GMRES and the Minimal Polynomial *

S. L. Campbell† I. C. F. Ipsen† C. T. Kelley†
C. D. Meyer †

Abstract.

We present a qualitative model for the convergence behaviour of the Generalised Minimal Residual (GMRES) method for solving nonsingular systems of linear equations $Ax = b$ in finite and infinite dimensional spaces. One application of our methods is the solution of discretised infinite dimensional problems, such as integral equations, where the constants in the asymptotic bounds are independent of the mesh size.

Our model provides simple, general bounds that explain the convergence of GMRES as follows: If the eigenvalues of A consist of a single cluster plus outliers then the convergence factor is bounded by the cluster radius, while the asymptotic error constant reflects the non-normality of A and the distance of the outliers from the cluster. If the eigenvalues of A consist of several close clusters, then GMRES treats the clusters as a single big cluster, and the convergence factor is the radius of this big cluster. We exhibit matrices for which these bounds are tight.

Our bounds also lead to a simpler proof of existing r-superlinear convergence results in Hilbert space.

AMS subject classification: 65F10, 65J10.

Key words: GMRES, superlinear convergence, minimal polynomial, eigenvalue index

1 Introduction

The generalised minimal residual algorithm (GMRES) [20] is a popular method for solving sparse non-Hermitian systems of linear equations $Ax = b$ [5, 16]. We present a qualitative model for the convergence behaviour of GMRES in exact arithmetic. Our results are valid for finite and infinite dimensions. One application is the solution of discretised infinite dimensional problems, where the constants in the asymptotic bounds are independent of the mesh size. For instance, application of our methods to the solution of integral equations shows that the convergence rate of GMRES does not depend on the discretisation, provided the discretisation is fine enough [2].

*Received September 1995. Revised ?.

†North Carolina State University, Center for Research in Scientific Computation and Department of Mathematics, Box 8205, Raleigh, N. C. 27695-8205, USA (`slc@math.ncsu.edu`, `ipsen@math.ncsu.edu`, `Tim_Kelley@ncsu.edu`, `meyer@math.ncsu.edu`). This research was partially supported by National Science Foundation grants DMS-9122745, DMS-9423705, CCR-9102853, CCR-9400921, DMS-9321938, DMS-9020915, and DMS-9403224.

Our model provides simple, general bounds that are easy to interpret and that explain the behaviour of GMRES as follows:

> Suppose the nonsingular operator (or matrix) A has P eigenvalue clusters whose relative radii are bounded by ρ. Let d be the sum of the indices[1] of the outlying eigenvalues that do not belong to any cluster; and let σ be a bound on the maximal relative distance between any two clusters. Then our model predicts that after an initial start-up of d iterations, GMRES requires a batch of P iterations for each further residual reduction by a factor of $\sigma^{P-1}\rho$.
>
> Formally, if $\|r_j\|$ is the residual norm at iteration j then
>
> $$\|r_{d+kP}\| \leq C(\sigma^{P-1}\rho)^k \, \|r_0\|.$$
>
> The constant C is independent of k. It reflects the relative distance of the outliers from the clusters and the non-normality of A.
>
> This bound is tight in the sense that there exist matrices for which it predicts the correct number of iterations.

In the context of our model, we interpret C as an asymptotic error constant [3, p 224] and $\sigma^{P-1}\rho$ as an asymptotic convergence factor [8, p 91]. Consequently, the convergence factor in case of a single cluster, $P = 1$, is solely determined by the cluster radius. The case of several clusters can be analysed in two ways: Either view the eigenvalues of A as P small clusters of radius ρ, or view them as one large cluster of radius σ. In the first case, P iterations reduce the residual by a factor of at least $\sigma^{P-1}\rho \leq \sigma^P$. In the second case, P iterations reduce the residual by a factor of at least σ^P. Unless the cluster radii are significantly smaller than the maximal distance between clusters, both interpretations provide similar bounds on the convergence factor after P iterations. Our model of GMRES treats several close clusters as a single big cluster, so the convergence factor is the radius of this big cluster.

