

24.6 Krylov Subspaces

Sections 4 and 5 both consist of multiplying by $\mathbf{A}$ over and over: the power method forms $\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots$ and keeps only the newest vector, and the QR algorithm, by Theorem 24.5.2, does the same thing to $\mathbf{e}_1$ inside its first column. In both cases everything computed before the last step is thrown away, and this section keeps it instead. The span of the vectors produced by k matrix-vector products is a subspace of dimension up to k , and the best approximations it contains are far better than its last vector alone. That subspace, the process that builds an orthonormal basis of it, and the eigenvalue estimates it yields are the subject here.

Throughout, $F = \mathbb{R}$ or $F = \mathbb{C}$, and F^n carries the standard inner product, linear in the first slot, with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$ and $\|\cdot\| = \|\cdot\|_2$.

24.6.1 The space that k products span

A large matrix is often not available as a table of n^2 numbers at all. What is available is a rule for computing $\mathbf{A}\mathbf{x}$ from $\mathbf{x}$ — a sparse pattern of neighbors on a grid, a graph, a differential operator discretized. For such a matrix, forming $\mathbf{A}\mathbf{x}$ is cheap and forming $\mathbf{A}^{-1}$, or even looking at an entry of $\mathbf{A}^2$, is not. So the honest question is: after k multiplications, what do we possess?

We possess $\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}$, and hence every linear combination of them. The power method uses one of these vectors. That is wasteful, and the waste has a name only because the span deserves one.

The Krylov space is everything that k matrix-vector products can reach.

Definition 24.6.1. Krylov Subspace

Let $\mathbf{A} \in M_n(F)$, let $\mathbf{b} \in F^n$, and let $k \geq 1$ be an integer. The k -th **Krylov subspace** of the pair $(\mathbf{A}, \mathbf{b})$ is

$$\mathcal{K}_k(\mathbf{A}, \mathbf{b}) := \text{span}(\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}) \subseteq F^n.$$

In words: we list the first k vectors of the sequence obtained by repeatedly applying $\mathbf{A}$ to $\mathbf{b}$ — **starting at $\mathbf{b}$ itself**, so that the last one is $\mathbf{A}^{k-1}\mathbf{b}$ and not $\mathbf{A}^k\mathbf{b}$ — and take their span. It is a subspace for the usual reason, being a span. The index k counts **vectors**, not powers; the off-by-one is a convention, and it is the one that makes $\dim \mathcal{K}_k \leq k$ with equality in the typical case.

Examples. Take $F = \mathbb{R}$ throughout.

- **The identity.** For $\mathbf{A} = \mathbf{I}_n$ and any $\mathbf{b} \neq \mathbf{0}$, every $\mathbf{A}^j\mathbf{b}$ equals $\mathbf{b}$, so $\mathcal{K}_k(\mathbf{I}_n, \mathbf{b}) = \text{span}(\mathbf{b})$ for every k . One product buys nothing, and neither do a hundred.
- **A companion matrix.** Let $\mathbf{A}\mathbf{e}_j = \mathbf{e}_{j+1}$ for $j \leq n-1$, as for the companion matrix of Definition 10.5.3, and let $\mathbf{b} = \mathbf{e}_1$. Then $\mathbf{A}^j\mathbf{e}_1 = \mathbf{e}_{j+1}$ for $j \leq n-1$, so $\mathcal{K}_k(\mathbf{A}, \mathbf{e}_1) = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_k)$ for $k \leq n$, of dimension exactly k . Here every product buys a new dimension.
- **A diagonal matrix.** Let $\mathbf{A} = \text{diag}(4, 3, 2, 1)$ and $\mathbf{b} = (1, 1, 1, 1)$. Then $\mathbf{A}^j\mathbf{b} = (4^j, 3^j, 2^j, 1)$, and these four vectors for $j = 0, 1, 2, 3$ form a Vandermonde matrix with distinct nodes, hence are independent (Theorem 7.8.1). So $\dim \mathcal{K}_k = k$ for $k \leq 4$, and $\mathcal{K}_4 = \mathbb{R}^4$.
- **Degenerate cases.** $\mathcal{K}_1(\mathbf{A}, \mathbf{b}) = \text{span}(\mathbf{b})$; and $\mathcal{K}_k(\mathbf{A}, \mathbf{0}) = \{\mathbf{0}\}$ for every k , which is why

$\mathbf{b} \neq \mathbf{0}$ is assumed from here on.

Nonexample by minimal change. Drop the middle vector of the list. Let $\mathbf{A} \in M_3(\mathbb{R})$ satisfy $\mathbf{A}\mathbf{e}_1 = \mathbf{e}_2$, $\mathbf{A}\mathbf{e}_2 = \mathbf{e}_3$, $\mathbf{A}\mathbf{e}_3 = \mathbf{0}$, and let $\mathbf{b} = \mathbf{e}_1$. The set $\text{span}(\mathbf{b}, \mathbf{A}^2\mathbf{b}) = \text{span}(\mathbf{e}_1, \mathbf{e}_3)$ is a perfectly good subspace, and it is not $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ for any k , since those are $\text{span}(\mathbf{e}_1)$, $\text{span}(\mathbf{e}_1, \mathbf{e}_2)$, and $\mathbb{R}^3$. The clause that fails is that the powers be **consecutive**, and what it costs is everything: $\mathbf{A}\text{span}(\mathbf{e}_1, \mathbf{e}_3) = \text{span}(\mathbf{e}_2)$, which meets $\text{span}(\mathbf{e}_1, \mathbf{e}_3)$ only at $\mathbf{0}$. The Krylov spaces, by contrast, are nested and almost invariant, which is the next proposition.

Why this definition. Chapter 10 already met the union of all of them. The cyclic subspace $Z(\mathbf{b}; \mathbf{A}) = \text{span}(\mathbf{b}, \mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \dots)$ of Definition 10.5.1 is the smallest $\mathbf{A}$ -invariant subspace containing $\mathbf{b}$ (Proposition 10.5.1), and Theorem 10.5.2 says that the list stops growing exactly at the degree of the $\mathbf{A}$ -annihilator $m_{\mathbf{A}, \mathbf{b}}$. The Krylov space is that construction **truncated at k** , and the truncation is the whole point: $Z(\mathbf{b}; \mathbf{A})$ is a theoretical object, while $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ is what k units of work actually deliver.

Proposition 24.6.1. Krylov Spaces: Nesting, Dimension, Invariance

Let $\mathbf{A} \in M_n(F)$ and $\mathbf{b} \in F^n$ with $\mathbf{b} \neq \mathbf{0}$, and let $d = \deg m_{\mathbf{A}, \mathbf{b}} \geq 1$ be the degree of the $\mathbf{A}$ -annihilator of $\mathbf{b}$ (Definition 10.5.2). Then, for every $k \geq 1$:

- (a) $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \{p(\mathbf{A})\mathbf{b} : p \in F[x], \deg p \leq k-1\}$, and $\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$ and $\mathbf{A}\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$;
- (b) $\dim \mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \min(k, d)$;
- (c) $\mathcal{K}_{k+1} = \mathcal{K}_k$ if and only if $k \geq d$, and in that case $\mathcal{K}_k = Z(\mathbf{b}; \mathbf{A})$ is $\mathbf{A}$ -invariant.

Idea. Everything is Theorem 10.5.2 read one index at a time. That theorem says $(\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{d-1}\mathbf{b})$ is a basis of $Z(\mathbf{b}; \mathbf{A})$; so the first d vectors of our list are independent, and from $\mathbf{A}^d\mathbf{b}$ on they add nothing.

Proof.

- (a) A linear combination $\sum_{j < k} a_j \mathbf{A}^j \mathbf{b}$ is $p(\mathbf{A})\mathbf{b}$ for $p = \sum_{j < k} a_j x^j$, and conversely, which is the first statement. The inclusion $\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$ holds because the spanning list of $\mathcal{K}_k$ is part of that of $\mathcal{K}_{k+1}$. And $\mathbf{A}\mathbf{A}^j \mathbf{b} = \mathbf{A}^{j+1} \mathbf{b} \in \mathcal{K}_{k+1}$ for $j \leq k-1$, so $\mathbf{A}$ maps a spanning list of $\mathcal{K}_k$ into $\mathcal{K}_{k+1}$, hence maps $\mathcal{K}_k$ into $\mathcal{K}_{k+1}$.
- (b) By Theorem 10.5.2 (a), the list $(\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{d-1}\mathbf{b})$ is a basis of $Z(\mathbf{b}; \mathbf{A})$. If $k \leq d$, the first k entries of that list are a sublist of a basis, hence independent, so $\dim \mathcal{K}_k = k$. If $k > d$, then $\mathcal{K}_k \subseteq Z(\mathbf{b}; \mathbf{A})$, because every $\mathbf{A}^j \mathbf{b}$ lies there (Proposition 10.5.1 (a)); and $\mathcal{K}_k \supseteq \mathcal{K}_d = Z(\mathbf{b}; \mathbf{A})$, since $\mathcal{K}_d$ already contains a basis of it. So $\mathcal{K}_k = Z(\mathbf{b}; \mathbf{A})$ and $\dim \mathcal{K}_k = d$. In both cases $\dim \mathcal{K}_k = \min(k, d)$.
- (c) ($\Leftarrow$) If $k \geq d$, then (b) gives $\dim \mathcal{K}_{k+1} = d = \dim \mathcal{K}_k$, and with $\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$ from (a) this forces equality; and the common space is $Z(\mathbf{b}; \mathbf{A})$, $\mathbf{A}$ -invariant by Proposition 10.5.1 (b). ($\Rightarrow$) If $k < d$, then (b) gives $\dim \mathcal{K}_{k+1} = k+1 > k = \dim \mathcal{K}_k$, so the two differ. This proves the proposition. ■

Warning. The list $\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}$ is a basis of $\mathcal{K}_k$ and is useless as one. By Section 4's reading of the power method (Theorem 24.4.1), $\mathbf{A}^j\mathbf{b}$ turns towards the dominant eigenvector as j grows, so the later vectors of the list are nearly parallel to each other. Even for $\mathbf{A} = \text{diag}(4, 3, 2, 1)$ and $\mathbf{b} = (1, 1, 1, 1)$, with its modest ratio $\lambda_1/\lambda_2 = 4/3$, the angle between consecutive vectors of the list falls from 24 degrees to 8.9 degrees across those four vectors, and the matrix with those four columns has 2-norm condition number about 1200 — for a 4×4 problem with one-digit data. Independent they may be; in the arithmetic of Definition 24.1.1 the coordinates of a vector in that basis are computed with errors magnified by that condition number, which grows quickly with k . **A Krylov method never uses this basis.** It builds an orthonormal one instead, and that is the next subsection.

