

ON THE OPTIMAL SOLUTION  
OF  
LARGE LINEAR SYSTEMS

J.F. Traub  
Department of Computer Science  
Columbia University  
New York, N.Y. 10027

H. Woźniakowski  
Institute of Informatics  
University of Warsaw  
Warsaw, Poland

and

Department of Computer Science  
Columbia University  
New York, N.Y. 10027

January 1980

Revised August 1982

This research was supported in part by the National Science Foundation under Grant MCS-7823676. One of us (JFT) was supported in part by the Advanced Research Projects Agency under contract N00039-82-C-0427.

## ABSTRACT

The information-based study of the optimal solution of large linear systems is initiated by studying the case of Krylov information. Among the algorithms which use Krylov information are minimal residual, conjugate gradient, Chebyshev, and successive approximation algorithms. A "sharp" lower bound on the number of matrix-vector multiplications required to compute an  $\epsilon$ -approximation is obtained for any orthogonally invariant class of matrices. Examples of such classes include many of practical interest such as symmetric matrices, symmetric positive definite matrices, and matrices with bounded condition number. It is shown that the minimal residual algorithm is within at most one matrix-vector multiplication of the lower bound. A similar result is obtained for the generalized minimal residual algorithm.

The lower bound is computed for certain classes of orthogonally invariant matrices. We show how the lack of certain properties (symmetry, positive definiteness) increases the lower bound. A conjecture and a number of open problems are stated.

## 1. INTRODUCTION

We study the approximate solution of large linear systems  $Ax = b$ , by algorithms using Krylov information  $b, Ab, \dots, A^k b$ . Examples of such algorithms are minimal residual, conjugate gradient, Chebyshev, and successive approximation algorithms.

We seek the optimal algorithm, that is, the algorithm with minimal complexity. In this paper we choose to minimize the number of matrix-vector multiplications to obtain an  $\epsilon$ -approximation. It is easy to translate our results on the minimum number of such multiplications into complexity results; see [6, Section 8].

We almost completely solve this problem for any matrix class having a certain property. The gap between the lower and upper bounds on the minimal number of matrix-vector multiplications is at most unity.

More precisely, we consider any class of matrices which is orthogonally invariant. Examples of such classes are symmetric matrices, symmetric positive definite matrices, and matrices with bounded condition number.

For any orthogonally invariant class of matrices we show that the minimal residual algorithm uses at most one more matrix-vector multiplication than the lower bound. Indeed, we show even more. For some classes  $F$  we know the minimum number of vector-matrix multiplications; for others there remains a gap of unity.

We contrast our approach with that which is typical in the approximate solution of large linear systems. One constructs an algorithm  $\phi$

which generates a sequence  $\{x_k\}$  approximating the solution  $\alpha = A^{-1}b$ ; the calculation of  $x_k$  requires  $k$  matrix-vector multiplications and  $x_k$  lies in the Krylov subspace spanned by  $b, Ab, \dots, A^k b$ . The algorithm  $\phi$  is often chosen to guarantee good approximation properties of the sequence  $\{x_k\}$ . In some cases,  $\phi$  is defined to minimize some measure of the error in a restrictive class of algorithms. For instance, let this class be defined as the class of "polynomial" algorithms, i.e.,  $\alpha - x_k = W_k(A)\alpha$ , where  $W_k$  is a polynomial of degree at most  $k$  and  $W_k(0) = 1$ . Then choosing  $W_k$  as the polynomial minimizing the  $k$ -th residual  $\|Ax_k - b\| = \|W_k(A)\alpha\|$ , we obtain the minimal residual algorithm,  $\phi^{mr}$ . If  $A$  is symmetric, positive definite and  $a = 1/\|A^{-1}\|$ ,  $b = \|A\|$  are known, then choosing  $W_k$  as the polynomial minimizing  $\max\{|W_k(t)| : t \in [a, b]\}$ , we obtain the Chebyshev algorithm,  $\phi^{ch}$ .

It seems to us that this procedure is unnecessarily restrictive. It is not clear, a priori, why an algorithm has to construct  $x_k$  of the form  $\alpha - x_k = W_k(A)\alpha$ .

Indeed, we show that for orthogonally invariant classes of matrices  $\phi^{mr}$  is within at most one matrix-vector multiplication of the lower bound without any restriction on the class of algorithms. However, if the class is not orthogonally invariant, the optimality property of  $\phi^{mr}$  may disappear.

We summarize the results of this paper. In Section 2 we define two types of optimality. The main result is established in Section 3. We give a lower bound on the number of matrix-vector multiplications for any orthogonally invariant class by showing that  $\phi^{mr}$  performs at most one multiplication more than necessary. A series of examples shows the

sharpness and applicability of the main result. In particular, it follows from the main theorem that the knowledge of all eigenvalues of  $A$  does not help. Furthermore, we show how the lack of symmetry and/or positive definiteness increases the lower bound.

In Section 4 we introduce a family of approximation criteria and generalize the previous optimality results (see Theorem 4.2). In particular, we show that the conjugate gradient algorithm performs at most one multiplication more than necessary and that the minimum error algorithm performs the minimal number of multiplications.

In the final section we pose some open problems concerning the optimality properties of the information studied in this paper.

The problems and proof techniques of this paper follow the information-based approach of the monographs [5] and [7]. There are many interesting relations between the optimality results of this paper and the general results of the monograph. For the reader's convenience we do not use the general terminology and results of [5] and [7].

For simplicity we consider only the real case, although the generalization to the complex case is straightforward.

This paper is a shortened version of [6]. The paper [6] contains a detailed discussion of the concepts presented here, all omitted proofs, as well as a complexity analysis.

## 2. BASIC CONCEPTS

Let  $F$  be a subclass of the class of  $n \times n$  nonsingular real matrices. Let  $b$  be a given  $n \times 1$  real vector such that  $\|b\| = \sqrt{(b,b)} = 1$ . For a given positive  $\epsilon$ ,  $\epsilon \leq 1$ , we seek a real vector  $x$  whose residual has norm less than  $\epsilon$ , i.e.,

$$(2.1) \quad \|Ax - b\| < \epsilon, \quad A \in F.$$

We call  $x$  an  $\epsilon$ -approximation. Since  $b$  is normalized to unity, (2.1) measures the relative error of the residual vector. In Section 4 we discuss the problem of finding  $x$  with relative error less than  $\epsilon$  in a variety of norms.