Residual reduction in our model starts only once the outlying eigenvalues have been processed. Thus residual reduction is delayed if the sum d of the indices of the outliers is large. The model also implies that GMRES treats the eigenvalues around zero as outliers and processes them one at a time. Moreover, when the outliers are close to zero the asymptotic error constant is much larger than when the outliers are far away from zero. When A has many eigenvalues close to zero, it may therefore take many iterations before a significant residual reduction sets in.

Our bounds are derived by constructing residual polynomials related to the minimal polynomial of A. Not only are they simpler than existing bounds based on resolvent integral methods [16, 17] and Ritz values [21], but they also lead to a new and short proof of the r-superlinear convergence result in [17].

[1] The *index* of an eigenvalue is the size of the largest Jordan block associated with the eigenvalue.

2 The Problem

Let A be a nonsingular operator on a separable Hilbert space or a nonsingular matrix, and let

$$Ax = b$$

denote a system of linear equations with solution $x^* = A^{-1}b$. In order to compute x^*, GMRES starts from an initial iterate x_0 and produces iterates x_k and residuals $r_k \equiv b - Ax_k$, $k \geq 1$. In exact arithmetic, an iterate x_k solves the linear least squares problem

$$\|r_k\| = \min_{x \in x_0 + \mathcal{K}_k} \|b - Ax\|, \qquad k \geq 1,$$

where the norm is the two-norm and

$$\mathcal{K}_k \equiv \mathrm{span}\{A^l r_0\}_{l=0}^{k-1}$$

is the Krylov space generated by the initial residual. That is, the kth iterate minimises its residual over the kth Krylov space.

As a consequence of this minimisation property [16, 20],

$$\|r_k\| = \min_{p \in \mathcal{P}_k} \|p(A) r_0\|,$$

where

$$\mathcal{P}_k \equiv \{p \;:\; p \text{ is a polynomial of degree } k \text{ and } p(0) = 1\}.$$

Hence for any polynomial $\bar{p} \in \mathcal{P}_k$,

$$\|r_k\| \leq \min_{p \in \mathcal{P}_k} \|p(A)\| \, \|r_0\| \leq \|\bar{p}(A)\| \, \|r_0\|. \tag{2.1}$$

Let $A = S \Lambda S^{-1}$, where Λ is a Jordan canonical form of A, and

$$\kappa \equiv \|S\| \, \|S^{-1}\|, \qquad \epsilon_k \equiv \min_{p \in \mathcal{P}_k} \|p(\Lambda)\|.$$

Then

$$\|r_k\| \leq \kappa \, \epsilon_k \, \|r_0\|.$$

Upper bounds have been derived for ϵ_k when A is diagonalisable, e.g. [6, 14, 19, 20, 22], and when A is defective, e.g. [9, 10].

In contrast, we construct polynomials $\bar{p} \in \mathcal{P}_k$ that are related to the minimal polynomial of A to bound $\|\bar{p}(A)\|$ in (2.1) directly. Since our bounds hold for any matrix, they may not be as accurate as bounds expressly derived for specific classes of matrices. As a trade-off, though, our bounds are much simpler and easy to interpret, so they provide a qualitative model for the behaviour of GMRES.

3 Exact Solution

We start by deriving the number of iterations required by GMRES to solve the system $Ax = b$ exactly for any b. Saad and Schultz [20] showed that, in exact arithmetic, GMRES requires at most d_{min} iterations to solve $Ax = b$, where d_{min} is the degree of the minimal polynomial of A. This means, the number of GMRES iterations to solve $Ax = b$ is small if the spectrum of A consists of a small number of non-defective eigenvalues of high multiplicity.

