-1}Z^T, \quad Z \in \mathbb{R}^{n\times r}, \tag{2}$$

where the column space of Z is the deflation subspace, i.e. the space to be projected out of the residual, and I is the identity matrix of appropriate size. We assume that $r \ll n$ and that Z has rank r. Under this assumption $E \equiv Z^TAZ$ may be easily computed and factored and is symmetric positive definite. Since $x = (I - P_D^T)x + P_D^Tx$ and because

$$(I - P_D^T)x = Z(Z^TAZ)^{-1}Z^TAx = ZE^{-1}Z^Tb \tag{3}$$

can be immediately computed, we need only compute P_D^Tx. In light of the identity $AP_D^T = P_DA$, we can solve the deflated system

$$P_DA\tilde{x} = P_Db \tag{4}$$

for $\tilde{x}$ using the Conjugate Gradient method, premultiply this by P_D^T and add it to (3).

Obviously (4) is singular. What consequences does the singularity of (4) imply for the Conjugate Gradient method? Kaasschieter [9] notes that a positive semidefinite system can be solved as long as the right-hand side is consistent (i.e. as long as $b = Ax$ for some x). This is certainly true for (4), where the same projection is applied to both sides of the nonsingular system. Furthermore, he notes (with reference to [23]) that because the null space never enters the iteration, the corresponding zero-eigenvalues do not influence the convergence. Motivated by this fact, we define the *effective condition number* of a positive semidefinite matrix $C \in \mathbb{R}^{n\times n}$ with r zero eigenvalues to be the ratio of its largest to smallest *positive* eigenvalues:

$$\kappa_{\text{eff}}(C) = \frac{\lambda_n}{\lambda_{r+1}}.$$

It is possible to combine both a standard preconditioning and preconditioning by deflation (for details see [7]). The convergence is then described by the effective condition number of $M^{-1}P_DA$.

The deflation technique has been exploited by several authors. For nonsymmetric systems, approximate eigenvectors can be extracted from the Krylov subspace produced by GMRES. Morgan [15] uses this approach to improve the convergence after a restart. In this case, deflation is not applied as a preconditioner, but the deflation vectors are augmented with the Krylov subspace and the minimization property of GMRES ensures that the deflation subspace is projected out of the residual (for related references we refer to [7]). A comparable approach for the CG method is described in [21]. Mansfield [12] shows how Schur complement-type domain decomposition methods can be seen as a series of deflations. Nicolaides [17] chooses Z to be a piecewise constant interpolation from a set of r subdomains and points out that deflation might be effectively used with a conventional preconditioner. Mansfield [13] uses the same "subdomain deflation" in combination with damped Jacobi smoothing, obtaining a preconditioner which is related to the two-grid method. In [11] Kolotilina uses a twofold deflation technique for simultaneously deflating the r largest and the r smallest eigenvalues using an appropriate deflating subspace of dimension r. Other authors have attempted to choose a subspace *a priori* that effectively represents the slowest modes. In [25] deflation is used to remove a few stubborn but *known* modes from the spectrum. This method is used in [3] to solve electromagnetic problems with large jumps in the coefficients. Thereafter this method has been generalized to other choices of the deflation vectors [26, 27]. Finally an analysis of the effective condition number and a parallel implementation is given in [7, 24].

We compare the deflation preconditioner with a well-known coarse grid correction preconditioner of the form

$$P_C = I + ZE^{-1}Z^T \tag{5}$$

and in the preconditioned case

$$P_{CM^{-1}} = M^{-1} + ZE^{-1}Z^T. \tag{6}$$

In the multigrid or domain decomposition language the matrices Z and Z^T are known as restriction and prolongation or interpolation operator. Moreover, the matrix $E = Z^TAZ$ is the Galerkin operator.

The above coarse grid correction preconditioner belongs to the class of additive Schwarz preconditioner. It is called the two level additive Schwarz preconditioner. If used in domain decomposition methods, typically, M^{-1} is the sum of the local (exact or inexact) solves in each domain. To speed up convergence a coarse grid correction $ZE^{-1}Z^T$ is added.

These methods are introduced by Bramble, Paschiak and Schatz [2] and Dryja and Widlund [5] [6] [4]. They show under mild conditions that the convergence rate of the PCG method is independent of the grid sizes.

For more details about this preconditioner we refer to the books of Quarteroni and Valli [19], and Smith, Bjørstad and Gropp [22]. A more abstract analysis of this preconditioner is given by Padiy, Axelsson and Polman [18], recently. To make the condition number of $P_{CM^{-1}}A$ smaller Padiy, Axelsson and Polman used a parameter $\sigma > 0$ and considered

$$P_C = I + \sigma ZE^{-1}Z^T \tag{7}$$

and

$$P_{CM^{-1}} = M^{-1} + \sigma ZE^{-1}Z^T. \tag{8}$$

If $M = I$, Z consists of eigenvectors and λ_{max} is known, then a good choice is $\sigma = \lambda_{max}$ which implies that $\kappa(P_CA) \leq \frac{2\lambda_{max}}{\lambda_{r+1}}$ [18]. If $M \neq I$ and/or Z consists of general vectors and λ_{max} is not known it is not clear how to choose σ.

More abstract results about Schwarz methods applied to nonsymmetric problems are given by Benzi, Frommer, Nabben and Szyld [1] and [16].

In this article we prove that the effective condition number of the deflated preconditioned system $M^{-1}P_DA$ is always below the condition number of the system preconditioned by the coarse grid correction $P_{CM^{-1}}A$. This implies that for all matrices $Z \in \mathbb{R}^{n\times r}$ and all positive definite preconditioners M^{-1} the Conjugate Gradient method applied to the deflated preconditioned system converges always faster than the Conjugate Gradient method applied to the system preconditioned by the coarse grid correction. These results are stated in Section 2. In Section 3 we compare other properties of the deflation and coarse grid preconditioner. These properties are scaling, approximation of E^{-1} and an estimate of the smallest eigenvalue. Section 4 contains our numerical results for porous media flows and parallel problems.