To find an  $\epsilon$ -approximation we need some information about the matrix  $A$  which belongs to the class  $F$ . We define an information operator  $N_k$  as

$$(2.2) \quad N_k(A,b) = [b, Ab, A^2b, \dots, A^k b]$$

for  $k = 0, 1, \dots$ .

(Note that  $N_k(A,b)$  is a basis of the Krylov subspace.)

Remark 2.1

Let  $z_0 = b$ ,  $z_i = Az_{i-1}$ , for  $i = 1, 2, \dots, k-1$ . Then (2.2) can be rewritten as

$$(2.3) \quad N_k(A,b) = [z_0, Az_0, Az_1, \dots, Az_{k-1}] .$$

Thus the computation of  $N_k(A,b)$  requires  $k$  matrix-vector multiplications. If  $A$  is sparse  $N_k(A,b)$  can be computed in time proportional to  $kn$  rather than  $kn^2$ . Usually, instead of computing  $N_k(A,b)$ , we compute

$$N'_k(A,b) = [b, Aw_1, Aw_2, \dots, Aw_k]$$

where  $w_i$  is a linear combination of  $b, Ab, \dots, A^{i-1}b$  for  $i = 1, 2, \dots, k$ . It is easy to show that all the results of this paper also hold for the information operator  $N'_k$ . ■

Knowing  $N_k(A,b)$  we seek an  $\epsilon$ -approximation by an algorithm  $\phi$ . We define an algorithm  $\phi = \{\phi_k\}$  as a sequence of mappings  $\phi_k : N_k(F, \mathbb{R}^n) \rightarrow \mathbb{R}^n$ . The algorithm  $\phi$  generates the sequence  $x_k = \phi_k(N_k(A,b))$  based on the information  $N_k(A,b)$ ,  $k = 0, 1, \dots$ . We are interested in the smallest value of  $k$  for which  $x_k$  satisfies (2.1), i.e.,  $\|Ax_k - b\| < \epsilon$ . In general, there exists many different matrices  $\tilde{A}$  from  $F$  which share the same information as  $A$ , i.e.,  $N_k(\tilde{A}, b) = N_k(A, b)$ . Thus  $x_k = \phi_k(N_k(A, b)) = \phi_k(N_k(\tilde{A}, b))$  must satisfy (2.1) for  $A$  and  $\tilde{A}$ . Define

$$(2.4) \quad V(y_k) = \{\tilde{A} : \tilde{A} \in F, N_k(\tilde{A}, b) = y_k\}, \quad y_k = N_k(A, b) .$$

Thus  $V(y_k)$  denotes those matrices belonging to  $F$  which are indistinguishable from  $A$  knowing the information  $N_k(A, b)$ .

Let

$$(2.5) \quad k(\phi, A) = \min \{k.: \|\tilde{A}x_k - b\| < \varepsilon, \forall \tilde{A} \in V(y_k)\}$$

be the matrix index of  $\phi$ . (If the set of  $k$  in (2.5) is empty, we set  $k(\phi, A) = +\infty$ .) Let

$$(2.6) \quad k(\phi, F) = \max_{A \in F} k(\phi, A)$$

be the class index of  $\phi$ .

Thus, the matrix index of  $\phi$  denotes the minimal number of steps required to find an  $\varepsilon$ -approximation using the algorithm  $\phi$  for all matrices  $\tilde{A}$  from  $F$  which share the same information  $N_k$  as  $A$ . The class index of  $\phi$  denotes the same concept for the hardest problem.

We seek algorithms with minimal indices. Let

$$(2.7) \quad k(A) = \min_{\phi} k(\phi, A)$$

be the optimal matrix index and let

$$(2.8) \quad k(F) = \max_{A \in F} k(A) (= \min_{\phi} k(\phi, F))$$

be the optimal class index.

### Remark 2.3

It is possible that  $k(A) \ll k(F)$ . For instance, assume that



$Ab = b$  . Then, of course, setting  $x_1 = \phi_1(y_1) = b$  we have  $\tilde{A}x_1 = b$  for  $\tilde{A} \in V(y_1)$  . Thus  $k(A) = 1$  for every  $\varepsilon$  . As we shall see later  $k(F)$  can be equal to  $n$  .

Thus even if the optimal class index is large it can happen, due to favorable properties of  $A$  with respect to  $b$  , that the optimal matrix index is small. The algorithms with small matrix index are therefore very useful for applications. This motivates our interest in algorithms with small matrix index. ■

We are ready to introduce two concepts of optimal algorithms. An algorithm  $\phi$  is called strongly optimal iff

$$(2.9) \quad k(\phi, A) = k(A), \quad \forall A \in F,$$

and is called optimal iff

$$(2.10) \quad k(\phi, F) = k(F).$$

We can sometimes establish that the matrix or class index of an algorithm is slightly larger than the optimal index. It is convenient to introduce the concepts of almost strongly optimal algorithm and almost optimal algorithm as follows. An algorithm  $\phi$  is almost strongly optimal iff

$$(2.11) \quad k(\phi, A) \leq k(A) + c, \quad \forall A \in F,$$

and is almost optimal iff

$$(2.12) \quad k(\phi, F) \leq k(F) + c$$

for some small integer  $c$ .

Thus an almost strongly optimal algorithm requires at most  $c$  more steps than a strongly optimal one. Usually  $k(A) \gg c$  and therefore an almost strongly optimal algorithm is as useful in practice as a strongly optimal one.

Remark 2.4

All concepts introduced in this section also depend on the size  $n$ , the information  $N_k$ , the vector  $b$  and  $\epsilon$ . To simplify notation and terminology we do not make this explicit but the reader should keep in mind that all the results are relative to  $n, N_k, b$  and  $\epsilon$ . ■

## 3. OPTIMALITY OF THE MR ALGORITHM

In this section we study optimality properties of the minimal residual algorithm defined as follows. Let

$$N_k(A,b) = [z_0, z_1, \dots, z_k] \quad \text{with } z_i = A^i b .$$

Knowing the vectors  $z_i$  we define  $c_1^*, \dots, c_k^*$  as the coefficients which minimize the norm of the residual in the space spanned by  $z_1, z_2, \dots, z_k$ . Thus

$$(3.1) \quad \| b - c_1^* z_1 - \dots - c_k^* z_k \| = \min_{c_i} \| b - c_1 z_1 - \dots - c_k z_k \| .$$

The minimal residual algorithm  $\phi^{\text{mr}}$ , briefly the mr algorithm, is defined as

$$(3.2) \quad x_0 = 0 ,$$

$$x_k = \phi_k^{\text{mr}}(N_k(A,b)) = c_1^* b + \dots + c_k^* A^{k-1} b , \quad k \geq 1 ,$$

see, for instance, [4].