Following [11], we express the spectral decomposition of a matrix so that the finite dimensional analysis can be immediately extended to Hilbert spaces. Denote the distinct eigenvalues of A by λ_j, $1 \le j \le J$, and the corresponding spectral projectors by

$$X_j \equiv \frac{1}{2\pi i}\int_{\Gamma_j} (zI - A)^{-1}\,dz, \qquad 1 \le j \le J, \tag{3.1}$$

where $i^2 = -1$, and Γ_j is any circle about λ_j that contains no other eigenvalue of A. Here we choose circles as curves of integration for the spectral projectors because we want to keep our bounds simple. In other situations, e.g. when constructing polynomial preconditioners, curves such as ellipses [1, 15] may be more effective.

The spectral projectors satisfy:

$$\begin{gathered} X_jA = AX_j \\ X_jX_j = X_j, \qquad X_jX_k = 0 \text{ for } j \ne k \\ \mathrm{Range}(X_j) = \mathrm{Ker}(\lambda_j I - A)^{k_j}, \end{gathered} \tag{3.2}$$

where k_j is the index of λ_j; i.e. k_j is the smallest positive number k such that

$$\mathrm{Ker}(\lambda_j I - A)^k = \mathrm{Ker}(\lambda_j I - A)^{k+1}.$$

Denote the minimal polynomial of A by

$$p_{min}(z) \equiv \prod_{j=1}^{J} (z - \lambda_j)^{k_j}$$

and its degree by

$$d_{min} \equiv \sum_{j=1}^{J} k_j \le N.$$

Since A is nonsingular, $p_{min}(0) \ne 0$. Hence some multiple of p_{min} belongs to $\mathcal{P}_{d_{min}}$. This, together with the fact that $p_{min}(A) = 0$ gives a tight bound on the number of GMRES iterations independent of the right-hand side:

PROPOSITION 3.1 (PROPOSITION 2 IN [20]). *For any b and x_0 GMRES terminates in at most d_{min} iterations with $x_{d_{min}} = x^*$.*

PROOF. If $\bar{p} \in \mathcal{P}_{d_{min}}$ is defined by

$$\bar{p}(z) \equiv \prod_{j=1}^{J} (1 - z/\lambda_j)^{k_j}$$

then $\bar{p}(A) = 0$, and (2.1) implies

$$\|r_{d_{min}}\| \leq \|\bar{p}(A)\| \, \|r_0\| = 0.$$

□

In the finite-dimensional case when A is a single Jordan block of order N, the index of its sole eigenvalue is N. In the absence of any knowledge about b, GMRES needs N iterations to solve $Ax = b$.

The following model of the convergence behaviour of GMRES distinguishes two cases. In the first case the eigenvalues of A consist of a single cluster plus outliers, and in the second case they consist of several clusters plus outliers.

4 A Single Eigenvalue Cluster

In finite precision arithmetic, eigenvalues of multiplicity greater than one are likely to be perturbed into clusters of close but distinct eigenvalues. This does not seem to disturb GMRES, though. It still converges fast if most of the eigenvalues make up a small number of clusters. We extend the results by Saad and Schultz [20] to explain the fast convergence in the presence of clustered but possibly distinct eigenvalues.

One way to express the effect of clustering on the convergence of GMRES is to give an asymptotic bound on the size of the residual. While asymptotic results are common for infinite dimensional problems [12, 13, 17], finite dimensional analysis has focused on finite termination [20], pseudospectral bounds [16], and Ritz values [21]. Asymptotic finite dimensional results, such as the following one, are meaningful in the context of discretised infinite dimensional problems where the constants in the asymptotic bounds are independent of the mesh size N.

