Helsinki University of Technology Institute of Mathematics Research Reports

Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja

Espoo 1999 A418

POLE ASSIGNMENT PROBLEMS FOR ERROR BOUNDS FOR GMRES

Marko Huhtanen

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Marko Huhtanen: Pole assignment problems for error bounds for GMRES ; Helsinki University of Technology Institute of Mathematics Research Reports A418 (1999).

Abstract: *We derive a family of new error bounds for solving a linear system* $Ax = b$ *using GMRES, where* $A \in \mathbb{C}^{n \times n}$ *and* $b \in \mathbb{C}^n$. *In the simplest case the estimates are based on "feedback" in the sense that running GMRES with* $Ax = b$ *is almost equivalent to running GMRES with* $(A + bv^*)y = b$ *for any vector* $v \in \mathbb{C}^n$. *Using this we obtain a generalization of the classical estimate*

$$\min_{p \in \mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(X) \min_{p \in \mathcal{P}_j(0)} \|p(\Lambda)\| \|b\|$$

for diagonalizable matrices $A = X \Lambda X^{-1}$ *by constructing bounds of similar type for every matrix* A*, regardless of* A *being diagonalizable or not. For an optimal choice of* v *there arises a pole assignment problem. Analogously, we obtain another family of error bounds by considering* $A(I + bv^*)z = b$ *aside from* $Ax = b$. *After considering these two particular cases it is straightforward to generalize the proposed technique to larger rank perturbations of* A.

AMS subject classifications: 65F10.

Keywords: GMRES, error bounds, pole assignment problem, nonnormality.

ISBN 951-22-4647-3
ISSN 0784-3143
Painotalo Libella Oy, Espoo, 1999

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics
P.O. Box 1100, 02015 HUT, Finland
email: *math@hut.fi*
downloadables: *http://www.math.hut.fi/*

author's email: *Marko.Huhtanen@@hut.fi*

1 Introduction

Assume $A \in \mathbb{C}^{n\times n}$ is a large, possibly sparse matrix and $b \in \mathbb{C}^n$ for which the linear system

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

is solved by using GMRES. Let $\mathcal{P}_j(0)$ denote the set of polynomials p of degree at most j normalized such that $p(0) = 1$. If A is diagonalizable, then the most standard bound for the error, already presented in the original paper by Saad and Schultz [11], is based on the spectrum of A such that

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(X) \min_{p\in\mathcal{P}_j(0)} \|p(\Lambda)\|\|b\|, \tag{2}$$

where $A = X\Lambda X^{-1}$ is an eigendecomposition of A. See also the book by Greenbaum [1] or [9] by Nachtigal, Reddy and Trefethen or the very nice short presentation by Hochbruck and Lubich [3]. Here $\kappa(X) = \|X\|\|X^{-1}\|$ denotes the condition number of X and, without any loss of generality, we have assumed that the initial guess for the solution is $x_0 = 0$. Hence, in the bound (2) the quantity obtained from the minimization problem on the spectrum of A is scaled by the condition number $\kappa(X)$. For nonnormal A this scaling can make the bound over-pessimistic which, especially, happens "in the limit". That is, when A is not diagonalizable this bound ceases to exist and then the scaling can be considered to be infinite. Consequently, then the error bounds for GMRES have to be derived by applying completely different techniques, see e.g. [1] or [10] by Nevanlinna. As nondiagonalizability is a rare incidence in practice, the most serious problem with (2) is that just a few ill-conditioned eigenvalues can completely ruin this estimate. In this paper we show how to overcome this and derive bounds of type (2) for *every* linear system, regardless of A being diagonalizable or not, and, in such a way that these bounds involve *only* properties of A with respect to b.

The key to new bounds for solving (1) using GMRES is to apply the approach initialized in [4, 5] based on small rank perturbations of the original matrix A. Now the difference is that we consider very carefully chosen small rank perturbations of A. In the simplest case assume solving, aside from (1), the linear system

$$(A + bv^*)y = b \tag{3}$$

for some vector $v \in \mathbb{C}^n$. That is, from the control theoretic point of view b is fed back to the system. In many ways, using GMRES with this linear system is almost equivalent to using GMRES with (1). The first preserved "invariants" are Krylov subspaces at b. More precisely, for any $j \geq 0$ holds $\mathcal{K}_j(A; b) = \mathcal{K}_j(A + bv^*; b)$, where, as usual,

$$\mathcal{K}_j(A; b) := \operatorname{span}\{b, Ab, ..., A^{j-1}b\}. \tag{4}$$

This simple property is very useful for our purposes since it allows to scale approximative solutions to the linear system (3) to those of $Ax = b$ and vice versa such that estimates are obtained when using GMRES. Still, even if Krylov subspaces at b are preserved, the spectra of A and $A + bv^*$ can be very different. In particular, if $\mathcal{K}_n(A; b) = \mathbb{C}^n$, then a nondiagonalizable A can always be made diagonalizable as $A + bv^*$ can be assigned any set of eigenvalues by choosing a vector $v \in \mathbb{C}^n$ appropriately. This is almost an elementary fact in control theory where it is utilized in the pole assignment problem, see e.g. the book by Wonham [14]. Thus, picking a $v \in \mathbb{C}^n$ such that $A + bv^*$ is diagonalizable with a well-conditioned eigenbasis we can use the bound (2) to give error estimates for the approximative solutions to (3). This can obviously be done regardless of A being diagonalizable or not. Then what remains is to scale this bound to apply to the original linear system (1). Consequently, in the obtained bound there appears an additional scaling factor depending on a moment. This, fortunately, does not lead to circular reasoning as this moment can also be estimated by using (2), see Theorem 2 for details. In all, we obtain an estimate that is more local than (2) in the sense that it involves an ideal GMRES problem for $A + bv^*$ instead of for A. In particular, corresponding to different choices of v, we obtain a family of error bounds for solving (1) using GMRES such that with the choice $v = 0$ we obtain (2) whenever A is diagonalizable. This family of bounds is readily available by simply varying $v \in \mathbb{C}^n$. Thus, choosing $v \in \mathbb{C}^n$ to make the resulting bound optimal as such, leads, in a natural way, to a robust pole assignment problem.

