Solving large scale eigenvalue problems

Lecture 8, April 18, 2018: Krylov spaces

http://people.inf.ethz.ch/arbenz/ewp/

Peter Arbenz
Computer Science Department, ETH Zürich
E-mail: arbenz@inf.ethz.ch

Survey of today's lecture

We are back at single vector iterations. But now we want to extract more information from the data we generate.

- Krylov (sub)spaces
- Orthogonal bases for Krylov spaces

Introduction

- In power method: we contruct sequence of the form (up to normalization
$$\mathbf{x}, A\,\mathbf{x}, A^2\,\mathbf{x}, \ldots$$
- Information at k-th iteration step: $\mathbf{x}^{(k)} = A^k\mathbf{x}/\|A^k\mathbf{x}\|$.
- All other information discarded!
- What about keeping all the information (vectors)?
More memory space required!
- Can we extract more information from all the vectors?
Less computational work!

Introductory example

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

- Initial vector $\mathbf{x} = [1, \ldots, 1]^*$.
- Compute first three iterates of IVI: $\mathbf{x}^{(1)} = \mathbf{x}$, $\mathbf{x}^{(2)} = T^{-1}\mathbf{x}$, and $\mathbf{x}^{(3)} = T^{-2}\mathbf{x}$.
- Compute Rayleigh quotients $\rho^{(i)} = {\mathbf{x}^{(i)}}^T T\mathbf{x}^{(i)}/\|\mathbf{x}^{(i)}\|^2$.
- Compute Ritz values $\vartheta_j^{(k)}$ obtained by Rayleigh-Ritz procedure with $\text{span}(\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(k)})$, $k = 1, 2, 3$,

Introductory example (cont.)

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

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.

Results show that cost of three matrix vector multiplications can be much better exploited than with plain (inverse) vector iteration – at the expense of more memory space.

Krylov spaces: definition and basic properties

Definition 1

Krylov matrix generated by vector $\mathbf{x} \in \mathbb{R}^n$ and A:

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

Krylov (sub)space:

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

We can also write $\mathcal{K}_m(A, \mathbf{x}) = \{p(A)\mathbf{x} \mid p \in \mathbb{P}_{m-1}\}$ where $\mathbb{P}_d$ denotes set of polynomials of degree at most d.

Krylov spaces: definition and basic properties (cont.)

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

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

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

$$Q^{(k)^*}AQ^{(k)}\mathbf{y} = \vartheta^{(k)}\mathbf{y}. \tag{3}$$

If $(\vartheta_j^{(k)}, \mathbf{y}_j)$ is an eigenpair of (3) then $(\vartheta_j^{(k)}, Q^{(k)}\mathbf{y}_j)$ is a Ritz pair of A in $\mathcal{K}^m(\mathbf{x})$.

Krylov spaces: definition and basic properties (cont.)

Simple properties of Krylov spaces [2, 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.

Dimension of $\mathcal{K}^k(\mathbf{x}, A)$

What is the dimension of $\mathcal{K}^k(\mathbf{x})$?

Clearly, $\dim(\mathcal{K}^k(\mathbf{x})) \leq k \leq n$.

There must be a m for which $\mathcal{K}_1 \subsetneqq \mathcal{K}_2 \subsetneqq \cdots \subsetneqq \mathcal{K}_m = \mathcal{K}_{m+1} = \cdots$
We have

$$A^m\mathbf{x} = \alpha_0\,\mathbf{x} + \alpha_1 A\mathbf{x} + \alpha_2 A^2\mathbf{x} + \cdots + \alpha_{p-1}A^{m-1}\mathbf{x}$$

Thus, $K^{m+1}(\mathbf{x})$ has linearly depending columns.

If we reach m there cannot be a further increase of dimension later.

Dimension of $\mathcal{K}^k(\mathbf{x}, A)$ (cont.)

Let A be diagonalizable and $\mathbf{x} = \sum_{i=1}^{m} \mathbf{q}_i$, where $A\mathbf{q}_i = \lambda_i \mathbf{q}_i, \mathbf{q}_i \neq \mathbf{0}$, with distinct λ_i. Then,

$$\underbrace{[\mathbf{x}, A\mathbf{x}, \cdots, A^k\mathbf{x}]}_{n\times(k+1)} = \underbrace{[\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_m]}_{n\times m} \underbrace{\left[\begin{array}{c|c|c|c} 1 & \lambda_1 & \cdots & \lambda_1^k \\ 1 & \lambda_2 & \cdots & \lambda_2^k \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_m & \cdots & \lambda_m^k \end{array}\right]}_{m\times(k+1)}$$

For $k < m$, the $m \times (k+1)$ matrix on the right has linearly independent columns. (Relation to *Vandermonde* matrices!)