With an eye toward the infinite dimensional analysis, we first state the bound for a single cluster of $J - M$ eigenvalues, where $M \leq J$, with center $z = 1$ and radius ρ. We refer to *outliers* as those M eigenvalues not associated with a cluster. Their distance to the cluster is δ. Let

$$d \equiv \sum_{j=1}^{M} k_j$$

be the degree of the minimal polynomial associated with the outlying eigenvalues $\{\lambda_j\}_{j=1}^{M}$, and Z_1 their spectral projector

$$Z_1 \equiv \sum_{j=1}^{M} X_j .$$

Then $Z_2 \equiv I - Z_1$ is the spectral projector associated with the clustered eigenvalues $\{\lambda_j\}_{j=M+1}^J$. We separate the outliers from the cluster by decomposing

$$A = A_1 + A_2, \qquad \text{where} \qquad A_1 \equiv Z_1 A, \quad A_2 \equiv Z_2 A.$$

When $M = 0$ then $d = 0$ and $A = A_2$.

PROPOSITION 4.1. *Given $\rho > 0$, determine $0 \le M \le J$ so that*

$$\{\lambda_j\}_{j=M+1}^J \subset \{z \,:\, |z-1| < \rho\}$$

are the clustered eigenvalues and

$$\{\lambda_j\}_{j=1}^M \subset \{z \,:\, |z-1| > \rho\}$$

are the outliers. Define the distance of the outliers from the cluster as

$$\delta \equiv \max_{|z-1|=\rho} \max_{1 \le j \le M} \frac{|\lambda_j - z|}{|\lambda_j|}.$$

Then for any b and x_0

$$\|r_{d+k}\| \le C\,\rho^k \|r_0\|, \tag{4.1}$$

where the constant

$$C \equiv \rho \delta^d \max_{|z-1|=\rho} \|(zI - A_2)^{-1}\|$$

is independent of k.

PROOF. Define $q \in \mathcal{P}_d$ and $\bar{p} \in \mathcal{P}_{d+k}$ by

$$q(z) \equiv \prod_{j=1}^M (1 - \lambda_j^{-1} z)^{k_j}, \qquad \bar{p}(z) \equiv (1-z)^k q(z),$$

where $q(z) = 1$ if $M = 0$. According to (3.2), $A_1 A_2 = A_2 A_1 = 0$. Hence,

$$A^j = (A_1 + A_2)^j = A_1^j + A_2^j = Z_1 A_1^j + Z_2 A_2^j, \qquad j \ge 1,$$

and $p(A) = Z_1 p(A_1) + Z_2 p(A_2)$ for any polynomial p. Now $q(A_1) = 0$ implies $Z_1 \bar{p}(A_1) = 0$, so

$$\bar{p}(A) = Z_1 \bar{p}(A_1) + Z_2 \bar{p}(A_2) = Z_2 \bar{p}(A_2).$$

Express Z_2 as a resolvent integral as in (3.1),

$$Z_2 = \frac{1}{2\pi i} \int_{|z-1|=\rho} (zI - A_2)^{-1}\, dz.$$

Hence

$$Z_2 \bar{p}(A_2) = \frac{1}{2\pi i} \int_{|z-1|=\rho} q(z)\,(1-z)^k\,(zI - A_2)^{-1}\, dz.$$

Taking norms and using the definition of δ completes the proof,

$$\|Z_2\bar{p}(A_2)\| \leq \rho\delta^d\rho^k \max_{|z-1|=\rho} \|(zI-A_2)^{-1}\|.$$

Note that we could have used A in place of A_2. $\square$

Proposition 4.1 says that our model of GMRES needs at least d iterations before the onset of residual reduction since GMRES has to process M outliers and go through k_j iterations for the jth outlier. Therefore residual reduction in our model is delayed if there are many outliers or outliers with high index. Since GMRES minimises the residual norm over the current Krylov space, our model's delay in residual reduction differs from the actual behaviour of GMRES. Therefore our model provides only a qualitative, rather than a quantitative description for the behaviour of GMRES.

The number of additional iterations for processing the cluster depends mainly on the cluster radius. Specifically, after d iterations, the residual norms in our model converge to zero as

$$\|r_{d+k}\|/\|r_0\| \sim \rho^k.$$

Convergence in our model thus requires that $\rho < 1$. This implies that zero eigenvalues are regarded as outliers.

