Chapter 8

Krylov subspaces

8.1 Introduction

In the power method or in the inverse vector iteration we computed, up to normalization, sequences of the form

$$\mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \ldots$$

The information available at the k-th step of the iteration is the single vector $\mathbf{x}^{(k)} = A^k\mathbf{x}/\|A^k\mathbf{x}\|$. One can pose the question if discarding all the previous information $\{\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(k-1)}\}$ is not a too big waste of information. This question is not trivial to answer. On one hand there is a big increase of memory requirement, on the other hand exploiting all the information computed up to a certain iteration step can give much better approximations to the searched solution. As an example, let us consider the symmetric matrix

$$T = \left(\frac{51}{\pi}\right)^2 \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{50\times 50}.$$

the lowest eigenvalue of which is around 1. Let us choose $\mathbf{x} = [1, \ldots, 1]^*$ and compute the first three iterates of inverse vector iteration, $\mathbf{x}$, $T^{-1}\mathbf{x}$, and $T^{-2}\mathbf{x}$. We denote their Rayleigh

k	$\rho^{(k)}$	$\vartheta_1^{(k)}$	$\vartheta_2^{(k)}$	$\vartheta_3^{(k)}$
1	10.541456	10.541456		
2	1.012822	1.009851	62.238885	
3	0.999822	0.999693	9.910156	147.211990

Table 8.1: Ritz values $\vartheta_j^{(k)}$ vs. Rayleigh quotients $\rho^{(k)}$ of inverse vector iterates.

quotients by $\rho^{(1)}$, $\rho^{(2)}$, and $\rho^{(3)}$, respectively. The Ritz values $\vartheta_j^{(k)}$, $1 \le j \le k$, obtained with the Rayleigh-Ritz procedure with $\mathcal{K}_k(\mathbf{x}) = \operatorname{span}(\mathbf{x}, T^{-1}\mathbf{x}, \ldots, T^{1-k}\mathbf{x})$, $k = 1, 2, 3$, are given in Table 8.1. The three smallest eigenvalues of T are 0.999684, 3.994943, and 8.974416. The approximation errors are thus $\rho^{(3)} - \lambda_1 \approx 0.000'14$ and $\vartheta_1^{(3)} - \lambda_1 \approx 0.000'009$, which is 15 times smaller.

These results immediately show that the cost of three matrix vector multiplications can be much better exploited than with (inverse) vector iteration. We will consider in this section a kind of space that is very often used in the iterative solution of linear systems as well as of eigenvalue problems.

8.2 Definition and basic properties

Definition 8.1 The matrix

$$K^m(\mathbf{x}) = K^m(\mathbf{x}, A) := [\mathbf{x}, A\mathbf{x}, \ldots, \mathbf{A}^{(m-1)}\mathbf{x}] \in \mathbb{F}^{n\times m} \tag{8.1}$$

generated by the vector $\mathbf{x} \in \mathbb{F}^n$ is called a **Krylov matrix**. Its columns span the **Krylov (sub)space**

$$\mathcal{K}^m(\mathbf{x}) = \mathcal{K}^m(\mathbf{x}, A) := \text{span}\left\{\mathbf{x}, A\mathbf{x}, \mathbf{A}^2\mathbf{x}, \ldots, A^{(m-1)}\mathbf{x}\right\} = \mathcal{R}\left(K^m(\mathbf{x})\right) \subset \mathbb{F}^n. \tag{8.2}$$

The **Arnoldi** and **Lanczos algorithms** are methods to compute an orthonormal basis of the Krylov space. Let

$$\left[\mathbf{x}, A\mathbf{x}, \ldots, A^{k-1}\mathbf{x}\right] = \mathbf{Q}^{(k)}\mathbf{R}^{(k)}$$

be the QR factorization of the Krylov matrix $K^m(\mathbf{x})$. The Ritz values and Ritz vectors of A in this space are obtained by means of the $k \times k$ eigenvalue problem

$$\mathbf{Q}^{(k)^*} A \mathbf{Q}^{(k)}\mathbf{y} = \vartheta^{(k)}\mathbf{y}. \tag{8.3}$$

If $(\vartheta_j^{(k)}, \mathbf{y}_j)$ is an eigenpair of (8.3) then $(\vartheta_j^{(k)}, \mathbf{Q}^{(k)}\mathbf{y}_j)$ is a Ritz pair of A in $K^m(\mathbf{x})$.