We now prove that the mr algorithm is an almost strongly optimal algorithm provided the class  $F$  is "orthogonally invariant". This concept is defined as follows. We say  $F$  is orthogonally invariant iff

$$(3.3) \quad A \in F \implies Q A Q^T \in F$$

for every orthogonal  $Q$ .

For example, the class of symmetric matrices, the class of symmetric positive definite matrices, the class of matrices with condition number bounded by a given constant, and the class of matrices with fixed eigenvalues are all orthogonally invariant.

The main result is

Theorem 3.1

If  $F$  is orthogonally invariant, then

$$(3.4) \quad k(\phi^{mr}, A) \geq k(A) \geq k(\phi^{mr}, A) - 1, \quad \forall A \in F.$$

Furthermore, both the upper and lower bounds can be achieved. ■

Proof

Let  $\phi = \{\phi_k\}$  be any algorithm. Let  $k = k(\phi, A) < +\infty$ .

This means that

$$(3.5) \quad \|A x'_k - b\| < \epsilon, \quad \tilde{A} \in V(y_k)$$

where  $x'_k = \phi(N_k(A, b))$ . Decompose  $x'_k = z_1 + z_2$

where  $z_1$  is a linear combination of  $b, Ab, \dots, A^k b$  and  $z_2$  is

orthogonal to  $b, Ab, \dots, A^k b$ . Define

$$\omega = \begin{cases} \frac{z_2}{\|z_2\|} & \text{if } z_2 \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

Clearly  $(\omega, A^i b) = 0$  for  $i = 0, 1, \dots, k$ . Let

$$\tilde{A} = QAQ \text{ with } Q = I - 2\omega\omega^T.$$

Then  $\tilde{A} \in F$  and

$$A^i b = QA^i Qb = QA^i b = A^i b, \quad i = 1, 2, \dots, k.$$

Thus,  $\tilde{A} \in V(y_k)$  and

$$\|\tilde{A}x'_k - b\| = \|Ax'_k - 2(\omega, x'_k)A\omega - b\|$$

Note that  $(\omega, x'_k)A\omega = Az_2$  which yields

$$\|\tilde{A}x'_k - b\| = \|Az_1 - b - Az_2\|.$$

Observe that

$$\begin{aligned} \|Az_1 - b\| &\leq \frac{1}{2}(\|Az_1 - b - Az_2\| + \|Az_1 - b + Az_2\|) \\ &= \frac{1}{2}(\|\tilde{A}x'_k - b\| + \|Ax'_k - b\|) < \epsilon \end{aligned}$$

due to (3.5)

Recall that  $x_{k+1} = \phi_{k+1}^{mr}(N_{k+1}(A, b))$  lies in the same subspace as  $z_1$  and

$$\|Ax_{k+1} - b\| \leq \|Az_1 - b\| < \epsilon.$$

Due to (3.2) we have  $\tilde{A}x_k - b = Ax_k - b$ ,  $\forall \tilde{A} \in V(y_k)$ . Thus

$$k(\phi^{mr}, A) \leq k + 1 = k(\phi, A) + 1.$$

Since  $\phi$  is an arbitrary algorithm we have

$$k(\phi^{mr}, A) \leq k(A) + 1.$$

On the other hand it is obvious that  $k(A) \leq k(\phi^{mr}, A)$ . This proves (3.4).

The fact that the lower and upper bounds in (3.4) can be achieved is established in Examples 3.2 and 3.4. ■

We illustrate Theorem 3.1 by a number of examples of classes  $F$  to exhibit the importance of this Theorem and the sharpness of both the assumption and the results.

#### Example 3.1 KNOWN EIGENVALUES

Suppose one knows all eigenvalues of the matrix  $A$  and asks what algorithm should be used for the approximate solution of  $Ax = b$ ? The surprising answer is that we still should use the  $mr$  algorithm

(although the  $mr$  algorithm does not make use of the eigenvalues of  $A$  ).  
The explanation is given by Theorem 3.1 applied to

$$F_1 = \{ \tilde{A} : \tilde{A} = Q A Q^T, Q \text{ is orthogonal} \} .$$

Note that all matrices in  $F_1$  have the same eigenvalues as  $A$  . Since  $F_1$  is orthogonally invariant, Theorem 3.1 guarantees that the  $mr$  algorithm performs at most one step more than is necessary. This shows that the knowledge of eigenvalues does not help for the approximate solution of linear systems. ■

Example 3.2      MATRICES OF THE FORM  $A = I - B$

Suppose one knows that  $A = I - B$  and  $\|B\|$  is known not to exceed  $\rho$  ,  $\rho < 1$  . This is a typical situation for the approximate solution of  $x = Bx + b$  by iterative algorithms. One asks what algorithm should be used in this case. The answer depends on what more is known on  $B$  . We report some results from Section 6 of [6]. See also [2] especially pp.19-20.

Consider first the symmetric case, i.e., let

$$F_2 = \{ A : A = I - B, B = B^T, \|B\| \leq \rho < 1 \} .$$

Since  $F_2$  is orthogonally invariant, Theorem 3.1 guarantees that the  $mr$  algorithm takes at most one step more than the minimum. In fact the matrix index of the  $mr$  algorithm is equal to

$$k(\phi^{mr}, F_2) = \min(n, q(\epsilon) + 1)$$

where

$$q(\epsilon) = \left\lfloor \ln \frac{1 + \sqrt{1-\epsilon^2}}{\epsilon} / \ln \frac{1 + \sqrt{1-\rho^2}}{\rho} \right\rfloor.$$

For the Chebyshev algorithm we have

$$k(\phi^{ch}, F_2) = q(\epsilon).$$

Thus

$$k(F_2) = \min(n, q(\epsilon)).$$

If  $q(\epsilon) < n$  then

$$(3.6) \quad k(F_2) = k(\phi^{mr}, F_2) - 1 = k(\phi^{ch}, F_2)$$

This shows that the  $mr$  algorithm takes exactly one step more than the optimal one. Thus the knowledge of  $\rho$ , which is not used by the  $mr$  algorithm, causes the loss of one step. This example shows that the lower bound can be achieved in Theorem 3.1 and hence the result is best possible.