In the context of our model, we interpret the cluster radius ρ as an asymptotic convergence factor [8, p 91], and the constant C as an asymptotic error constant [3, p 224]. If we write

$$C = (\rho\delta^M)\,(\delta^{d-M} \max_{|z-1|=\rho} \|(zI-A_2)^{-1}\|),$$

then $\rho\delta^M$ reflects the cluster size and the distance of the outliers from the cluster, while the remaining terms are indicators for the non-normality of A ($d = M$ when A is normal).

Corollary 4.2. *When A is normal, C can be chosen so that $C \leq \delta^M$; and when A is diagonalisable, C can be chosen so that $C \leq \kappa\delta^M$.*

When A is normal then δ is bounded above by the condition number of A. In this case C can be large when A is ill-conditioned.

Saad and Schultz [20, Theorem 5] derive the following result for diagonalisable matrices with eigenvalues with non-positive real parts.

Proposition 4.3 (Theorems 4 and 5 in [20]). *If A is diagonalisable then*

$$\|r_{k+d}\| \leq \kappa\, \epsilon_{k+d}\, \|r_0\|,$$

and $\epsilon_{d+k} = \min_{p\in\mathcal{P}_{d+k}} \max_{1\leq j\leq J} |p(\lambda_j)|$.

If $\{\lambda_j\}_{j=1}^M$ are the eigenvalues of A with non-positive real parts, and all other eigenvalues $\{\lambda_j\}_{j=M+1}^J$ are situated in a circle with center $z = 1$ and radius $\rho < 1$, then for any b and x_0

$$\epsilon_{k+d} \leq \rho^{k+d-M} \max_{M+1\leq j\leq J} \prod_{l=1}^{M} \frac{|\lambda_l - \lambda_j|}{|\lambda_l|}.$$

Since

$$\max_{M+1\leq j\leq J}\prod_{l=1}^{M}\frac{|\lambda_l-\lambda_j|}{|\lambda_l|}\leq\delta^M,$$

and $d=M$ for diagonalisable matrices, Proposition 4.3 implies

$$\|r_{k+d}\|\leq C\rho^k\|r_0\|,$$

where $C=\kappa\delta^M$. But this is exactly Corollary 4.2. Hence Saad and Schultz's result can be considered a special case of our Proposition 4.1.

In summary, our model of GMRES says that in the single cluster case the asymptotic convergence factor of GMRES is proportional to the radius of the cluster; while the asymptotic error constant reflects the non-normality of A and the distance of the outliers from the cluster.

4.1 Examples

We consider the two extreme cases of a maximally defective matrix and a normal matrix, both with a single eigenvalue equal to one. Then $M=0$, $J=1$, $A_2=A$, $\lambda_{M+1}=\lambda_1=1$.

In the case of a maximally defective matrix, Proposition 4.1 predicts that residual reduction occurs only after the maximal number of iterations. Suppose A is a single Jordan block of order N with eigenvalue one. Then $d=N$. For any $\rho>0$, Proposition 4.1 implies

$$\|r_{N+k}\|\leq C\rho^k\|r_0\|,$$

with $C=\rho\delta^N\max_{|z-1|=\rho}\|(zI-A)^{-1}\|$, so residual reduction occurs only after N iterations. This is in agreement with Proposition 3.1, which implies that in this case GMRES needs N iterations to solve $Ax=b$ for any b.

Moreover, $C\geq\delta^N/\rho^{N-1}$ for any $\rho>0$, as the magnitude of the $(1,N)$ element in $(zI-A)^{-1}$ is $1/|z-1|^N$ [7, §6.2.13]. This illustrates that C can be arbitrarily large for highly non-normal matrices.