2 Spectral properties

In this section we compare the effective condition number for the deflation and coarse grid correction preconditioned matrices. In Section 2.1 we give some definitions and preliminary results. Thereafter a comparison is made if the projection vectors are equal to eigenvectors in Section 2.2 and for general projection vectors in Section 2.3.

2.1 Notations and Preliminary Results

In the following we denote by $\lambda_i(M)$ the eigenvalues of a matrix M. If the eigenvalues are real the $\lambda_i(M)'s$ are ordered increasingly.

For two Hermitian $n \times n$ matrices A and B we write $A \succeq B$, if $A - B$ is positive semidefinite. Next we mention well-known properties of the eigenvalues of Hermitian matrices.

Lemma 2.1 *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian. For each $k = 1, 2, \ldots, n$ we have*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B)$$

From the above lemma we easily obtain the next lemma.

Lemma 2.2 *If $A, B \in \mathbb{C}^{n\times n}$ are positive semidefinite with $A \succeq B$, then $\lambda_i(A) \geq \lambda_i(B)$.*

Moreover, we will use

Lemma 2.3 *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian and suppose that B has rank at most r. Then*

- $\lambda_k(A+B) \leq \lambda_{k+r}(A), \quad k = 1, 2, \cdots n - r,$
- $\lambda_k(A) \leq \lambda_{k+r}(A+B), \quad k = 1, 2, \cdots n - r.$

Lemma 2.1, Lemma 2.2 and Lemma 2.3 can be found e.g. as Theorem 4.3.1, Corollary 7.7.4. and Theorem 4.3.6, respectively, in [20].

2.2 Projection vectors are eigenvectors

In this section we compare the effective condition number of $P_D A$ and $P_C A$ if the projection vectors are equal to eigenvectors of A.

Definition 2.4 *Choose the eigenvectors v_k of A such that $v_k^T v_j = \delta_{kj}$, and define $Z = [v_1 \ldots v_r]$.*

Theorem 2.5 *Using Z as given in Definition 2.4 the spectra of $P_D A$ and $P_C A$ given in (2) and (7) are:*

$$spectrum(P_D A) = \{0, \ldots, 0, \lambda_{r+1}, \ldots, \lambda_n\}, \text{ and}$$

$$spectrum(P_C A) = \{\sigma + \lambda_1, \ldots, \sigma + \lambda_r, \lambda_{r+1}, \ldots, \lambda_n\}.$$

Proof: For this choice of Z it appears that

$$E = Z^T A Z = diag(\lambda_1, \ldots, \lambda_r). \tag{9}$$

To proof the first part we note that (9) implies $P_D = I - AZE^{-1}Z^T = I - ZZ^T$. Consider $P_D A v_k = (I - ZZ^T)\lambda_k v_k$ for $k = 1, \ldots, n$. Since $ZZ^T v_k = v_k$, for $k = 1, \ldots, r$ and $ZZ^T v_k = 0$ for $k = r+1, \ldots, n$ it is easy to show that

$$P_D A v_k = 0, \text{ for } k = 1, \ldots, r, \text{ and } P_D A v_k = \lambda_k v_k, \text{ for } k = r+1, \ldots, n,$$

which proofs the first part.

Secondly we consider $P_C A v_k$. For $k = 1, \ldots, r$ we obtain

$$P_C A v_k = (I + \sigma Z \, diag(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_r}) \, Z^T)\lambda_k v_k = (\sigma + \lambda_k) v_k,$$

whereas for $k = r+1, \ldots, n$ it appears that

$$P_C A v_k = (I + \sigma Z \, diag(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_r}) \, Z^T)\lambda_k v_k = \lambda_k v_k,$$

since $Z^T v_k = 0$ for $k = r+1, \ldots, n$. This proofs the second part (cf. Theorem 2.6 in [18]). □

In order to compare both approaches we note that

$$\kappa_{eff}(P_D A) = \frac{\lambda_n}{\lambda_{r+1}}, \tag{10}$$

and

$$\kappa(P_C A) = \frac{\max\{\sigma + \lambda_r, \lambda_n\}}{\min\{\sigma + \lambda_1, \lambda_{r+1}\}}. \tag{11}$$

From (10) and (11) it follows that $\kappa(P_C A) \geq \kappa_{eff}(P_D A)$, so the convergence bound based on the effective condition number implies that Deflated CG converges faster than CG combined with coarse grid correction if both methods use the eigenvectors corresponding to the r smallest eigenvalues as projection vectors.

2.3 Projection vectors are general vectors

In the last section we showed that the deflation technique is better then a coarse grid correction, if eigenvectors are used. However, computing the r smallest eigenvalues is mostly very expensive. Moreover, in multigrid methods and domain decomposition methods special interpolation and prolongation matrices are used to obtain grid independent convergence rates. So a comparison only for eigenvectors is not enough. But in this section we generalize the results of Section 2.2. We prove that the effective condition number of the deflated preconditioned system is always, i.e. for all matrices $Z \in \mathbb{R}^{n \times r}$, below the condition number of the system preconditioned by the coarse grid correction.

Theorem 2.6 *Let $A \in \mathbb{R}^{n \times n}$ be symmetric positive definite. Let $Z \in \mathbb{R}^{n \times r}$ with $rank Z = r$. Then the preconditioner defined in* (2) *and* (7) *satisfies*

$$\lambda_1(P_D A) = \cdots = \lambda_r(P_D A) \;=\; 0 \tag{12}$$

$$\lambda_n(P_D A) \;\leq\; \lambda_n(P_C A) \tag{13}$$

$$\lambda_{r+1}(P_D A) \;\geq\; \lambda_1(P_C A). \tag{14}$$

Proof: Obviously all eigenvalues of $P_C A$ are real and positive. With Lemma 2.1 of [7] $P_D A$ is positive semidefinite. Thus, all eigenvalues of $P_D A$ are real and nonnegative. Since $P_D A Z = 0$ statement (12) holds.