Furthermore, (3.6) shows that the Chebyshev algorithm is optimal whenever  $q(\epsilon) < n$ . This holds if  $\epsilon$  is not too small and  $\rho$  not too close to unity. However, the Chebyshev algorithm is not strongly optimal.



Consider now the non-symmetric case, i.e., let

$$F_3 = \{A : A = I - B, \quad \|B\| \leq \rho < 1\}.$$

Theorem 3.1 still applies since  $F_3$  is orthogonally invariant. We have

$$(3.7) \quad k(\phi^{mr}, F_3) = \min(n, \left\lfloor \frac{\ln \frac{1}{\epsilon}}{\ln \rho} (1 - \delta) \right\rfloor + 1)$$

where  $\delta \in [0, \ln(1 - \rho^2 + \rho^2 \epsilon^2) / \ln \epsilon^2]$ .

For small  $\epsilon$ ,  $\rho$  close to unity and  $n$  so large that the minimum in (3.7) is obtained for the second argument, we have

$$k(F_3) = k(\phi^{mr}, F_3)(1 + o(1)) = \frac{\ln \frac{1}{\epsilon}}{1 - \rho} (1 + o(1)),$$

$$k(F_2) = k(\phi^{mr}, F_2) - 1 = \frac{\ln \frac{1}{\epsilon}}{\sqrt{2(1-\rho)}} (1 + o(1)).$$

From this we get

$$\frac{k(F_3)}{k(F_2)} = \sqrt{\frac{2}{1-\rho}} (1 + o(1)).$$

This shows how the lack of symmetry in  $B$  increases the optimal class index. ■

### Example 3.3      BOUNDED CONDITION NUMBER

Suppose one knows a bound  $M$  on the condition number  $\text{cond}(A) = \|A\| \|A^{-1}\|$  and asks how this bound influences the optimal class index. We report some results from Section 5 of [6] for three orthogonally invariant classes of matrices. See also [2], especially pp.19-20.

$$F_4 = \{A : A = A^T > 0, \text{ cond}(A) \leq M\},$$

$$F_5 = \{A : A = A^T, \text{ cond}(A) \leq M\},$$

$$F_6 = \{A : \text{ cond}(A) \leq M\}.$$

Thus,  $F_4$  is the class of symmetric positive definite matrices with condition number bounded by  $M$ ,  $F_5$  differs from  $F_4$  by the lack of positive definiteness and  $F_6$  differs from  $F_5$  by the lack of symmetry.

We have

$$k(F_4) - \alpha_1 = k(\phi^{mr}, F_4) = \min(n, \left\lfloor \ln \frac{1 + \sqrt{1-\epsilon^2}}{\epsilon} / \ln \frac{\sqrt{M+1}}{\sqrt{M-1}} \right\rfloor + 1),$$

$$k(F_5) - \alpha_2 = k(\phi^{mr}, F_5) = \min(n, 2 \left\lfloor \ln \frac{1 + \sqrt{1-\epsilon^2}}{\epsilon} / \ln \frac{M+1}{M-1} \right\rfloor + 2),$$

$$k(F_6) = k(\phi^{mr}, F_6) = n$$

where  $\alpha_i = 0$  or  $\alpha_i = -1$  for  $i = 1, 2$ . For small  $\epsilon$  and large  $M$  we can simplify to

$$k(F_4) = \min(n, \frac{\sqrt{M}}{2} \ln \frac{2}{\epsilon} (1 + o(1))) + \alpha_1$$

$$k(F_5) = \min(n, M \ln \frac{2}{\epsilon} (1 + o(1))) + \alpha_2$$

If  $n$  is large then

$$\frac{k(F_5)}{k(F_4)} = 2\sqrt{M} (1 + o(1)) .$$

This shows that the lack of positive definiteness increases the optimal class index by a factor of about  $2\sqrt{M}$  . For large  $M$  , which arise frequently in practice, this is a very significant difference.

For the class  $F_6$  we have  $k(F_6) = n$  . Thus if fewer than  $n$  matrix-vector multiplications are permitted it is impossible to find an  $\epsilon$ - approximation no matter what algorithm is used. Note that this result holds for arbitrary  $\epsilon$  and  $M$  , i.e.,  $\epsilon$  and  $M$  can even be equal to unity. It is the lack of symmetry which causes the increase of the optimal class index to its maximum possible value. ■

#### Example 3.4      UNBOUNDED CONDITION NUMBER

Suppose one does not know a bound on the condition number and agrees to enlarge the class  $F_4$  to

$$F_7 = \{A : A = A^T > 0\} .$$

Then, as was observed in a short note [1], the  $mr$  algorithm is strongly optimal, i.e.,

$$(3.8) \quad k(\phi^{mr}, A) = k(A) \quad \forall A \in F ,$$

and this shows that the upper bound can be achieved in Theorem 3.1.

To show this we use a slightly different proof than the proof of

Theorem 3.1. Let  $x'_k = \phi_k(N_k(A, b))$  for some algorithm  $\phi$ . Suppose that  $x'_k$  does not lie in  $\text{span}(b, \dots, A^{k-1}b)$ . Then  $\sup \{ \|\tilde{A}x'_k - b\| : \tilde{A} \in V(y_k) \} = +\infty$ . Indeed, take  $\tilde{A} = A + cuu^T$  where  $u$  is a projection of  $x'_k$  onto  $\text{span}(b, \dots, A^{k-1}b)^\perp$  and  $c$  is a positive constant. Then  $\|\tilde{A}x'_k - b\| \geq c|u^T x'_k| \|u\| - \|Ax'_k - b\|$  goes to infinity with  $c$  as claimed. Thus  $\|\tilde{A}x'_k - b\| < \epsilon$ ,  $\forall A \in V(y_k)$  implies that  $x'_k$  belongs to  $\text{span}(b, \dots, A^{k-1}b)$ . Due to the definition of the  $mr$  algorithm we have  $k(\phi^{mr}, A) \leq k(\phi, A)$  which proves (3.8).