In the other extreme case suppose that $A=I$ is the identity matrix. Then $d=0$. For any $\rho>0$, Proposition 4.1 and Corollary 4.2 imply

$$\|r_k\|\leq C\rho^k\|r_0\|,$$

with $C\leq\rho\delta/\rho=\delta$. Thus $\|r_1\|\leq\delta\rho\|r_0\|$ for any $\rho>0$, which implies that GMRES converges in one iteration. Again, this agrees with the predicted number of iterations from Proposition 3.1.

To get the same result, i.e. convergence in one iteration, for $A=\gamma I$ when γ is not necessarily equal to one, we have to define *relative clusters*

$$\{\lambda_j\}_{j=1}^M\subset\{z\,:\,|z-\gamma|<\rho|\gamma|\}.$$

This is the motivation for the definition of clusters in the next section. Note that the notion of *relative* cluster rules out a cluster of eigenvalues around zero. Therefore, our model of GMRES treats eigenvalues around zero as outliers and processes them one at a time. Consequently, residual reduction in our model is delayed if many eigenvalues are clustered around zero.

5 Several Clusters of Eigenvalues

We extend Proposition 4.1 to the case where the eigenvalues of A belong to more than one cluster. In contrast to Theorems 3.4.9 and 5.11.1 in [17], our results, although less general, have simple statements and short and elementary proofs.

In addition to the bound ρ on the relative cluster radii and the relative distance δ of the outliers from the clusters, the bound now also contains a factor σ describing the interaction among clusters: it is the maximal relative distance between any two clusters. Suppose A has M_1 outlying eigenvalues and P non-intersecting clusters of eigenvalues. Let

$$d \equiv \sum_{j=1}^{M_1} k_j$$

be the degree of the minimal polynomial associated with the outliers $\{\lambda_j\}_{j=1}^{M_1}$, and Z_1 their spectral projector

$$Z_1 \equiv \sum_{j=1}^{M_1} X_j.$$

The clusters are centered at distinct non-zero points γ_p, where the cluster around γ_p contains eigenvalues $\{\lambda_j\}_{j=M_{p-1}+1}^{M_p}$, $2 \le p \le P+1$. The associated spectral projectors are

$$Z_p \equiv \sum_{q=M_{p-1}+1}^{M_p} X_q, \qquad 2 \le p \le P+1.$$

We separate the clusters and the outliers by decomposing

$$A = A_1 + \sum_{p=2}^{P+1} A_p, \qquad \text{where} \qquad A_1 = Z_1 A, \quad A_p = Z_p A.$$

Proposition 5.1. *Given $\rho > 0$ and distinct non-zero complex numbers $\{\gamma_p\}_{p=2}^{P+1}$, determine*

$$0 \le M_1 \le M_2 \le \ldots \le M_{P+1} \equiv J$$

so that the non-intersecting sets

$$\{\lambda_j\}_{j=M_{p-1}+1}^{M_p} \subset \{z \,:\, |z-\gamma_p| < \rho|\gamma_p|\}, \qquad 2 \le p \le P+1,$$

are the clusters and

$$\{\lambda_j\}_{j=1}^{M_1} \subset \{z \,:\, |z-\gamma_p| > \rho|\gamma_p| \quad \text{for} \quad 2 \le p \le P+1\}$$

are the outliers. Define the distance of the outliers from the clusters as

$$\delta \equiv \max_{2\le p\le P+1} \; \max_{|z-\gamma_p|=\rho|\gamma_p|} \; \max_{1\le j\le M} \frac{|\lambda_j - z|}{|\lambda_j|},$$

and the maximal distance between clusters as

$$\sigma \equiv \max_{2 \le p \le P+1} \max_{|z-\gamma_p|=\rho|\gamma_p|} \max_{q \ne p} \frac{|\gamma_q - z|}{|\gamma_q|}.$$

Then for any b and x_0

$$\|r_{d+kP}\| \le C\,(\sigma^{P-1}\rho)^k\,\|r_0\|, \tag{5.1}$$

where the constant

$$C \equiv P\rho\delta^d \max_{2 \le p \le P+1} \max_{|z-\gamma_p|=\rho|\gamma_p|} \|(zI - A_p)^{-1}\| \tag{5.2}$$

is independent of k.

