ON THE SPECTRUM OF DEFLATED MATRICES WITH APPLICATIONS TO THE DEFLATED SHIFTED LAPLACE PRECONDITIONER FOR THE HELMHOLTZ EQUATION

LUIS GARCÍA RAMOS AND REINHARD NABBEN

Abstract. The deflation technique to accelerate Krylov subspace iterative methods for the solution of linear systems has been known for a long time. The first landmark papers are due to Nicolaides [21] and Dostál [3] in the late eighties where deflation is used for Hermitian positive definite (HPD) linear systems. In the last decade deflation was used and analyzed in combination with domain decomposition and multigrid methods, which results in very effective algorithms. Examples are the multilevel Krylov methods introduced in [6, 7], see also [24, 25] where multilevel deflation techniques are presented. Although these algorithms work very well in practice for non-Hermitian problems, not much theoretical results are known so far in this direction. Here, we show inclusion regions for the spectrum of an arbitrary deflated matrix based on the field of values, which generalize known results for HPD systems. Moreover, for deflated GMRES we show a residual bound based on the field of values. We apply our results to linear systems arising from the Helmholtz equation. We focus on the combination of the complex shifted Laplace (CSL) preconditioner [8] with the multilevel Krylov technique. Numerical examples indicate that the eigenvalues of the deflated CSL-preconditioned system lie on exact the same circles as the CSL-preconditioned linear systems and are shifted away from zero. Here we are able to prove these surprising results for any wavenumber and any dimension using Möbius transformations and the Spectral Mapping Theorem. Our new results help to explain the good performance of multilevel Krylov methods for the Helmholtz equation.

1. Introduction

The solution of linear systems of equations of the form

$$Ax = b,$$

where $A \in \mathbb{C}^{n\times n}$ is nonsingular and $x, b \in \mathbb{C}^n$, is a major component in numerical simulations arising in scientific and engineering applications. When the matrix A is sparse, iterative methods based on Krylov subspaces are often used. The convergence of Krylov subspace methods for Hermitian matrices (as e.g., CG [14]) depends on the eigenvalues of A, while for non-Hermitian matrices the behaviour of Krylov subspace methods (as, e.g., GMRES [23]) can only be directly linked to the eigenvalues when the matrix is normal or close to normal (with the condition number of its matrix of eigenvectors close to 1), see [18]. In these cases, if the eigenvalues of the matrix A are too spread or near zero the convergence of the iteration can be slow. To obtain a better clustered spectrum, one can solve the equivalent system of equations

$$M^{-1}Ax = M^{-1}b, \tag{1}$$

with M an easily invertible matrix that approximates A in some sense. In this setting the matrix M is known as a preconditioner. Alternatively, if the spectrum of A contains very small eigenvalues these can be eliminated by deflating the corresponding eigenspaces. More precisely, suppose we want to solve (1) using a Krylov subspace method, and the spectrum of A can be decomposed as

$$\Lambda(A) = \{\lambda_1, \ldots, \lambda_r\} \cup \{\lambda_{r+1}, \ldots, \lambda_n\}$$

where the eigenvalues $\lambda_1, \ldots, \lambda_r$ are close to zero. If $U \in \mathbb{C}^{n\times r}$ is a matrix whose columns span the eigenspace $\mathcal{U}$ corresponding to the small eigenvalues, we can form the following *deflation* operator

$$P = I - AU(U^H AU)^{-1}U^H, \tag{2}$$

provided that $U^H AU$ is nonsingular, where U^H denotes the Hermitian transpose of U. It can be shown that the operator P is a projection with $\mathrm{im}(P) = \mathcal{U}^\perp$ and $\ker(P) = A\mathcal{U}$. Since the spectrum of PA

satisfies

$$\Lambda(PA) = \{0\} \cup \{\lambda_{r+1}, \dots, \lambda_n\},$$

the problematic small eigenvalues are eliminated from the spectrum and a Krylov iteration to solve

$$PAx = Pb \tag{3}$$

is expected to converge faster. In practice, exact eigenvectors are expensive to compute and the deflation operator P in (2) is formed with a full-rank matrix U that contains approximate eigenvectors of A. In this case we have

$$\Lambda(PA) = \{0\} \cup \{\tilde{\lambda}_{r+1}, \dots, \tilde{\lambda}_n\},$$

where the nonzero eigenvalues $\tilde{\lambda}_j$ of PA may differ from the eigenvalues of A. It is then important to quantify the difference between the nonzero spectrum of the deflated matrix PA and the spectrum of the original matrix A and determine if the nonzero eigenvalues of PA can be shifted near zero or grow increasingly large. Deflation for linear systems was introduced by Nicolaides [21] and Dostál [3], see also [1, 2, 9, 13, 17, 20]. The deflation technique is also related to multigrid methods, where the columns of U are formed from the interpolation operator, see [27, 28] for a detailed comparison. Further, deflation can be combined with standard preconditioning. One option is to precondition the deflated system (3), it is also possible to first precondition the linear system and then construct the deflation operator from the preconditioned matrix, i.e. first build $\hat{A} = M^{-1}A$, and then use the projection

$$\hat{P} = I - \hat{A}U(U^H\hat{A}U)^{-1}U^H.$$

In the special case where A is Hermitian positive semidefinite (HPD) the deflated matrix PA is Hermitian positive semidefinite, and it has been shown by Nicolaides [21] that for an HPD matrix A and any operator P of the form (2) we have

$$\lambda_{\min}(A) \leq \lambda_{\min}(PA), \quad \text{and} \quad \lambda_{\max}(PA) \leq \lambda_{\max}(A). \tag{4}$$

where $\lambda_{\min}(PA), \lambda_{\max}(PA)$ denote the minimum nonzero eigenvalues of PA. Since the standard worst-case residual bounds for the convergence of Krylov methods for HPD problems are based on the condition number $\kappa = \lambda_{\min}/\lambda_{\max}$, this property provides some motivation for the use of deflation techniques in this case.