The following properties of Krylov spaces are easy to verify [1, p.238]

1. *Scaling.* $\mathcal{K}^m(\mathbf{x}, A) = \mathcal{K}^m(\alpha\mathbf{x}, \beta A), \quad \alpha, \beta \neq 0.$
2. *Translation.* $\mathcal{K}^m(\mathbf{x}, A - \sigma\mathbf{I}) = \mathcal{K}^m(\mathbf{x}, A).$
3. *Change of basis.* If U is unitary then $U\mathcal{K}^m(U^*\mathbf{x}, U^*AU) = \mathcal{K}^m(\mathbf{x}, A)$.

   In fact,

$$\begin{aligned} K^m(\mathbf{x}, A) &= [\mathbf{x}, A\mathbf{x}, \ldots, A^{(m-1)}\mathbf{x}] \\ &= U[U^*\mathbf{x}, (U^*AU)U^*\mathbf{x}, \ldots, (U^*AU)^{m-1}U^*\mathbf{x}], \\ &= UK^m(U^*\mathbf{x}, U^*AU). \end{aligned}$$

Notice that the scaling and translation invariance hold only for the Krylov subspace, not for the Krylov matrices.

What is the dimension of $\mathcal{K}^m(\mathbf{x})$? It is evident that for $n \times n$ matrices A the columns of the Krylov matrix $K^{n+1}(\mathbf{x})$ are linearly dependent. (A subspace of $\mathbb{F}^n$ cannot have a dimension bigger than n.) On the other hand if $\mathbf{u}$ is an eigenvector corresponding to the eigenvalue λ then $A\mathbf{u} = \lambda\mathbf{u}$ and, by consequence, $\mathcal{K}^2(\mathbf{u}) = \text{span}\{\mathbf{u}, A\mathbf{u}\} = \text{span}\{\mathbf{u}\} = \mathcal{K}^1(\mathbf{u})$. So, there is a smallest m, $1 \leq m \leq n$, *depending on* $\mathbf{x}$ such that

$$\mathcal{K}^1(\mathbf{x}) \subsetneq \mathcal{K}^2(\mathbf{x}) \subsetneq \cdots \subsetneq \mathcal{K}^m(\mathbf{x}) = \mathcal{K}^{m+1}(\mathbf{x}) = \cdots$$

For this number m,

$$K_{m+1}(\mathbf{x}) = [\mathbf{x}, A\mathbf{x}, \ldots, A^m\mathbf{x}] \in \mathbb{F}^{n\times m+1} \tag{8.4}$$

has linearly dependant columns, i.e., there is a *nonzero* vector $\mathbf{a} \in \mathbb{F}^{m+1}$ such that

$$K_{m+1}(\mathbf{x})\mathbf{a} = p(A)\mathbf{x} = \mathbf{0}, \qquad p(\lambda) = a_0 + a_1\lambda + \cdots + a_m\lambda^m. \tag{8.5}$$

The polynomial $p(\lambda)$ is called the minimal polynomial of A relativ to $\mathbf{x}$. By construction, the highest order coefficient $a_m \neq 0$.

If A is *diagonalizable*, then the degree of the minimal polynomial relativ to $\mathbf{x}$ has a simple geometric meaning (which does not mean that it is easily checked). Let

$$\mathbf{x} = \sum_{i=1}^{m} \mathbf{u}_i = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

where the $\mathbf{u}_i$ are eigenvectors of A, $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$, and $\lambda_i \neq \lambda_j$ for $i \neq j$. Notice that we have arranged the eigenvectors such that the coefficients in the above sum are all unity. Now we have

$$A^k\mathbf{x} = \sum_{i=1}^{m} \lambda_i^k \mathbf{u}_i = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \begin{pmatrix} \lambda_1^k \\ \vdots \\ \lambda_m^k \end{pmatrix},$$

and, by consequence,

$$K^j\mathbf{x} = \underbrace{[\mathbf{u}_1, \ldots, \mathbf{u}_m]}_{\in\mathbb{C}^{n\times m}} \underbrace{\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{j-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{j-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \cdots & \lambda_m^{j-1} \end{bmatrix}}_{\in\mathbb{C}^{m\times j}}.$$

Since matrices of the form

$$\begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{s-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{s-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_s & \cdots & \lambda_m^{s-1} \end{bmatrix} \in \mathbb{F}^{s\times s}, \quad \lambda_i \neq \lambda_j \text{ for } i \neq j,$$

so-called *Vandermonde matrices*, are *nonsingular* if the λ_i are different (their determinant equals $\prod_{i\neq j}(\lambda_i - \lambda_j)$) the Krylov matrices $K^j(\mathbf{x})$ are nonsingular for $j \leq m$. Thus for diagonalizable matrices A we have

$$\dim \mathcal{K}^j(\mathbf{x}, A) = \min\{j, m\}$$

where m is the number of eigenvectors needed to represent $\mathbf{x}$. The subspace $\mathcal{K}^m(\mathbf{x})$ is the smallest invariant space that contains $\mathbf{x}$.