Since  $k(F_4)$  increases with  $M$  to the value  $n$  it comes as no surprise that

$$k(F_7) = n.$$

This shows that the class  $F_7$  is too large and one has to decrease the class  $F_7$  to find an  $\epsilon$ -approximation in fewer than  $n$  matrix-vector multiplications for all matrices  $A$  from a given class. ■

#### Example 3.5 NOT ORTHOGONALLY INVARIANT

We end this section by an example of a class  $F$  which is not orthogonally invariant. Then none of the optimality properties of the  $mr$  algorithm hold. More precisely we present an example of  $F$  for which the  $mr$  algorithm can be arbitrarily far from optimal.

Let  $TRI$  be the class of  $n \times n$  symmetric tridiagonal matrices whose diagonal elements are equal to unity. Thus  $A \in TRI$  implies

$$A = \begin{pmatrix} 1 & a_1 & & & \\ a_1 & 1 & a_2 & & \\ & & \dots & & \\ & & & a_{n-1} & \\ & & & a_{n-1} & 1 \end{pmatrix}, \quad a_i \in \mathbb{R}.$$

Let

$$F_8 = \{A : A \in \text{TRI} ; \quad \text{cond}(A) \leq M\}$$

for a given  $M, M > 1$ . The class  $F_8$  is not orthogonally invariant since the matrix  $QAQ^T$  with orthogonal  $Q$  is not necessarily tri-diagonal.

Assume that

$$b = [1/\sqrt{n}, 1/\sqrt{n}, \dots, 1/\sqrt{n}]^T$$

Then knowing  $z = Ab = [z_1, \dots, z_n]^T$  we get

$$1 + a_1 = z_1 \sqrt{n},$$

$$a_{i-1} + 1 + a_i = z_i \sqrt{n}, \quad i = 2, \dots, n-1.$$

From this we find the coefficients  $a_i$

$$a_1 = z_1 \sqrt{n} - 1,$$

$$a_i = z_i \sqrt{n} - 1 - a_{i-1}, \quad i = 2, \dots, n-1.$$

Since we know the matrix  $A$ , the algorithm

$$x_1' = \phi_1(b, Ab) = A^{-1}b$$

is well defined and  $\|Ax_1' - b\| = 0$ . Thus

$$k(F_8) = 1, \forall \varepsilon \in (0, 1].$$

It can be verified that for sufficiently small  $\varepsilon$ , the algorithm  $\phi^{mr}$  has to use the information  $N_n(A, b)$  which means that

$$k(\phi^{mr}, F_8) = n.$$

Hence we get the smallest possible value of  $k(F_8)$  and the largest possible value of  $k(\phi^{mr}, F_8)$ . ■

## 4. GENERALIZED CRITERIA

In this section we introduce a family of approximation criteria depending on a parameter  $p$ . The criterion used in Sections 2 and 3 corresponds to  $p = 1$ . The values of  $p$  of greatest practical importance are  $p = 0, 1/2, 1$ .

In (2.1) we defined an  $\epsilon$ -approximation as a vector whose residual has norm less than  $\epsilon$ . Here we assume that the  $\epsilon$ -approximation  $x$  satisfies the inequality

$$(4.1) \quad \frac{\|A^p(x - \alpha)\|}{\|A^p \alpha\|} < \epsilon$$

where  $\alpha = A^{-1}b$  and  $p$  is a nonnegative real. Note that for  $p = 1$ , (4.1) coincides with (2.1). For  $p = 0$ , (4.1) means that the relative error of  $x$  is less than  $\epsilon$ .

If  $p$  is not an integer we assume that  $A$  is symmetric and positive definite to guarantee the existence of  $A^p$ .

We generalize the concept of the matrix index of  $\phi$  to

$$(4.2) \quad k(\phi, A) = \min \{k : \|\tilde{A}^p(x_k - \tilde{A}^{-1}b)\| / \|\tilde{A}^{p-1}b\| < \epsilon, \forall \tilde{A} \in V(y_k)\}$$

where  $\phi = \{\phi_k\}$ ,  $x_k = \phi_k(N_k(A, b))$  and  $V(y_k)$  is given by (2.4). (If the set of  $k$  is empty, we set  $k(\phi, A) = +\infty$ .) Then all concepts introduced in Section 2 may be generalized in an obvious way using the new definition of the matrix index of  $\phi$ .

For given  $A$  and  $m$  define the coefficients  $c_0^*, c_1^*, \dots, c_m^*$  and the error  $e(A, m)$  as

$$(4.3) \quad e(A, m) = \|A^P(\alpha - c_0^*b - \dots - c_m^*A^m b)\| = \min_{c_j} \|A^P(\alpha - c_0 b - \dots - c_m A^m b)\|.$$

Let

$$m(A) = \min \{m : e(\tilde{A}, m) / \|\tilde{A}^P \alpha\| < \epsilon, \forall \tilde{A} \in V(y_m)\}$$

where  $\tilde{\alpha} = \tilde{A}^{-1}b$ . We prove

#### Theorem 4.1

If  $F$  is orthogonally invariant then

$$(4.4) \quad k(A) \geq m(A), \quad \forall A \in F. \quad \blacksquare$$

#### Proof

As in the proof of Theorem 3.1 let  $\phi = \{\phi_k\}$  be any algorithm such that  $k = k(\phi, A) < +\infty$ . This means

$$(4.5) \quad \|\tilde{A}^P(x'_k - \tilde{\alpha})\| / \|\tilde{A}^P \tilde{\alpha}\| < \epsilon, \quad \forall \tilde{A} \in V(y_k),$$

where  $x'_k = \phi_k(N_k(A, b))$ . Decompose

$$x'_k = z_1 + z_2$$



where  $z_1 \in \text{span}(b, Ab, \dots, A^k b)$  and  $z_2$  is orthogonal to  $b, Ab, \dots, A^k b$ . Define  $\tilde{A}_1 = Q\tilde{A}Q$  with  $Q = I - 2\omega\omega^T$  and  $\omega = z_2/\|z_2\|$  for a nonzero  $z_2$  and  $\omega = 0$  for  $z_2 = 0$ . Then  $\tilde{A}_1 \in F$  and  $\tilde{A}_1^i b = A^i b$ ,  $i = 1, 2, \dots, k$ . Thus  $\tilde{A}_1 \in V(y_k)$ . Observe that

$$(4.6) \quad \tilde{\alpha}_1 = Q\tilde{\alpha} \quad \text{and} \quad \|\tilde{A}_1^P \tilde{\alpha}_1\| = \|\tilde{A}^P \tilde{\alpha}\|.$$