In addition to the perturbed systems (3) one can consider multiplicative rank-one modifications of A leading to systems of the form

$$A(I + bv^*)z = b \tag{5}$$

with a vector $v \in \mathbb{C}^n$. Obviously this is not of standard feedback type in the control theoretic sense. Still, we obtain another family error bounds for GMRES as with (3) but without additional scaling, see Theorem 6 for details. And, after considering the cases (3) and (5) it is straightforward to generalize the proposed technique to larger rank perturbations of similar type. That is, introducing the linear system

$$(A + \sum_{k=0}^{p} A^k b v_k^*)w = b \tag{6}$$

with $v_k \in \mathbb{C}^n$, for $0 \le k \le p$, aside from the original system $Ax = b$ yields a very general family of bounds for GMRES for soving $Ax = b$, see Theorem 7.

The paper is organized as follows. In Section 2 we derive new error bounds for GMRES and we illustrate the bounds with several examples. In Section 3 we consider the arisen robust pole assignment problems and discuss how the algorithms derived by numerical analyst working in control theory apply to our problems.

2 Error bounds for GMRES

Let $A \in \mathbb{C}^{n\times n}$ be a large, possibly sparse matrix and $b \in \mathbb{C}^n$. The most standard error bound for solving a diagonalizable linear system

$$Ax = b, \tag{7}$$

using GMRES with the initial guess $x_0 = 0$ is

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(X) \min_{p\in\mathcal{P}_j(0)} \|p(\Lambda)\|\|b\|, \tag{8}$$

where $A = X\Lambda X^{-1}$ is an eigendecomposition of A. Here $\mathcal{P}_j(0)$ denotes the set of polynomials p of degree at most j normalized such that $p(0) = 1$. To simplify the notation, we have taken the initial guess $x_0 = 0$ but the changes are straightforward for nontrivial initial guesses.

The bound (8) is not much of use if the scaling factor, i.e., the condition number $\kappa(X) = \|X\|\|X^{-1}\|$ is large. In particular, if A is not diagonalizable, then the scaling can be considered to be infinite and the bound (8) does not even exist. In what follows we demonstrate how to improve the bound (8) with large (or infinite) scaling $\kappa(X)$ by using a particular type of small rank perturbations of A.

A natural rank-one perturbation of A is obtained by using b in the simplest possible way. That is, consider, aside from $Ax = b$, the linear system

$$(A + bv^*)y = b \tag{9}$$

with $v \in \mathbb{C}^n$. From the control theoretic point of view the vector b is just "fed back" to the original linear system. As to using GMRES the linear systems (7) and (9) are nearly equivalent. In particular, for Krylov subspaces at b the following should be obvious.

Lemma 1 *Let $A \in \mathbb{C}^{n\times n}$ and $b \in \mathbb{C}^n$. Then for any vector $v \in \mathbb{C}^n$ holds*

$$\mathcal{K}_j(A; b) = \mathcal{K}_j(A + bv^*; b)$$

for all $j \geq 0$.

Although this lemma is simple, it is the key to the following bound where, for a diagonalizable $A+bv^*$, we denote by $A+bv^* = X_v\Lambda_v X_v^{-1}$ its diagonalization (some). That is, the following theorem shows that by using completely different eigenvalues than those of A we can bound the error for GMRES for the linear system (7).

Theorem 2 *Assume $A + bv^*$ is invertible and diagonalizable and, at the j^{th} step, y_j is the approximation to (9) generated by using GMRES with $x_0 = 0$. Then*

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(X_v) \min_{p\in\mathcal{P}_j(0)} \|p(\Lambda_v)\| \frac{\|b\|}{|1 - v^*y_j|}. \tag{10}$$

While j grows, $\frac{1}{|1-v^*y_j|}$ converges to $|1+v^*A^{-1}b|$ with the speed that can be estimated by using

$$\|(A+bv^*)^{-1}b - y_j\| \leq \|(A+bv^*)^{-1}\|\kappa(X_v) \min_{p\in\mathcal{P}_j(0)} \|p(\Lambda_v)\|\|b\|.$$

Proof. Let $(A+bv^*)y_j = b+\epsilon_j$ with $\epsilon_j = (A+bv^*)y_j - b$. Thus, we have $Ay_j = (A+bv^*)y_j - bv^*y_j = b(1-v^*y_j)+\epsilon_j$ so that

$$A\frac{y_j}{1-v^*y_j} = b + \frac{\epsilon_j}{1-v^*y_j},$$

provided that $1-v^*y_j \neq 0$. Being not necessarily the norm minimizing approximative solution to the linear system $Ax=b$, we have by the definition of GMRES

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \|A\frac{y_j}{1-v^*y_j} - b\| = \|\frac{\epsilon_j}{1-v^*y_j}\|.$$

This holds since $y_j \in \mathcal{K}_j(A;b)$ by Lemma 1, that is, at this point the same Krylov subspaces are necessary for inequality. In particular, applying now the bound (8) to $A+bv^*$ to estimate ϵ_j gives the first part of the claim.