In this paper we give a simple inclusion region for the eigenvalues of a deflated matrix PA that reduces to (4) in the case of HPD matrices. We also give necessary and sufficient conditions on a matrix that imply that the maximum eigenvalue of PA is bounded independently of the deflation subspace $\mathcal{U}$. For a normal matrix, the inclusion regions depend only on its eigenvalues. We will use a result that describes the nonzero spectrum of PA shown by Gaul in [12], and by Kahl and Rittich [16] for HPD matrices. We also present a field of values bound for deflated GMRES, which can be useful for non-normal problems where the convergence cannot be linked to the eigenvalues. We then apply our results to linear systems that arise from the discretization of the Helmholtz equation. We will focus here on the multiplicative combination of a deflation operator with the discrete Helmholtz operator preconditioned by the shifted Laplacian [8], one of the most efficient preconditioners currently in use. This operator is then the basis of a multilevel Krylov method introduced in [6, 7] which leads to convergence factors that are just mildly dependent of the dimension of the linear system and the wave number. Using results of [29] we will show that for a class of model problems the spectrum of the deflated deflated CSL-preconditioned system lie on exact the same circles as the CSL-preconditioned linear systems and are farther away from zero. This partially explains the improved performance of methods based on the deflated shifted Laplacian in contrast to the shifted Laplace preconditioner.

2. Projections and Deflation

In this section we switch to the language of linear operators to present simpler matrix-free proofs of the results that follow. Recall that a linear operator P is a *projection* if $P^2 = P$. A projection can be completely characterized by its range and kernel. More precisely, if a projection $P : \mathbb{C}^n \to \mathbb{C}^n$ is given then $\mathbb{C}^n = \text{im}(P) \oplus \ker(P)$, and further, if $\mathcal{V}$, $\mathcal{W}$ are subspaces of $\mathbb{C}^n$ such that $\mathbb{C}^n = \mathcal{V} \oplus \mathcal{W}$ it can be shown that there exists a unique projection operator $P_{\mathcal{V},\mathcal{W}}$ such that $\mathrm{im}(P)=\mathcal{V}$ and $\ker(P)=\mathcal{W}$. We first define some notation.

Definition 2.1. *For a pair of closed subspaces $\mathcal{V},\mathcal{W}\subset\mathbb{C}^n$ with $\mathbb{C}^n=\mathrm{im}(P)\oplus\ker(P)$ the operator $P_{\mathcal{V},\mathcal{W}}$ is defined as the unique projection with $\mathrm{im}(P)=\mathcal{V}$ and $\ker(P)=\mathcal{W}$. The projection $P_{\mathcal{V},\mathcal{W}}$ is called the projection onto $\mathcal{V}$ along $\mathcal{W}$. The orthogonal projection $P_{\mathcal{V},\mathcal{V}^\perp}$ onto $\mathcal{V}$ is denoted by $P_{\mathcal{V}}$.*

Note that for a given projection $P_{\mathcal{V},\mathcal{W}}$ the complementary projection is $P_{\mathcal{W},\mathcal{V}}=I-P_{\mathcal{V},\mathcal{W}}$. To use deflation techniques for solving linear systems of equations, we will consider subspaces $\mathcal{U}$ with dimension r and of the form $A\mathcal{U}$, where A is the system matrix of the linear system. If a basis of the subspace $\mathcal{U}$ is given, the projection $P_{A\mathcal{U},\mathcal{U}^\perp}$ can be represented as described in the following theorem. For proof of this theorem and other facts on projections we refer the reader to [12, 22].

Theorem 2.2. *Let $\mathcal{U}$ be a subspace of $\mathbb{C}^n$ of dimension r. Let U be a matrix whose columns form a basis of $\mathcal{U}$, and $A\in\mathbb{C}^{n\times n}$. Then the following are equivalent:*

1. $\mathbb{C}^n=A\mathcal{U}\oplus\mathcal{U}^\perp$.
2. $U^HAU\in\mathbb{C}^{r\times r}$ *is nonsingular and the projection* $P_{A\mathcal{U},\mathcal{U}^\perp}$ *can be represented by*
$$P_{A\mathcal{U},\mathcal{U}^\perp}=AU(U^HAU)^{-1}U^H.$$
3. $U^HAU\in\mathbb{C}^{r\times r}$ *is nonsingular and the projection* $P_{\mathcal{U}^\perp,A\mathcal{U}}$ *can be represented by*
$$P_{\mathcal{U}^\perp,A\mathcal{U}}=I-P_{A\mathcal{U},\mathcal{U}^\perp}=I-AU(U^HAU)^{-1}U^H.$$
4. $A\mathcal{U}\cap\mathcal{U}^\perp=\{0\}$.

Thus, with the help of a basis of $\mathcal{U}$ we have
$$P_{\mathcal{U}^\perp,A\mathcal{U}}=I-AU(U^HAU)^{-1}U^H=P,$$
which is the deflation operator given in (2). In the following we will use $P_{\mathcal{U}^\perp,A\mathcal{U}}$ and P to denote the basis free and the matrix (basis) oriented version of the projection, since the matrix version is much more used in the literature of deflation techniques. But we point out that $\mathcal{U}$ is just a subspace of $\mathbb{C}^n$, not necessarily an invariant subspace, as is often associated with deflation.

The next theorem describes the spectrum of a deflated operator and is shown in [12], and in the special case of Hermitian positive definite matrices in [16]. We include here our own proof for completeness.

