University of Maryland College Park

Institute for Advanced Computer Studies TR–91–???
Department of Computer Science TR–????

# Backward Error Bounds for Approximate Krylov Subspaces*

G. W. Stewart†

May 2001

ABSTRACT

Let $A$ be a matrix of order $n$ and let $\mathcal{U} \subset \mathbb{C}^n$ be a subspace of dimension $k$. In this note we determine a matrix $E$ of minimal norm such that $\mathcal{U}$ is a Krylov subspace of $A + E$.

*This report is available by anonymous ftp from thales.cs.umd.edu in the directory pub/reports or on the web at http://www.cs.umd.edu/~stewart/.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (stewart@cs.umd.edu). Work supported by the National Science Foundation under Grant No. 970909-8562. Part of this work was done during the author's weekly visits at the Mathematics and Computer Science Division of the National Institute for Standards and Technology.

# Backward Error Bounds for Approximate Krylov Subspaces

G. W. Stewart

ABSTRACT

Let $A$ be a matrix of order $n$ and let $\mathcal{U} \subset \mathbb{C}^n$ be a subspace of dimension $k$. In this note we determine a matrix $E$ of minimal norm such that $\mathcal{U}$ is a Krylov subspace of $A + E$.

## 1. Introduction

Let $A$ be a matrix of order $n$. Given a starting vector $u$, we say that the sequence

$$u, Au, A^2u, \ldots$$

is the *Krylov sequence* associated with $A$ and $u$. The subspace

$$\mathcal{K}_k(A, u) = \mathrm{span}(u, Au, A^2u, \ldots, A^{k-1}u)$$

is called a *Krylov subspace.*

Krylov subspaces arise in many applications. They are especially important in algorithms for the iterative solution of linear systems [2] and for approximating eigenpairs of large matrices [4, 6]. Since bases for Krylov subspaces are sometimes computed inaccurately, it is desirable to have some way of assessing their quality. There are two approaches. Given a Krylov subspace $\mathcal{U}$, we can

1. give bounds on the angle between $\mathcal{U}$ and the nearest Krylov subspace of $A$,

2. determine a matrix $E$ of minimal norm such that $\mathcal{U}$ is a Krylov subspace of $A+E$.

The first approach leads to a seemingly difficult and currently unsolved problem. The purpose of this note is to show that the second approach has a simple, constructive solution.

To solve the problem we will use a characterization of Krylov subspaces called a Krylov decomposition [5]. Accordingly, in the next section we will discuss these decompositions and their relation to the widely used Arnoldi decompositions. In Section 3 we will present our results and comment on them.

Throughout this note $\|\cdot\|$ will denote a family of consistent unitarily invariant norms. The special cases of the spectral 2-norm and the Frobenius norm will be denoted by $\|\cdot\|_2$ and $\|\cdot\|_\mathrm{F}$. For more on unitarily invariant norms see [7].

## 2. Arnoldi and Krylov decompositions

As a rule, the vectors in a Krylov sequence $u, Au, A^2u, \ldots$ become increasingly dependent. To circumvent this problem we can construct orthonormal bases $u_1, u_2, \ldots$ for the Krylov subspaces $\mathcal{K}_k(A, u_1)$ by successively orthogonalizing $Au_j$ against $u_1, \ldots, u_j$ and normalizing the result — a process known as the Arnoldi algorithm [1]. If we set $U_k = (u_1, \ldots, u_k)$, then the results of the Arnoldi algorithm can be summarized by the relation

$$AU_{k-1} = U_k H_k,$$

where

$$H_k = U_k^{\mathrm{H}} A U_{k-1}$$

is an $k\times(k-1)$ upper Hessenberg matrix — that is, it is zero below its first subdiagonal. We call such a relation an *Arnoldi decomposition.*

In general, all the subdiagonal elements of $H_k$ will be nonzero, in which case the Arnoldi decomposition is uniquely determined by the starting vector $u_1$. If, however, $b_{j,j-1}$ is zero, then $Au_{j-1}$ is exactly dependent on $u_1, \ldots, u_{j-1}$, so that that one must restart the Arnoldi process with some vector $u_j$ that is orthogonal to $u_1, \ldots, u_{j-1}$. It this case we will say that the corresponding Krylov subspace is restarted. Although our results will apply to restarted Krylov subspaces, it should be kept in mind that the unrestarted case is the norm.[1]

The essential uniqueness of the Arnoldi decomposition is a drawback when we wish to consider different bases for a particular Krylov subspace. To circumvent this problem we introduce *Krylov decompositions,* which have the form

$$AU_{k-1} = U_k B_k, \tag{2.1}$$

where $U_k$ has independent columns and $B_k$ is arbitrary. We call the column space of $U_k$ the space of the decomposition. Any Arnoldi decomposition is, of course, a Krylov decomposition. Conversely, it can be shown [5] that corresponding to any Krylov decomposition there is an Arnoldi decomposition with the same space. Thus Krylov decompositions are a general characterization of Krylov subspaces. In what follows we will assume that the matrices $U_k$ in our Krylov decompositions are orthonormal.