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

Polynomial basis for $\mathcal{K}^m$

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

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

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

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

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

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

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

Polynomial basis for $\mathcal{K}^m$ (cont.)

Let $Q \in \mathbb{R}^{n\times j}$ be matrix with orthonormal basis of $\mathcal{K}^j(\mathbf{x})$

Let $\tilde{A} = Q^*AQ$. The spectral decomposition

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

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

By construction the Ritz vectors are mutually orthogonal.

Furthermore,

$$A\mathbf{y}_i - \vartheta_i\mathbf{y}_i \perp \mathcal{K}^j(\mathbf{x}) \tag{4}$$

Polynomial basis for $\mathcal{K}^m$ (cont.)

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

Lemma 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*

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

Polynomial basis for $\mathcal{K}^m$ (cont.)

Proof. "$\Longleftarrow$"

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

$$\begin{aligned} \mathbf{y}_k^* \omega(A)\mathbf{x} &= \mathbf{y}_k^*(A - \vartheta_k \mathbf{I})\pi(A)\mathbf{x}, \qquad \text{here we use that } A = A^* \\ &= (A\mathbf{y}_k - \vartheta_k \mathbf{y}_k)^* \pi(A)\mathbf{x} \stackrel{(4)}{=} 0. \end{aligned} \tag{5}$$

"$\Longrightarrow$"

Let $\mathcal{S}_k \subset \mathcal{K}^j(\mathbf{x})$ be defined by

$$\mathcal{S}_k := (A - \vartheta_k I)\,\mathcal{K}^{j-1}(\mathbf{x}) = \{\tau(A)\mathbf{x} \mid \tau \in \mathbb{P}_{j-1}, \tau(\vartheta_k) = 0\},$$

(5) $\Longrightarrow$ $\mathbf{y}_k$ is orthogonal to $\mathcal{S}_k$. $\mathcal{S}_k$ has dimension $j-1$.
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 $\mathcal{S}_k$. □

Polynomial basis for $\mathcal{K}^m$ (cont.)

We define the polynomials

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

(Normalized) Ritz vector $\mathbf{y}_k$ can be represented in the form

$$\mathbf{y}_k = \frac{\pi_k(A)\mathbf{x}}{\|\pi_k(A)\mathbf{x}\|}, \tag{6}$$

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

Polynomial basis for $\mathcal{K}^m$ (cont.)

By the first part of Lemma 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 $\psi(A)\mathbf{x} \in \mathcal{K}^j(\mathbf{x})$.

Let $\mathbf{u}_1, \cdots, \mathbf{u}_m$ be the eigenvectors of A corresponding to $\lambda_1 < \cdots < \lambda_m$ that span $\mathcal{K}^m(\mathbf{x})$.

Let $\|\mathbf{x}\| = 1$. Let $\varphi := \angle(\mathbf{x}, \mathbf{u}_1)$. Then

$$\|\mathbf{u}_1\mathbf{u}_1^*\mathbf{x}\| = \cos\varphi, \qquad \|(\mathbf{I} - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x}\| = \sin\varphi.$$

Polynomial basis for $\mathcal{K}^m$ (cont.)

Let $\mathbf{h} := (I - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x}/\|(I - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x}\|$.

Lemma 3 (Parlett [2])

For each $\pi \in \mathbb{P}_{j-1}$ *the Rayleigh quotient*

$$\rho(\pi(A)\mathbf{x}; A-\lambda_1 I) = \rho(\pi(A)\mathbf{x}; A)-\lambda_1 = \frac{(\pi(A)\mathbf{x})^*(A - \lambda_1 I)(\pi(A)\mathbf{x})}{\|\pi(A)\mathbf{x}\|^2}$$

satisfies the inequality

$$\rho(\pi(A)\mathbf{x}; A - \lambda_1 I) \leq (\lambda_m - \lambda_1)\left[\tan\varphi \, \frac{\|\pi(A)\mathbf{h}\|}{\pi(\lambda_1)}\right]^2.$$

Polynomial basis for $\mathcal{K}^m$ (cont.)