As to the claim concerning the numerator, there holds $(A+bv^*)y = b$ at the limit, so that $Ay = b(1-v^*y)$ and $y = (1-v^*y)A^{-1}b$. Thus $(A+bv^*)^{-1}b = (1-v^*y)A^{-1}b$ from which we obtain $|1-v^*y| = \|A(A+bv^*)^{-1}\frac{b}{\|b\|}\|$. Now the claim follows after using the Sherman-Morrison-Woodbury formula with $(A+bv^*)^{-1}$ and multiplying with A from the left and then performing obvious computations.

As to the approximation concerning $\frac{1}{|1-v^*y_j|}$, this is readily accomplished by using (8) with the linear system $(A+bv^*)y = b$. $\square$

First, note that the term $\frac{1}{|1-v^*y_j|}$ can be guaranteed to converge to the constant $|1+v^*A^{-1}b|$ with the bound depending on the minimization problem on the spectrum of $A+bv^*$. Thus, as soon as $\min_{p\in\mathcal{P}_j(0)}\|p(\Lambda_v)\|$ can be made small, the convergence of GMRES for solving $Ax=b$ takes place as $\frac{1}{|1-v^*y_j|}$ cannot vary any longer. Second, this bound is more local than (8) where only the norm of b appears on the right-hand side. In the bound (10) there is now, in addition to the factor $\frac{\|b\|}{|1-v^*y_j|}$, an ideal GMRES bound for $A+bv^*$. This obviously depends strongly on the vector b and, especially, on the properties of A with respect to b. Third, eigenvalues *do* matter as opposed to some negative results by Greenbaum and Strakos [2]. They need not be the eigenvalues of A that count but those of $A+bv^*$ for an appropriate choice of v. In particular, A and $A+bv^*$ do not, in general, generate the same Krylov residual spaces as defined in [2] so that the analysis of Greenbaum and Strakos is not applicable in this case. Fourth, this bound generalizes (8) and reduces to the same bound with the choice $v=0$ whenever A is diagonalizable. Since this really yields, for different choices of $v \in \mathbb{C}^n$, a family of bounds for $\min_{p\in\mathcal{P}_j(0)}\|p(A)b\|$, there is no reason to expect (8) to be optimal. This, especially, happens in the case of nondiagonalizable A. In addition to the choice $v=0$, the numerator disappears also in the following case.

Corollary 3 *Assume v lies in the orthogonal complement of $\mathcal{K}_l(A,b)$. Then*

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(X_v) \min_{p\in\mathcal{P}_j(0)} \|p(\Lambda_v)\|\|b\|$$

for $j \leq l$.

Proof. This follows immediately from the fact that $y_j \in \mathcal{K}_l(A,b)$ for every $j \leq l$ so that $v^* y_j = 0$. □

Only very seldom bounds of this type do not exist.

Corollary 4 *If $\mathcal{K}(A;b) = \mathbb{C}^n$, then there exists bounds of type (10).*

Proof. This follows from the so-called pole assignment theorem which is a widely used property in control theory. That is, if $\mathcal{K}(A;b) = \mathbb{C}^n$, then any set of eigenvalues can be assigned to $A + bv^*$ with an appropriate choice of $v \in \mathbb{C}^n$, see e.g. [14]. In particular, choosing v such that $A + bv^*$ is semi-simple gives the claim . □

Thus, if A is cyclic (but not necessarily diagonalizable!) and b is a cyclic vector for A, i.e., $\mathcal{K}(A;b) = \mathbb{C}^n$, then one can *always* bound GMRES via solving a minimization problem on the spectrum of a diagonalizable matrix $A + bv^*$ for a somehow chosen vector v. The scaling is then the condition number $\kappa(X_v)$ multiplied by a number close to reciprocal of "1 plus the moment $v^* A^{-1} b$". Note that, by Corollary 3, if one had a procedure of choosing $v \in \mathbb{C}^n$ orthogonal to $\mathcal{K}_l(A,b)$, where l is the maximum number of steps to be performed (bounded by memory requirements), then this constant would disappear. We illustrate this in Example 1.

If $\mathcal{K}(A;b) \neq \mathbb{C}^n$, then it may not be possible to make $A + bv^*$ diagonalizable with a vector $v \in \mathbb{C}^n$. However, there still exists a bound of type given in Theorem 2 and, for the sake of completeness, we present it to show that GMRES can *always* be bounded by using the spectrum of a diagonalizable matrix depending on the properties of A and b. Namely, A is unitary similar to $\begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$, where A_1 is the Hessenberg's matrix constructed by using the Arnoldi method with b. For the Arnoldi method, see e.g. [1]. In control theory this is a canonical form for A decomposing $\mathbb{C}^n$ into controllable subspace plus its orthogonal complement. Let b_1 denote the vector b in the coordinates of this basis of $\mathcal{K}(A;b)$. We denote by k the dimension of $\mathcal{K}_n(A;b)$ and by $A_1 + b_1 v^* = \hat{X}_v \hat{\Lambda}_v \hat{X}_v^{-1}$ a diagonalization of $A_1 + b_1 v^*$. Note that A_1 is cyclic with cyclic vector b_1, so that there exists a vector $v \in \mathbb{C}^k$ such that $A_1 + b_1 v^*$ is diagonalizable by the pole assignment theorem.