## 3. The results

Given a subspace $\mathcal{U}$, our object is to show it is a Krylov subspace of a perturbation of $A$ and to bound the perturbation. We proceed indirectly. First we show that there

[1] This state of affairs is due to the law of perversity of nature. In applications, a restarting represents the convergence of an iterative method or the isolation of an eigenspace — something to be happy about.

is a basis for for $\mathcal{U}$ that satisfies an approximate Krylov relation for $A$ with a minimal residual. We then use standard techniques to throw the residual back onto $A$.

The following lemma is the starting point for the first part of our program.

**Lemma 3.1.** *Let $U = (U_1\ \ u_2)$ orthonormal. Then the* Krylov residual

$$R = AU_1 - UB$$

*is minimized in any unitarily invariant norm when*

$$B = U^{\mathrm{H}} AU_1, \tag{3.1}$$

*in which case*

$$U^{\mathrm{H}} R = 0. \tag{3.2}$$

**Proof.** Let $\hat{U} = (U\ \ U_3)$ be unitary. Then by unitary invariance, $R$ has the same norm as

$$\hat{U}^{\mathrm{H}} R = \begin{pmatrix} U^{\mathrm{H}} AU_1 - B \\ U_3^{\mathrm{H}} R \end{pmatrix}.$$

Since $U_3^{\mathrm{H}} R$ is independent of $B$, the norm of $R$ is minimized when when $B = U^{\mathrm{H}} AU_1$. The orthogonality condition $U^{\mathrm{H}} R = 0$ can be verified directly. ∎

Given a subspace $\mathcal{U} \subset \mathbb{C}^n$ of dimension $k$, this theorem suggests that we proceed with our program by choosing an orthonormal basis $U$ for $\mathcal{U}$ and use (3.1) to compute an optimal Krylov residual. Unfortunately, this residual is optimal only for the specific choice of $U$. The reason is that not every basis for a Krylov subspace corresponds to a Krylov decomposition, so that Lemma 3.1 is likely to give us large Krylov residuals, even when $\mathcal{U}$ is itself a Krylov subspace. To optimize globally over all bases, we must try to determine a $k\times k$ unitary matrix $V$ such that $UV$ has a Krylov residual that is as small as possible.

To do this, partition $V = (V_1\ \ v_2)$. Let

$$S = AU - U(U^{\mathrm{H}} AU).$$

If we postmultiply $S$ by $V_1$ we get

$$SV_1 = A(UV_1) - UV[(UV)^{\mathrm{H}} A(UV_1)].$$

It follows that $SV_1$ is the optimal Krylov residual for the particular basis $UV$. Thus we wish to minimize the norm of $SV_1$ as $V_1$ varies over the set of $k\times(k-1)$ orthonormal matrices.

This is easily done. Let $\sigma_1 \geq \cdots \geq \sigma_k \geq 0$ be the singular values of $S$ and let $\tau_1 \geq \cdots \geq \tau_{k-1} \geq 0$ be the singular values of $SV_1$. Then by the interleaving theorem for singular values [3, Lemma 3.3.1], $\tau_i \geq \sigma_{i+1}$ $(i = 1, \ldots, k-1)$. Since a unitarily invariant norm of a matrix is a nondecreasing function of its singular values, the norm of $SV_1$ is minimized when $\tau_i = \sigma_{i+1}$ $(i = 1, \ldots, k-1)$. These equalities can be attained if we take $V_1$ to be the right singular vectors of $S$ corresponding to $\sigma_2, \ldots, \sigma_k$. The vector $v_2$ is necessarily the right singular vector corresponding to $\sigma_1$, and this choice of $V = (V_1\ v_2)$ gives us a globally optimal Krylov residual for $\mathcal{U}$.

The second step in our program is to project the Krylov residual back on $A$. Let $\hat{U} = UV$, where $V = (V_1\ v_2)$ is as in the last paragraph. Then $R = A\hat{U} - \hat{U}(\hat{U}^{\mathrm{H}}A\hat{U}_1)$ is a globally optimal Krylov residual. If we set $E = RU_1^{\mathrm{H}}$, then $\|E\| = \|R\|$, and it follows from (3.2) that

$$(A + E)\hat{U}_1 = \hat{U}[\hat{U}^{\mathrm{H}}(A + E)\hat{U}_1]$$

is a Krylov decomposition of $A + E$.

Moreover, $E$ is the smallest possible such backward error. For if $(A + F)\hat{U}_1 = \hat{U}[\hat{U}^{\mathrm{H}}(A + F)\hat{U}_1]$, then

$$R = A\hat{U}_1 - \hat{U}(\hat{U}A\hat{U}_1) = (\hat{U}\hat{U}^{\mathrm{H}} - I)F\hat{U}_1.$$

But $\hat{U}\hat{U}^{\mathrm{H}} - I$ and $\hat{U}_1$ both have 2-norm one, so that $\|E\| = \|R\| \leq \|F\|$.