Quick check. Let $\mathbf{A} \in M_n(F)$ be invertible with $\mathbf{A}^2 = \mathbf{I}_n$, and let $\mathbf{b} \neq \mathbf{0}$ be a vector that is not an eigenvector of $\mathbf{A}$. What is $\dim \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ for each $k \geq 1$?

Solution. It is 1 for $k = 1$ and 2 for every $k \geq 2$. The polynomial $x^2 - 1$ kills $\mathbf{b}$, so $m_{\mathbf{A}, \mathbf{b}}$ divides it and has degree at most 2 (Definition 10.5.2). Degree 1 would mean $m_{\mathbf{A}, \mathbf{b}} = x - \lambda$, that is $\mathbf{A}\mathbf{b} = \lambda\mathbf{b}$, making $\mathbf{b}$ an eigenvector, which is excluded; and $\deg m_{\mathbf{A}, \mathbf{b}} \geq 1$ because $\mathbf{b} \neq \mathbf{0}$. So $d = 2$, and Proposition 24.6.1 (b) gives $\min(k, 2)$.

24.6.2 Arnoldi: Gram–Schmidt with a purpose

We want an orthonormal basis $\mathbf{q}_1, \dots, \mathbf{q}_k$ of $\mathcal{K}_k$, with the nesting preserved: $\text{span}(\mathbf{q}_1, \dots, \mathbf{q}_j) = \mathcal{K}_j$ for every $j \leq k$. Running Theorem 11.2.2 on the list $\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b}$ would do it, and would inherit exactly the defect the warning describes, since it needs those vectors first.

The repair is one line long, and it is the idea of the whole subsection. Do not orthogonalize $\mathbf{A}^j\mathbf{b}$. Orthogonalize $\mathbf{A}\mathbf{q}_j$. The vector $\mathbf{A}\mathbf{q}_j$ lies in $\mathcal{K}_{j+1}$ by Proposition 24.6.1 (a), and subtracting its components along $\mathbf{q}_1, \dots, \mathbf{q}_j$ leaves a vector in $\mathcal{K}_{j+1}$ orthogonal to $\mathcal{K}_j$ — which is what the next basis vector must be. The powers of $\mathbf{A}$ are never formed.

Algorithm. **The Arnoldi process.** Given $\mathbf{A} \in M_n(F)$ and $\mathbf{b} \neq \mathbf{0}$, set $\mathbf{q}_1 = \mathbf{b}/\|\mathbf{b}\|$. For $j = 1, 2, \dots$:

$$h_{ij} = \langle \mathbf{A}\mathbf{q}_j, \mathbf{q}_i \rangle \quad (1 \leq i \leq j), \quad \mathbf{w}_j = \mathbf{A}\mathbf{q}_j - \sum_{i=1}^j h_{ij}\mathbf{q}_i,$$

$$h_{j+1,j} = \|\mathbf{w}_j\|, \quad \mathbf{q}_{j+1} = \mathbf{w}_j/h_{j+1,j} \quad \text{if } h_{j+1,j} \neq 0.$$

If $h_{j+1,j} = 0$ the process stops; this is called a **breakdown**, and the theorem below says it is good news.

Write $\mathbf{Q}_j = (\mathbf{q}_1 \mid \dots \mid \mathbf{q}_j) \in M_{n \times j}(F)$, let $\mathbf{H}_k \in M_k(F)$ have entries h_{ij} for $i \leq j+1$ and zeros elsewhere, and let $\tilde{\mathbf{H}}_k \in M_{(k+1) \times k}(F)$ be $\mathbf{H}_k$ with the extra row $h_{k+1,k}\mathbf{e}_k^\top$ appended at the bottom.

Theorem 24.6.2. The Arnoldi Process

Let $\mathbf{A} \in M_n(F)$, let $\mathbf{b} \in F^n$ be nonzero, let $d = \deg m_{\mathbf{A}, \mathbf{b}}$, and run the Arnoldi process. Then the **Arnoldi relation**

$$\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k\mathbf{H}_k + h_{k+1,k} \mathbf{q}_{k+1} \mathbf{e}_k^\top \quad (1 \leq k \leq d) \quad (24.6.1)$$

holds, the last term being $\mathbf{0}$ when $k = d$. Moreover:

- (a) For every $j \leq d$, the list $(\mathbf{q}_1, \dots, \mathbf{q}_j)$ is an orthonormal basis of $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$; moreover $h_{j+1,j} > 0$ for $j < d$, and $h_{d+1,d} = 0$, so the process breaks down at step d and at no earlier step.
- (b) For $1 \leq k \leq d$ we have $\mathbf{Q}_k^* \mathbf{Q}_k = \mathbf{I}_k$, and $\mathbf{H}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k$ is **upper Hessenberg**. For $k < d$, Eq. 24.6.1 can also be written $\mathbf{A} \mathbf{Q}_k = \mathbf{Q}_{k+1} \tilde{\mathbf{H}}_k$.
- (c) A breakdown means an invariant subspace: if $h_{k+1,k} = 0$, then $\mathbf{A} \mathcal{K}_k \subseteq \mathcal{K}_k$, and every eigenvalue of $\mathbf{H}_k$ is an eigenvalue of $\mathbf{A}$.

Idea. Only one thing has to be checked, and it is the nesting: if $\mathbf{q}_1, \dots, \mathbf{q}_j$ is an orthonormal basis of $\mathcal{K}_j$, then $\mathbf{A} \mathbf{q}_j$ lies in $\mathcal{K}_{j+1}$ but generally not in $\mathcal{K}_j$, and what the step subtracts is exactly its $\mathcal{K}_j$ -component. The matrix identity is then just the definition of $\mathbf{w}_j$ rearranged, column by column, and the Hessenberg shape is the statement that $\mathbf{A} \mathbf{q}_j$ needs only $\mathbf{q}_1, \dots, \mathbf{q}_{j+1}$ — never $\mathbf{q}_{j+2}$ — because $\mathbf{A} \mathcal{K}_j \subseteq \mathcal{K}_{j+1}$.

Proof.

- (a) Induction on j . For $j = 1$, $\mathbf{q}_1 = \mathbf{b}/\|\mathbf{b}\|$ is a unit vector spanning $\mathcal{K}_1 = \text{span}(\mathbf{b})$. Let $1 \leq j$ and suppose $(\mathbf{q}_1, \dots, \mathbf{q}_j)$ is an orthonormal basis of $\mathcal{K}_j$. The subtraction in the process makes $\mathbf{w}_j$ orthogonal to each $\mathbf{q}_i$ with $i \leq j$: for such i ,

$$\langle \mathbf{w}_j, \mathbf{q}_i \rangle = \langle \mathbf{A} \mathbf{q}_j, \mathbf{q}_i \rangle - \sum_{l=1}^j h_{lj} \langle \mathbf{q}_l, \mathbf{q}_i \rangle = h_{ij} - h_{ij} = 0,$$

using orthonormality. Also $\mathbf{A} \mathbf{q}_j \in \mathbf{A} \mathcal{K}_j \subseteq \mathcal{K}_{j+1}$ by Proposition 24.6.1 (a), and each $\mathbf{q}_i \in \mathcal{K}_j \subseteq \mathcal{K}_{j+1}$, so $\mathbf{w}_j \in \mathcal{K}_{j+1}$.

Claim. $\mathcal{K}_{j+1} = \mathcal{K}_j + \text{span}(\mathbf{A} \mathbf{q}_j)$.

Proof. $(\supseteq)$ Both summands lie in $\mathcal{K}_{j+1}$, by Proposition 24.6.1 (a). $(\subseteq)$ Since $\mathcal{K}_{j+1} = \mathcal{K}_j + \text{span}(\mathbf{A}^j \mathbf{b})$, it is enough to place $\mathbf{A}^j \mathbf{b}$ on the right. Write $\mathbf{A}^{j-1} \mathbf{b} = \sum_{i \leq j} c_i \mathbf{q}_i$, possible because $\mathbf{A}^{j-1} \mathbf{b} \in \mathcal{K}_j$ and the $\mathbf{q}_i$ are a basis of it. Applying $\mathbf{A}$,

$$\mathbf{A}^j \mathbf{b} = \sum_{i \leq j} c_i \mathbf{A} \mathbf{q}_i,$$

and for $i \leq j-1$ we have $\mathbf{A} \mathbf{q}_i \in \mathbf{A} \mathcal{K}_i \subseteq \mathcal{K}_{i+1} \subseteq \mathcal{K}_j$. So only the term $c_j \mathbf{A} \mathbf{q}_j$ leaves $\mathcal{K}_j$. ■

Now $\mathbf{w}_j = \mathbf{0}$ if and only if $\mathbf{A} \mathbf{q}_j \in \mathcal{K}_j$, and by the Claim that happens if and only if $\mathcal{K}_{j+1} = \mathcal{K}_j$. Suppose first that $j < d$. Then $\dim \mathcal{K}_{j+1} = j+1 > j = \dim \mathcal{K}_j$ by Proposition 24.6.1 (b),

so $\mathcal{K}_{j+1} \neq \mathcal{K}_j$ and $\mathbf{w}_j \neq \mathbf{0}$. Hence $h_{j+1,j} > 0$, and $\mathbf{q}_{j+1} = \mathbf{w}_j/h_{j+1,j}$ is a unit vector in $\mathcal{K}_{j+1}$ orthogonal to $\mathbf{q}_1, \dots, \mathbf{q}_j$. An orthogonal list of nonzero vectors is independent, so $(\mathbf{q}_1, \dots, \mathbf{q}_{j+1})$ is an orthonormal list of $j+1$ vectors in the $(j+1)$ -dimensional space $\mathcal{K}_{j+1}$, hence a basis of it.

If instead $j = d$, then $\mathcal{K}_{d+1} = \mathcal{K}_d$ by [Proposition 24.6.1](#) (c), so $\mathbf{A}\mathbf{q}_d \in \mathbf{A}\mathcal{K}_d \subseteq \mathcal{K}_{d+1} = \mathcal{K}_d$ and $\mathbf{w}_d = \mathbf{0}$. This gives (a).