8.3 Polynomial representation of Krylov subspaces

In this section we assume A to be Hermitian. Let $\mathbf{s} \in \mathcal{K}^j(\mathbf{x})$. Then

$$\mathbf{s} = \sum_{i=0}^{j-1} c_i A^i \mathbf{x} = \pi(A)\mathbf{x}, \qquad \pi(\mathbf{x}) = \sum_{i=0}^{j-1} c_i \mathbf{x}^i. \tag{8.6}$$

Let $\mathbb{P}_j$ be the space of polynomials of degree $\leq j$. Then (8.6) becomes

$$\mathcal{K}^j(\mathbf{x}) = \{\pi(A)\mathbf{x} \mid \pi \in \mathbb{P}_{j-1}\}. \tag{8.7}$$

Let m be the smallest index for which $\mathcal{K}^m(\mathbf{x}) = \mathcal{K}^{m+1}(\mathbf{x})$. Then, for $j \leq m$ the mapping

$$\mathcal{P}^{j-1} \ni \sum c_i \xi^i \to \sum c_i A^i \mathbf{x} \in \mathcal{K}^j(\mathbf{x})$$

is bijective, while it is only surjective for $j > m$.

Let $Q \in \mathbb{F}^{n \times j}$ be a matrix with orthonormal columns that span $\mathcal{K}^j(\mathbf{x})$, and let $A' = Q^*AQ$. The spectral decomposition

$$A'X' = X'\Theta, \qquad X'^*X' = I, \quad \Theta = \operatorname{diag}(\vartheta_i, \ldots, \vartheta_j),$$

of A' provides the Ritz values of A in $\mathcal{K}^j(\mathbf{x})$. The columns $\mathbf{y}_i$ of $Y = QX'$ are the Ritz vectors.

By construction the Ritz vectors are mutually orthogonal. Furthermore,

$$A\mathbf{y}_i - \vartheta_i \mathbf{y}_i \perp \mathcal{K}^j(\mathbf{x}) \tag{8.8}$$

because

$$Q^*(AQ\mathbf{x}'_i - Q\mathbf{x}'_i\vartheta_i) = Q^*AQ\mathbf{x}'_i - \mathbf{x}'_i\vartheta_i = A'\mathbf{x}'_i - \mathbf{x}'_i\vartheta_i = \mathbf{0}.$$

It is easy to represent a vector in $\mathcal{K}^j(\mathbf{x})$ that is orthogonal to $\mathbf{y}_i$.

Lemma 8.2 *Let $(\vartheta_i, \mathbf{y}_i)$, $1 \leq i \leq j$ be Ritz values and Ritz vectors of A in $\mathcal{K}^j(\mathbf{x}), j \leq m$. Let $\omega \in \mathbb{P}_{j-1}$. Then*

$$\omega(A)\mathbf{x} \perp \mathbf{y}_k \quad \Longleftrightarrow \quad \omega(\vartheta_k) = 0. \tag{8.9}$$

Proof. "$\Longrightarrow$" Let first $\omega \in \mathbb{P}_j$ with $\omega(x) = (x - \vartheta_k)\pi(x), \pi \in \mathbb{P}_{j-1}$. Then

$$\begin{aligned} \mathbf{y}_k^*\omega(A)\mathbf{x} &= \mathbf{y}_k^*(A - \vartheta_k\mathbf{I})\pi(A)\mathbf{x}, \qquad \text{here we use that } A = A^* \\ &= (A\mathbf{y}_k - \vartheta_k\mathbf{y}_k)^*\pi(A)\mathbf{x} \overset{(8.8)}{=} 0, \end{aligned} \tag{8.10}$$

whence (8.9) is sufficient.