Proof. Following the proof of Proposition 4.1, define $q \in \mathcal{P}_d$ and $\bar{p} \in \mathcal{P}_{d+kP}$ by

$$q(z) \equiv \prod_{j=1}^{M_1}(1 - \lambda_j^{-1}z)^{k_j}, \qquad \bar{p}(z) \equiv q(z)\prod_{p=2}^{P+1}(1 - \gamma_p^{-1}z)^k,$$

where $q(z) = 1$ if $M_1 = 0$. As before, $q(A_1) = 0$ implies $Z_1\bar{p}(A_1) = 0$, so

$$\bar{p}(A) = \sum_{p=1}^{P+1} Z_p\bar{p}(A_p) = \sum_{p=2}^{P+1} Z_p\bar{p}(A_p).$$

Expressing the projectors Z_p as resolvent integrals yields

$$\begin{aligned}\bar{p}(A) \;=\; \sum_{p=2}^{P+1}\frac{1}{2\pi i}\int_{|z-\gamma_p|=\rho|\gamma_p|} (1-\gamma_p^{-1}z)^k \prod_{j=1}^{M_1}(1-\lambda_j^{-1}z)^{k_j} \\ \prod_{q \ne p}(1-\gamma_q^{-1}z)^k (zI - A_p)^{-1}dz.\end{aligned}$$

Taking norms and using the definitions of δ and σ completes the proof,

$$\|\bar{p}(A)\| \le P\rho\rho^k\delta^d\sigma^{(P-1)k} \max_{2 \le p \le P+1} \max_{|z-\gamma_p|=\rho|\gamma_p|} \|(zI - A_p)^{-1}\|.$$

□

In the special case of a single cluster, $P = 1$, Proposition 5.1 reduces to Proposition 4.1. To understand the expression for the convergence factor in case of several clusters, we pursue two approaches for analysing the convergence in our model of GMRES. Either use Proposition 5.1 and view the eigenvalues as P small clusters of radius ρ, or else use Proposition 4.1 and view the eigenvalues as one large cluster of radius σ. In the first approach, P iterations reduce the residual by a factor of $\sigma^{P-1}\rho \le \sigma^P$. In the second approach, one iteration reduces the residual by a factor of σ, hence P iterations reduce the residual by a factor of σ^P. If the radii of the individual clusters are much smaller than the distances among the clusters then $\sigma^{P-1}\rho < \sigma^P$. Hence over the course of P iterations the first approach results in a somewhat smaller bound on the convergence factor. However, unless $\sigma^{P-1}\rho << \sigma^P$, the difference in convergence factors is small and our model of GMRES treats a conglomerate of several clusters as one giant cluster. Thus, after every batch of P iterations, progress of GMRES on several small clusters is similar to progress of GMRES on one giant cluster.

6 Infinite Dimensional Analysis

We extend the analysis of the previous section to the following Hilbert space setting. Denote by H a separable Hilbert space, by K a compact operator on H so that $A \equiv I - K$ is nonsingular, and by $\|\cdot\|$ the norm on H. Let the inner product on H define orthogonality.

According to the spectral theorem for compact operators [4], A has a countable sequence of eigenvalues with one as the only accumulation point. This means the set of eigenvalues λ_j for which $|\lambda_j - 1| > \rho$ is finite. Since the algebraic multiplicity of an eigenvalue $\lambda_j \neq 1$ is finite, the spectral projectors for $\lambda_j \neq 1$ satisfy (3.1) and (3.2). Therefore, Proposition 4.1 and its proof hold in infinite dimensions.