We obtain

$$A^{\frac{1}{2}} P_C A^{\frac{1}{2}} - P_D A = AZE^{-1}Z^T A + \sigma A^{\frac{1}{2}} Z E^{-1} Z^T A^{\frac{1}{2}}.$$

The right hand side is positive semidefinite. Thus, we have with Lemma 2.2

$$\lambda_i(P_C A) = \lambda_i(A^{\frac{1}{2}} P_C A^{\frac{1}{2}}) \geq \lambda_i(P_D A).$$

Hence, (13) holds. Next consider

$$\begin{aligned} P_C A P_C - P_D A &= A + \sigma Z E^{-1} Z^T A + \sigma A Z E^{-1} Z^T + \sigma^2 Z E^{-1} Z^T A Z E^{-1} Z^T \\ &\quad - A + A Z E^{-1} Z^T A \\ &= \sigma Z E^{-1} Z^T A + \sigma A Z E^{-1} Z^T + \sigma^2 Z E^{-1} Z^T + A Z E^{-1} Z^T A \\ &= (A + \sigma I) Z E^{-1} Z^T (A + \sigma I). \end{aligned}$$

Thus, $P_CAP_C - P_DA$ is symmetric and of rank r. Using Lemma 2.3 we obtain

$$\lambda_{r+1}(P_DA) \geq \lambda_1(P_CAP_C) = \lambda_1(P_C^2A).$$

But since $P_C - I$ is positive semidefinite, $P_C^2 - P_C$ and $A^{\frac{1}{2}}P_C^2A^{\frac{1}{2}} - A^{\frac{1}{2}}P_CA^{\frac{1}{2}}$ are positive semidefinite also. Hence,

$$\lambda_i(P_C^2A) = \lambda_i(A^{\frac{1}{2}}P_C^2A^{\frac{1}{2}}) \geq \lambda_i(A^{\frac{1}{2}}P_CA^{\frac{1}{2}}) = \lambda_i(P_CA).$$

Thus,

$$\lambda_{r+1}(P_DA) \geq \lambda_1(P_C^2A) \geq \lambda_1(P_CA).$$

□

It follows from Theorem 2.6 that

$$\kappa(P_CA) \geq \kappa_{eff}(P_DA)$$

so the convergence bound based on the effective condition number implies that Deflated CG converges faster than CG combined with coarse grid correction for arbitrary matrices $Z \in \mathbb{R}^{n\times r}$.

In Theorem 2.11 we will extend this result to the preconditioned versions of the deflation and coarse grid correction preconditioners.

Before that, we will show how the deflated preconditioner behave if we increase the number of deflation vectors. In detail we will show that the effective condition number decrease if we use a matrix Z_2 in (2) satisfying $ImZ \subseteq ImZ_2$ rather than Z. To do so we need several Lemmata.

The first Lemma is probably well-known, but for completeness we give the proof here.

Lemma 2.7 *Let $A \in \mathbb{R}^{n\times n}$ be nonsingular and be partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $A_{11} \in M_r(R)$ and $A_{22} \in M_{n-r}(R)$. Assume that A_{11} is nonsingular. Define

$$\tilde{A}_{11}^{-1} := \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Then, $rank(A^{-1} - \tilde{A}_{11}^{-1}) = n - r$.

Proof: The inverse of A is given by

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix},$$

where $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Hence

$$\begin{aligned} A^{-1} - \tilde{A}_{11}^{-1} &= \begin{bmatrix} A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1} \end{bmatrix} \left[A_{21}A_{11}^{-1}, -I\right]. \end{aligned}$$

Since S and the $n-r \times n-r$ identity matrix I have rank $n-r$ we get $rank(A^{-1} - \tilde{A}_{11}^{-1}) = n-r$. □

In the next lemma we compare the preconditioned matrices if a different number of deflation vectors is used.

Lemma 2.8 *Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite. Let $Z_1 \in \mathbb{R}^{n\times r}$ and $Z_2 \in \mathbb{R}^{n\times s}$ with $rank Z_1 = r$ and $rank Z_2 = s$. Define $E_1 := Z_1^T A Z_1$ and $E_2 := Z_2^T A Z_2$. If $Im Z_1 \subseteq Im Z_2$, then*

$$(I - AZ_1E_1^{-1}Z_1^T)A \succeq (I - AZ_2E_2^{-1}Z_2^T)A.$$

Proof: It suffices to prove that

$$Z_2E_2^{-1}Z_2^T \succeq Z_1E_1^{-1}Z_1^T.$$

Since $Im Z_1 \subseteq Im Z_2$, there exists a matrix $T \in M_{s\times r}(R)$ such that

$$Z_1 = Z_2T.$$

Therefore

$$\begin{aligned} Z_2E_2^{-1}Z_2^T - Z_1E_1^{-1}Z_1^T &= Z_2(E_2^{-1} - TE_1^{-1}T^T)Z_2^T \\ &= Z_2E_2^{-\frac{1}{2}}(I - E_2^{\frac{1}{2}}TE_1^{-1}T^TE_2^{\frac{1}{2}})E_2^{-\frac{1}{2}}Z_2^T. \end{aligned}$$

But $E_2^{\frac{1}{2}}TE_1^{-1}T^TE_2^{\frac{1}{2}}$ is an orthogonal projection. Thus $E_2^{\frac{1}{2}}TE_1^{-1}T^TE_2^{\frac{1}{2}}$ has the only eigenvalues 0 and 1. Hence $I - E_2^{\frac{1}{2}}TE_1^{-1}T^TE_2^{\frac{1}{2}}$ is positive semidefinite. Therefore

$$Z_2E_2^{-1}Z_2^T \succeq Z_1E_1^{-1}Z_1^T.$$

□

In the next lemma we show that $P_{D_1}A = P_{D_2}A$, if $Im Z_1 = Im Z_2$.