(b), and [Eq. 24.6.1](#). The columns of $\mathbf{Q}_k$ are orthonormal by (a), which is $\mathbf{Q}_k^* \mathbf{Q}_k = \mathbf{I}_k$. Rearranging the definition of $\mathbf{w}_j$,

$$\mathbf{A}\mathbf{q}_j = \sum_{i=1}^j h_{ij} \mathbf{q}_i + h_{j+1,j} \mathbf{q}_{j+1} = \sum_{i=1}^{j+1} h_{ij} \mathbf{q}_i,$$

for each $j \leq k$, where for $j = d$ the last term is $\mathbf{0}$. For $j < k$ the right-hand side is $\mathbf{Q}_k \mathbf{H}_k \mathbf{e}_j$, since column j of $\mathbf{H}_k$ has entries $h_{1j}, \dots, h_{j+1,j}$ and zeros below and $j+1 \leq k$; for $j = k$ it is $\mathbf{Q}_k \mathbf{H}_k \mathbf{e}_k + h_{k+1,k} \mathbf{q}_{k+1}$. Collecting the columns gives [Eq. 24.6.1](#), and the form $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1} \tilde{\mathbf{H}}_k$ is the same identity with the extra column of $\mathbf{Q}_{k+1}$ and the extra row of $\tilde{\mathbf{H}}_k$ absorbing the last term. The entries h_{ij} with $i > j+1$ are zero by definition, which is exactly upper Hessenberg form. Finally, multiplying [Eq. 24.6.1](#) on the left by $\mathbf{Q}_k^*$ and using $\mathbf{Q}_k^* \mathbf{Q}_k = \mathbf{I}_k$ and $\mathbf{Q}_k^* \mathbf{q}_{k+1} = \mathbf{0}$,

$$\mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k = \mathbf{H}_k + h_{k+1,k} \mathbf{Q}_k^* \mathbf{q}_{k+1} \mathbf{e}_k^\top = \mathbf{H}_k.$$

- (c) If $h_{k+1,k} = 0$, then [Eq. 24.6.1](#) reads $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k \mathbf{H}_k$, so $\mathbf{A}\mathbf{q}_j \in \mathcal{K}_k$ for every $j \leq k$ and hence $\mathbf{A}\mathcal{K}_k \subseteq \mathcal{K}_k$. If $\mathbf{H}_k \mathbf{y} = \theta \mathbf{y}$ with $\mathbf{y} \neq \mathbf{0}$, put $\mathbf{x} = \mathbf{Q}_k \mathbf{y}$, which is nonzero because the columns of $\mathbf{Q}_k$ are independent; then $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{Q}_k \mathbf{y} = \mathbf{Q}_k \mathbf{H}_k \mathbf{y} = \theta \mathbf{Q}_k \mathbf{y} = \theta \mathbf{x}$. This proves the theorem. ■

The cost of k steps is k matrix-vector products plus the orthogonalizations, which use about $4nj$ flops at step j — j inner products and j subtractions, about $2n$ flops each — so $O(nk^2)$ in total; and all $k+1$ vectors must be stored. Both grow with k , which is why Arnoldi is restarted in practice. For Hermitian $\mathbf{A}$ neither grows, and that is the next theorem.

24.6.3 Lanczos: the recurrence collapses to three terms

Theorem 24.6.3. The Lanczos Recurrence

Let $\mathbf{A} \in M_n(\mathbb{C})$ be **Hermitian**, let $\mathbf{b} \neq \mathbf{0}$, and run the Arnoldi process. Then $\mathbf{H}_k$ is Hermitian for every $k \leq d$, hence **tridiagonal**, with real diagonal entries; writing $\alpha_j = h_{jj} \in \mathbb{R}$ and $\beta_j = h_{j+1,j}$, which is > 0 for $j < d$ and 0 for $j = d$, we have $h_{j,j+1} = \beta_j$ and

$$\mathbf{A}\mathbf{q}_j = \beta_{j-1} \mathbf{q}_{j-1} + \alpha_j \mathbf{q}_j + \beta_j \mathbf{q}_{j+1} \quad (1 \leq j \leq d), \quad (24.6.2)$$

with $\beta_0 \mathbf{q}_0$ read as $\mathbf{0}$ and $\beta_d \mathbf{q}_{d+1}$ as $\mathbf{0}$. Consequently $\mathbf{H}_k = \mathbf{T}_k$ is the real symmetric tridiagonal

matrix with diagonal $(\alpha_1, \dots, \alpha_k)$ and off-diagonal $(\beta_1, \dots, \beta_{k-1})$, and each step of the process needs only $\mathbf{q}_{j-1}$ and $\mathbf{q}_j$.

Idea. Two facts collide. The matrix $\mathbf{H}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k$ is Hermitian because $\mathbf{A}$ is, and it is upper Hessenberg by Theorem 24.6.2. A Hermitian matrix with zeros below the subdiagonal has zeros above the superdiagonal too, by conjugating. Everything else is bookkeeping. The reason is worth naming separately: $h_{ij} = \langle \mathbf{A} \mathbf{q}_j, \mathbf{q}_i \rangle$ lets $\mathbf{A}$ move to the other slot for free, and $\mathbf{A} \mathbf{q}_i$ lives one index further along than $\mathbf{q}_i$ does.

Proof. By Theorem 24.6.2 (b), $\mathbf{H}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k$, so $\mathbf{H}_k^* = \mathbf{Q}_k^* \mathbf{A}^* \mathbf{Q}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k = \mathbf{H}_k$ and $\mathbf{H}_k$ is Hermitian; in particular its diagonal entries are real.

Let $i \leq j - 2$. Since $\mathbf{A}^* = \mathbf{A}$ is self-adjoint,

$$h_{ij} = \langle \mathbf{A} \mathbf{q}_j, \mathbf{q}_i \rangle = \langle \mathbf{q}_j, \mathbf{A} \mathbf{q}_i \rangle.$$

Now $\mathbf{q}_i \in \mathcal{K}_i$, so $\mathbf{A} \mathbf{q}_i \in \mathcal{K}_{i+1} \subseteq \mathcal{K}_{j-1}$ by Proposition 24.6.1 (a) and $i + 1 \leq j - 1$, while $\mathbf{q}_j$ is orthogonal to $\mathcal{K}_{j-1}$ by Theorem 24.6.2 (a). Hence $h_{ij} = 0$ whenever $i \leq j - 2$. Together with $h_{ij} = 0$ for $i > j + 1$, this leaves only $i \in \{j - 1, j, j + 1\}$, so $\mathbf{H}_k$ is tridiagonal. Being Hermitian, $h_{j,j+1} = \overline{h_{j+1,j}} = \beta_j$, which is real and positive. The displayed recurrence Eq. 24.6.2 is then the expansion $\mathbf{A} \mathbf{q}_j = \sum_i h_{ij} \mathbf{q}_i$ of Theorem 24.6.2 (b) with only three surviving terms; and $\mathbf{H}_k$ has real entries, is symmetric and is tridiagonal, as claimed. Since the step needs only $h_{j-1,j} = \beta_{j-1}$, $h_{jj} = \alpha_j$, and $\mathbf{q}_{j-1}, \mathbf{q}_j$, this proves the theorem. ■

This is Chapter 11's three-term recurrence, and for the same reason. Chapter 11 §10 built the monic orthogonal polynomials for an inner product on $\mathbb{R}[x]$ and found (Theorem 11.10.3) that $p_{k+1} = (x - a_k)p_k - b_k p_{k-1}$, with only three terms; the reason given there was Proposition 11.10.1, “multiplication by x moves from one slot to the other for free”. The two statements are the same statement. Given Hermitian $\mathbf{A}$ and $\mathbf{b} \neq \mathbf{0}$, define on $\mathbb{R}[x]$

$$\langle p, q \rangle_{\mathbf{b}} := \langle p(\mathbf{A})\mathbf{b}, q(\mathbf{A})\mathbf{b} \rangle.$$

This is symmetric and bilinear, and on $\mathbb{R}[x]_{\leq d-1}$ it is positive definite: $\langle p, p \rangle_{\mathbf{b}} = \|p(\mathbf{A})\mathbf{b}\|^2$ vanishes only when $p(\mathbf{A})\mathbf{b} = \mathbf{0}$, which by Definition 10.5.2 means $m_{\mathbf{A},\mathbf{b}} \mid p$, impossible for a nonzero p of degree below d . Multiplication by x is symmetric for it, since $\mathbf{A}$ is Hermitian:

$$\langle xp, q \rangle_{\mathbf{b}} = \langle \mathbf{A} p(\mathbf{A})\mathbf{b}, q(\mathbf{A})\mathbf{b} \rangle = \langle p(\mathbf{A})\mathbf{b}, \mathbf{A} q(\mathbf{A})\mathbf{b} \rangle = \langle p, xq \rangle_{\mathbf{b}}.$$

Now Proposition 24.6.1 (a) says $\mathbf{q}_{j+1}$ is $r(\mathbf{A})\mathbf{b}$ for some r of degree at most j , and orthogonality to $\mathcal{K}_j$ says $\langle r, s \rangle_{\mathbf{b}} = 0$ for every s of degree below j . So for $j \leq d - 1$ the polynomial r is a scalar multiple of the j -th monic orthogonal polynomial p_j for $\langle \cdot, \cdot \rangle_{\mathbf{b}}$, and Eq. 24.6.2, at those indices, is the recurrence of Theorem 11.10.3 rewritten in terms of the vectors $p_j(\mathbf{A})\mathbf{b}$ — read, as that theorem's node inner product is read, only in the range of degrees where the form is positive definite. The last step $j = d$ lies outside that range: there the monic polynomial of degree d orthogonal to everything of lower degree is $m_{\mathbf{A},\mathbf{b}}$ itself, and $\langle m_{\mathbf{A},\mathbf{b}}, m_{\mathbf{A},\mathbf{b}} \rangle_{\mathbf{b}} = 0$. A third appearance of the same three terms is Chapter 7's: by Theorem 7.8.3, the characteristic polynomials $D_k(x) = \det(x\mathbf{I}_k - \mathbf{T}_k)$ of the leading blocks satisfy

$$D_k = (x - \alpha_k)D_{k-1} - \beta_{k-1}^2 D_{k-2},$$

which is Theorem 11.10.3 with $a_{k-1} = \alpha_k$ and $b_{k-1} = \beta_{k-1}^2 > 0$. The Lanczos coefficients, the orthogonal polynomials, and the determinants of the tridiagonal blocks are one object seen three ways.