"$\Longleftarrow$" Let now

$$\begin{aligned} \mathcal{K}^j(\mathbf{x}) \supset \mathcal{S}_k &:= \{\tau(A)\mathbf{x} \mid \tau \in \mathbb{P}_{j-1}, \tau(\vartheta_k) = 0\}, \qquad \tau(\vartheta_k) = (x - \vartheta_k)\psi(x), \psi \in \mathbb{P}_{j-2} \\ &= (A - \vartheta_k\mathbf{I})\{\psi(A)\mathbf{x} \mid \psi \in \mathbb{P}_{j-2}\} \\ &= (A - \vartheta_k\mathbf{I})\mathcal{K}^{j-1}(\mathbf{x}). \end{aligned}$$

$\mathcal{S}_k$ has dimension $j-1$ and $\mathbf{y}_k$, according to (8.10) is orthogonal to $\mathcal{S}_k$. As the dimension of a subspace of $\mathcal{K}^j(\mathbf{x})$ that is orthogonal to $\mathbf{y}_k$ is $j-1$, it must coincide with $\mathcal{S}_k$. These elements have the form that is claimed by the Lemma. ∎

Next we define the polynomials

$$\mu(\xi) := \prod_{i=1}^{j}(\xi - \vartheta_i) \in \mathbb{P}_j, \quad \pi_k(\xi) := \frac{\mu(\xi)}{(\xi - \vartheta_k)} = \prod_{\substack{i=1 \\ i \neq k}}^{j}(\xi - \vartheta_i) \in \mathbb{P}_{j-1}.$$

Then the Ritz vector $\mathbf{y}_k$ can be represented in the form

$$\mathbf{y}_k = \frac{\pi_k(A)\mathbf{x}}{\|\pi_k(A)\mathbf{x}\|}, \tag{8.11}$$

as $\pi_k(\xi) = 0$ for all $\vartheta_i, i \neq k$. According to Lemma 8.2 $\pi_k(A)\mathbf{x}$ is perpendicular to all $\mathbf{y}_i$ with $i \neq k$. Further,

$$\beta_j := \|\mu(A)\mathbf{x}\| = \min\{\|\omega(\mathbf{A})\mathbf{x}\| \mid \omega \in \mathbb{P}_j \text{ monic}\}. \tag{8.12}$$

(A polynomial in $\mathbb{P}_j$ is monic if its highest coefficients $a_j = 1$.) By the first part of Lemma 8.2 $\mu(A)\mathbf{x} \in \mathcal{K}^{j+1}(\mathbf{x})$ is orthogonal to $\mathcal{K}^j(\mathbf{x})$. As each monic $\omega \in \mathbb{P}_j$ can be written in the form

$$\omega(\xi) = \mu(\xi) + \psi(\xi), \qquad \psi \in \mathbb{P}_{j-1},$$

we have

$$\|\omega(A)\mathbf{x}\|^2 = \|\mu(A)\mathbf{x}\|^2 + \|\psi(A)\mathbf{x}\|^2,$$

as $\psi(A)\mathbf{x} \in \mathcal{K}^j(\mathbf{x})$. Because of property (8.12) μ is called the *minimal polynomial* of $\mathbf{x}$ of degree j. (In (8.5) we constructed the minimal polynomial of degree m in which case $\beta_m = 0$.)

Let $\mathbf{u}_1, \cdots, \mathbf{u}_m$ be the eigenvectors of A corresponding to $\lambda_1 < \cdots < \lambda_m$ that span $\mathcal{K}^m(\mathbf{x})$. We collect the first i of them in the matrix $U_i := [\mathbf{u}_1, \ldots, \mathbf{u}_i]$. Let $\|\mathbf{x}\| = 1$. Let $\varphi := \angle(\mathbf{x}, \mathbf{u}_i)$ and $\psi := \angle(\mathbf{x}, U_iU_i^*\mathbf{x})$ $(\leq \varphi)$. (Remember that $U_iU_i^*\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ on $\mathcal{R}(U_i)$.)

Let

$$\mathbf{g} := \frac{U_iU_i^*\mathbf{x}}{\|U_iU_i^*\mathbf{x}\|} \qquad \text{and} \qquad \mathbf{h} := \frac{(I - U_iU_i^*)\mathbf{x}}{\|(I - U_iU_i^*)\mathbf{x}\|}.$$

Then we have

$$\|U_iU_i^*\mathbf{x}\| = \cos\psi, \qquad \|(\mathbf{I} - U_iU_i^*)\mathbf{x}\| = \sin\psi.$$

The following Lemma will be used for the estimation of the difference $\vartheta_i^{(j)} - \lambda_i$ of the desired eigenvalue and its approximation from the Krylov subspace.