Lemma 2.9 *Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite. Let $Z_1 \in \mathbb{R}^{n\times r}$ and $Z_2 \in \mathbb{R}^{n\times r}$ with $rank Z_1 = rank Z_2 = r$. Define $E_1 := Z_1^T A Z_1$ and $E_2 := Z_2^T A Z_2$. If $Im Z_1 = Im Z_2$, then*

$$(I - AZ_1E_1^{-1}Z_1^T)A = (I - AZ_2E_2^{-1}Z_2^T)A.$$

Proof: We can follow the proof of Lemma 2.8. Since $Im Z_1 = Im Z_2$, the matrix T is nonsingular. Hence

$$\begin{aligned} Z_2E_2^{-1}Z_2^T - Z_1E_1^{-1}Z_1^T &= Z_2(E_2^{-1} - TE_1^{-1}T^T)Z_2^T \\ &= Z_2E_2^{-\frac{1}{2}}(I - E_2^{\frac{1}{2}}TE_1^{-1}T^TE_2^{\frac{1}{2}})E_2^{-\frac{1}{2}}Z_2^T. \\ &= Z_2E_2^{-\frac{1}{2}}(I - E_2^{\frac{1}{2}}T(T^TE_2T)^{-1}T^TE_2^{\frac{1}{2}})E_2^{-\frac{1}{2}}Z_2^T \\ &= Z_2E_2^{-\frac{1}{2}}(I - E_2^{\frac{1}{2}}TT^{-1}E_2^{-1}T^{-T}T^TE_2^{\frac{1}{2}})E_2^{-\frac{1}{2}}Z_2^T \\ &= 0. \end{aligned}$$

□

Using the above lemmata, we can prove the following theorem

Theorem 2.10 *Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite. Let $Z_1 \in \mathbb{R}^{n\times r}$ and $Z_2 \in \mathbb{R}^{n\times s}$ with $rank Z_1 = r$ and $rank Z_2 = s$. Let $E_1 := Z_1^T A Z_1$ and $E_2 := Z_2^T A Z_2$. If $Im Z_1 \subseteq Im Z_2$, then*

$$\lambda_n((I - AZ_1E_1^{-1}Z_1^T)A) \geq \lambda_n((I - AZ_2E_2^{-1}Z_2^T)A) \tag{15}$$

$$\lambda_{r+1}((I - AZ_1E_1^{-1}Z_1^T)A) \leq \lambda_{s+1}((I - AZ_2E_2^{-1}Z_2^T)A) \tag{16}$$

Proof: With Lemma 2.2 and Lemma 2.8 we obtain inequality (15).

Next we will prove (16). Observe that $Z_1E_1^{-1}Z_1^T$ and $Z_2E_2^{-1}Z_2^T$ are invariant under permutations of the columns of Z_1 and Z_2 respectively.

Thus, using Lemma 2.9, we can assume without loss of generality that $Z_2 = [Z1, D]$, with $D \in \mathbb{R}^{n\times s-r}$.

Moreover, define the $s \times s$ matrix

$$\tilde{E}_1^{-1} = \begin{bmatrix} E_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Obviously, we then obtain

$$Z_1E_1^{-1}Z_1^T = Z_2\tilde{E}_1^{-1}Z_2^T.$$

Thus,

$$\begin{aligned}(I - AZ_2E_2^{-1}Z_2^T)A - (I - AZ_1E_1^{-1}Z_1^T)A &= A(Z_1E_1^{-1}Z_1^T - Z_2E_2^{-1}Z_2^T)A \\ &= A(Z_2\tilde{E}_1^{-1}Z_2 - Z_2E_2^{-1}Z_2^T)A \\ &= AZ_2(\tilde{E}_1^{-1} - E_2^{-1})Z_2^TA.\end{aligned}$$

But since E_1 is the leading principal $r \times r$ submatrix of E_2 we can apply Lemma 2.7. Thus $(I - AZ_2E_2^{-1}Z_2^T)A - (I - AZ_1E_1^{-1}Z_1^T)A$ is of rank $s - r$. Hence with Lemma 2.3

$$\lambda_{r+1}((I - AZ_1E_1^{-1}Z_1^T)A) \le \lambda_{s+1}((I - AZ_2E_2^{-1}Z_2^T)A).$$

□

Next we include an additional symmetric positive definite preconditioner M^{-1}. We then consider the coarse grid preconditioner

$$P_{CM^{-1}} := M^{-1} + \sigma ZE^{-1}Z^T. \tag{17}$$

This type of preconditioner includes a lot of well-known preconditioners. It belongs to the class of additive Schwarz preconditioner and is called the two level additive Schwarz preconditioner. If used in domain decomposition methods, typically, M^{-1} is the sum of the local (exact or inexact) solves in each domain. To speed up convergence a coarse grid correction $ZE^{-1}Z^T$ is added. Notice, that the BPS preconditioner introduced by Bramble, Paschiak and Schatz [2] and Dryja and Widlund [5] [6] [4] are of the same type. They show under mild conditions that the convergence rate of the PCG method is independent of the grid sizes.

We compare the preconditioner (17) with the corresponding deflated preconditioner

$$M^{-1}P_D. \tag{18}$$

We obtain

Theorem 2.11 *Let $A \in \mathbb{R}^{n\times n}$ and $M \in \mathbb{R}^{n\times n}$ be symmetric positive definite. Let $Z \in \mathbb{R}^{n\times r}$ with $rankZ = r$. Then*

$$\lambda_n(M^{-1}P_DA) \le \lambda_n(P_{CM^{-1}}A), \tag{19}$$

$$\lambda_{r+1}(M^{-1}P_DA) \ge \lambda_1(P_{CM^{-1}}A). \tag{20}$$

Proof: First observe that Theorem 2.6 still holds if we replace A everywhere by $L^{-1}AL^{-T}$ with an arbitrary nonsingular matrix L. Here, we will consider $M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$. The idea is to transform P_D and P_C to this form. We start with

$$M^{-1}P_DA = M^{-1}(A - AZE^{-1}Z^TA).$$

The eigenvalues of this matrix are the same as the eigenvalues of

$$M^{-\frac{1}{2}}P_D AM^{-\frac{1}{2}} = M^{-\frac{1}{2}}(A - AZE^{-1}Z^TA)M^{-\frac{1}{2}}.$$