Theorem 2.3 (Theorem 3.24 in [12])**.** *Let $A:\mathbb{C}^n\to\mathbb{C}^n$ be a nonsingular linear operator and $\mathcal{U}$ a subspace of $\mathbb{C}^n$ such that $\mathbb{C}^n=\mathcal{U}^\perp\oplus A\mathcal{U}$. Then, the operator $P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp}:\mathcal{U}^\perp\to\mathcal{U}^\perp$ is nonsingular and*
$$P_{\mathcal{U}^\perp,A\mathcal{U}}A=(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})^{-1}P_{\mathcal{U}^\perp}. \tag{5}$$
In particular, the spectrum satisfies
$$\Lambda(P_{\mathcal{U}^\perp,A\mathcal{U}}A)=\{0\}\cup\Lambda\left((P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})^{-1}\right). \tag{6}$$

Proof. Since $P_{\mathcal{U}^\perp,A\mathcal{U}}A|_{\mathcal{U}}=0$, to prove (5) it is sufficient to show
$$P_{\mathcal{U}^\perp,A\mathcal{U}}A|_{\mathcal{U}^\perp}\left(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp}\right)=I.$$
For every $x\in\mathbb{C}^n$ we have
$$P_{\mathcal{U}^\perp}A^{-1}x=A^{-1}x-P_{\mathcal{U}}A^{-1}x,$$
Applying $P_{\mathcal{U}^\perp,A\mathcal{U}}A$ on both sides we get
$$\begin{aligned}(P_{\mathcal{U}^\perp,A\mathcal{U}}A)P_{\mathcal{U}^\perp}A^{-1}x&=(P_{\mathcal{U}^\perp,A\mathcal{U}}A)(A^{-1}x-P_{\mathcal{U}}A^{-1}x)\\&=P_{\mathcal{U}^\perp,A\mathcal{U}}AA^{-1}x=P_{\mathcal{U}^\perp,A\mathcal{U}}x.\end{aligned}$$
In particular, if $x\in\mathcal{U}^\perp$
$$(P_{\mathcal{U}^\perp,A\mathcal{U}}A|_{\mathcal{U}^\perp})(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})x=x,$$

so the assertion is proved. The statement (6) follows from the fact that both $\mathcal{U}$ and $\mathcal{U}^\perp$ are $P_{\mathcal{U}^\perp,A\mathcal{U}}A$-invariant subspaces, since $\mathcal{U}$ is the kernel of $P_{\mathcal{U}^\perp,A\mathcal{U}}A$ and $\mathcal{U}^\perp$ is the range of $P_{\mathcal{U}^\perp,A\mathcal{U}}A$. Hence the eigenvalues of

$$P_{\mathcal{U}^\perp,A\mathcal{U}}A = (P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})^{-1}P_{\mathcal{U}^\perp}$$

are either zero or the eigenvalues of $(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})^{-1}$. □

Corollary 2.4. *Let $A : \mathbb{C}^n \to \mathbb{C}^n$ be a nonsingular linear operator and $\mathcal{U}$ a subspace of $\mathbb{C}^n$ such that $\mathbb{C}^n = \mathcal{U}^\perp \oplus A\mathcal{U}$. Moreover, let V be a matrix whose columns form an orthonormal basis of $\mathcal{U}^\perp$. Then $V^HA^{-1}V$ is nonsingular and*

$$\Lambda(P_{\mathcal{U}^\perp,A\mathcal{U}}A) = \Lambda(PA) = \{0\} \cup \Lambda((V^HA^{-1}V)^{-1}). \tag{7}$$

Proof. The matrix representation of $P_{\mathcal{U}^\perp}$ is just $P_{\mathcal{U}^\perp} = VV^H$. If $T = [(T)_{ij}]$ is the matrix representation of $P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp}$ in the orthonormal basis of $\mathcal{U}^\perp$ formed by the columns of V, then the ij-th entry of T is given by

$$\begin{aligned}(T)_{ij} &= v_i^H(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp}v_j) = v_i^H(VV^HA^{-1}v_j)\\ &= (VV^Hv_i)^HA^{-1}v_j = v_i^HA^{-1}v_j = (V^HA^{-1}V)_{ij}.\end{aligned}$$

This implies by Theorem 2.3 that $V^HA^{-1}V$ is nonsingular and with equation (6) we obtain (7). □

Note that for symmetric positive semidefinite matrices the well-known fact that PA is positive semidefinite follows immediately from Corollary 2.4.

3. Deflation and the Field of Values

To set the stage for the following results, recall that for a matrix $A \in \mathbb{C}^{n\times n}$, the field of values of A is the set

$$W(A) = \{x^HAx : x \in \mathbb{C}^n, \|x\| = 1\}.$$

Clearly $\Lambda(A) \subset W(A)$ for any matrix A. The Toeplitz-Hausdorff theorem states that field of values is a convex set [15]. In particular, if A is normal and $\Lambda(A) = \{\lambda_1, \dots, \lambda_n\}$, the field of values of A is the convex hull of its spectrum:

$$W(A) = \operatorname{conv}(\Lambda(A)) = \left\{\sum_{i=1}^n \alpha_i\lambda_i :\ \alpha_i \geq 0,\ i = 1,\dots,n,\ \sum_{i=1}^n \alpha_i = 1,\ \lambda_i \in \Lambda(A)\right\}$$

The next lemma is stated as an exercise in [15].

Lemma 3.1. *Let $A \in \mathbb{C}^{n\times n}$ and $Q \in \mathbb{C}^{n\times r}$ with $Q^HQ = I \in \mathbb{C}^{r\times r}$. Then*

$$W(Q^HAQ) \subset W(A).$$

Theorem 2.3 and Lemma 3.1 can be combined to give an inclusion set for the eigenvalues of a deflated matrix.