Lemma 8.3 ([1, p.241]) *For* each $\pi \in \mathbb{P}_{j-1}$ *and each* $i \leq j \leq m$ *the Rayleigh quotient*

$$\rho(\pi(A)\mathbf{x}; A - \lambda_i I) = \frac{(\pi(A)\mathbf{x})^*(A - \lambda_i I)(\pi(A)\mathbf{x})}{\|\pi(A)\mathbf{x}\|^2} = \rho(\pi(A)\mathbf{x}; A) - \lambda_i$$

satisfies the inequality

$$\rho(\pi(A)\mathbf{x}; A - \lambda_i I) \leq (\lambda_m - \lambda_i)\left[\frac{\sin\psi}{\cos\varphi}\,\frac{\|\pi(A)\mathbf{h}\|}{\pi(\lambda_i)}\right]^2. \tag{8.13}$$

Proof. With the definitions of $\mathbf{g}$ and $\mathbf{h}$ from above we have

$$\mathbf{x} = U_iU_i^*\mathbf{x} + (I - U_iU_i^*)\mathbf{x} = \cos\psi\,\mathbf{g} + \sin\psi\,\mathbf{h}.$$

which is an orthogonal decomposition. As $\mathcal{R}(U_i)$ is *invariant* under A,

$$\mathbf{s} := \pi(A)\mathbf{x} = \cos\psi\,\pi(A)\mathbf{g} + \sin\psi\,\pi(A)\mathbf{h}$$

is an orthogonal decomposition of $\mathbf{s}$. Thus,

$$\rho(\pi(A)\mathbf{x}; A - \lambda_i I) = \frac{\cos^2\psi\,\mathbf{g}^*(A - \lambda_i I)\mathbf{g} + \sin^2\psi\,\mathbf{h}^*(A - \lambda_i I)\mathbf{h}}{\|\pi(A)\mathbf{x}\|^2}. \tag{8.14}$$

Since $\lambda_1 < \lambda_2 < \cdots < \lambda_m$, we have

(i) $\mathbf{v}^*(A - \lambda_i I)\mathbf{v} \le 0$ for all $\mathbf{v} \in \mathcal{R}(U_i)$,

(ii) $\mathbf{w}^*(A - \lambda_i I)\mathbf{w} \le (\lambda_m - \lambda_i)\|\mathbf{w}\|^2$ for all $\mathbf{w} \in \mathcal{R}(U_i)^\perp$.

Setting $\mathbf{v} = \pi(A)\mathbf{g}$ and $\mathbf{w} = \pi(A)\mathbf{h}$ we obtain from (8.14)

$$\rho(\mathbf{s}; A - \lambda_i I) \le \sin^2\psi\,(\lambda_m - \lambda_i)\frac{\|\pi(A)\mathbf{h}\|^2}{\|\pi(A)\mathbf{x}\|^2}.$$

With

$$\|\mathbf{s}\|^2 = \|\pi(A)\mathbf{x}\|^2 = \sum_{l=1}^{m} \pi^2(\lambda_l)(\mathbf{x}^*\mathbf{u}_l)^2 \ge \pi^2(\lambda_i)\cos^2\varphi$$

we obtain the claim. ■

8.4 Error bounds of Saad

The error bounds to be presented have been published by Saad [2]. We follow the presentation in Parlett [1]. The error bounds for $\vartheta_i^{(j)} - \lambda_i$ are obtained by carefully selecting the polynomial π in Lemma 8.3. Of course we would like $\pi(A)$ to be as small as possible and $\pi(\lambda_i)$ to be as large as possible. First, by the definition of $\mathbf{h}$, we have

$$\begin{aligned}\|\pi(A)\mathbf{h}\|^2 &= \frac{\|\pi(A)(I - U_iU_i^*)\mathbf{x}\|^2}{\|(I - U_iU_i^*)\mathbf{x}\|^2} = \frac{\|\pi(A)\sum_{l=i+1}^m(\mathbf{u}_l^*\mathbf{x})\mathbf{u}_l\|^2}{\|\sum_{l=i+1}^m(\mathbf{u}_l^*\mathbf{x})\mathbf{u}_l\|^2}\\ &= \frac{\sum_{l=i+1}^m(\mathbf{u}_l^*\mathbf{x})^2\pi^2(\lambda_l)}{\sum_{l=i+1}^m(\mathbf{u}_l^*\mathbf{x})^2} \le \max_{i<l\le m}\pi^2(\lambda_l) \le \max_{\lambda_{i+1}\le\lambda\le\lambda_m}\pi^2(\lambda).\end{aligned}$$