Corollary 5 *Assume $A_1 + b_1 v^* \in \mathbb{C}^{k\times k}$ is invertible and diagonalizable and, at the j^{th} step, y_j is the approximation to (9) generated by using GMRES with $x_0 = 0$ represented in the basis of $\mathcal{K}(A;b)$. Then*

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| \leq \kappa(\hat{X}_v) \min_{p\in\mathcal{P}_j(0)} \|p(\hat{\Lambda}_v)\| \frac{\|b\|}{|1 - v^* \hat{y}_j|}.$$

While j grows, $\frac{1}{|1-v^*\hat{y}_j|}$ converges to $|1+v^*A_1^{-1}b_1|$ with the speed that can be estimated by using

$$\|(A_1+b_1v^*)^{-1}b_1-\hat{y}_j\| \le \|(A_1+b_1v^*)^{-1}\|\kappa(\hat{X}_v)\min_{p\in\mathcal{P}_j(0)}\|p(\hat{\Lambda}_v)\|\|b\|.$$

Proof. This follows readily after restricting A to the invariant subspace $\mathcal{K}_n(A;b)$ of A. □

If v is generic, then $v^*A^{-1}b$ is likely to be small compared with the norm of v as long as A is not nearly singular. In particular, if v lies in the orthogonal complement, or almost, of $A^{-1}b$, then this constant has practically no effect on the bounds. Let us demonstrate this with the following example which also shows that the disappearance of the numerator in Corollary 3 is not an artificial trick. All the computations are performed with MATLAB.

EXAMPLE 1. Let $A\in\mathbb{C}^{n\times n}$ be the standard stumbling block for the bound (8), that is, let A be the translated nilpotent shift

$$A=\begin{bmatrix} 2 & 0 & \cdots & 0 & 0\\ 1 & 2 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ 0 & 0 & \cdots & 2 & 0\\ 0 & 0 & \cdots & 1 & 2\end{bmatrix} \tag{11}$$

and consider using GMRES for $Ax=b$ with b being the first standard basis vector. The bound (8) is now useless, since A is not even diagonalizable. Taking v to be the n^{th} basis vector $A+bv^*$ becomes even normal, that is, $\kappa(X_v)=1$ and the spectrum of $A+bv^*$ is equal to the set $2+\{z\in\mathbb{C}: z^n=1\}$. Now v is in the orthogonal complement of $\mathcal{K}_j(A,b)$ for $1\le j\le n-1$. Also, $|1+v^*A^{-1}b|$ converges to 1 very quickly while n grows as shown in Table 2.1.

	$n=5$	$n=10$	$n=15$	$n=20$	$n=25$
$1+v^*A^{-1}b$	1.031250	0.9990234	1.000031	0.9999990	1.000000

Table 1: The behavior of $1+v^*A^{-1}b$ as a function of dimension n for Example 1.

Thus, in this case the bound is essentially an ideal GMRES bound for $A+bv^*$. Note that $\frac{1}{|1-v^*y_j|}$ can even ameliorate the bounds as the numerator $|1+v^*A^{-1}b|$ can be larger than 1.

EXAMPLE 2. $A\in\mathbb{C}^{25\times 25}$ is as in Example 1 but, to get a more "realistic example", b is a random vector. Without *any* optimality considerations, we take v also randomly. In Figure 2.1 we have plotted $\kappa(X_v)$ for 20 different choices (in a row). The smallest $\kappa(X_v)$ was approximatively 19. This might give some order of magnitude for "good" choice of v.

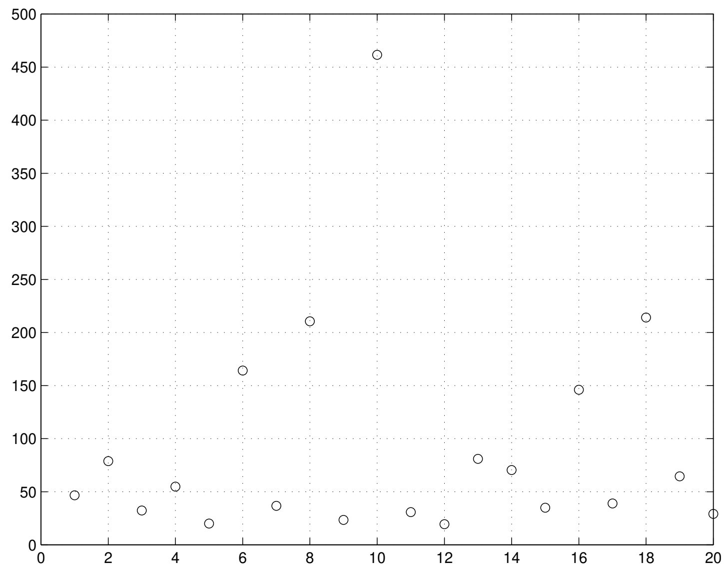

Figure 1: For Example 2 $\kappa(X_v)$ for 20 different randomly taken b's and v's.

EXAMPLE 3. $A \in \mathbb{C}^{100\times 100}$ consists of 4 Jordan blocks of size 25 each such that the eigenvalues of A are 2, 5, -2 and 2-4i. Again we take b a random vector and, as in Example 3, without *any* optimality considerations, we take v also randomly. In Figure 2.2 we have plotted $\kappa(X_v)$ for 20 different choices (in a row). The smallest $\kappa(X_v)$ was approximatively 45.