Define the matrix G such that $G = M^{\frac{1}{2}}Z$ and thus $Z = M^{-\frac{1}{2}}G$. Substituting this in the previous matrix leads to $E = Z^TAZ = G^TM^{-\frac{1}{2}}AM^{-\frac{1}{2}}G$ and

$$M^{-\frac{1}{2}}P_D AM^{-\frac{1}{2}} = M^{-\frac{1}{2}}(A - AM^{-\frac{1}{2}}GE^{-1}G^TM^{-\frac{1}{2}}A)M^{-\frac{1}{2}} =$$

$$(I - M^{-\frac{1}{2}}AM^{-\frac{1}{2}}GE^{-1}G^T)M^{-\frac{1}{2}}AM^{-\frac{1}{2}},$$

which is in the required form.

In the same way we can transform $P_{CM^{-1}}A = (M^{-1} + \sigma ZE^{-1}Z^T)A$ to

$$P_{CM^{-1}}A = M^{-1}A + \sigma M^{-\frac{1}{2}}GE^{-1}G^TM^{-\frac{1}{2}}A$$

which has the same eigenvalues as:

$$M^{-\frac{1}{2}}AM^{-\frac{1}{2}} + \sigma GE^{-1}G^TM^{-\frac{1}{2}}AM^{-\frac{1}{2}} = (I + \sigma GE^{-1}G^T)M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$$

which is also in the required form.

Thus, Theorem 2.6 gives the desired result. □

For the case $L^{-1}AL^{-T}$ the same result can be proved if one chooses $G = L^TZ$.

Theorem 2.11 describes the most general case. Arbitrary vectors or matrices $Z \in \mathbb{R}^{n\times r}$ combined with arbitrary preconditioners are considered. The effective condition number of the deflated CG method is always below the condition number of the CG method preconditioned by the coarse grid correction. Thus, the interpolation or prolongation matrices Z used e.g. in the BPS method give a better preconditioner if used in a deflation technique.

At the end of this section we generalize Theorem 2.10.

Theorem 2.12 *Let $A, M \in \mathbb{R}^{n\times n}$ be symmetric positive definite. Let $Z_1 \in \mathbb{R}^{n\times r}$ and $Z_2 \in \mathbb{R}^{n\times s}$ with $rank Z_1 = r$ and $rank Z_2 = s$. Let $E_1 := Z_1^TAZ_1$ and $E_2 := Z_2^TAZ_2$. If $Im Z_1 \subseteq Im Z_2$, then*

$$\begin{aligned} \lambda_n(M^{-1}(I - AZ_1E_1^{-1}Z_1^T)A) &\geq \lambda_n(M^{-1}(I - AZ_2E_2^{-1}Z_2^T)A) \\ \lambda_{r+1}(M^{-1}(I - AZ_1E_1^{-1}Z_1^T)A) &\leq \lambda_{s+1}(M^{-1}(I - AZ_2E_2^{-1}Z_2^T)A) \end{aligned}$$

Proof: The proof is almost the same as the proof of Theorem 2.10

3 Other properties of Deflation and Coarse Grid Correction

In this section we compare other properties of deflation and coarse grid correction. These properties are scaling, approximation of E^{-1} and an estimate of the smallest eigenvalue.

Scaling
Note that P_DA is scaling invariant whereas P_CA is not scaling invariant. This means that if deflation is applied to a system $\gamma Ax = \gamma b$ the effective condition number of $P_D\gamma A$ is independent of the scalar γ, whereas the condition number of $P_C\gamma A$ depends on the choice of γ.

Inaccurate solution
If the matrix E becomes large (so many projection vectors are used) it seems to be good to compute E^{-1} approximately (by an iterative method/or to do the procedure recursively). It appears that the coarse grid correction operator is insensitive for this approach, whereas deflation is sensitive for the accuracy of the approximation of E^{-1}. A proof of this property if the projection vectors are eigenvectors is given in the next lemma.

Lemma 3.1 *Use Z as given in Definition 2.4, and assume that*

$$\tilde{E}^{-1} = diag(\frac{1}{\lambda_1}(1+\epsilon_1), \ldots, \frac{1}{\lambda_r}(1+\epsilon_r))$$

is an approximation of E^{-1}, where $|\epsilon_i|$ is small. The spectra of $\tilde{P}_D A$ and $\tilde{P}_C A$ given in (2) and (5), where E^{-1} is replaced by $\tilde{E}^{-1}$ are:

$$spectrum(\tilde{P}_D A) = \{\lambda_1\epsilon_1, \ldots, \lambda_r\epsilon_r, \lambda_{r+1}, \ldots, \lambda_n\}, \text{ and}$$

$$spectrum(\tilde{P}_C A) = \{\sigma + \lambda_1 + \epsilon_1, \ldots, \sigma + \lambda_r + \epsilon_r, \lambda_{r+1}, \ldots, \lambda_n\}.$$

Proof: The proof of this lemma is almost the same as the proof of Theorem 2.5 □

For general vectors a similar situation appears. Assume that $\tilde{E}^{-1} = (I-F)E^{-1}(I-F)$ is a symmetric approximation ($F = F^T$) of E^{-1}. Let $H := -FE^{-1} - E^{-1}F + FE^{-1}F$. Then we have

$$\tilde{P}_D A = P_D A + AZHZ^T A.$$

Hence, using Lemma 2.1 we obtain

$$\lambda_k(P_D A) + \lambda_1(AZHZ^T A) \le \lambda_k(\tilde{P}_D A) \le \lambda_k(P_D A) + \lambda_n(AZHZ^T A).$$

Since the first r eigenvalues of $\lambda_k(P_D A)$ are 0, we get for $i = 1, \ldots, r$

$$\lambda_1(AZHZ^T A) \le \lambda_i(\tilde{P}_D A) \le \lambda_n(AZHZ^T A).$$

If all eigenvalues of $AZHZ^T A$ are small, the first r eigenvalues $\lambda_i(\tilde{P}_D A)$ are also very small. Observe that $\lambda_1(\tilde{P}_D A)$ can be negative if the perturbation H is negative definite.