Furthermore,

$$(4.7) \quad \begin{aligned} \|\tilde{A}_1^P(x'_k - \tilde{\alpha}_1)\| &= \|\tilde{A}^P Q(z_1 - \tilde{\alpha} + z_2 + 2(\omega, \tilde{\alpha})\omega)\| = \\ &\|\tilde{A}^P(z_1 - \tilde{\alpha}) + \tilde{A}^P z_2 + 2((\omega, \tilde{\alpha}) - (\omega, z_1 - \tilde{\alpha} + z_2 + 2(\omega, \tilde{\alpha})\omega))\tilde{A}^P \omega\| \\ &\|\tilde{A}^P(z_1 - \tilde{\alpha}) + \tilde{A}^P z_2 - 2(\omega, z_2)\tilde{A}^P \omega\| = \|\tilde{A}^P(z_1 - \tilde{\alpha}) - \tilde{A}^P z_2\|. \end{aligned}$$

From (4.5), (4.6), and (4.7) we get

$$\begin{aligned} \frac{e(\tilde{A}, k)}{\|\tilde{A}^P \tilde{\alpha}\|} &\leq \frac{\|\tilde{A}^P(z_1 - \tilde{\alpha})\|}{\|\tilde{A}^P \tilde{\alpha}\|} \leq \frac{1}{2} \left( \frac{\|\tilde{A}^P(z_1 - \tilde{\alpha}) - \tilde{A}^P z_2\|}{\|\tilde{A}^P \tilde{\alpha}\|} + \frac{\|\tilde{A}^P(z_1 - \tilde{\alpha}) + \tilde{A}^P z_2\|}{\|\tilde{A}^P \tilde{\alpha}\|} \right) \\ &\leq \frac{1}{2} \left( \frac{\|\tilde{A}_1^P(x'_k - \tilde{\alpha}_1)\|}{\|\tilde{A}_1^P \tilde{\alpha}_1\|} + \frac{\|\tilde{A}^P(x'_k - \tilde{\alpha})\|}{\|\tilde{A}^P \tilde{\alpha}\|} \right) < \epsilon. \end{aligned}$$

Thus  $k \geq m(A)$ . Since  $\phi$  is an arbitrary algorithm we conclude  $k(A) \geq m(A)$ . Hence (4.4) is proven.  $\square$

Theorem 4.1 provides a lower bound on the optimal matrix index. The next part of this section is devoted to finding algorithms whose class indices are close to this lower bound. As we shall see, this can only be done for certain values of  $p$ .

We check when the coefficients  $c_i^*$  defined by (4.3) can be computed in terms of the information  $N_k(A, b)$ . From (4.3) it follows that  $c^* = [c_0^*, c_1^*, \dots, c_m^*]^T$  satisfies the linear equations

$$(4.8) \quad Hc^* = h$$

where  $H = ((A^{i+p}b, A^{j+p}b))$ ,  $i, j = 0, 1, \dots, m$ , and  $h = [A^p b, A^{p-1}b], \dots, (A^{m+p}b, A^{p-1}b)]^T$ .

We consider two cases.

- (i)  $A = A^T$ . Then if  $2p$  is integer,  $2p \geq 1$  and  $m = k - \lceil p \rceil$ , the vector  $c^*$  depends only on  $N_k(A, b)$ .
- (ii)  $A \neq A^T$ . Then if  $p$  is integer,  $p \geq 1$  and  $m = k - p$ , the vector  $c^*$  depends only on  $N_k(A, b)$ .

If either (i) or (ii) holds then the algorithm  $\phi^{mr} = \{\phi_k^{mr}\}$ ,

$$(4.9) \quad x_0 = 0, x_k = \phi_k^{mr}(N_k(A, b)) = c_0^* b + \dots + c_{k-\lceil p \rceil}^* A^{k-\lceil p \rceil} b, \quad k \geq 1,$$

is well defined and is called the generalized minimal residual algorithm.

Note that for  $p = 1$ , (4.9) coincides with (3.2). Assuming that  $A = A^T > 0$  we can set  $p = 1/2$  and the algorithm  $\phi^{mr}$  is known as the classical conjugate gradient algorithm. See for instance [4].

For  $p = 0$  and  $A = A^T$  the first component, which is the inner product  $(b, \alpha)$ , of the vector  $h$  is in general unknown. If, however, one considers the consistent system  $Mx = g$  and if one agrees to multiply this system by  $M^T$  then  $A = M^T M$ ,  $b = M^T g$ , and  $(b, \alpha) = (g, g)$  is computable. Then the generalized minimal residual algorithm is well defined and is known as the minimum error algorithm. In this case we can compute  $x_k$  as follows. Let  $x_0 = 0$ . For  $i = 0, 1, \dots, k-1$  define

$$(4.10) \quad x_{i+1} = x_i + \frac{1}{q_i} \{f_{i-1}(x_i - x_{i-1}) - r_i\}, \quad r_i = Ax_i - b,$$

where

$$(4.11) \quad q_i = \frac{(r_i, r_i)}{\|Mx_i - g\|^2} - f_{i-1},$$

$$f_{-1} = 0, \quad f_{i-1} = \frac{\|Mx_i - g\|^2}{\|Mx_{i-1} - g\|^2} q_{i-1}.$$

We are ready to show that the generalized minimal residual algorithm is almost strongly optimal.

#### Theorem 4.2

Let  $F$  be orthogonally invariant. Suppose that the following two conditions hold:

- (i) If  $A \in F$  implies  $A = A^T$ ,  $\forall A \in F$ , then  $2p$  is an integer, otherwise  $p$  is an integer.