Example. A breakdown at step two

Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ and $\mathbf{b} = \mathbf{e}_1$. Run the Lanczos recurrence of Theorem 24.6.3, and identify what it finds.

Solution. $\mathbf{q}_1 = \mathbf{e}_1$, and $\mathbf{A}\mathbf{q}_1 = (3, 1, 1)$. So $\alpha_1 = \langle \mathbf{A}\mathbf{q}_1, \mathbf{q}_1 \rangle = 3$ and

$$\mathbf{w}_1 = (3, 1, 1) - 3(1, 0, 0) = (0, 1, 1), \quad \beta_1 = \sqrt{2}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}}(0, 1, 1).$$

Next, $\mathbf{A}\mathbf{q}_2 = \frac{1}{\sqrt{2}}(2, 4, 4)$, so $\alpha_2 = \langle \mathbf{A}\mathbf{q}_2, \mathbf{q}_2 \rangle = \frac{1}{2}(0 + 4 + 4) = 4$ and

$$\mathbf{w}_2 = \frac{1}{\sqrt{2}}(2, 4, 4) - 4 \cdot \frac{1}{\sqrt{2}}(0, 1, 1) - \sqrt{2}(1, 0, 0) = \frac{1}{\sqrt{2}}(2, 0, 0) - \sqrt{2}(1, 0, 0) = \mathbf{0}.$$

The process breaks down at step 2, so $d = 2$ and $\mathcal{K}_2$ is $\mathbf{A}$ -invariant, by Theorem 24.6.2 (a), (c). Then

$$\mathbf{T}_2 = \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 4 \end{pmatrix}, \quad p_{\mathbf{T}_2}(x) = (x - 3)(x - 4) - 2 = x^2 - 7x + 10,$$

with roots 5 and 2. And indeed $\text{spec}(\mathbf{A}) = \{5, 2\}$, with 5 simple and 2 of multiplicity 2: the two steps have produced both eigenvalues exactly. This is Theorem 24.6.2 (c) in action, and it is also Proposition 24.6.1 (b) with $d = \deg m_{\mathbf{A}, \mathbf{e}_1} = 2$, since $m_{\mathbf{A}} = (x - 5)(x - 2)$ and neither factor alone kills $\mathbf{e}_1$.

24.6.4 Ritz values

An invariant subspace would give eigenvalues exactly. A Krylov space is an **almost** invariant subspace — by Eq. 24.6.1 it fails to be invariant only through the single vector $h_{k+1,k}\mathbf{q}_{k+1}$ — so it should give almost-eigenvalues, and it should be possible to say how almost.

Definition 24.6.2. Ritz Values and Ritz Vectors

Let $\mathbf{A} \in M_n(F)$, let $\mathbf{Q}_k \in M_{n \times k}(F)$ have orthonormal columns, and put $\mathbf{H}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k$. The eigenvalues $\theta_1, \dots, \theta_k$ of $\mathbf{H}_k$ are the **Ritz values** of $\mathbf{A}$ from the subspace $\text{span}(\mathbf{q}_1, \dots, \mathbf{q}_k)$, and for an eigenvector $\mathbf{y}$ of $\mathbf{H}_k$ with $\mathbf{H}_k \mathbf{y} = \theta \mathbf{y}$ and $\|\mathbf{y}\| = 1$, the vector $\mathbf{x} = \mathbf{Q}_k \mathbf{y}$ is the corresponding **Ritz vector**; the pair $(\theta, \mathbf{x})$ is a **Ritz pair**.

In words: project the problem onto the subspace, solve the small problem exactly, and lift the answer back. The Ritz values depend on the subspace and not on which orthonormal basis of it we chose: replacing $\mathbf{Q}_k$ by $\mathbf{Q}_k \mathbf{V}$ with $\mathbf{V}$ unitary replaces $\mathbf{H}_k$ by $\mathbf{V}^* \mathbf{H}_k \mathbf{V}$, which has the same eigenvalues. Note that $\|\mathbf{x}\| = \|\mathbf{Q}_k \mathbf{y}\| = \|\mathbf{y}\| = 1$, since $\mathbf{Q}_k$ has orthonormal columns.

Theorem 24.6.4. Ritz Values Interlace, and Their Residual Is One Number

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian with eigenvalues $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(\mathbf{A})$, let $1 \leq k \leq d$, and let $\theta_1 \geq \cdots \geq \theta_k$ be the Ritz values from $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$, that is, the eigenvalues of $\mathbf{T}_k$.

- (a) $\lambda_i(\mathbf{A}) \geq \theta_i \geq \lambda_{i+n-k}(\mathbf{A})$ for $1 \leq i \leq k$.
- (b) If $\mathbf{T}_k \mathbf{y} = \theta \mathbf{y}$ with $\|\mathbf{y}\| = 1$ and $\mathbf{x} = \mathbf{Q}_k \mathbf{y}$, then

$$\mathbf{A}\mathbf{x} - \theta\mathbf{x} = \beta_k y_k \mathbf{q}_{k+1}, \quad \text{so} \quad \|\mathbf{A}\mathbf{x} - \theta\mathbf{x}\| = \beta_k |y_k|,$$

where y_k is the last coordinate of $\mathbf{y}$ and $\beta_k = h_{k+1,k}$, with $\beta_d \mathbf{q}_{d+1}$ read as $\mathbf{0}$.

Idea. For (a), the obstacle is that $\mathbf{T}_k = \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k$ is a compression, not a principal submatrix, and Chapter 17's interlacing is about principal submatrices. The repair is to complete $\mathbf{Q}_k$ to a unitary matrix: then $\mathbf{T}_k$ is the leading principal submatrix of a matrix unitarily similar to $\mathbf{A}$, and Corollary 17.4.3 applies unchanged. For (b), just apply Eq. 24.6.1 to $\mathbf{y}$ and watch the two $\mathbf{Q}_k \mathbf{T}_k \mathbf{y}$ terms cancel.

Proof.

- (a) Extend $\mathbf{q}_1, \dots, \mathbf{q}_k$ to an orthonormal basis $\mathbf{q}_1, \dots, \mathbf{q}_n$ of $\mathbb{C}^n$, which is possible by Theorem 11.2.2 applied to any extension of the list to a basis (Theorem 1.6.5), and let $\mathbf{U} = (\mathbf{q}_1 \mid \cdots \mid \mathbf{q}_n)$, which is unitary. Then $\mathbf{B} = \mathbf{U}^* \mathbf{A} \mathbf{U}$ is Hermitian and unitarily similar to $\mathbf{A}$, so $\lambda_i(\mathbf{B}) = \lambda_i(\mathbf{A})$ for every i . Its entries are $b_{ij} = \langle \mathbf{A} \mathbf{q}_j, \mathbf{q}_i \rangle$, and for $i, j \leq k$ these are exactly the entries of $\mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k = \mathbf{T}_k$. So $\mathbf{T}_k = \mathbf{B}_{I,I}$ with $I = \{1, \dots, k\}$, a principal submatrix of $\mathbf{B}$ obtained by deleting $m = n - k$ rows and the same columns. Now Corollary 17.4.3 gives

$$\lambda_i(\mathbf{B}) \geq \lambda_i(\mathbf{T}_k) \geq \lambda_{i+m}(\mathbf{B}) \quad (1 \leq i \leq k),$$

which is the assertion, since $\lambda_i(\mathbf{T}_k) = \theta_i$ and $\lambda_i(\mathbf{B}) = \lambda_i(\mathbf{A})$.

- (b) By Eq. 24.6.1, $\mathbf{A} \mathbf{Q}_k = \mathbf{Q}_k \mathbf{T}_k + \beta_k \mathbf{q}_{k+1} \mathbf{e}_k^\top$. Applying both sides to $\mathbf{y}$ and using $\mathbf{T}_k \mathbf{y} = \theta \mathbf{y}$ and $\mathbf{e}_k^\top \mathbf{y} = y_k$,

$$\mathbf{A}\mathbf{x} = \mathbf{A} \mathbf{Q}_k \mathbf{y} = \mathbf{Q}_k \mathbf{T}_k \mathbf{y} + \beta_k y_k \mathbf{q}_{k+1} = \theta \mathbf{Q}_k \mathbf{y} + \beta_k y_k \mathbf{q}_{k+1} = \theta \mathbf{x} + \beta_k y_k \mathbf{q}_{k+1}.$$

Subtracting $\theta \mathbf{x}$ gives the displayed identity, and taking norms gives the second form, since $\|\mathbf{q}_{k+1}\| = 1$ and $\beta_k \geq 0$. (For $k = d$ the right-hand side is $\mathbf{0}$, and the Ritz pair is an exact eigenpair, as Theorem 24.6.2 (c) already said.) This proves the theorem. ■

Part (b) is where the chapter's perturbation theory pays off, and it pays off for free. The residual $\|\mathbf{A}\mathbf{x} - \theta\mathbf{x}\|$ is the one quantity Chapter 20 §06 asks for, and here it is available **without forming $\mathbf{A}\mathbf{x}$** : it is the product of the number β_k the process has just computed and one coordinate of the small eigenvector $\mathbf{y}$. So Theorem 20.6.1 applies verbatim, and says that $\mathbf{A}$ has an eigenvalue within $\beta_k |y_k|$ of θ . Better still, $\theta = \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$ is the Rayleigh quotient of $\mathbf{x}$ — because $\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \mathbf{y}^* \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k \mathbf{y} = \mathbf{y}^* \mathbf{T}_k \mathbf{y} = \theta$ — so Theorem 20.6.2 applies too, and turns the same number into a bound of size $(\beta_k |y_k|)^2$ divided by the distance to the rest of the spectrum.

Nothing in this section had to be proved twice.