EXAMPLE 4. Suppose $\mathcal{K}_n(A;b) = \mathbb{C}^n$. As to the minimization problem on the spectrum, the most optimal choice for v is to make the spectrum of $A+bv^*$ equal a point (different from zero). This choice, however, would make $A+bv^*$ nondiagonalizable since $\mathcal{K}_n(A;b) = \mathbb{C}^n$. This nondiagonalizability must also happen because of the bound (10) as it is a lower bound for this minimization problem. That is, the condition number *must* be infinite if v makes the spectrum equal a point. Thus, using (10) "inversely" shows that choosing v such that $A+bv^*$ has an "unnaturally" concentrated spectrum forces $\kappa(X_v)$ to become large. An opposite of this problem arose in [5].

Using multiplicative perturbations of rank-one A we obtain another family of error bounds for GMRES. That is, aside from (7) consider

$$A(I+bv^*)z = b \tag{12}$$

for a vector $v \in \mathbb{C}^n$. If $A(I+bv^*)$ is diagonalizable, we denote by $A(I+bv^*) = X_v\Lambda_v X_v^{-1}$ its diagonalization.

Theorem 6 *Assume $A(I+bv^*)$ is invertible and diagonalizable with $A(I+bv^*) = X_v\Lambda_v X_v^{-1}$. Then*

$$\min_{p\in\mathcal{P}_j(0)} ||p(A)b|| \leq \kappa(X_v) \min_{p\in\mathcal{P}_j(0)} ||p(\Lambda_v)||||b||. \tag{13}$$

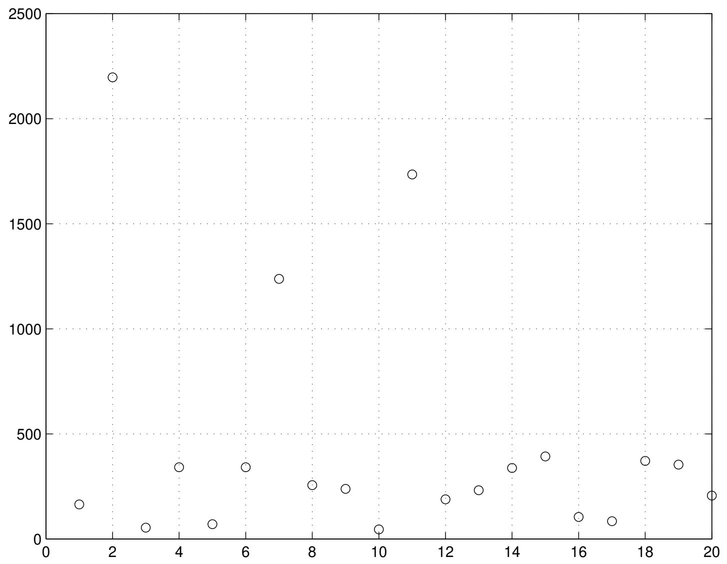

Figure 2: For Example 3 $\kappa(X_v)$ for 20 different randomly taken b's and v's.

Proof. Again we need the invariance of Krylov subspaces in this particular type of perturbations. That is, since $A(I+bv^*)$ is invertible, there holds

$$\mathcal{K}_j(A;b) = \mathcal{K}_j(A(I+bv^*);b) \tag{14}$$

for every j. Namely, for $j=2$ we have $\mathcal{K}_2(A(I+bv^*);b) = \mathcal{K}_2(A;b)$ as $0 \neq A(I+bv^*)b = (1+v^*b)Ab$ and for $j \geq 2$ we have $(A+bv^*)^j b = A^j b + q_{j-1}(A)b$ for a polynomial q_{j-1} of degree at most $j-1$.

If z_j is generated using GMRES for solving (12) with the initial guess $x_0 = 0$, then $A(I+bv^*)z_j = b + \epsilon_j$. Now, this ϵ_j is also the residual for (7) as $x_j = (I+bv^*)z_j$ is the approximation for (7) while using GMRES. The reason for this is that $(I+bv^*)\mathcal{K}_j(A;b) = \mathcal{K}_j(A;b)$ for every j, i.e., the subspaces $\mathcal{K}_j(A;b)$ are invariant for $I+bv^*$ and getting smaller residual for $Ax=b$ than for $A(I+bv^*)z=b$ is not possible since a multiplication of x_j by $(I+bv^*)^{-1}$ yields the same residual for $A(I+bv^*)z=b$ (and vice versa). Consequently,

$$\min_{p\in\mathcal{P}_j(0)} \|p(A)b\| = \min_{p\in\mathcal{P}_j(0)} \|p(A(I+bv^*))b\|$$

and the claim follows after applying (8) with (12). $\square$

This approach leads to a different family of bounds compared with using the perturbation $A+bv^*$. To see this, we consider our standard example.

Example 5. Consider our model problem $Ax=b$ with A from (11) and with b being the first standard basis vector of $\mathbb{C}^n$. Now, by choosing v appropriately, $A(I+bv^*)$ can be made diagonalizable although not normal. Namely, the eigenvalue 2 of A cannot be relocated even though it can be made simple.