- (ii) If  $(b, \alpha)$  is known and  $A \in F$  implies  $A = A^T, \forall A \in F$ , then  $p \geq 0$ , otherwise  $p > 0$ .

Then the generalized minimal residual algorithm is almost strongly optimal,

$$(4.14) \quad k(\phi^{mr}, A) \geq k(A) \geq k(\phi^{mr}, A) - \lceil p \rceil, \quad \forall A \in F,$$

where

$$k(\phi^{mr}, A) = m(A) + \lceil p \rceil.$$

#### Proof

Note first that (i) and (ii) guarantee that the algorithm  $\phi^{mr}$  is well defined. From (4.3) and (4.9) we have

$$\| A^p(\alpha - x_k) \| = e(A, k - \lceil p \rceil).$$

Thus

$$k(\phi^{mr}, A) = m(A) + \lceil p \rceil.$$

Obviously  $k(\phi^{mr}, A) \geq k(A)$  which due to (4.4) yields

$$0 \leq k(\phi^{mr}, A) - k(A) \leq \lceil p \rceil.$$

This proves (4.14).

Observe that for  $p = 1$ , the conditions (i) and (ii) are always satisfied and Theorem 4.2 coincides with Theorem 3.1.

For  $p = 1/2$ , Theorem 4.2 states that the classical conjugate gradient algorithm is almost strongly optimal and the matrix index of the classical conjugate gradient differs by at most unity from the optimal matrix index.

If  $p$  can be set equal to zero, then (4.14) states that

$$k(A) = k(\phi^{mr}, A) = m(A) ,$$

Thus, the minimum error algorithm is strongly optimal.

The optimal class index  $k(F)$  for the class  $F = F_2$  for arbitrary  $p$ , and for the class  $F = F_3$  with  $p = 0$  is found in [6]. Recall that

$$q(\epsilon) = \left\lfloor \ln \frac{\epsilon}{1 - \sqrt{1 - \epsilon^2}} / \ln \frac{1 + \sqrt{1 - \rho^2}}{\rho} \right\rfloor .$$

Then

$$k(F_2) = \min(n, q(\epsilon)) .$$

For the Chebyshev algorithm  $\phi^{ch}$  we have

$$k(\phi^{ch}, F_2) = q(\epsilon) .$$

Thus, if  $q(\epsilon) \leq n$  then the Chebyshev algorithm is optimal (but not strongly optimal).

For the class  $F = F_3$  with  $p = 0$  we have

$$k(F_3) = \min(n, \lfloor \ln \epsilon / \ln \rho \rfloor) .$$

The successive approximation algorithm (i.e.,  $x_{i+1} = Bx_i + b$  with  $x_0 = b$ ) has the class index given by

$$k(\phi^{sa}, F_3) = \lfloor \ln \epsilon / \ln \rho \rfloor .$$

Thus, if  $\lfloor \ln \epsilon / \ln \rho \rfloor \leq n$  then the successive approximation algorithm is optimal (but not strongly optimal).

## 5. OPEN PROBLEMS

In this paper we studied optimal algorithms for the solution of  $Ax = b$  using the information operator  $N_k(A, b) = [b, Ab, \dots, A^k b]$ . We have focused on this information operator because it is widely used in practice and because it is susceptible to a very thorough analysis. It would of course be desirable to generalize results of this paper to more general information operators. Until this is accomplished we won't know if  $N_k(A, b)$  is "optimal" information.

For instance, let

$$(5.1) \quad N_k(A, b) = [b, Az_1, Az_2, \dots, Az_k]$$

where  $z_i = z_i(b, Az_1, \dots, Az_{i-1})$  for  $i = 1, 2, \dots, k$ . That is, we still compute the matrix-vector multiplications but now the vector  $z_i$  is an arbitrary function of the previously computed information. For information (5.1) we can generalize the definition of the optimal matrix and class indices in an obvious way. We ask what is the optimal choice of the  $z_i$ , i.e., for which  $z_i$  are the optimal indices minimized. We propose

Conjecture 5.1

If  $F$  is orthogonally invariant then the optimal matrix and class indices are minimized for the vectors  $z_i = A^{i-1}b$ ,  $i = 1, 2, \dots, k$ . That is, the information  $N_k(A, b) = [b, Ab, \dots, A^k b]$  is optimal in the class of information operators of the form (5.1). □

We now consider more general information operators than

(5.1) Let

$$(5.2) \quad N_s(A, b) = [b, L_1(A; b), L_2(A; b, u_1), \dots, L_s(A; b, u_1, \dots, u_{s-1})]$$

where  $u_i = L_i(A; b, u_1, \dots, u_{i-1})$ ,  $i = 1, 2, \dots, s-1$ , and  $L_i$  is a functional which depends linearly on the first argument. The  $L_i$  can depend nonlinearly on  $b$  and on the previously computed information  $u_1, u_2, \dots, u_{i-1}$ . Note that (5.2) is the general form of adaptive linear information and (5.1) as well as (2.2) are special examples of (5.2). We ask what is the optimal adaptive linear information, i.e., what functionals  $L_i$  minimize the optimal matrix and class indices. It would also be interesting to know the minimal value of  $s$  for which we can find the exact solution of a linear system. From [3] we can conclude that  $s \leq (n+1)(n+2)/2 - 1$  with no restriction on the class  $F$ .

We also want to pose a complexity problem. It is known that for the information  $N_k(A, b) = [b, Ab, \dots, A^k b]$ , where  $A = A^T > 0$ , there exist algorithms which are optimal (or almost optimal) and which have linear combinatorial complexity. These two properties guarantee finding an  $\epsilon$ -approximation with minimal (or almost minimal) complexity.

Let  $N_s(A, b)$  be an optimal adaptive linear information of the form (5.2). Does there exist an almost optimal algorithm using  $N_s(A, b)$  with linear combinatorial complexity? Or conversely, is it true that if an information operator is



better than  $N_k(A, b) = [b, Ab, \dots, A^k b]$ , then the combinatory complexity of an almost optimal algorithm cannot be linear?