For the coarse grid correction

$$\tilde{P}_C A = P_C A + ZHZ^T A.$$

we obtain

$$\lambda_k(P_C A) + \lambda_1(ZHZ^T A) \le \lambda_k(\tilde{P}_C A) \le \lambda_k(P_C A) + \lambda_n(ZHZ^T A).$$

Thus, if all eigenvalues of $ZHZ^T A$ are small, the perturbation has not much effect.

Hence the coarse grid correction operator is insensitive for the accuracy of the approximation, whereas deflation is sensitive.

To illustrate this we consider two problems. The first one is motivated by a porous media flow with large contrasts in the coefficients (ratio 10^{-6}, see Example 1 in Section 4) and the second one is a Poisson problem. In both examples $r(=7)$ *algebraic* projection vectors are used (see [27], Definition 4). We replace E^{-1} by $\tilde{E}^{-1} = (I+\epsilon R)E^{-1}(I+\epsilon R)$, where R is a symmetric $r \times r$ matrix with random elements chosen from the interval $[-\frac{1}{2}, \frac{1}{2}]$. From Figure 1 (porous media flow) it follows that the convergence of the error remains good for $|\epsilon| < 10^{-12}$. For larger values of $|\epsilon|$ we see that the convergence stagnates. For the Poisson problem it appears that the convergence is good as long as $|\epsilon| < 10^{-6}$ (Figure 2). For the coarse grid correction operator there is no difference in the convergence behavior. Using the coarse grid correction operator we need 75 iterations for the porous media flow problem and 70 iterations for the Poisson problem.

Estimate of smallest eigenvalue

In this paragraph we restrict ourselves to the case that the deflation vectors approximate the eigenvectors corresponding to the smallest eigenvalues. For a robust iterative solver it is important to have a good estimate of the smallest and largest eigenvalues of the (preconditioned) matrix. From the CG method an approximation of the extreme eigenvalues can be obtained from the

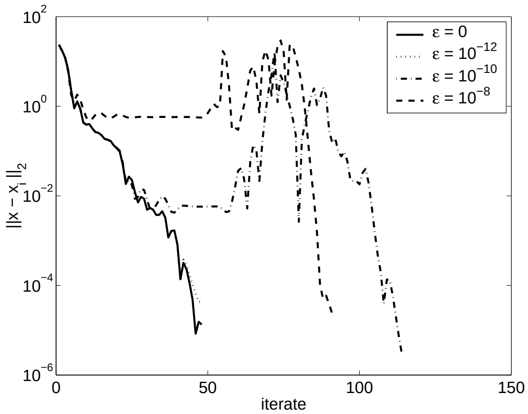

Figure 1: Convergence behavior of DICCG for the straight layers problem

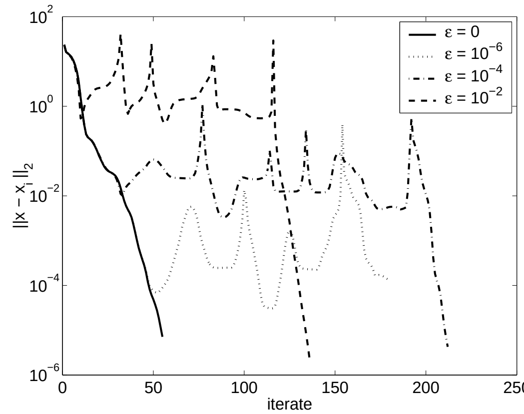

Figure 2: Convergence behavior of DICCG for the Poisson problem

Ritzvalues [10]. However using $M^{-1}P_DA$ the smallest eigenvalues of $M^{-1}A$ are removed. To obtain an estimate of these eigenvalues, we can use the matrix E as follows:

$$\lambda_{min}(M^{-\frac{1}{2}}AM^{-\frac{1}{2}}) \leq \min_{y\in\mathbb{R}^r} \frac{y^TG^TM^{-\frac{1}{2}}AM^{-\frac{1}{2}}Gy}{y^TG^TGy} = \min_{y\in\mathbb{R}^r} \frac{y^TZ^TAZy}{y^TZ^TMZy}.$$

This means that the smallest eigenvalue μ_{min} of the generalized eigenvalue problem

$$Ey = \mu Z^TMZy,$$

is an upper bound for the smallest eigenvalue of $M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$, whereas the smallest eigenvalue μ_{min} of the generalized eigenvalue problem

$$Ey = \mu Z^TZy,$$

is an upper bound for the smallest eigenvalue of A. From experiments for the porous media flow problem, it appears that the estimates are reasonable sharp see Table 1.

matrix	λ_{min}	$\lambda_{min}(estimated)$
$M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$	$0.7 \cdot 10^{-8}$	$3.1 \cdot 10^{-8}$
A	$3.3 \cdot 10^{-9}$	$9.9 \cdot 10^{-9}$

Table 1: The estimated smallest eigenvalue using matrix E

4 Numerical experiments

All numerical experiments are done by using the SEPRAN FEM package developed at the TU Delft. The multiplication $y = E^{-1}b$ is always done by solving y from $Ey = b$, where E is decomposed in its Cholesky factor. In this section, coarse grid correction is abbreviated as CGC. The choice of the boundary conditions is such that all problems have as exact solution the vector with compononents equal to 1. In order to make the convergence behavior representative for general problems we chose a random vector as starting solution, in stead of the zero start vector.

4.1 Porous media flows

In this section we consider problems motivated by porous media flow [25]. Our first problem is a simple 2 dimensional model problem, whereas our second problem mimics the flow of oil in a reservoir. In both problems *physical* projection vectors are used (see [27], Definition 2), which approximate the eigenvectors corresponding to the small eigenvalues.