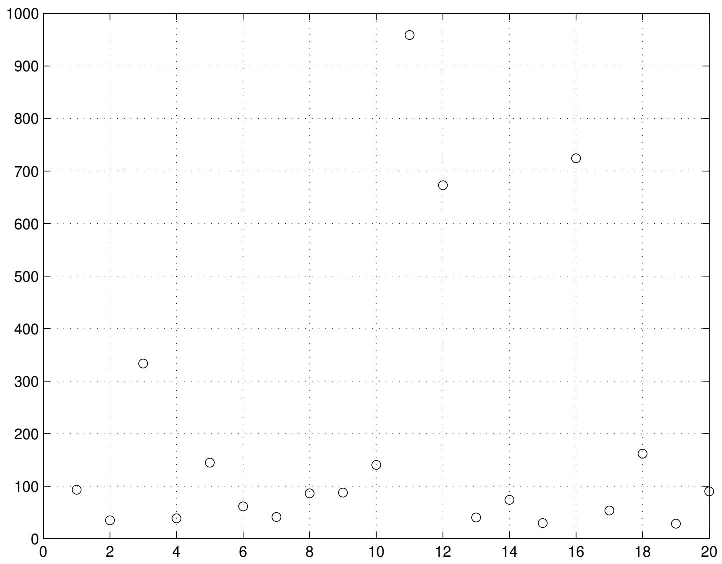

Figure 3: For Example 6 $\kappa(X_v)$ for 20 different randomly taken b's and v's.

The following example demonstrates how using (12) can drastically improve the conditioning in the standard bound.

EXAMPLE 6. [2] Let $A \in \mathbb{C}^{20\times 20}$ be of the form $Z\Lambda Z^{-1}$, where

$$Z = \begin{bmatrix} 1 & \sqrt{1-\delta} & 0 & \cdots & 0 \\ 0 & \sqrt{\delta} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \ \delta \ll 1, \ \Lambda = \mathrm{diag}(20, 10, 5, ..., 1).$$

A has two large well-separated eigenvalues, 20 and 10, corresponding to the first block in Z. The remaining eigenvalues of A are uniformly distributed in the interval $[1, 5]$. As in [2], with $\delta = 10^{-8}$ we have $\kappa(Z) = 2 * 10^4$ because of this one ill-conditioned block and therefore the standard bound (8) is not much of use. Also the pseudospectra [13] are not useful for this problem as explained in [2]. In Figure 2.3 we have plotted the condition numbers for X_v in $A(I + bv^*) = X_v\Lambda_v X_v^{-1}$ for 20 randomly taken b's and v's again without any optimality considerations. As we see, typically there is a huge drop in condition number.

At this point it should be clear how to generalize the bounds (10) and (13) to larger rank perturbations. Namely, one can consider the linear system

$$(A + \sum_{k=0}^{p} A^k b v_k^*)w = b \tag{15}$$

with $v_k \in \mathbb{C}^n$, for $0 \le k \le p$, aside from the original system $Ax = b$. The bounds will then be essentially the same as in the rank-one case except that non-preservation of Krylov subspaces causes a "false start". That is, because $\mathcal{K}_j(A+\sum_{k=0}^p A^k b v_k^*; b) \subset \mathcal{K}_{j+p-1}(A; b)$, for $p \geq 1$, one can derive bounds with a false start in the number of steps taken using GMRES for the system $Ax = b$. The reason for this is that then one obtains from (15) an approximation

$$(A + \sum_{k=1}^{p} A^k b v_k^*) \frac{w_j}{1 - v_0^* w_j} =$$

$$A(I + \sum_{k=1}^{p} A^{k-1} b v_k^*) \frac{w_j}{1 - v_0^* w_j} = b + \frac{\epsilon_j}{1 - v_0^* w_j}, \tag{16}$$

where $w_j \in \mathcal{K}_j(A + \sum_{k=1}^p A^k b v_k^*; b)$ is obtained for (15) using GMRES at the step j. Thus, the approximation $(I + \sum_{k=1}^p A^{k-1} b v_k^*) \frac{w_j}{1 - v_0^* w_j} \in \mathcal{K}_{j+p-1}(A; b)$ is not better than the GMRES approximant after $j+p-1$ steps for $Ax = b$. This establishes again the desired inequality. Note that whenever $v_0 = 0$ (like in Theorem 6) there is no additional scaling caused by the numerator $1 - v_0^* w_j$. For simplicity we assume this in the following, where $A + \sum_{k=0}^p A^k b v_k^* = X_V \Lambda_V X_V^{-1}$ is a diagonalization of $A + \sum_{k=0}^p A^k b v_k^*$.

Theorem 7 *If $A + \sum_{k=1}^p A^k b v_k^*$ is invertible and diagonalizable, then*

$$\min_{p \in \mathcal{P}_{j+p-1}(0)} \|p(A)b\| \leq \kappa(X_V) \min_{p \in \mathcal{P}_j(0)} \|p(\Lambda_V)\| \|b\|. \tag{17}$$

If A is cyclic and b is a cyclic vector for A, then one extreme is to content with A and the other is to take $A + \sum_{k=0}^{n-1} A^k b v_k^*$ which can be *any* matrix with an appropriate choice of v_k's. Thus, it is possible to change A radically (i.e. to make $\kappa(X_V)$ very reasonable!) with perturbations of the form $\sum_{k=0}^p A^k b v_k^*$. Of course in practice p should be small.

Finally we want to state a result we did not notice while prepairing [5] where we used arbitrary small rank perturbations of A for lower bounds for ideal GMRES for A. The key here is that using pertubations of type (15) means that the dimension of a certain auxiliary Block-Krylov subspace achieves a very mild growth [5]. As a consequence, essentially the spectrum of Λ_V (scaled again by the condition number of X_V) bounds from below the behavior of ideal GMRES for A. Here $\lambda_j(M)$ denotes the eigenvalues of a matrix M arranged in decreasing order in modulus.