Theorem 3.2. *Let $A \in \mathbb{C}^n$, and $\mathcal{U} \subset \mathbb{C}^n$ a subspace such that $\mathbb{C}^n = \mathcal{U} \oplus A\mathcal{U}$. Let $P = P_{\mathcal{U}^\perp,A\mathcal{U}}$ be the projection onto $\mathcal{U}^\perp$ along $A\mathcal{U}$. Then the spectrum of PA satisfies*

$$\Lambda(PA) \setminus \{0\} \subset W(A^{-1})^{-1}$$

Proof. Let $r = \dim(\mathcal{U})$, $m = n - r$ and $V \in \mathbb{C}^{n\times m}$ be a matrix with orthonormal columns that form a basis for $\mathcal{U}^\perp$. Since $V^HV = I$, we have, by Lemma 3.1 and Corollary 2.4

$$\Lambda(P_{\mathcal{U}^\perp,A\mathcal{U}}A) \setminus \{0\} \subset \Lambda\left((P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp})^{-1}\right) = \Lambda\left(P_{\mathcal{U}^\perp}A^{-1}|_{\mathcal{U}^\perp}\right)^{-1} = \Lambda((V^HA^{-1}V))^{-1} \subset W(A^{-1})^{-1}.$$

□

Theorem 3.2 shows that every nonzero element of $\Lambda(PA)$ is the reciprocal of an element of the field of values of A^{-1}. When A is normal with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, every $z \in W(A^{-1})^{-1}$ is a weighted harmonic mean of the eigenvalues of A,

$$z = \left[\sum_{i=1}^{n} \alpha_j \lambda_j^{-1}\right]^{-1}$$

with $\alpha_j \geq 0$ and $\sum_{j=}^{n} \alpha_j = 1$. Therefore, for a normal matrix A, the inclusion region from Theorem 3.2 depends only on the eigenvalues of A.

Corollary 3.3. *Let $A \in \mathbb{C}^n$ be normal, and $\mathcal{U} \subset \mathbb{C}^n$ a subspace as in the hypothesis of Theorem 3.2. Then the spectrum of $P = P_{\mathcal{U}^\perp, A\mathcal{U}}A$ satisfies*

$$\Lambda(P_{\mathcal{U}^\perp, A\mathcal{U}}A) \setminus \{0\} \subset \mathrm{conv}(\Lambda(A^{-1}))^{-1} = \{z^{-1} : z \in \mathrm{conv}(\Lambda(A^{-1}))\}.$$

In the HPD case, we have

$$\mathrm{conv}(\Lambda(A^{-1})) = [\lambda_{\max}(A)^{-1}, \lambda_{\min}(A)^{-1}]$$

hence $\mathrm{conv}(\Lambda(A^{-1}))^{-1} = [\lambda_{\min}(A), \lambda_{\max}(A)]$, so Theorem 3.2 is a generalization of (4). The connection between the spectrum of a deflated matrix PA and the field of values of A^{-1} shown in Theorem 3.2 allows us to give bounds for the modulus of the eigenvalues of a deflated matrix. For a matrix A, we denote by $\nu(A)$ the distance of the field of values to 0 and $\mu(A)$ the numerical radius, defined as

$$\nu(A) = \min\{|z| : z \in W(A)\}, \text{ and } \mu(A) = \max\{|z| : z \in W(A)\}.$$

It is clear that $\nu(A) > 0$ if and only if $0 \notin W(A)$. In view of the equality

$$W(A^{-1}) = \left\{\frac{x^H A^{-1} x}{x^H x} : x \neq 0\right\} = \left\{\frac{y^H A^H y}{y^H A^H A y} : y \neq 0\right\},$$

the condition $0 \notin W(A)$ is equivalent to $0 \notin W(A^{-1})$. The following corollary of Theorem 3.2 gives bounds for the minimum and maximum eigenvalues of A that are independent of the deflation subspace.

Corollary 3.4. *Let $A \in \mathbb{C}^n$ and $\mathcal{U} \subset \mathbb{C}^n$ a subspace such that $\mathbb{C}^n = \mathcal{U} \oplus A\mathcal{U}$. Let $P = P_{\mathcal{U}^\perp, A\mathcal{U}}$ be the projection onto $\mathcal{U}^\perp$ along $A\mathcal{U}$. Let $\lambda_{\min}(PA), \lambda_{\max}(PA)$ denote the nonzero eigenvalues of PA of minimum and maximum modulus. Then*

$$|\lambda_{\min}(PA)| \geq \mu(A^{-1})^{-1},$$

and, if $\nu(A) > 0$, we also have

$$|\lambda_{\max}(PA)| \leq \nu(A^{-1})^{-1}. \tag{8}$$

Proof. By Theorem 3.2, we have, for every $\lambda \in \Lambda(PA) \setminus \{0\}$,

$$|\lambda^{-1}| \leq \mu(A^{-1}),$$

therefore $|\lambda_{\min}(PA)| \geq \mu(A^{-1})^{-1}$. The inequality for $\lambda_{\max}(PA)$ is proved similarly. □