The last inequality is important! In this step the search of a maximum in a few selected points $(\lambda_{i+1}, \ldots, \lambda_m)$ is replaced by a search of a maximum in a *whole interval* containing these points. Notice that λ_i is *outside* of this interval. Among all polynomials of a given degree that take a given fixed value $\pi(\lambda_i)$ the Chebyshev polynomial have the smallest maximum. As $\vartheta_i^{(j)}$ is a Ritz value, we know from the monotonicity principle 2.17 that

$$0 \le \vartheta_i^{(j)} - \lambda_i.$$

Further, from the definition of $\vartheta_i^{(j)}$ (as an eigenvalue of A in the subspace $\mathcal{K}^j(\mathbf{x})$),

$$\vartheta_i^{(j)} - \lambda_i \le \rho(\mathbf{s}, A - \lambda_i I) \quad \text{provided that } \mathbf{s} \perp \mathbf{y}_l,\ 1 \le l \le i-1.$$

According to Lemma 8.2 $\mathbf{s} = \pi(A)\mathbf{x}$ is orthogonal on $\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}$, if π has the form

$$\pi(\xi) = (\xi - \vartheta_1^{(j)})\cdots(\xi - \vartheta_{i-1}^{(j)})\omega(\xi), \quad \omega \in \mathbb{P}_{j-i}.$$

With this choice of π we get

$$\frac{\|\pi(A)\mathbf{h}\|}{\pi(\lambda_i)} \le \frac{\|(A - \vartheta_1^{(j)}I)\cdots(A - \vartheta_{i-1}^{(j)}I)\|\cdot\|\omega(A)\mathbf{h}\|}{|(\lambda_i - \vartheta_1^{(j)})|\cdots|(\lambda_i - \vartheta_{i-1}^{(j)})|\cdot|\omega(\lambda_i)|} \le \prod_{l=1}^{i-1}\frac{\lambda_m - \vartheta_l^{(j)}}{\lambda_i - \vartheta_l^{(j)}}\max_{\lambda_{i+1}\le\lambda\le\lambda_m}\frac{\omega(\lambda)}{\omega(\lambda_i)}.$$

This expression should be as small as possible. Now we have

$$\min_{\omega\in\mathbb{P}_{j-1}} \max_{\lambda_{i+1}\le\lambda\le\lambda_m} \frac{|\omega(\lambda)|}{|\omega(\lambda_i)|} = \frac{\max\limits_{\lambda_{i+1}\le\lambda\le\lambda_m} T_{j-i}(\lambda;[\lambda_{i+1},\lambda_m])}{T_{j-i}(\lambda_i;[\lambda_{i+1},\lambda_m])} = \frac{1}{T_{j-i}(\lambda_i;[\lambda_{i+1},\lambda_m])} = \frac{1}{T_{j-i}(1+2\gamma)}, \qquad \gamma = \frac{\lambda_{i+1}-\lambda_i}{\lambda_m-\lambda_{i+1}}.$$

$T_{j-i}(1+2\gamma)$ is the value of the Chebyshev polynomial corresponding to the normal interval $[-1,\ 1]$. The point $1+2\gamma$ is obtained if the affine transformation

$$[\lambda_{i+1},\lambda_m] \ni \lambda \longrightarrow \frac{2\lambda-\lambda_{i+1}-\lambda_m}{\lambda_i-\lambda_{i+1}} \in [-1,\ 1]$$

is applied to λ_i.

Thus we have proved the first part of the following

Theorem 8.4 [2] *Let $\vartheta_1^{(j)},\ldots,\vartheta_j^{(j)}$ be the Ritz values of A in $\mathcal{K}^j(\mathbf{x})$ and let $(\lambda_l,\mathbf{u}_l)$, $l=1,\ldots,m$, be the eigenpairs of A (in $\mathcal{K}^m(\mathbf{x})$). Then for all $i\le j$ we have*