Proposition 6.1. *Let the eigenvalues of A be numbered so that*

$$|\lambda_j - 1| \geq |\lambda_{j+1} - 1|, \qquad j \geq 1.$$

Given $\rho > 0$, determine $0 \leq M < \infty$ so that

$$\{\lambda_j\}_{j=1}^{M} \subset \{z \,:\, |z-1| > \rho\},$$

are the outliers and

$$\{\lambda_j\}_{j \geq M+1} \subset \{z \,:\, |z-1| < \rho\}$$

is the cluster. Define the distance of the outliers from the cluster as

$$\delta \equiv \max_{|z-1|=\rho} \max_{1 \leq j \leq M} \frac{|\lambda_j - z|}{|\lambda_j|}.$$

Then for any b and x_0

$$\|r_{d+k}\| \leq C\, \rho^k\, \|r_0\|, \tag{6.1}$$

where the constant

$$C \equiv \rho \delta^d \max_{|z-1|=\rho} \|(zI - A_2)^{-1}\|$$

is independent of k.

The superlinear convergence theorem in [17] says that under the assumptions of Proposition 6.1

$$\lim_{k \to \infty} \|r_k\|^{1/k} = 0. \tag{6.2}$$

The result in [17] is an r-superlinear result and is stated in a slightly different, but completely equivalent [18], way. Our proof of (6.2) is straightforward and requires only the observation that (6.1) implies

$$\mathrm{limsup}_{k \to \infty} \|r_k\|^{1/k} \leq \rho$$

for any $\rho > 0$.

Acknowledgements

We thank Stan Eisenstat and Olavi Nevanlinna for helpful discussions.

REFERENCES

1. S. F. Ashby, T. A. Manteuffel, and P. E. Saylor, *Adaptive polynomial preconditioning for Hermitian indefinite linear systems*, BIT, 29 (1989), pp. 583–609.
2. S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, C. D. Meyer, and Z. Q. Xue, *Convergence estimates for solution of integral equations with GMRES*, Tech. Report CRSC-TR95-13, North Carolina State University, Center for Research in Scientific Computation, March 1995.
3. G. Dahlquist and Å. Björck, *Numerical Methods*, Prentice Hall, 1974.
4. N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory*, Wiley Interscience, New York, 1957.
5. R. Freund, G. Golub, and N. Nachtigal, *Iterative solution of linear systems*, in Acta Numerica 1992, Cambridge University Press, Cambridge, 1992, pp. 57–100.
6. G. H. Golub and R. S. Varga, *Chebyshev semiiterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods*, Numer. Math., 3 (1961), pp. 147–168.
7. R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.
8. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods*, John Wiley &Sons, New York, 1966.
9. Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*. 1993.
10. ——, *The convergence of Krylov subspace methods for large unsymmetric linear systems*. 1994.
11. T. Kato, *Perturbation Theory for Linear Operators*, no. 132 in Grundlehren der mathematischen Wissenschaften, Springer Verlag, New York, second ed., 1980.
12. C. T. Kelley and Z. Q. Xue, *GMRES and integral operators*, SIAM J. on Sci. Comp., 17 (1996).
13. T. Kerkhoven and Y. Saad, *On acceleration methods for coupled nonlinear elliptic systems*, Numer. Math., 60 (1992), pp. 525–548.
14. T. A. Manteuffel, *The Tchebychev iteration for solving nonsymmetric linear systems*, Numer. Math., 28 (1977), pp. 307–327.
15. ——, *Adaptive procedure for estimating parameters for the nonsymmetric Tchebychev iteration*, Numer. Math., 31 (1978), pp. 183–208.

16. N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 778–795.

17. O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

18. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

19. Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comput., 37 (1981), pp. 105–126.

20. Y. Saad and M. Schultz, *GMRES a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comp., 7 (1986), pp. 856–869.

21. H. A. van der Vorst and C. Vuik, *The superlinear convergence behaviour of GMRES*, Journal of computational and applied mathematics, 48 (1993), pp. 327–341.

22. O. Widlund, *A Lanczos method for a class of nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 15 (1978), pp. 801–812.