Corollary 3.5. *Let $A \in \mathbb{C}^n$ be normal and $\mathcal{U} \subset \mathbb{C}^n$ a subspace such that $\mathbb{C}^n = \mathcal{U} \oplus A\mathcal{U}$. Let $P_{\mathcal{U}^\perp, A\mathcal{U}}$ be the projection onto $\mathcal{U}^\perp$ along $A\mathcal{U}$. If $\lambda_{\min}(PA), \lambda_{\max}(PA)$ denote the nonzero eigenvalues of PA of minimum and maximum modulus, then*

$$|\lambda_{\min}(PA)| \geq \lambda_{\min}(A), \tag{9}$$

and, if $\nu(A) > 0$, we also have

$$|\lambda_{\max}(PA)| \leq \frac{\lambda_{\max}(A)^2}{\nu(A)}. \tag{10}$$

Proof. To show (9), note that for a normal matrix A, we have $\mu(A^{-1}) = |\lambda_{\max}(A^{-1})| = |\lambda_{\min}(A)^{-1}|$, combining this with the first inequality from the previous corollary then (9) follows. For the second part, we recall an inequality for $\nu(A^{-1})$ shown in [4]:

$$\nu(A^{-1}) = \min_{0\neq x\in\mathbb{C}^n} \frac{x^H A^{-1} x}{x^H x} = \min_{0\neq y\in\mathbb{C}^n} \frac{y^H A^H y}{y^H y}\frac{y^H y}{y^H A^H A y} \geq \frac{\nu(A)}{\|A\|^2}. \tag{11}$$

For a normal matrix A we have $\|A\|^2 = \lambda_{\max}(A)^2$, so from (11) and the previous corollary we obtain inequality (10). □

The condition $\nu(A) > 0$ is necessary for the maximum nonzero eigenvalue $\lambda_{\max}(P_{\mathcal{U}^\perp,A\mathcal{U}}A)$ to be bounded above independently of the deflation subspace $\mathcal{U}$. This follows from the fact that every $z = (x^H A^{-1} x)^{-1} \in W(A^{-1})^{-1}$ can be attained as an eigenvalue of a deflated matrix $P_{\mathcal{U}^\perp,A\mathcal{U}}A$ by choosing the deflation subspace $\mathcal{U} = \{x\}^\perp$. When $\nu(A) = 0$ we have $0 \in W(A)$, which implies that the set $W(A^{-1})^{-1}$ is unbounded.

4. A residual bound for deflated GMRES

As remarked in the introduction, eigenvalues alone may not be sufficient to determine the convergence of GMRES for non-normal matrices. The field of values has been studied as an alternative set that can provide information on GMRES convergence. In particular, if r_k is the k-th residual of a GMRES iteration, the following bound holds:

$$\frac{\|r_k\|}{\|r_0\|} \leq (1 - \nu(A)\nu(A^{-1}))^{k/2}. \tag{12}$$

This bound is a generalization of a result due to Elman [5] for the residuals of the GCR method. It was first proved by Starke in [26] for matrices with a positive-definite Hermitian part, and later shown to hold in the general case by Eiermann and Ernst in [4], see also [19]. Since the bound is only informative in the case where $0 \notin W(A)$, it is not directly applicable to an iteration with a (singular) deflated matrix PA. In order to overcome this problem we show now using Theorem 2.3 that a GMRES iteration with PA is equivalent to an iteration with a smaller nonsingular matrix M. Further, we show that the quantities $\nu(M)$ and $\nu(M^{-1})$ are larger than the corresponding quantities for A, thus improving the bound (12).

Theorem 4.1. *Let $A \in \mathbb{C}^{n\times n}$, $\mathcal{U}$ a subspace of $\mathbb{C}^n$ and suppose that the deflation operator $P = P_{\mathcal{U}^\perp,A\mathcal{U}}$ is well defined. Moreover, let $V \in \mathbb{C}^{m\times m}$ a matrix with orthonormal columns that form a basis for $\mathcal{U}^\perp$, and $M = (V^H A^{-1} V)^{-1}$. Then GMRES applied to the singular system $PAx = b$ with a zero initial guess generates the sames residuals as GMRES applied to the nonsingular system $My = V^H b$ with a zero initial guess.*

Proof. From Corollary 2.4 we have that M is well defined and is the matrix representation of $(P_{\mathcal{U}^\perp} A^{-1}|_{\mathcal{U}})^{-1}$ in the basis formed by the columns of V. Let U be a matrix with orthonormal columns that form a basis for $\mathcal{U}$, and let $Y = [U\ V]$. Then Y is unitary, and Theorem 2.3 implies that PA can be block-diagonalized in the basis given by the columns of Y,

$$PA = [U\ V]\begin{bmatrix} 0 & 0 \\ 0 & (V^H A^{-1} V)^{-1}\end{bmatrix}[U\ V]^H = [U\ V]\begin{bmatrix} 0 & 0 \\ 0 & M\end{bmatrix}[U\ V]^H. \tag{13}$$

Therefore, for any polynomial p we have

$$p(PA) = [U\ V]\begin{bmatrix} 0 & 0 \\ 0 & p(M)\end{bmatrix}[U\ V]^H.$$

Let Π_k denote the set of polynomials p of degree at most k such that $p(0) = 1$. The GMRES method applied to $PAx = b$ with a zero initial guess leads to a residual minimization problem in step k:

$$\min_{p\in\Pi_k} \|p(PA)b\| = \min_{p\in\Pi_k} \|Vp(M)V^H b\| = \min_{p\in\Pi_k} \|p(M)V^H b\|,$$

where we have used in the last step that V has orthonormal columns. This is the same residual minimization problem that results from applying GMRES to the system $My = V^H b$. completing the proof. □

We remark that the matrix M from the previous theorem is only introduced for theoretical purposes and is never formed explicitly, since the deflated GMRES iteration uses only the matrix PA. The next theorem gives a bound for the residuals of the deflated GMRES method.