$$0 \le \vartheta_i^{(j)} - \lambda_i \le (\lambda_m-\lambda_i)\left[\frac{\sin\psi}{\cos\varphi}\cdot\frac{\prod_{l=1}^{i-1}\frac{\lambda_m-\vartheta_l^{(j)}}{\lambda_i-\vartheta_l^{(j)}}}{T_{j-i}(1+2\gamma)}\right]^2, \qquad \gamma=\frac{\lambda_{i+1}-\lambda_i}{\lambda_m-\lambda_{i+1}}, \tag{8.15}$$

and

$$\tan\angle(\mathbf{u}_i, \text{projektion of } \mathbf{u}_i \text{ on } \mathcal{K}^j) \le \frac{\sin\psi}{\cos\varphi}\cdot\frac{\prod_{l=1}^{i-1}\frac{\lambda_m-\lambda_l}{\lambda_i-\lambda_l}}{T_{j-i}(1+2\gamma)}. \tag{8.16}$$

Proof. For proving the second part of the Theorem we write

$$\mathbf{x} = \mathbf{g}\cos\angle(\mathbf{x},U_{i-1}U_{i-1}^*\mathbf{x}) + \mathbf{u}_i\cos\underbrace{\angle(\mathbf{x},\mathbf{u}_i)}_{\varphi} + \mathbf{h}\sin\underbrace{\angle(\mathbf{x},U_iU_i^*\mathbf{x})}_{\psi}.$$

We choose π such that $\pi(\lambda_1)=\cdots=\pi(\lambda_{i-1})=0$. Then.

$$\mathbf{s} = \pi(A)\mathbf{x} = \pi(\lambda_i)\mathbf{u}_i\cos\varphi + \pi(A)\mathbf{h}\sin\psi$$

is an orthogonal decomposition of $\mathbf{s}$. By consequence,

$$\tan\angle(\mathbf{s},\mathbf{u}_i) = \frac{\sin\psi\ \|\pi(A)\mathbf{h}\|}{\cos\varphi\ |\pi(\lambda_i)|}.$$

The rest is similar as above. ■

Remark 8.1. Theorem 8.4 does *not* give bounds for the angle between $\angle(\mathbf{u}_i,\mathbf{y}_i)$, an angle that would be more interesting than the abstract angle between $\mathbf{u}_i$ and its projection on $\mathcal{K}^j(\mathbf{x})$. It is possible however to show that [1, p. 246]

$$\sin\angle(\mathbf{u}_i,\mathbf{y}_i) \le \sqrt{1+\frac{\beta_j^2}{{\gamma_i^{(j)}}^2}}\ \sin\angle(\mathbf{u}_i, \text{projection of } \mathbf{u}_i \text{ onto } \mathcal{K}^j)$$

β_j is the number that appeared earlier in the discussion after Lemma 8.2, and

$$\gamma_i^{(j)} = \min_{s \neq i} |\lambda_i - \vartheta_s^{(j)}|$$

□

Theorem 8.4 can easily be rewritten to give error bounds for $\lambda_m - \vartheta_j^{(j)}$, $\lambda_{m-1} - \vartheta_{j-1}^{(j)}$, etc.

We see from this Theorem that the eigenvalues at the beginning and at the end of the spectrum are approximated the quickest. For the first eigenvalue the bound (8.15) simplifies a little,

$$(8.17) \quad 0 \leq \vartheta_1^{(j)} - \lambda_1 \leq (\lambda_m - \lambda_1) \frac{\tan^2 \varphi_1}{T_{j-i}(1+2\gamma_1)^2}, \qquad \gamma_1 = \frac{\lambda_2 - \lambda_1}{\lambda_m - \lambda_2}, \quad \varphi_1 = \angle(\mathbf{x}, \mathbf{u}_1).$$

Analogously, for the largest eigenvalue we have

$$(8.18) \qquad 0 \leq \lambda_m - \vartheta_j^{(j)} \leq (\lambda_m - \lambda_1)\ \tan^2 \varphi_m \frac{1}{T_{j-i}(1+2\gamma_m)^2},$$

with

$$\gamma_m = \frac{\lambda_m - \lambda_{m-1}}{\lambda_{m-1} - \lambda_1}, \quad \text{and} \quad \cos \varphi_m = \mathbf{x}^* \mathbf{u}_m.$$

If the Lanczos algorithmus is applied with $(A - \sigma I)^{-1}$ as with the shifted and inverted vector iteration then we form Krylov spaces $\mathcal{K}^j(\mathbf{x}, (A - \sigma I)^{-1})$. Here the largest eigenvalues are $\frac{1}{\hat{\lambda}_1} \geq \frac{1}{\hat{\lambda}_2} \geq \cdots \geq \frac{1}{\hat{\lambda}_j}$, $\hat{\lambda}_i = \lambda_i - \sigma$.