7 layer problem

We solve the following equation

$$\text{div } (\sigma \nabla p) = 0,$$

with p the fluid pressure and σ the permeability. At the earth's surface the fluid pressure is prescribed. At the other boundaries we use homogeneous Neumann conditions. In this two dimensional problem we consider 7 horizontal layers. We use linear triangular elements and the number of grid points is equal to 22680. The top layer is sandstone then a shale layer etc. We assume that σ in sandstone is equal to 1 and σ in shale is equal to 10^{-7}. From [26] it follows that the IC preconditioned matrix has 3 eigenvalues of order 10^{-7}, whereas the remaining eigenvalues are of order 1. Computing the solution with 3 projection vectors we observe that in every iteration the norm of the residual using deflation or CGC is comparable. In Figure 3 the norm of the error for both methods is given. To our surprise the error using deflation stagnates at a lower level than that of CGC.

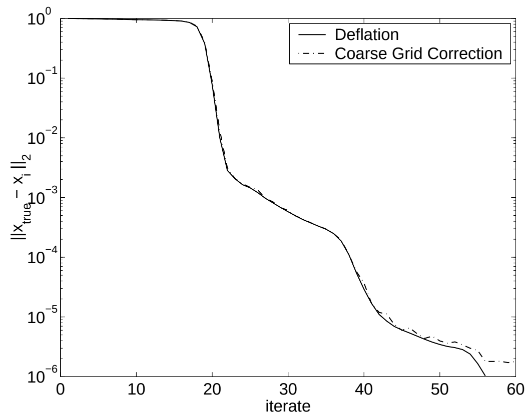

Figure 3: The norm of the error for projected ICCG for the 7 layer problem

An oil flow problem

In this paragraph we simulate a porous media flow in a 3 dimensional layered geometry, where the layers vary in thickness and orientation (see figures 4 and 5 for a 4 layer problem). Figure 4 shows a part of the earth's crust. The depth of this part varies between 3 and 6 kilometers, whereas horizontally its dimensions are 40 x 60 kilometers. The upper layer is a mixture of sandstone and shale and has a permeability of 10^{-4}. Below this layer, shale and sandstone layers are present with permeabilities of 10^{-7} and 10 respectively. We consider a problem with 9 layers. Five sandstone layers are separated by four shale layers. At the top of the first sandstone/shale layer a Dirichlet boundary condition is posed, so the IC preconditioned matrix has 4 small eigenvalues. We use 4 *physical* projection vectors and stop if $\|r_k\|_2 \leq 10^{-5}$. Trilinear hexahedral elements are used and the total number of gridpoints is equal to 148185. The results are given in Table 2 and correspond well with our theoretical results.

method	deflation	CGC
iterations	36	47
CPU time	5.9	8.2

Table 2: The results for the oil flow problem

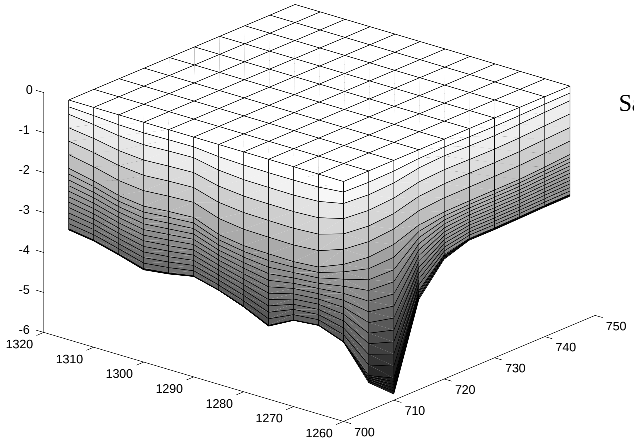

Figure 4: The geometry of an oil flow problem

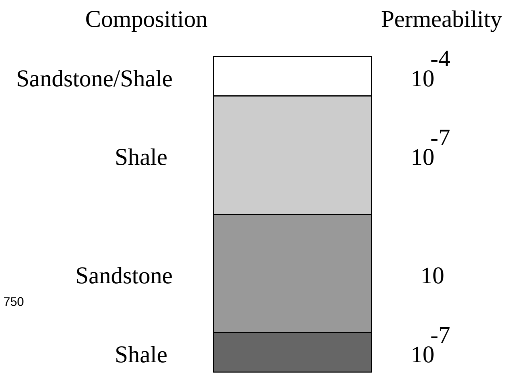

Figure 5: Permeabilities for each layer

4.2 Parallel problems

In this section we consider a Poisson equation on a 2 dimensional rectangular domain. On top a Dirichlet boundary condition is posed, whereas at the other boundaries a homogeneous Neumann condition is used. We use linear triangular elements. We stop the iteration if $\|r_k\|_2 \leq 10^{-8}$.

As a first test we solve a problem, where the grid is decomposed into 7 layers with various gridsizes per layer. The results are given in Table 3. In this table the symbol 'no' means that there is no projection method used. Note that in the parallel case we use a block IC preconditioner. Deflation needs again less iterations than CGC. However both projection methods lead to a considerable gain in the number of iterations. Note that the number of iterations increases if the gridsize per layer increases.

	sequential			parallel		
grid points	deflation	CGC	no	deflation	CGC	no
10×10	21	29	35	25	38	50
20×20	36	48	65	42	61	90
40×40	62	82	125	80	103	168
80×80	106	131	244	128	161	321

Table 3: The effect of the gridsize per layer

Secondly we consider the parallel performance for an increasing number of layers or blocks. The gridsize per layer is 80×80 and per block 100×100. This implies that the total number of grid points increases proportionally to the number of layers/blocks. In figures 6 and 7 the results are given. Note that initially both projection methods show a small increase in the number of iterations if the number of layers/blocks increases but thereafter the number of iterations is constant (scalable). If no projection method is used the number of iterations keep increasing.