We can establish one result for  $N_s(A, b)$ . The functionals  $L_i$  in (5.2) must depend on  $b$ . Otherwise the information  $N_s(A, b)$  does not supply enough knowledge to find an  $\epsilon$ -approximation. To show this assume that

$$(5.3) \quad N_s(A, b) = [b, L_1(A), L_2(A; u_1), \dots, L_s(A; u_1, \dots, u_{s-1})]$$

where  $u_i = L_i(A; u_1, \dots, u_{i-1})$  is independent of  $b$ . As in (2.8), let  $k(F)$  be the minimal value of  $s$  such that there exists an algorithm which uses  $N_s(A, b)$  and finds an  $\epsilon$ -approximation in the sense of (4.1).

For simplicity we establish the desired result only for the class  $F_2$ . Without loss of generality we assume that  $\epsilon \leq \rho$ . (Otherwise the algorithm  $\phi_s(N_s(A, b)) = b$  yields an  $\epsilon$ -approximation.)

#### Theorem 5.1

Let  $\epsilon \leq \rho$ ,  $F = F_2$  and  $p$  be arbitrary. There exists a vector  $b$  such that

$$k(F_2) \geq \frac{n(n+1)}{2} .$$

□

#### Proof

Let  $A = I + B$  where

$$(5.4) \quad L_i(B, u_1, \dots, u_{i-1}) = 0, \quad i = 1, 2, \dots, s,$$

and  $u_i = L_i(I, u_1, \dots, u_{i-1})$ . Note that (5.4) corresponds to  $s$  homogenous linear equations in coefficients of  $B$ . Since  $B$  is an  $n \times n$  symmetric matrix, we have  $n(n+1)/2$  unknowns.

If  $s < n(n+1)/2$  then there exists a nonzero matrix  $B$  satisfying (5.4). We can normalize  $B$  such that  $\|B\| = \rho$ . Define a vector  $b$ ,  $\|b\| = 1$ , such that  $Bb = cb$  with  $c = \pm\rho$ . Let  $\tilde{A} = I - B$ . Then  $\tilde{A} \in F_2$  and  $N_s(\tilde{A}, b) = N_s(A, b)$ . Let  $\phi = \{\phi_k\}$  be an algorithm and  $x_k = \phi_k(N_k(A, b))$ . Let

$$a = \max \left( \frac{\|\tilde{A}^p(x_k - \tilde{\alpha})\|}{\|\tilde{A}^p \tilde{\alpha}\|}, \frac{\|A^p(x_k - \alpha)\|}{\|A^p \alpha\|} \right)$$

where  $\alpha = A^{-1}b$  and  $\tilde{\alpha} = \tilde{A}^{-1}b$ . Then  $\alpha = \frac{1}{1+c}b$ ,

$\|A^p \alpha\| = (1+c)^{p-1}$ ,  $\tilde{\alpha} = \frac{1}{1-c}b$  and  $\|\tilde{A}^p \tilde{\alpha}\| = (1-c)^{p-1}$ . Let

$$x_k = c_1 b + x$$

where  $c_1 = (x_k, b)$  and  $x$  is orthogonal to  $b$ . Then

$$((I \pm B)^p x, b) = 0 \quad \text{and}$$

$$\|(I \pm B)^p(x_k - \frac{1}{1 \pm c}b)\| \geq |c_1(1 \pm c)^p - (1 \pm c)^{p-1}|.$$

Thus

$$a \geq \max(|c_1(1+\rho) - 1|, |c_1(1-\rho) - 1|) \geq \rho \geq \varepsilon.$$

Since  $\phi$  is arbitrary, this proves that it is impossible to find an  $\varepsilon$ -approximation for  $s < n(n+1)/2$ . This completes the proof.  $\square$

Note that for the class  $F_2$ , we can recover the matrix  $A = (a_{kj})$  knowing a suitable chosen  $N_s(A, b)$  with  $s = n(n+1)/2$ .

Indeed, it is enough to define

$$L_1(A) = a_{11}, \dots, L_n(A, u_1, \dots, u_{n-1}) = a_{1n}, L_{n+1}(A, u_1, \dots, u_n) = a_{22}, \dots, L_{2n-1}(A, u_1, \dots, u_{2n-2}) = a_{2n}, \dots, L_s(A, u_1, \dots, u_{s-1}) = a_{nn}.$$

Knowing  $A$ , we can define  $\phi_s(N_s(A, b)) = A^{-1}b$ . This and Theorem 5.1 shows that  $n(n+1)/2$  evaluations of linear functionals are necessary to find an  $\varepsilon$ -approximation for any  $\varepsilon \in [0, \rho]$ .

Thus, even for very moderate values of  $\varepsilon$ , the information operators (5.3) do not supply enough information with  $s$  less than  $n(n+1)/2$  for  $F = F_2$ . This is in sharp contrast with the information  $N_k(A, b) = [b, Ab, \dots, A^k b]$  (where all evaluations depend on  $b$ ) and where only a few evaluations of  $A^i b$  are sufficient to find an  $\varepsilon$ -approximation for moderate  $\varepsilon$ .

#### ACKNOWLEDGEMENT

We are grateful to A. Kielbasiński and J. Kuczyński for careful checking of the manuscript.

## REFERENCES

1. Kuczyński [80] Kuczyński, J., "Optimality of MR Algorithm", Polish Academy of Sciences Report, 1980.
2. Meinardus [63] Meinardus, G., "Über eine Verallgemeinerung einer Ungleichung von L.V. Kantorovitsch", Numer. Math., V.5, 1963, 14-23.
3. Rabin [72] Rabin, M.O., "Solving Linear Equations by Means of Scalar Products", in Complexity of Computer Computations, edited by R.E. Miller and J.W. Thatcher, Plenum Press, New York, 1972, 11-20.
4. Stiefel [58] Stiefel, E., "Kernel Polynomials in Linear Algebra and Their Numerical Applications", NBS. Appl. Math., 43, 1958, 1-22.
5. Traub and Woźniakowski [80] Traub, J.F. and Woźniakowski, H., A General Theory of Optimal Algorithms, Academic Press, New York, 1980.
6. Traub and Woźniakowski [80] Traub, J.F. and Woźniakowski, H., "On the Optimal Solution of Large Linear Systems", Dept. of Computer Science Report, Columbia University, 1980.
7. Traub, Wasilkowski, [83] Traub, J.F., Wasilkowski, G.W., and and Woźniakowski Woźniakowski, H., Information, Uncertainty, Complexity, Addison-Wesley, Reading, Mass., 1983.