We summarize these results in the following theorem, in which we recapitulate our notation and constructions.

**Theorem 3.2.** *Let $A$ be of order $n$ and let $U = (U_1\ u_2) \in \mathbb{C}^{n\times k}$ be orthonormal. Let*

$$S = AU - U(U^{\mathrm{H}}AU) \tag{3.3}$$

*and let $\sigma_1 \geq \cdots \geq \sigma_k \geq 0$ be the singular values of $S$. Let $V = (V_1\ v_2)$ be unitary with the columns $V_1$ being the right singular vectors of $S$ corresponding to $\sigma_2, \ldots, \sigma_k$. Set*

$$\hat{U} = UV = (\hat{U}_1\ \hat{u}_2) \quad \text{and} \quad R = SV_1.$$

*Then the approximate Krylov decomposition*

$$A\hat{U}_1 = \hat{U}(\hat{U}^{\mathrm{H}}A\hat{U}_1) + R,$$

*has minimal residual norm in any unitarily invariant norm. If we set*

$$E = RU^{\mathrm{H}}, \tag{3.4}$$

*then*

$$\|E\| = \|R\|$$

*and* $A + E$ *has the Krylov decomposition*

$$(A+E)\hat{U}_1 = \hat{U}[\hat{U}^{\mathrm{H}}(A+E)\hat{U}_1]. \tag{3.5}$$

*Of all matrices* $E$ *satisfying* (3.5)*, the matrix* (3.4) *has minimal norm.*

There are several comments to be made about this theorem.

First, our results are independent of the initial choice of a basis $U$ for $\mathcal{U}$. Specifically, if we replace $U$ by $UQ$, where $Q$ is unitary, $S$ in (3.3) is replaced by $SQ$, $V$ is replaced by $Q^{\mathrm{H}}V$, and hence $\hat{U}$ does not change.

Second, we can give explicit expressions for $\|R\|$ in the 2- and Frobenius norms. Namely

$$\|R\|_2 = \sigma_2 \quad \text{and} \quad \|R\|_{\mathrm{F}} = \sqrt{\sigma_2^2 + \cdots + \sigma_k^2}.$$

Third, the process is constructive. Given a basis for $\mathcal{U}$, we can actually construct the backward error.

Fourth, $\mathcal{U}$ is actually a Krylov subspace, then $R$ must be zero. This means that only the singular value $\sigma_1$ of $S$ can be nonzero. Thus we have an alternate characterization of what it means to be a Krylov subspace.

**Corollary 3.3.** *An orthonormal matrix* $U$ *spans a Krylov subspace of* $A$ *if and only the matrix* $S = AU - U(U^{\mathrm{H}}AU)$ *has rank not greater than one.*

In fact, this characterization can be derived in another way. Write $S = (I - UU^{\mathrm{H}})(AU)$. Because $U$ is a basis for a Krylov sequence, $AU$ can have at most one vector that is orthogonal to the column space of $U$. Since $I - UU^{\mathrm{H}}$ is the projection onto the orthogonal complement of the column space of $U$, the column space of $S$ can contain at most one vector.

Fifth, our first candidate for assessing an approximate Krylov subspace — namely, finding the nearest Krylov subspace — is more direct than the approach taken here — namely, finding an optimal backward perturbation. But in applications the latter is often more useful. For the implication of backward error analyses for eigenproblems see [6, Theorem II.1.3].

Finally, if $A$ is Hermitian, it is natural to require that the backward error $E$ also be Hermitian. This can be done by setting

$$E = RU^{\mathrm{H}} + UR^{\mathrm{H}}.$$

It is easily verified $\|E\|_2 = \|R\|_2$, so that $E$ is optimal in the 2-norm. But $\|E\|_{\mathrm{F}} = \sqrt{2}\|R\|_{\mathrm{F}}$, so that $E$ might not be optimal in the Frobenius norm. But it can be off by no more than a factor of $\sqrt{2}$.

**Acknowledgement**

Part of this work was done during a pleasant stay at the Mathematics Department of the University of Utrecht.

**References**

[1] W. E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9:17–29, 1951.

[2] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM, Philadelphia, 1997.

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

[4] Y. Saad. *Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms.* John Wiley, New York, 1992.

[5] G. W. Stewart. A Krylov–Schur algorithm for large eigenproblems. Technical Report TR–4127, Department of Computer Science, University of Maryland, 2000.

[6] G. W. Stewart. *Matrix Algorithms II: Eigensystems.* SIAM, Philadelphia, 2001. To appear August, 2001. Preprint available at `http://www.cs.umd.edu/~stewart`.

[7] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, New York, 1990.