5 Conclusions

We have compared various preconditioners used in the numerical solution of partial differential equations. On one hand we considered a coarse grid correction preconditioner. On the other hand a so-called deflation preconditioner was studied. It turned out that the effective condition number of the deflated preconditioned system is always, i.e. for all deflation vectors and all restrictions and prolongations, below the condition number of the system preconditioned by the

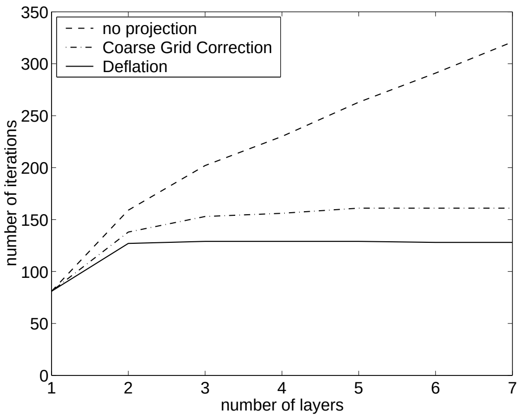

Figure 6: The number of iterations for a layered domain decomposition

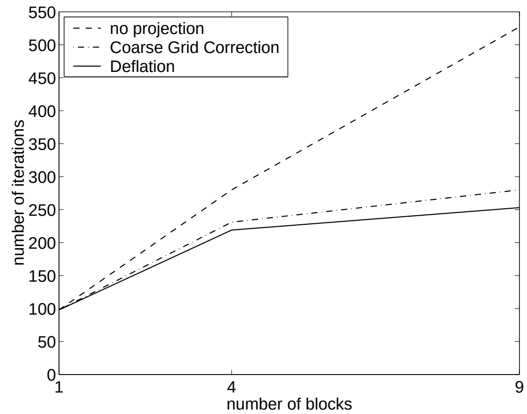

Figure 7: The number of iterations for a block domain decomposition

coarse grid correction. This implies that the Conjugate Gradient method applied to the deflated preconditioned system converges always faster than the Conjugate Gradient method applied to the system preconditioned by the coarse grid correction. Numerical results for porous media flows and parallel preconditioners emphasized the theoretical results.

References

[1] M. Benzi, A. Frommer, R. Nabben, and D.B. Szyld. Algebraic theory of multiplicative Schwarz methods. *Numerische Mathematik*, 89:606–639, 2001.

[2] J.H. Bramble, J.E. Pasciak, and A.H. Schatz. The construction of preconditioners for elliptic problems by substructuring. I. *Math. Comput.*, 47:103–134, 1986.

[3] 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.*, 13:45–49, 2000.

[4] M. Dryja. An additive Schwarz algorithm for two- and three dimensional finite element elliptic problems. In *Domain Decomposition methods*, pages 168–172. SIAM Philadelphia PA, 1989.

[5] M. Dryja and O.B. Widlund. Towards a unified theory of domain decomposition algorithms for elliptic problems. In *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 3–21. SIAM Philadelphia PA, 1990.

[6] M. Dryja and O.B. Widlund. Multilevel additive methods for elliptic finite element problems in three dimensions. In *Parallel Algorithms for PDEs*, pages 58–69. Vieweg Braunschweig Germany, 1991.

[7] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. *SIAM Journal on Scientific Computing*, 23:442–462, 2001.

[8] G. H. Golub and C. F. van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, third edition, 1996.

[9] E .F. Kaasschieter. Preconditioned Conjugate Gradients for solving singular systems. *J. Comput. Appl. Math.*, 24:265–275, 1988.

[10] E.F. Kaasschieter. A practical termination criterion for the conjugate gradient method. *BIT*, 28:308–322, 1988.

[11] L.Yu. Kolotilina. Preconditioning of systems of linear algebraic equations by means of twofold deflation. I. Theory. *J. Math. Sci.*, 89:1652–1689, 1998.

[12] L. Mansfield. On the Conjugate Gradient solution of the Schur complement system obtained from domain decomposition. *SIAM J. Numer. Anal.*, 27(6):1612–1620, 1990.

[13] L. Mansfield. Damped Jacobi preconditioning and coarse grid deflation for Conjugate Gradient iteration on parallel computers. *SIAM J. Sci. Stat. Comput.*, 12(6):1314–1323, 1991.

[14] 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. *Mathematics of Computation*, 31:148–162, 1977.

[15] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16(4):1154–1171, 1995.

[16] R. Nabben. Comparisons between additive and multiplicative Schwarz iterations in domain decomposition methods. *Numerische Mathematik*. to appear, Online publication: February 18, 2003.

[17] R. A. Nicolaides. Deflation of Conjugate Gradients with applications to boundary value problems. *SIAM J. Numer. Anal.*, 24(2):355–365, 1987.

[18] 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.*, 22:793–818, 2000.

[19] A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Scinece Publications, Oxford, 1999.

[20] C.R. Johnson R.A. Horn. *Matrix Analysis*. Cambridge University Press, Cambrige, 1985.

[21] Y. Saad, M. Yeung, J. Ehrel, and F. Guyomarc'h. A deflated version of the Conjugate Gradient algorithm. *SIAM Journal on Scientific Computing*, 21:1909–1926, 2000.

[22] B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition*. Cambridge University Press, Cambridge, 1996.

[23] A. van der Sluis and H. A. van der Vorst. The rate of convergence of Conjugate Gradients. *Numer. Math.*, 48:543–560, 1986.

[24] F.J. Vermolen, C. Vuik, and A. Segal. Deflation in preconditioned Conjugate Gradient methods for finite element problems. *Journal of Computational Methods in Sciences and Engineering*, 2003. to appear.

[25] 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.*, 152:385–403, 1999.

[26] C. Vuik, A. Segal, J.A. Meijerink, and G.T. Wijma. The construction of projection vectors for a Deflated ICCG method applied to problems with extreme contrasts in the coefficients. *Journal of Computational Physics*, 172:426–450, 2001.

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