Theorem 4.2. *Let $A \in \mathbb{C}^{n\times n}$, $\mathcal{U}$, P, V and M as in the hypothesis of Theorem 4.1. Then*

$$\nu(M^{-1}) \geq \nu(A^{-1}), \quad \text{and} \quad \nu(M) \geq \nu(A).$$

and for the k-th residual of GMRES applied to $PAx = b$, we have the bound

$$\frac{\|r_k\|}{\|r_0\|} \leq (1 - \nu(M)\nu(M^{-1}))^{k/2}.$$

Proof. The first inequality follows from the inclusion

$$W(M^{-1}) = W(V^H A^{-1} V) \subset W(A^{-1}).$$

For the second inequality, let $\frac{z^H M z}{\|z\|} \in W(M)$, then for $y = (V^H A^{-1} V)^{-1} z$ and $x = A^{-1} V y$ we have

$$\frac{|z^H M z|}{\|z\|} = \frac{|z^H (V^H A^{-1} V)^{-1} z|}{\|z\|} = \frac{|y^H (V^H A^{-1} V) y|}{\|y\|}$$
$$= \frac{|x^H A x|}{\|V^H x\|} = \frac{|x^H A x|}{\|V V^H x\|} \geq \frac{|x^H A x|}{\|x\|} \in W(A),$$

where in the two last steps we have used that V has orthonormal columns and $\|VV^H x\| \leq \|x\|$ since VV^H is the matrix form of the orthogonal projection $P_{\mathcal{U}^\perp}$. This shows that $\nu(M) \geq \nu(A)$. The second part of the theorem follows from the first, combined with Theorem 4.1 and the residual bound (12). □

5. The spectrum of the deflated shifted Laplace preconditioner for the Helmholtz equation

In this section we use the previous results to analyze the spectrum of a deflated matrix arising from the discretization of the Helmholtz equation. The Helmholtz equation is a fundamental physical model for the propagation and scattering of waves, with applications arising in acoustics, seismics and medical imaging, among other areas. For $\Omega = (0,1)^d$ and $k \in \mathbb{R}$, we consider the model boundary value problem:

$$-\Delta u - k^2 u = f \quad \text{and } u = 0 \text{ on } \partial\Omega.$$

Equation (14) is known as the Helmholtz equation and k is the wavenumber. In practical applications the dimension d is equal to 2 or 3 and the wavenumber can be large. After discretizing using finite differences on a uniform grid G_h with grid size h, we obtain a system of equations of the form

$$Ax = b, \quad \text{where} \quad A = -\Delta_h - k^2 I, \tag{14}$$

with $-\Delta_h$ the discretized Laplacian. For accuracy of the solution the quantity kh must be kept constant for increasing wave numbers k. This leads to a large linear system of equations that has to be solved with (preconditioned) iterative methods. The design and analysis of iterative methods and preconditioners for the discrete Helmholtz equation with $d = 2$ or $d = 3$ and high or non-constant wavenumber is an active area of research, and no standard method exists at the moment. Since the matrix $-\Delta_h$ is HPD, the spectrum $\Lambda(-\Delta_h)$ is real and positive. Moreover, the spectrum of $-\Delta_h$ contains small eigenvalues, so the matrix A will have positive and negative eigenvalues with some of them close to zero. The sparse linear system (14) is solved using Krylov subspace methods with a preconditioner to accelerate the convergence. A widely used preconditioner for the Helmholtz equation is the complex shifted Laplacian

$$M = -\Delta_h - k^2(\beta_1 - i\beta_2)I, \quad 0 \leq \beta_1, \beta_2 \leq 1. \tag{15}$$

The complex shifted Laplace (CSL) preconditioner was introduced in [8] and it is currently ranked among the most efficient preconditioners for the Helmholtz equation. The resulting preconditioned matrix is given by $\hat{A} = M^{-1}A$. The CSL problem in (15) can be interpreted physically as a damped Helmholtz problem, and this damping allows the preconditioner to be inverted efficiently with a multigrid method. The resulting preconditioned system has a more favorable eigenvalue distribution for Krylov methods. One of the nice properties of the shifted Laplacian preconditioner is that the eigenvalues of $\hat{A}$ lie on the boundary of a circle for our class of model problems. This was proved in [29], here we give a shorter proof.

Theorem 5.1. *Let $A \in \mathbb{C}^n$ and $M \in \mathbb{C}^n$ be given as in* (14) *and* (15)*. Then the spectrum of $M^{-1}A$ lies on the boundary of the circle*

$$|z - c|^2 = r^2 \tag{16}$$

with center $c = \left(\frac{1}{2}, \frac{1-\beta_1}{2\beta_2}\right)$ and radius $r = \frac{1}{2}\sqrt{1 + \frac{(1-\beta_1)^2}{\beta_2^2}} = |c|$.

Proof. Observe that $\hat{A} = T(-\Delta)$, where T is the Möbius transformation

$$T(z) = \frac{z - k^2}{z - k^2(\beta_1 - i\beta_2)}.$$

The Spectral Mapping Theorem implies that $\Lambda(\hat{A}) = T(\Lambda(-\Delta_h))$. Since a Möbius transformation maps (generalized) circles to (generalized) circles in the complex plane, the image of the real line under T is a circle. A straightforward computation shows that $\mathbb{R}$ is mapped to the circle

$$|z - c|^2 = r^2 \tag{17}$$

with center $c = \left(\frac{1}{2}, \frac{1-\beta_1}{2\beta_2}\right)$ and radius $r = \frac{1}{2}\sqrt{1 + \frac{(1-\beta_1)^2}{\beta_2^2}} = |c|$. Therefore, the eigenvalues of the preconditioned matrix $\hat{A}$ lie also on this circle. □