Warning. A Ritz value that happens to equal an eigenvalue is not a converged one, and for a non-Hermitian $\mathbf{A}$ there is no interlacing to fall back on. Take

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \mathbf{e}_3.$$

Two Arnoldi steps give $\mathbf{q}_1 = \mathbf{e}_3$, $\mathbf{q}_2 = \frac{1}{\sqrt{2}}(1, 1, 0)$, and $\mathbf{H}_2 = \begin{pmatrix} 1 & 0 \\ \sqrt{2} & \frac{5}{2} \end{pmatrix}$, hence the Ritz values $\frac{5}{2}$ and 1; [Exercise \(B2: Two Arnoldi steps\)](#) works the details. The second is an exact eigenvalue of $\mathbf{A}$ — and its Ritz vector $\frac{1}{\sqrt{17}}(-2, -2, 3)$ is not an eigenvector of $\mathbf{A}$, the residual being $\sqrt{34}/17 \approx 0.343$, nowhere near zero. What certifies convergence is the residual, never the agreement of a number with a number. Part (a) of [Theorem 24.6.4](#) needs $\mathbf{A}$ Hermitian; without that hypothesis there is no interlacing and, by Theorem 20.4.4, a residual bound carries the factor $\kappa_2(\mathbf{X})$.

Example. Ritz values closing in on the extremes

Let $\mathbf{A} = \text{diag}(4, 3, 2, 1)$ and $\mathbf{b} = (1, 1, 1, 1)$. Compute the Lanczos coefficients, the Ritz values for $k = 1, 2, 3$, and the residual of the largest Ritz pair at $k = 3$.

Solution. Here $\mathbf{q}_1 = \frac{1}{2}(1, 1, 1, 1)$ and $\mathbf{A}\mathbf{q}_1 = \frac{1}{2}(4, 3, 2, 1)$, so $\alpha_1 = \frac{1}{4}(4 + 3 + 2 + 1) = \frac{5}{2}$. Then

$$\mathbf{w}_1 = \frac{1}{2}(4, 3, 2, 1) - \frac{5}{2} \cdot \frac{1}{2}(1, 1, 1, 1) = \frac{1}{4}(3, 1, -1, -3), \quad \beta_1 = \frac{\sqrt{20}}{4} = \frac{\sqrt{5}}{2},$$

and $\mathbf{q}_2 = \frac{1}{\sqrt{20}}(3, 1, -1, -3)$. Continuing in the same way gives $\alpha_2 = \alpha_3 = \alpha_4 = \frac{5}{2}$ as well, together with $\beta_2 = \frac{2}{\sqrt{5}}$ and $\beta_3 = \frac{3}{2\sqrt{5}}$.

The repetition of $\frac{5}{2}$ is a symmetry, not an accident. Let $\mathbf{S}$ be the reversal permutation matrix, $\mathbf{S}\mathbf{e}_j = \mathbf{e}_{5-j}$, and put $\mathbf{B} = \mathbf{A} - \frac{5}{2}\mathbf{I} = \text{diag}(\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2})$. Then $\mathbf{S}$ is orthogonal with $\mathbf{S}^2 = \mathbf{I}$, and $\mathbf{SBS} = -\mathbf{B}$, $\mathbf{Sb} = \mathbf{b}$. We show by induction, simultaneously, that $\mathbf{S}\mathbf{q}_j = \varepsilon_j\mathbf{q}_j$ with $\varepsilon_j = (-1)^{j-1}$ and that $\alpha_j = \frac{5}{2}$. The first holds at $j = 1$. Given it at j ,

$$\alpha_j - \frac{5}{2} = \langle \mathbf{B}\mathbf{q}_j, \mathbf{q}_j \rangle = \langle \mathbf{SB}\mathbf{q}_j, \mathbf{S}\mathbf{q}_j \rangle = \langle \mathbf{SBS}(\mathbf{S}\mathbf{q}_j), \mathbf{S}\mathbf{q}_j \rangle = -\varepsilon_j^2 \langle \mathbf{B}\mathbf{q}_j, \mathbf{q}_j \rangle,$$

and $\varepsilon_j^2 = 1$, so this number equals its own negative and $\alpha_j = \frac{5}{2}$. The Lanczos step then reads $\mathbf{w}_j = \mathbf{B}\mathbf{q}_j - \beta_{j-1}\mathbf{q}_{j-1}$, and applying $\mathbf{S}$ with $\varepsilon_{j-1} = -\varepsilon_j$ gives $\mathbf{S}\mathbf{w}_j = -\varepsilon_j\mathbf{w}_j$, so $\mathbf{S}\mathbf{q}_{j+1} = \varepsilon_{j+1}\mathbf{q}_{j+1}$. The Ritz values are the eigenvalues of the leading blocks of $\mathbf{T}_4$:

k	Ritz values	check against 4, 3, 2, 1
1	$\frac{5}{2}$	$4 \geq 2.5 \geq 1$
2	$\frac{5 \pm \sqrt{5}}{2} \approx 3.6180, 1.3820$	$4 \geq 3.618 \geq 2; 3 \geq 1.382 \geq 1$
3	$\frac{5}{2} \pm \frac{\sqrt{205}}{10} \approx 3.9318, 1.0682,$ and $\frac{5}{2}$	$4 \geq 3.932 \geq 3; 3 \geq 2.5 \geq 2;$ $2 \geq 1.068 \geq 1$
4	4, 3, 2, 1	exact

Every row obeys [Theorem 24.6.4](#) (a), with $m = n - k$ shifting the index on the right. The extreme Ritz values close in on 4 and 1 fast, and the interior ones lag: at $k = 3$ the outer two are within 0.07 of 4 and 1, while the single interior value $\frac{5}{2}$ is half a unit from each of 3 and 2.

The residual at $k = 3$. The eigenvector of $\mathbf{T}_3$ for $\theta_1 = \frac{5}{2} + \frac{\sqrt{205}}{10}$ is, normalized,

$$\mathbf{y} = \left(\frac{5}{\sqrt{82}}, \frac{1}{\sqrt{2}}, \frac{4}{\sqrt{82}} \right),$$

so [Theorem 24.6.4](#) (b) gives $\|\mathbf{Ax} - \theta_1 \mathbf{x}\| = \beta_3 |y_3| = \frac{3}{2\sqrt{5}} \cdot \frac{4}{\sqrt{82}} = \frac{3\sqrt{410}}{205} \approx 0.2963$. By [Theorem 20.6.1](#), $\mathbf{A}$ has an eigenvalue within 0.2963 of 3.9318; the truth is 4, at distance 0.0682, so the certificate costs a factor of about 4. Suppose we also know that $\mathbf{A}$ has no eigenvalue in the open interval $(3, 4)$. Then [Theorem 20.6.2](#) (b) with $\alpha = 3$ improves the upper end to

$$\theta_1 + \frac{(0.2963)^2}{3.9318 - 3} = 3.9318 + 0.0942 = 4.0260,$$

and the lower end is θ_1 itself by interlacing. The bracket $[3.9318, 4.0260]$ has width 0.094 where the residual bound gave 0.593. That is the quadratic gain of Chapter 20 §06, obtained here from one number the process had already produced.

24.6.5 How fast the extreme Ritz values converge

The example suggests the extremes converge quickly. Here is why, and how much of it we can prove.

Theorem 24.6.5. A Rate for the Largest Ritz Value

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian with eigenvalues $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ and an orthonormal eigenbasis $\mathbf{v}_1, \dots, \mathbf{v}_n$. Let $\mathbf{b}$ be a unit vector with $\mathbf{b} = \sum_i c_i \mathbf{v}_i$ and $c_1 \neq 0$, and put

$$\tau^2 = \frac{\sum_{i \geq 2} |c_i|^2}{|c_1|^2}, \quad \rho = \frac{\lambda_2 - \lambda_n}{2\lambda_1 - \lambda_2 - \lambda_n} \in [0, 1).$$

Then the largest Ritz value $\theta_1^{(k)}$ from $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ satisfies, for $1 \leq k \leq d$,

$$0 \leq \lambda_1 - \theta_1^{(k)} \leq (\lambda_1 - \lambda_n) \tau^2 \rho^{2(k-1)},$$

and $\theta_1^{(k)}$ is nondecreasing in k .

Idea. Three moves. ① The largest Ritz value is the largest Rayleigh quotient **over the Krylov space**, by the Rayleigh principle transported through $\mathbf{Q}_k$. ② Every vector of the Krylov space is $p(\mathbf{A})\mathbf{b}$ for a polynomial of degree below k , so we may choose p and get a bound for free: whichever p makes p large at λ_1 and small at the other eigenvalues gives a good one. ③ Take the simplest such p , the $(k-1)$ -st power of x minus the midpoint of $[\lambda_n, \lambda_2]$. That is not the best choice, and the best one is the subject of the next section.

Proof. **Step 1.** $\theta_1^{(k)} = \max\{R_{\mathbf{A}}(\mathbf{x}) : \mathbf{x} \in \mathcal{K}_k(\mathbf{A}, \mathbf{b}), \mathbf{x} \neq \mathbf{0}\}$. Indeed $\theta_1^{(k)} = \lambda_1(\mathbf{T}_k) = \max_{\mathbf{y} \neq \mathbf{0}} R_{\mathbf{T}_k}(\mathbf{y})$ by Proposition 17.1.1, and the map $\mathbf{y} \mapsto \mathbf{Q}_k \mathbf{y}$ is a bijection from $\mathbb{C}^k \setminus \{\mathbf{0}\}$ onto $\mathcal{K}_k \setminus \{\mathbf{0}\}$, because the columns of $\mathbf{Q}_k$ are a basis of $\mathcal{K}_k$ (Theorem 24.6.2 (a)). Under it, with $\mathbf{x} = \mathbf{Q}_k \mathbf{y}$,

$$\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \mathbf{y}^* \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k \mathbf{y} = \mathbf{y}^* \mathbf{T}_k \mathbf{y}, \quad \langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{y}^* \mathbf{Q}_k^* \mathbf{Q}_k \mathbf{y} = \mathbf{y}^* \mathbf{y},$$

so $R_{\mathbf{A}}(\mathbf{x}) = R_{\mathbf{T}_k}(\mathbf{y})$ and the two maxima agree. Monotonicity in k follows at once from $\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$ (Proposition 24.6.1 (a)), and $\theta_1^{(k)} \leq \lambda_1$ from Proposition 17.1.1 applied on all of $\mathbb{C}^n$ — which is also Theorem 24.6.4 (a) at $i = 1$.