Eq. (8.18) then becomes

$$0 \leq \frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\vartheta}_j^{(j)}} \leq (\frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\lambda}_j}) \frac{\tan^2 \varphi_1}{T_{j-1}(1+2\hat{\gamma}_1)^2}, \quad \hat{\gamma}_1 = \frac{\frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\lambda}_2}}{\frac{1}{\hat{\lambda}_2} - \frac{1}{\hat{\lambda}_j}}.$$

Now, we have

$$1 + 2\hat{\gamma}_1 = 2(1+\hat{\gamma}_1) - 1 = 2\left(\frac{\frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\lambda}_j}}{\frac{1}{\hat{\lambda}_2} - \frac{1}{\hat{\lambda}_j}}\right) - 1 = 2\frac{\hat{\lambda}_2}{\hat{\lambda}_1} \underbrace{\left(\frac{1 - \frac{\hat{\lambda}_1}{\hat{\lambda}_j}}{1 - \frac{\hat{\lambda}_2}{\hat{\lambda}_j}}\right)}_{>1} - 1 \geq 2\frac{\hat{\lambda}_2}{\hat{\lambda}_1} - 1 > 1.$$

Since $|T_{j-1}(\xi)|$ grows rapidly and monotonically outside $[-1, 1]$ we have

$$T_{j-1}(1+2\hat{\gamma}_1) \geq T_{j-1}(2\frac{\hat{\lambda}_2}{\hat{\lambda}_1} - 1),$$

and thus

$$(8.19) \qquad \frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\vartheta}_1^{(j)}} \leq c_1 \left(\frac{1}{T_{j-1}(2\frac{\hat{\lambda}_2}{\hat{\lambda}_1} - 1)}\right)^2$$

With the simple inverse vector iteration we had

$$(8.20) \qquad \frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\lambda}_1^{(j)}} \leq c_2 \left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2}\right)^{2(j-1)}$$

In Table 8.2 the numbers

$$\left(\frac{1}{T_{j-1}(2\frac{\hat{\lambda}_2}{\hat{\lambda}_1}-1)}\right)^2$$

are compared with

$$\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2}\right)^{2(j-1)}$$

for $\hat{\lambda}_2/\hat{\lambda}_1 = 2,\ 1.1,\ 1.01$. If this ratio is large both methods quickly provide the desired results. If however the ratio tends to 1 then a method that computes the eigenvalues by means of Ritz values of Krylov spaces shows an acceptable convergence behaviour whereas vector iteration hardly improves with j. Remebre that j is the number of matrix-vector multiplications have been executed, or, with the shift-and-invert spectral transformation, how many systems of equations have been solved.

$\hat{\lambda}_2/\hat{\lambda}_1$	$j=5$	$j=10$	$j=15$	$j=20$	$j=25$
2.0	$\frac{3.0036e-06}{3.9063e-03}$	$\frac{6.6395e-14}{3.8147e-06}$	$\frac{1.4676e-21}{3.7253e-09}$	$\frac{3.2442e-29}{3.6380e-12}$	$\frac{7.1712e-37}{3.5527e-15}$
1.1	$\frac{2.7152e-02}{4.6651e-01}$	$\frac{5.4557e-05}{1.7986e-01}$	$\frac{1.0814e-07}{6.9343e-02}$	$\frac{2.1434e-10}{2.6735e-02}$	$\frac{4.2482e-13}{1.0307e-02}$
1.01	$\frac{5.6004e-01}{9.2348e-01}$	$\frac{1.0415e-01}{8.3602e-01}$	$\frac{1.4819e-02}{7.5684e-01}$	$\frac{2.0252e-03}{6.8515e-01}$	$\frac{2.7523e-04}{6.2026e-01}$

Table 8.2: Ratio $\frac{(1/T_{j-1}(2\hat{\lambda}_2/\hat{\lambda}_1-1))^2}{(\hat{\lambda}_1/\hat{\lambda}_2)^{2(j-1)}}$ for varying j and ratios $\hat{\lambda}_2/\hat{\lambda}_1$.

Bibliography

[1] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, NJ, 1980. (Republished by SIAM, Philadelphia, 1998.).

[2] Y. Saad, *On the rates of convergence of the Lanczos and the block Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.