For low wave numbers $k \in \mathbb{R}$ the spectrum of the matrix lies on an arc close to $(1, 0)$ but for increasing wave numbers the spectrum of $\hat{A}$ contains very small eigenvalues that slow down the convergence of a Krylov solver, see also [10]. It was proposed in [7] to accelerate the convergence using a coarse grid correction operator scheme that resembles a deflation operator. This technique can be described as follows. We consider the coarse grid $G_{2h} \subset G_h$ and let Z and Y be the standard linear interpolation and full-weight restriction operators. The intergrid operators satisfy $Z = cY^T$ for a constant c that depends on the dimension d. The resulting two-grid correction matrix

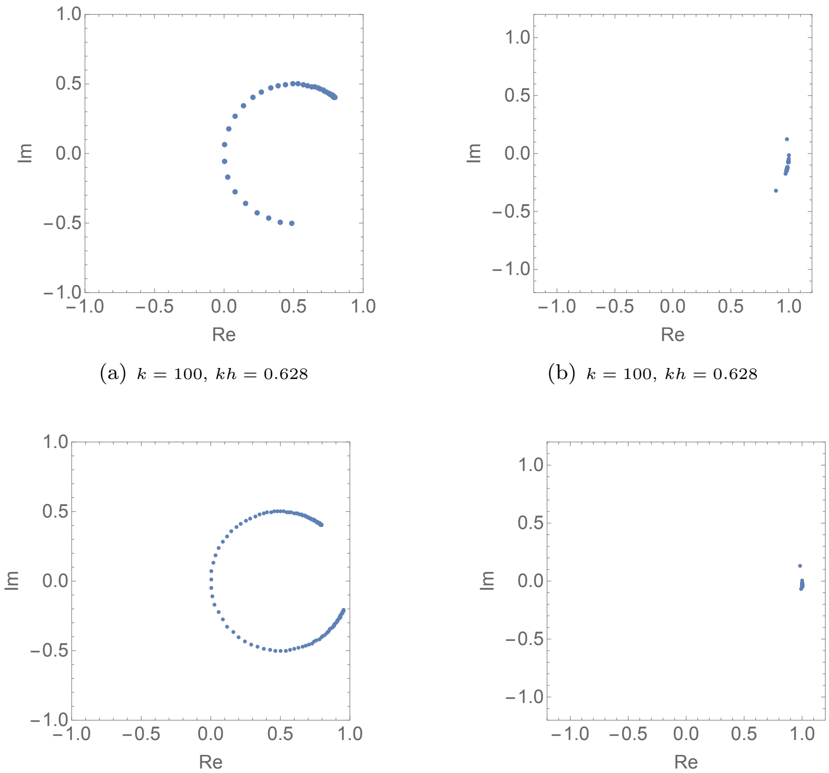
$$P_{TG} = I - \hat{A}Z(Z\hat{A}Z^T)^{-1}Z^T \tag{18}$$

is a deflation operator, and is the basis of a multilevel Krylov method introduced in [7], which leads to convergence factors that are just mildly dependent of the dimension of the linear system and the wave number.

Note that since the eigenvalues of the (normal) matrix $\hat{A}$ lie on the right half of the complex plane, which is a convex set, we have $0 \notin W(\hat{A}) = \mathrm{conv}(\Lambda(\hat{A}))$, and this implies that the coarse grid system $Z\hat{A}Z^T$ is nonsingular and the projection P_{TG} is well-defined. It has been observed in numerical experiments that the eigenvalues of $P_{TG}\hat{A}$ lie on the same circles as the eigenvalues of $M^{-1}A$ for one dimensional problems, and that the smallest eigenvalue (in modulus) is larger than the corresponding eigenvalue of $M^{-1}A$. Explicit formulas for the eigenvalues of $P_{TG}\hat{A}$ for the one-dimensional Helmholtz equation are given in [11]. Here we are able to prove that the above mentioned observations hold not only for 1D but for larger dimensions as well.

Theorem 5.2. *Let $A, M \in \mathbb{C}^n$ be given as in* (14) *and* (15)*. Let $P_{TG} \in \mathbb{C}^n$ be given as in* (18)*. Then the spectrum of P_{TG} lies on the boundary of the same circle as the spectrum of $M^{-1}A$, i.e. the spectrum of P_{TG} lies on the boundary of the circle*

$$|z - c|^2 = r^2 \tag{19}$$



(a) $k = 100,\ kh = 0.628$ (b) $k = 100,\ kh = 0.628$

(c) $k = 200,\ kh = 0.418$ (d) $k = 200,\ kh = 0.418$

Figure 1. Nonzero spectrum of $\hat{A}$ (left) and $P_{TG}\hat{A}$ (right) for $\beta_1 = 1$, $\beta_2 = 0.5$ and various values of k, and kh.

with center $c = \left(\frac{1}{2}, \frac{1-\beta_1}{2\beta_2}\right)$ *and radius* $r = \frac{1}{2}\sqrt{1 + \frac{(1-\beta_1)^2}{\beta_2^2}} = |c|$. *Moreover,*

$$|\lambda_{\min}(\hat{A})| \leq |\lambda_{\min}(P_{TG}\hat{A})|. \tag{20}$$

Proof. By Theorem 3.2, we have

$$\Lambda(P_{TG}A) \setminus \{0\} \subset \operatorname{conv}(\Lambda(\hat{A}^{-1}))^{-1}.$$

Since $\hat{A}^{-1} = \hat{T}(-\Delta)$ for the Möbius transformation $\hat{T}(z) = (z - k^2(\beta_1 - i\beta_2)/(z - k^2)$, we have $\Lambda(\hat{A}^{-1}) = \hat{T}(-\Lambda(\Delta)) \subset T(\mathbb{R})$. The image of the real line under $\hat{T}$ is the line on the complex plane that goes through the points $p_1 = (\beta_1, -\beta_2)$ and $p_2 = (1, 0)$. Therefore, the set $\operatorname{conv}(\Lambda(\hat{A})^{-1})$ is the line segment bounded by $\hat{T}(0) = (\beta_1, -\beta_2)$ and $\hat{T}(\infty) = (1, 0)$. The image of this line segment under the inversion $z \mapsto 1/z$ is contained in the circle described by equation (19). Equation (20) follows immediately from Corollary 3.5 □

Thus we proved the surprising fact that the eigenvalues of the shifted Laplace preconditioned system stay on the the same circles after applying deflation. However, they are much better clustered. Figure 1 illustrates the situation for one-dimensional Helmholtz problems.