Step 2. Let $p \in \mathbb{R}[x]$ have $\deg p \leq k - 1$ and $p(\lambda_1) \neq 0$, and put $\mathbf{x} = p(\mathbf{A})\mathbf{b}$. Then $\mathbf{x} \in \mathcal{K}_k$ by Proposition 24.6.1 (a), and $\mathbf{x} = \sum_i c_i p(\lambda_i) \mathbf{v}_i$ since $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$. The list $(\mathbf{v}_i)$ being orthonormal, Theorem 11.2.1 gives

$$R_{\mathbf{A}}(\mathbf{x}) = \frac{\sum_i |c_i|^2 p(\lambda_i)^2 \lambda_i}{\sum_i |c_i|^2 p(\lambda_i)^2},$$

and $\mathbf{x} \neq \mathbf{0}$ because its first coefficient $c_1 p(\lambda_1)$ is nonzero. Subtracting from λ_1 ,

$$\lambda_1 - R_{\mathbf{A}}(\mathbf{x}) = \frac{\sum_{i \geq 2} |c_i|^2 p(\lambda_i)^2 (\lambda_1 - \lambda_i)}{\sum_i |c_i|^2 p(\lambda_i)^2},$$

the term $i = 1$ having vanished from the numerator. Every factor $\lambda_1 - \lambda_i$ is at most $\lambda_1 - \lambda_n$; and the denominator is at least its single term $|c_1|^2 p(\lambda_1)^2$, all terms being nonnegative. Hence

$$\begin{aligned} \lambda_1 - R_{\mathbf{A}}(\mathbf{x}) &\leq (\lambda_1 - \lambda_n) \frac{\sum_{i \geq 2} |c_i|^2 p(\lambda_i)^2}{|c_1|^2 p(\lambda_1)^2} \\ &\leq (\lambda_1 - \lambda_n) \tau^2 \max_{i \geq 2} \left(\frac{p(\lambda_i)}{p(\lambda_1)} \right)^2. \end{aligned} \quad (*)$$

By Step 1, $\lambda_1 - \theta_1^{(k)} \leq \lambda_1 - R_{\mathbf{A}}(\mathbf{x})$.

Step 3. Take $\mu = \frac{1}{2}(\lambda_2 + \lambda_n)$, the midpoint of the interval holding the unwanted eigenvalues, and $p(x) = (x - \mu)^{k-1}$, of degree $k-1$. For $i \geq 2$ we have $\lambda_n \leq \lambda_i \leq \lambda_2$, so $|\lambda_i - \mu| \leq \frac{1}{2}(\lambda_2 - \lambda_n)$. And

$$\lambda_1 - \mu = \frac{1}{2}(\lambda_2 - \lambda_n) + (\lambda_1 - \lambda_2) > \frac{1}{2}(\lambda_2 - \lambda_n) \geq 0,$$

using $\lambda_1 > \lambda_2$; in particular $p(\lambda_1) = (\lambda_1 - \mu)^{k-1} \neq 0$. Therefore

$$\max_{i \geq 2} \left| \frac{p(\lambda_i)}{p(\lambda_1)} \right| \leq \left(\frac{(\lambda_2 - \lambda_n)/2}{\lambda_1 - \mu} \right)^{k-1} = \rho^{k-1},$$

since $2(\lambda_1 - \mu) = 2\lambda_1 - \lambda_2 - \lambda_n$. Substituting into (*) gives the bound, and $\rho < 1$ because $\lambda_1 > \lambda_2$. This proves the theorem. ■

For $\mathbf{A} = \text{diag}(4, 3, 2, 1)$ and $\mathbf{b} = (1, 1, 1, 1)/2$ the constants are $\tau^2 = 3$, $\lambda_1 - \lambda_n = 3$, and $\rho = \frac{1}{2}$, so the bound reads $9 \cdot 4^{-(k-1)}$: the values 9, 2.25, 0.5625 for $k = 1, 2, 3$, against the true gaps 1.5, 0.382, 0.068 of Example (Ritz values closing in on the extremes). The bound is correct and loose, and the looseness is in Step 3.

It is worth naming what ρ is. Assume for the rest of this subsection that $\lambda_2 > \lambda_n$, which the theorem does not require; if instead $\lambda_2 = \lambda_n$, then $\rho = 0$ and the bound of [Theorem 24.6.5](#) is already as strong as this route can make it. Put

$$\gamma = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n},$$

the size of the gap above λ_2 measured against the spread of the rest. Then $2\lambda_1 - \lambda_2 - \lambda_n = 2(\lambda_1 - \lambda_2) + (\lambda_2 - \lambda_n) = (\lambda_2 - \lambda_n)(1 + 2\gamma)$, so

$$\rho = \frac{1}{1 + 2\gamma}.$$

(For $\mathbf{A} = \text{diag}(4, 3, 2, 1)$, $\gamma = \frac{1}{2}$ and $\rho = \frac{1}{2}$.) So Step 3 buys a factor $(1 + 2\gamma)^{-(k-1)}$ per unit of k , and a small gap makes it nearly worthless.

What is not proved here. The polynomial of Step 3 is not the best one. Minimizing $\max_{i \geq 2} |p(\lambda_i)/p(\lambda_1)|$ over the polynomials of degree at most $k-1$ with $p(\lambda_1) = 1$ is a question about how small a polynomial can be kept on an interval while staying large at a point outside it, and its answer is the Chebyshev polynomial T_{k-1} , shifted and scaled to that interval. The resulting estimate, due to Kaniel, Paige, and Saad, is

$$0 \leq \lambda_1 - \theta_1^{(k)} \leq \frac{(\lambda_1 - \lambda_n) \tau^2}{(T_{k-1}(1 + 2\gamma))^2}.$$

The next section defines the Chebyshev polynomials and proves the two facts this bound rests on: that T_k stays within 1 on $[-1, 1]$, and that it grows geometrically just outside it; the substitution into (*) is then immediate, but **we do not carry it out here, and nothing in this section depends on it**. The reason it is worth stating is the shape of the improvement. For small γ , $T_{k-1}(1 + 2\gamma)$ grows roughly like $(1 + 2\sqrt{\gamma})^{k-1}$, where Step 3 managed only $(1 + 2\gamma)^{k-1}$. A small gap costs a square root, not a full factor. That is the same square root that the next section extracts for conjugate gradients, and it is the reason Krylov methods are used at all.

24.6.6 Losing orthogonality

Everything above is exact arithmetic. In the floating-point model of Definition 24.1.1, one feature of the Lanczos recurrence fails, and it fails in a way that is worth describing honestly rather than proving badly.

The three-term recurrence never re-examines $\mathbf{q}_1, \dots, \mathbf{q}_{j-2}$. In exact arithmetic it does not need to: [Theorem 24.6.3](#) proves that $\mathbf{w}_j$ is automatically orthogonal to them. In floating-point arithmetic the computed vectors satisfy the recurrence to within a small multiple of $u\|\mathbf{A}\|$, and nothing removes the small components along the earlier vectors that rounding introduces. Those components do not stay small.

The mechanism is specific, and it is not random drift. The component of the computed $\mathbf{q}_{k+1}$ along a Ritz vector $\mathbf{x}$ grows in inverse proportion to that Ritz pair's residual $\beta_k|y_k|$. So orthogonality is lost first, and most, in exactly the direction of a Ritz pair that has **converged** — the better the approximation, the faster the basis is corrupted in its direction. The visible symptom is that the computed $\mathbf{T}_k$ acquires a second copy of a Ritz value that converged several steps earlier, and later a third: the spurious duplicates, which correspond to no extra multiplicity in $\mathbf{A}$. An algorithm that counts eigenvalues by counting Ritz values will count wrong.

This paragraph is a description of observed behavior, not a theorem, and this book does not prove it. The analysis that makes it precise — that the loss of orthogonality is

confined to the converged directions, and that the computed $\mathbf{T}_k$ is the exact tridiagonal matrix of a larger problem — is due to Paige, and it is outside the scope of this chapter. Nothing above depends on it. The standard remedies are to reorthogonalize each new vector against all the earlier ones, which restores the guarantee and abandons the three-term economy, or to reorthogonalize only against the converged Ritz vectors, which is cheaper and rests on precisely the analysis we are not giving. Both are engineering, not theorems, and we label them as such.

24.6.7 Exercises

24.6.7.1 A. Check your understanding

Exercise. A1

- (a) Define $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ and state its dimension in terms of $m_{\mathbf{A}, \mathbf{b}}$.
- (b) State the Arnoldi identity of [Theorem 24.6.2](#) (b), naming the shape of every matrix in it.
- (c) True or false: for Hermitian $\mathbf{A}$, the Arnoldi process needs all of $\mathbf{q}_1, \dots, \mathbf{q}_j$ to produce $\mathbf{q}_{j+1}$. Justify your answer.
- (d) What does a breakdown $h_{k+1,k} = 0$ mean, and why is it good news?
- (e) Give the formula for the residual of a Ritz pair, and say why it costs nothing to evaluate.

Solution.

- (a) $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) = \text{span}(\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^{k-1}\mathbf{b})$, and $\dim \mathcal{K}_k = \min(k, d)$ with $d = \deg m_{\mathbf{A}, \mathbf{b}}$ ([Proposition 24.6.1](#)).
- (b) $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_{k+1}\tilde{\mathbf{H}}_k$, where $\mathbf{Q}_k \in M_{n \times k}$ and $\mathbf{Q}_{k+1} \in M_{n \times (k+1)}$ have orthonormal columns and $\tilde{\mathbf{H}}_k \in M_{(k+1) \times k}$ is upper Hessenberg; equivalently $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k\mathbf{H}_k + h_{k+1,k}\mathbf{q}_{k+1}\mathbf{e}_k^\top$ with $\mathbf{H}_k = \mathbf{Q}_k^*\mathbf{A}\mathbf{Q}_k \in M_k$ upper Hessenberg.
- (c) False. By [Theorem 24.6.3](#) the recurrence has three terms, so only $\mathbf{q}_{j-1}$ and $\mathbf{q}_j$ are needed. The reason is that $\mathbf{H}_k$ is both Hermitian and upper Hessenberg, hence tridiagonal.
- (d) It means $\mathcal{K}_k$ is $\mathbf{A}$ -invariant ([Theorem 24.6.2](#) (c)), equivalently $k = d$. It is good news because every eigenvalue of $\mathbf{H}_k$ is then an exact eigenvalue of $\mathbf{A}$: the method has finished, not failed.
- (e) $\|\mathbf{A}\mathbf{x} - \theta\mathbf{x}\| = h_{k+1,k}|y_k|$ ([Theorem 24.6.4](#) (b)). Both factors are already known — $h_{k+1,k}$ from the last step of the process, y_k from the small eigenvector — so no multiplication by $\mathbf{A}$ is needed.