Theorem 8 *If $A + \sum_{k=0}^p A^k b v_k^*$ is invertible and diagonalizable, then*

$$\min_{p \in \mathcal{P}_j(0)} \|p(A)\| \geq \min_{p \in \mathcal{P}_j(0)} \frac{1}{\kappa(X_V)} |\lambda_{j+p+2}(p(\Lambda_V))|. \tag{18}$$

Proof. The advantage of using perturbations of type (15) of A is that the difference of A and its perturbation is $\sum_{k=0}^p A^k b v_k^*$. The range of this matrix obviously belongs to span$\{b, Ab, ..., A^p b\}$. In particular, then the dimension of the Block-Krylov subspace generated from span$\{b, Ab, ..., A^p b\}$ with A grows as fast as the number of steps taken and the claim follows from [5]. □

As to a block equation $AX = B$ with $B \in \mathbb{C}^{n\times k}$ with some $k \ll n$, the used techniques can be generalized in an obvious way. Then, for instance, for (7) the analogous perturbation of A will be of the form $A + BV^*$ with $V \in \mathbb{C}^{n\times k}$ and the multiplicative perturbation (12) of the form $A(I + BV^*)$.

3 Pole assignment problems for GMRES

As the bound (10) shows (or (13) which is also a pole assignment problem but of the form $A + Abv^*$), v should be chosen in such a way that the spectrum of $A + bv^*$ becomes small, and, at the same time, $A + bv^*$ ought to have a well-conditioned eigendecomposition. But, as described in Example 4, these are mutually competing goals. Still, the problem of choosing $v \in \mathbb{C}^n$ to make $A+bv^*$ to have a well-conditioned eigendecomposition is actually close to the robust pole assignment problem in control theory. See e.g. [6] by Kautsky, Nichols and Van Dooren and [7] by Mehrmann and Xu and references therein for this extensively studied problem. In [7] the authors consider the problem of finding poles that belong to a subset of $\mathbb{C}$ such that $A + bv^*$ has a well-conditioned eigendecomposition. This formulation is closer to our task in the sense that in our problem, instead having to match some strictly preassigned set of eigenvalues in the left half-plane, the restriction is by far more modest. That is, assuming A can be made diagonalizable with a rank-one perturbation of the form bv^*, the spectrum has to avoid just one point as $A+bv^*$ needs to remain invertible. This means that we have more degrees of freedom to choose v in such a way that the resulting $A + b^*v$ has a well-conditioned eigenbasis. In particular, we do not need to relocate well-conditioned eigenvalues at all which is obviously a huge advantage. Thus, although the dimension n is very large, the point is that the bound (8) can be improved significantly with a rank-one correction bv^* of A if the amount of ill-conditioned eigenvalues is not large. It is obviously a serious disadvantage of (8) that just one ill-conditioned eigenvalue can ruin this estimate. If the number of ill-conditioned eigenvalues is very large, then one can derive approximations starting from (15). This, in turn, leads in a natural way to a multi-input pole assignment problem. For the multi-input pole assignment problem, see [8].

Unfortunately the robust pole assignment problem is difficult. The existing algorithms for the single-input case (which $A + bv^*$ is) are not quite applicable to our problem because the first starting point is that A is diagonalizable. Also, the bounds for a very ill-conditioned but diagonalizable case are practically vacuous [7]. Basically the problem is simple to formulate. Namely, if $A - bv^* = X_v \Lambda_v X_v^{-1}$, then multiplying by X_v from the right we have, after rearranging, $AX_v - X_v\Lambda_v = bv^*X_v$. Thus, if $x_{v,i}$ denotes the i^{th} column of X_v, then writing this componentvise we obtain

$$Ax_{v,i} - \lambda_{v,i}x_{v,i} = c_{v,i}b \tag{19}$$

for some $c_{v,i} \in \mathbb{C}$. On the other hand, if there exists linearly independent vectors $x_{v,i}$ for which (19) is true, then obviously a vector v can be found such that $AX_v - X_v\Lambda_v = bv^*X_v$ holds. For more general statements, see [8]. Thus, (19) could be a starting point for a nonlinear optimization problem minimizing the "energy" $\kappa(X_v)$. This problem, however, seems very difficult.

Next we describe a coarse way of constructing a perturbation of A. What we propose can be applied to the multi-input case (15) with obvious changes. Namely there holds, as shown by R. Smith [12], whenever $M \in \mathbb{C}^{n\times n}$ is diagonalizable

$$\kappa(V) \geq (1 + \frac{||[M, M^*]||_{\mathcal{F}}}{2||M^2||_{\mathcal{F}}})^{1/4}. \tag{20}$$

Here $[M, M^*] := MM^* - M^*M$ is the self-commutator of M and $M = V\Lambda V^{-1}$ is a diagonalization of M. Thus, (20) yields a necessary condition for smallness of $\kappa(V)$ (but of course always not less than 1), that is, $||[M, M^*]||_{\mathcal{F}}$ should be small. However, this is not a sufficient condition and it is easy to construct examples for which the right-hand side of (20) is arbitrarily close to 1 while the left-hand side is arbitrarily large. A sufficient condition is the following, where $\kappa_{\mathcal{F}}(V)$ denotes the condition number of V in the Frobenius norm.