References

[1] K. Burrage, J. Erhel, B. Pohl, and A. Williams, *A deflation technique for linear systems of equations*, SIAM Journal of Scientific Computing, 19 (1998), pp. 1245–1260.

[2] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numerical Linear Algebra with Applications, 4 (1997), pp. 43–66.

[3] Z. Dostál, *Conjugate gradient method with preconditioning by projector*, International Journal of Computer Mathematics, 23 (1988), pp. 315–323.

[4] M. Eiermann and O. G. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numerica, 10 (2003), pp. 251–312.

[5] H. H. Elman, *Iterative Methods for Large, Sparse, Nonsymmetric Systems of Linear Equations*, PhD thesis, Yale University, 1982.

[6] Y. A. Erlangga and R. Nabben, *Multilevel projection-based nested Krylov iteration for boundary value problems*, SIAM Journal on Scientific Computing, 30 (2008), pp. 1572–1595.

[7] Y. A. Erlangga and R. Nabben, *On a multilevel Krylov method for the Helmholtz equation preconditioned by shifted Laplacian*, Electronic Transactions on Numerical Analysis, 31 (2008), pp. 403–424.

[8] Y. A. Erlangga, C. Vuik, and C. W. Oosterlee, *A novel multigrid-based preconditioner for heterogeneous Helmholtz problems*, SIAM Journal on Scientific Computing, 27 (2006), pp. 1471–1492.

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

[10] M. J. Gander, I. G. Graham, and E. A. Spence, *Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: what is the largest shift for which wavenumber-independent convergence is guaranteed?*, Numerische Mathematik, 131 (2015), pp. 1–48.

[11] L. García Ramos, *Fourier Analysis of the Preconditioned Helmholtz Equation for Scattering Problems* , Master thesis, Technische Universität Berlin, 2013.

[12] A. Gaul, *Recycling Krylov Subspace Methods for Sequences of Linear Systems*, PhD thesis, Technische Universität Berlin, 2014.

[13] A. Gaul, M. H. Gutknecht, J. Liesen, and R. Nabben, *A framework for deflated and augmented Krylov subspace methods*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 495–518.

[14] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the National Bureau of Standards, 49 (1952), pp. 409–436 (1953).

[15] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[16] K. Kahl and H. Rittich, *The deflated conjugate gradient method: convergence, perturbation and accuracy*, arXiv.org, (2014), arXiv:1209.1963v3 [math.NA].

[17] S. A. Kharchenko and A. Y. Yeremin, *Eigenvalue translation based preconditioners for the GMRES(k) method.*, Numerical Linear Algebra with Applications, 2 (1995), pp. 51–77.

[18] J. Liesen and Z. Strakoš, *Krylov Subspace Methods*, Principles and Analysis, Oxford University Press, 2012.

[19] J. Liesen and P. Tichý, *The field of values bound on ideal GMRES*, arXiv.org, (2012), arXiv:1211.5969v2 [math.NA].

[20] R. B. Morgan, *GMRES with deflated restarting*, SIAM Journal on Scientific Computing, 24 (2002), pp. 20–37 (electronic).

[21] R. A. Nicolaides, *Deflation of Conjugate Gradients with Applications to Boundary Value Problems*, SIAM Journal on Numerical Analysis, 24 (1987), pp. 355–365.

[22] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.

[23] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 856–869.

[24] A. H. Sheikh, D. Lahaye, L. García Ramos, R. Nabben, and C. Vuik, *Accelerating the Shifted Laplace Preconditioner for the Helmholtz Equation by Multilevel Deflation*, May 2015.

[25] A. H. Sheikh, D. Lahaye, and C. Vuik, *On the convergence of shifted Laplace preconditioner combined with multilevel deflation*, Numerical Linear Algebra with Applications, 20 (2013), pp. 645–662.

[26] G. Starke, *Field-of-values analysis of preconditioned iterative methods for nonsymmetric elliptic problems*, Numerische Mathematik, 78 (1997), pp. 103–117.

[27] J. Tang, S. P. MacLachlan, R. Nabben, and C. Vuik, *A comparison of two-level preconditioners based on multigrid and deflation*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 1715–1739.

[28] J. Tang, R. Nabben, C. Vuik, and Y. A. Erlangga, *Comparison of two-level preconditioners derived from deflation, domain decomposition and multigrid methods*, Journal of Scientific Computing, 39 (2009), pp. 340–370.

[29] M. B. van Gijzen, Y. A. Erlangga, and C. Vuik, *Spectral analysis of the discrete Helmholtz operator preconditioned with a shifted Laplacian*, SIAM Journal on Scientific Computing, 29 (2007), pp. 1942–1958.

Technische Universität Berlin, Institut für Mathematik, Strasse des 17. Juni 136, 10623 Berlin, Germany

E-mail address: garcia@math.tu-berlin.de

Technische Universität Berlin, Institut für Mathematik, Strasse des 17. Juni 136, 10623 Berlin, Germany

E-mail address: nabben@math.tu-berlin.de