24.6.7.2 B. Practice

Exercise. B1: Dimensions

For each pair $(\mathbf{A}, \mathbf{b})$, determine $\dim \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ for every $k \geq 1$. Justify your answer.

- (a) $\mathbf{A} = \text{diag}(2, 2, 5)$, $\mathbf{b} = (1, 1, 1)$.
- (b) $\mathbf{A} = \text{diag}(2, 2, 5)$, $\mathbf{b} = (1, -1, 0)$.

(c) $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{b} = \mathbf{e}_2$.

(d) $\mathbf{A}$ the 4×4 matrix with $\mathbf{A}\mathbf{e}_j = \mathbf{e}_{j+1}$ for $j \leq 3$ and $\mathbf{A}\mathbf{e}_4 = \mathbf{0}$, $\mathbf{b} = \mathbf{e}_1$.

Solution. By [Proposition 24.6.1](#) (b) the answer is always $\min(k, d)$ with $d = \deg m_{\mathbf{A}, \mathbf{b}}$.

- (a) $\mathbf{b}$ has nonzero components in both eigenspaces, so no single factor of $(x - 2)(x - 5)$ kills it, while the product does: $d = 2$, and the dimensions are $1, 2, 2, 2, \dots$
- (b) $\mathbf{A}\mathbf{b} = (2, -2, 0) = 2\mathbf{b}$, so $\mathbf{b}$ is an eigenvector and $m_{\mathbf{A}, \mathbf{b}} = x - 2$, $d = 1$. The dimensions are $1, 1, 1, \dots$: every further product is wasted.
- (c) $\mathbf{A}\mathbf{e}_2 = \mathbf{e}_1 \neq \mathbf{0}$ and $\mathbf{A}^2\mathbf{e}_2 = \mathbf{0}$, so $m_{\mathbf{A}, \mathbf{e}_2} = x^2$ and $d = 2$: dimensions $1, 2, 2, \dots$
- (d) $\mathbf{A}^j\mathbf{e}_1 = \mathbf{e}_{j+1}$ for $j \leq 3$ and $\mathbf{A}^4\mathbf{e}_1 = \mathbf{0}$, so $m_{\mathbf{A}, \mathbf{e}_1} = x^4$ and $d = 4$: dimensions $1, 2, 3, 4, 4, \dots$. Here the Krylov spaces grow as long as they possibly can.

Exercise. B2: Two Arnoldi steps

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{b} = \mathbf{e}_3.$$

Carry out two Arnoldi steps. Write down $\mathbf{Q}_2$, $\tilde{\mathbf{H}}_2$, and $\mathbf{H}_2$, verify [Eq. 24.6.1](#), and compute the two Ritz values with their residuals. Hence say which, if either, is a converged eigenvalue estimate.

Solution. $\mathbf{q}_1 = \mathbf{e}_3$ and $\mathbf{A}\mathbf{q}_1 = (1, 1, 1)$, so $h_{11} = \langle \mathbf{A}\mathbf{q}_1, \mathbf{q}_1 \rangle = 1$ and

$$\mathbf{w}_1 = (1, 1, 1) - (0, 0, 1) = (1, 1, 0), \quad h_{21} = \sqrt{2}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}}(1, 1, 0).$$

Next $\mathbf{A}\mathbf{q}_2 = \frac{1}{\sqrt{2}}(3, 2, 0)$, so $h_{12} = \langle \mathbf{A}\mathbf{q}_2, \mathbf{q}_1 \rangle = 0$ and $h_{22} = \frac{1}{2}(3 + 2) = \frac{5}{2}$, and

$$\mathbf{w}_2 = \frac{1}{\sqrt{2}}(3, 2, 0) - \frac{5}{2} \cdot \frac{1}{\sqrt{2}}(1, 1, 0) = \frac{1}{2\sqrt{2}}(1, -1, 0), \quad h_{32} = \frac{1}{2},$$

with $\mathbf{q}_3 = \frac{1}{\sqrt{2}}(1, -1, 0)$. Hence

$$\mathbf{Q}_2 = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \end{pmatrix}, \quad \tilde{\mathbf{H}}_2 = \begin{pmatrix} 1 & 0 \\ \sqrt{2} & \frac{5}{2} \\ 0 & \frac{1}{2} \end{pmatrix}, \quad \mathbf{H}_2 = \begin{pmatrix} 1 & 0 \\ \sqrt{2} & \frac{5}{2} \end{pmatrix}.$$

Check [Eq. 24.6.1](#) column by column: $\mathbf{A}\mathbf{q}_1 = (1, 1, 1) = 1 \cdot \mathbf{q}_1 + \sqrt{2} \mathbf{q}_2$, and $\mathbf{A}\mathbf{q}_2 = \frac{1}{\sqrt{2}}(3, 2, 0) = 0 \cdot \mathbf{q}_1 + \frac{5}{2} \mathbf{q}_2 + \frac{1}{2} \mathbf{q}_3$, both correct.

$\mathbf{H}_2$ is lower triangular as well as upper Hessenberg, so its eigenvalues are $\theta = 1$ and $\theta = \frac{5}{2}$. For $\theta = \frac{5}{2}$: solving $\mathbf{H}_2\mathbf{y} = \frac{5}{2}\mathbf{y}$ gives $y_1 = 0$, so $\mathbf{y} = (0, 1)$, $y_2 = 1$, and the residual is $h_{32}|y_2| = \frac{1}{2}$. For $\theta = 1$: $\sqrt{2}y_1 + \frac{3}{2}y_2 = 0$, so $\mathbf{y} = \frac{1}{\sqrt{17}}(3, -2\sqrt{2})$ and the residual is $\frac{1}{2} \cdot \frac{2\sqrt{2}}{\sqrt{17}} = \frac{\sqrt{34}}{17} \approx 0.343$.

Neither is converged. The eigenvalues of $\mathbf{A}$ are 2, 2, 1, so the Ritz value 1 is exactly an eigenvalue — yet its residual is 0.343, not 0, and its Ritz vector $\mathbf{Q}_2\mathbf{y} = \frac{1}{\sqrt{17}}(-2, -2, 3)$ is not an eigenvector: the eigenvectors for 1 are the multiples of $(0, 1, -1)$, which is not in $\mathcal{K}_2 = \text{span}(\mathbf{e}_3, (1, 1, 0))$. The agreement is a coincidence of two numbers, and the residual is what says so.

Exercise. B3: Lanczos on a tridiagonal matrix

Let $\mathbf{A} \in M_n(\mathbb{R})$ be symmetric tridiagonal with every subdiagonal entry nonzero, and let $\mathbf{b} = \mathbf{e}_1$. Prove that the Lanczos process produces $\mathbf{q}_j = \pm \mathbf{e}_j$ for every j , and hence that $\mathbf{T}_n = \mathbf{A}$ up to signs. What does this say about the cost of Lanczos on a matrix that is already tridiagonal?

Solution. We show by induction on j that

$$\mathbf{A}^j \mathbf{e}_1 = \sum_{i \leq j+1} c_i^{(j)} \mathbf{e}_i \quad \text{with} \quad c_{j+1}^{(j)} = a_{21} a_{32} \cdots a_{j+1,j} \neq 0,$$

for $0 \leq j \leq n-1$. At $j=0$ the sum is $\mathbf{e}_1$ and the coefficient is the empty product 1. Suppose it holds at $j-1$. Since $\mathbf{A}$ is tridiagonal, $\mathbf{A}\mathbf{e}_i \in \text{span}(\mathbf{e}_{i-1}, \mathbf{e}_i, \mathbf{e}_{i+1})$, so applying $\mathbf{A}$ to $\mathbf{A}^{j-1}\mathbf{e}_1 = \sum_{i \leq j} c_i^{(j-1)} \mathbf{e}_i$ lands in $\text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{j+1})$; and the only term contributing to $\mathbf{e}_{j+1}$ is $c_j^{(j-1)} \mathbf{A}\mathbf{e}_j$, whose $\mathbf{e}_{j+1}$ -coefficient is $c_j^{(j-1)} a_{j+1,j} \neq 0$, both factors being nonzero. This proves the induction.

Consequently the change-of-basis matrix from $(\mathbf{e}_1, \dots, \mathbf{e}_j)$ to $(\mathbf{e}_1, \mathbf{A}\mathbf{e}_1, \dots, \mathbf{A}^{j-1}\mathbf{e}_1)$ is triangular with nonzero diagonal, so $\mathcal{K}_j(\mathbf{A}, \mathbf{e}_1) = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_j)$ for $1 \leq j \leq n$.

By [Theorem 24.6.2](#) (a), $\mathbf{q}_j$ is a unit vector in $\mathcal{K}_j$ orthogonal to $\mathcal{K}_{j-1} = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{j-1})$, hence a unit multiple of $\mathbf{e}_j$; the sign is fixed by $\beta_{j-1} > 0$. Therefore $(\mathbf{T}_n)_{ij} = \langle \mathbf{A}\mathbf{q}_j, \mathbf{q}_i \rangle = \pm a_{ij}$, and since $\mathbf{q}_j = \varepsilon_j \mathbf{e}_j$ with $\varepsilon_j = \pm 1$, $\mathbf{T}_n = \mathbf{D}\mathbf{A}\mathbf{D}$ with $\mathbf{D} = \text{diag}(\varepsilon_1, \dots, \varepsilon_n)$ — a similarity, so the two have the same eigenvalues.

The cost lesson is that Lanczos does no work at all here beyond rediscovering the matrix: on a tridiagonal matrix with $\mathbf{b} = \mathbf{e}_1$ it reproduces $\mathbf{A}$ and finds nothing early. The method earns its keep when n is large and only the first few extreme eigenvalues are wanted, so that $k \ll n$.

24.6.7.3 C. Going deeper

Exercise. C1: Ritz values from the other end

Let $\mathbf{A} \in M_n(\mathbb{C})$ be Hermitian and let $\theta_1^{(k)} \geq \dots \geq \theta_k^{(k)}$ be the Ritz values from $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$, $1 \leq k \leq d$.