Proposition 9 *[5] Suppose M is diagonalizable and let $\{\lambda_j\}_{j=1}^n$ denote the eigenvalues of M and $\delta_j := \min\{|\lambda_i - \lambda_j| : i \neq j\}$. Then for an optimal diagonalization $M = V\Lambda V^{-1}$ holds*

$$\kappa_{\mathcal{F}}(V) \leq \sum_{j=1}^{n} \big(1 + (\frac{n(n+1)}{12(n-1)})^{1/2} \frac{||[M, M^*]||_{\mathcal{F}}}{\delta_j^2}\big)^{(n-1)/2}.$$

Hence, a way to pick $v \in \mathbb{C}^n$ is to first use the criterion

$$\min_{v\in\mathbb{C}^n} ||[A + bv^*, (A + bv^*)^*]||_{\mathcal{F}} \tag{21}$$

by applying, for instance, the method of steepest descend. After this has been accomplished, one has $A + bv^*$. However, at this point there is no guarantee that $A + bv^*$ is not nearly singular. Thus, another step may be needed to move away those eigenvalues of $A + bv^*$ that are close to zero. But, since now $A + bv^*$ ought to have a well-conditioned eigendecomposition, this can be done by using the algorithms in [7]. Assuming w is obtained in such a way, then, combining this with v from (21), we finally have $A + b(v + w)^*$. Obviously this two-step approach could also be applied to the robust pole assignment problem.

4 Conclusions

In this paper we have considered error bounds for GMRES for solving $Ax = b$. We have shown that, by considering very particular type of small rank perturbations of A, it is possible to generalize the most classical bound also to nondiagonalizable matrices. The estimates involve properties of A with respect to b in the sense that the used perturbations are of the form $\sum_{k=0}^{p} A^k b v_k^*$ for vectors $v_0, ..., v_p \in \mathbb{C}^n$. To our mind, the main contribution of the paper is that also for nonnormal A eigenvalues are the most important indicator of convergence of GMRES. The main point is that they need not be the eigenvalues of A that count but the eigenvalues of an appropriately perturbed A. As a result, there arises pole assignment problems that can, however, be very difficult to solve optimally. We have shortly discussed their relation to existing pole assignment algorithms and briefly outlined a possible way of constructing perturbations of A.

References

[1] A. GREENBAUM, Iterative Methods for Solving Linear Systems, *Frontiers in Applied Mathematics, SIAM, Philadelphia,* 1997.

[2] A. GREENBAUM AND Z. STRAKOS, *Matrices that generate same Krylov residual spaces,* Recent Advances in Iterative Methods, G. Golub, A. Greenbaum and M. Luskin eds. IMA volumes in mathematics and its applications, Vol 60, Springer-Verlag, New York, 1994.

[3] M. HOCHBRUCK AND C. LUBICH, *Error analysis of Krylov methods in a nutshell,* SIAM J. Sci. Comput. 19 (1998), pp. 695-701.

[4] M. HUHTANEN AND O. NEVANLINNA, *Minimal decompositions and iterative methods,* Centro Int. Math., Portugal, Preprint 3, 1998. Also to appear in Numer. Math.

[5] M. HUHTANEN, *Ideal GMRES can be bounded from below by three factors,* Helsinki Univ. Tech., Inst. Math. Res. Rep. A412, 1999. Also to appear in Lin. Alg. Appl.

[6] J. KAUTSKY, N. K. NICHOLS AND P. VAN DOOREN, *Robust pole assignment in linear state feedback,* Int. J. Contr. vol. 41, no. 5, pp. 1129-1155, 1985.

[7] V. MEHRMANN AND H. XU, *Choosing poles so that the single-input pole placement problem is well conditioned,* SIAM J. Matr. Anal. Vol. 19, No. 3, pp. 644-681, 1998.

[8] V. MEHRMANN AND H. XU, *An analysis of the pole palcement problem II. The multi-input case*, Electr. Trans. Num. Anal., Vol. 5, pp. 77-97, 1997.

[9] N. M. NACHTIGAL, S. REDDY AND L. N. TREFETHEN *"How fast are nonsymmetric matrix iterations?"*, SIAM J. on Matrix Anal. Appl., 13 (1992), pp. 796-825

[10] O. NEVANLINNA, Convergence of Iterations for Linear Equations, *Lectures in Mathematics ETH Zürich, Birkäuser Verlag*, 1993.

[11] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimum residual algorithm for solving nonsymmetric linear systems,* SIAM J. Sci. Stat. Comp., 7 (1986), pp. 856-869.

[12] R. A. SMITH, *The condition numbers of the matrix eigenvalue problem,* Numer. Math., 10, pp. 232-240, 1967.

[13] L. N. TREFETHEN, *Pseudospectra of matrices,* in Numerical Analysis 1991, D. F. Griffiths and G. A. Watson, eds., Longman, Harlow, London, 1992.

[14] W. M. WONHAM, Linear Multivariable Control: A Geometric Approach, *Springer-Verlag, New-York*, 1979.

(continued from the back cover)

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/* .

A417 Kirsi Peltonen
Examples of uniformly quasiregular mappings, Jul 1999

A416 Saara Hyvönen
On the Iterative Solution of Nonnormal Problems, Jul 1999

A415 Ville Turunen
A smooth operator valued symbol analysis, Aug 1999

A414 Marko Huhtanen
A stratification of the set of normal matrices, May 1999

A413 Clément Ph., Gripenberg G. and Londen S-O
Regularity properties of solutions of functional evalution equations, May 1999

ISBN 951-22-4647-3
ISSN 0784-3143
Painotalo Libella Oy, Espoo, 1999