- Prove that $\theta_k^{(k)}$ is nonincreasing in k and that $\theta_k^{(k)} \geq \lambda_n(\mathbf{A})$.
- Deduce a bound for $\theta_k^{(k)} - \lambda_n(\mathbf{A})$ from [Theorem 24.6.5](#), with no new argument.
- Show that at $k=d$ the Ritz values are d of the eigenvalues of $\mathbf{A}$, and say which.

Solution.

- (a) The argument of Step 1 in the proof of [Theorem 24.6.5](#) applies verbatim with minima in place of maxima, using the second half of Proposition 17.1.1: $\theta_k^{(k)} = \min\{R_{\mathbf{A}}(\mathbf{x}) : \mathbf{x} \in \mathcal{K}_k, \mathbf{x} \neq \mathbf{0}\}$. A minimum over a larger set is no larger, and $\mathcal{K}_k \subseteq \mathcal{K}_{k+1}$, so the sequence is nonincreasing; and the minimum over all of $\mathbb{C}^n$ is $\lambda_n(\mathbf{A})$, so it is bounded below by that. (Alternatively, both statements are [Theorem 24.6.4](#) (a) at $i = k$.)
- (b) Apply [Theorem 24.6.5](#) to $-\mathbf{A}$. Its eigenvalues are $-\lambda_n \geq -\lambda_{n-1} \geq \dots \geq -\lambda_1$, with the same orthonormal eigenbasis in reversed order. Its Krylov spaces from $\mathbf{b}$ are the same subspaces, because $(-\mathbf{A})^j \mathbf{b} = (-1)^j \mathbf{A}^j \mathbf{b}$ and scaling the vectors of a list does not change its span. And $R_{-\mathbf{A}} = -R_{\mathbf{A}}$, so by Step 1 of the proof the largest Ritz value of $-\mathbf{A}$ from $\mathcal{K}_k$ is $-\theta_k^{(k)}$. The theorem gives

$$\theta_k^{(k)} - \lambda_n \leq (\lambda_1 - \lambda_n) \tilde{\tau}^2 \tilde{\rho}^{2(k-1)}, \quad \tilde{\rho} = \frac{\lambda_1 - \lambda_{n-1}}{\lambda_1 + \lambda_{n-1} - 2\lambda_n},$$

with $\tilde{\tau}^2 = \sum_{i \leq n-1} |c_i|^2 / |c_n|^2$, provided $\lambda_{n-1} > \lambda_n$ and $c_n \neq 0$.

- (c) At $k = d$ the process breaks down, so $\mathcal{K}_d$ is $\mathbf{A}$ -invariant and every Ritz value is an eigenvalue of $\mathbf{A}$ by [Theorem 24.6.2](#) (c). Which ones: writing $\mathbf{b} = \sum_i c_i \mathbf{v}_i$ in an orthonormal eigenbasis and grouping the eigenvectors by eigenvalue, $p(\mathbf{A})\mathbf{b} = \sum_i c_i p(\lambda_i) \mathbf{v}_i$, so $\mathcal{K}_d$ is contained in the span of those $\mathbf{v}_i$ with $c_i \neq 0$. The Ritz values are therefore the distinct eigenvalues λ of $\mathbf{A}$ for which $\mathbf{b}$ has a nonzero component in $E_\lambda(\mathbf{A})$, each occurring once; and d is their number, as [Example \(A breakdown at step two\)](#) illustrates with $d = 2$ for a matrix with eigenvalues 5, 2, 2.

Exercise. C2: Shift invariance, and what it costs

Let $\mathbf{A} \in M_n(F)$, $\mathbf{b} \neq \mathbf{0}$, and let $\sigma, c \in F$ with $c \neq 0$.

- (a) Prove that $\mathcal{K}_k(c\mathbf{A} + \sigma\mathbf{I}, \mathbf{b}) = \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ for every k .
- (b) Deduce that the Arnoldi vectors $\mathbf{q}_j$ are the same for $\mathbf{A}$ and for $c\mathbf{A} + \sigma\mathbf{I}$, up to a scalar of modulus one, and that the Ritz values transform by $\theta \mapsto c\theta + \sigma$.
- (c) Contrast this with Section 5: explain in two sentences why shifting accelerates the QR algorithm but does **not** accelerate a Krylov method, and what one must do instead to get the effect.

Solution.

- (a) Put $\mathbf{B} = c\mathbf{A} + \sigma\mathbf{I}$. By [Proposition 24.6.1](#) (a), $\mathcal{K}_k(\mathbf{B}, \mathbf{b}) = \{p(\mathbf{B})\mathbf{b} : \deg p \leq k-1\}$ and $p(\mathbf{B}) = p(c\mathbf{A} + \sigma\mathbf{I}) = \tilde{p}(\mathbf{A})$ where $\tilde{p}(x) = p(cx + \sigma)$ has the same degree, $c \neq 0$. The map $p \mapsto \tilde{p}$ is a bijection of $F[x]_{\leq k-1}$ onto itself, with inverse $q \mapsto q((x - \sigma)/c)$. So the two sets of vectors coincide.
- (b) The Arnoldi vectors are determined by the flag $\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \dots$ alone, up to a scalar of modulus one: $\mathbf{q}_j$ is a unit vector in $\mathcal{K}_j$ orthogonal to $\mathcal{K}_{j-1}$ ([Theorem 24.6.2](#) (a)), and the unit vectors of $\mathcal{K}_j \cap \mathcal{K}_{j-1}^\perp$ are exactly the modulus-one multiples of one of them. By (a) the flag is the same for $\mathbf{B}$ as for $\mathbf{A}$, so the run on $\mathbf{B}$ produces $\mathbf{Q}_k^{\mathbf{B}} = \mathbf{Q}_k^{\mathbf{A}} \mathbf{D}$ with $\mathbf{D} = \text{diag}(d_1, \dots, d_k)$ and $|d_j| = 1$. The scalars need not all be 1: each run fixes its own by $h_{j,j-1} > 0$, and for $c = -1$, $\sigma = 0$, $\mathbf{A} = \text{diag}(4, 3, 2, 1)$, and $\mathbf{b} = (1, 1, 1, 1)$ the

two runs give $\mathbf{q}_j^{\mathbf{B}} = (-1)^{j-1} \mathbf{q}_j$. Consequently

$$(\mathbf{Q}_k^{\mathbf{B}})^* \mathbf{B} \mathbf{Q}_k^{\mathbf{B}} = \mathbf{D}^* (c \mathbf{Q}_k^* \mathbf{A} \mathbf{Q}_k + \sigma \mathbf{Q}_k^* \mathbf{Q}_k) \mathbf{D} = \mathbf{D}^* (c \mathbf{H}_k + \sigma \mathbf{I}_k) \mathbf{D},$$

whose eigenvalues, $\mathbf{D}$ being unitary, are those of $c \mathbf{H}_k + \sigma \mathbf{I}_k$, namely $c\theta_i + \sigma$.

- (c) The QR algorithm converges at a rate governed by ratios $|\lambda_{i+1} - \mu|/|\lambda_i - \mu|$, and those ratios genuinely change with μ ; a Krylov space does not change at all under a shift, by (a), so nothing it produces can change either. To get the analogous effect one must change the space, not the matrix: apply the method to $(\mathbf{A} - \mu \mathbf{I})^{-1}$, whose Krylov spaces are different, which is the Krylov version of the inverse iteration of Theorem 24.4.3 and costs a factorization of $\mathbf{A} - \mu \mathbf{I}$ rather than a matrix-vector product.

Exercise. C3: When the Krylov space is everything

Let $\mathbf{A} \in M_n(F)$.

- Prove that there exists $\mathbf{b}$ with $\mathcal{K}_n(\mathbf{A}, \mathbf{b}) = F^n$ if and only if $m_{\mathbf{A}} = p_{\mathbf{A}}$.
- Prove that in that case the Arnoldi process run to $k = n$ produces a unitary $\mathbf{Q}_n$ with $\mathbf{Q}_n^* \mathbf{A} \mathbf{Q}_n$ upper Hessenberg, all of whose subdiagonal entries are nonzero.
- Hence give a second proof of Theorem 11.8.5 over $\mathbb{C}$ for such $\mathbf{A}$, and explain why it does not replace Chapter 11's.

Hint for (a): Chapter 10 §05 calls such a $\mathbf{b}$ a cyclic vector.

Solution.

- $\mathcal{K}_n(\mathbf{A}, \mathbf{b}) = F^n$ means $\dim \mathcal{K}_n = n$, which by Proposition 24.6.1 (b) means $\deg m_{\mathbf{A}, \mathbf{b}} = n$, that is, $Z(\mathbf{b}; \mathbf{A}) = F^n$ and $\mathbf{b}$ is a cyclic vector for $\mathbf{A}$ (Definition 10.5.4). By Theorem 10.5.6, such a vector exists exactly when $m_{\mathbf{A}} = p_{\mathbf{A}}$.
- With $d = n$, Theorem 24.6.2 (a) says that $h_{j+1,j} \neq 0$ for every $j < n$ and that $(\mathbf{q}_1, \dots, \mathbf{q}_n)$ is an orthonormal basis of F^n ; so $\mathbf{Q}_n$ is unitary, and $\mathbf{Q}_n^* \mathbf{A} \mathbf{Q}_n = \mathbf{H}_n$ is upper Hessenberg by Theorem 24.6.2 (b). Its subdiagonal entries are the $h_{j+1,j}$, all nonzero. (A Hessenberg matrix with this property is called **unreduced**, and Proposition 24.5.5 is about exactly these.)
- Part (b) exhibits a unitary $\mathbf{Q}_n$ with $\mathbf{Q}_n^* \mathbf{A} \mathbf{Q}_n$ upper Hessenberg, which is the conclusion of Theorem 11.8.5. It does not replace that theorem for two reasons. First, it needs $m_{\mathbf{A}} = p_{\mathbf{A}}$ by (a), while Theorem 11.8.5 holds for every square matrix; when no cyclic vector exists, the process breaks down at some $d < n$ and delivers an invariant subspace instead of a full basis. Second, Theorem 11.8.5 is proved by reflections, each applied to the whole matrix, and its $O(n^3)$ cost is what Section 5 assumed; the Arnoldi route needs n matrix-vector products and $O(n^3)$ orthogonalization work as well, and in floating-point arithmetic the orthogonality it produces is exactly what the last subsection says can be lost. The two constructions agree in exact arithmetic when both apply, and the reflection route is the one that is run.