Proof.

With **h** from above we have the orthogonal decompositions

$$\mathbf{x} = \mathbf{u}_1\mathbf{u}_1^*\mathbf{x} + (I - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x} = \cos\varphi\,\mathbf{u}_1 + \sin\varphi\,\mathbf{h}$$

and

$$\mathbf{s} := \pi(A)\mathbf{x} = \cos\varphi\,\pi(A)\mathbf{u}_1 + \sin\varphi\,\pi(A)\mathbf{h}.$$

Thus,

$$\rho(\mathbf{s}; A - \lambda_1 I) = \frac{\cos^2\varphi\,\mathbf{u}_1^*(A-\lambda_1 I)\pi^2(A)\mathbf{u}_1 + \sin^2\varphi\,\mathbf{h}^*(A-\lambda_1 I)\pi^2(A)\mathbf{h}}{\|\mathbf{s}\|^2}$$

$$\overset{(A\mathbf{u}_1 = \lambda_1\mathbf{u}_1)}{=} \frac{\sin^2\varphi\,\mathbf{h}^*(A-\lambda_1 I)\pi^2(A)\mathbf{h}}{\|\mathbf{s}\|^2}.$$

Polynomial basis for $\mathcal{K}^m$ (cont.)

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

$$\mathbf{w}^*(A - \lambda_1 I)\mathbf{w} \le (\lambda_m - \lambda_1)\|\mathbf{w}\|^2 \text{ for all } \mathbf{w} \in \mathcal{R}(\mathbf{u}_1)^\perp.$$

Setting $\mathbf{w} = \pi(A)\mathbf{h}$ we obtain

$$\rho(\mathbf{s}; A - \lambda_1 I) \le \sin^2\varphi\,(\lambda_m - \lambda_1)\frac{\|\pi(A)\mathbf{h}\|^2}{\|\pi(A)\mathbf{x}\|^2}.$$

With

$$\|\mathbf{s}\|^2 = \|\pi(A)\mathbf{x}\|^2 = \sum_{\ell=1}^{m} \pi^2(\lambda_\ell)(\mathbf{x}^*\mathbf{u}_\ell)^2 \ge \pi^2(\lambda_1)\cos^2\varphi$$

we obtain the claim.

Error bounds by Saad

- For simplicity we consider convergence of Ritz values $\vartheta_1^{(j)}$ to λ_1.
- The error bounds to be presented have been published by Saad [3]. We follow the presentation in Parlett [3].
- The error bounds for $\vartheta_1^{(j)} - \lambda_1$ are obtained by carefully selecting the polynomial π in Lemma 3.
- Of course we would like $\pi(A)$ to be as small as possible and $\pi(\lambda_1)$ to be as large as possible.

Error bounds by Saad (cont.)

First, by the definition of **h**, we have

$$\|\pi(A)\mathbf{h}\|^2 = \frac{\|\pi(A)(I - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x}\|^2}{\|(I - \mathbf{u}_1\mathbf{u}_1^*)\mathbf{x}\|^2} = \frac{\|\pi(A)\sum_{\ell=2}^{m}(\mathbf{u}_\ell^*\mathbf{x})\mathbf{u}_\ell\|^2}{\|\sum_{\ell=2}^{m}(\mathbf{u}_\ell^*\mathbf{x})\mathbf{u}_\ell\|^2}$$

$$= \frac{\sum_{\ell=2}^{m}(\mathbf{u}_\ell^*\mathbf{x})^2\pi^2(\lambda_\ell)}{\sum_{\ell=2}^{m}(\mathbf{u}_\ell^*\mathbf{x})^2} \le \max_{2\le\ell\le m}\pi^2(\lambda_\ell) \le \max_{\lambda_2\le\lambda\le\lambda_m}\pi^2(\lambda).$$

The last inequality is important! In this step the search of a maximum in a few selected points $(\lambda_2, \ldots, \lambda_m)$ is replaced by a search of a maximum in a *whole interval* containing these points.

Notice that λ_1 is *outside* of this interval.

Error bounds by Saad (cont.)

Among all polynomials of given degree that take a fixed value $\pi(\lambda_1)$ the Chebyshev polynomial has the smallest maximum.

$$
\begin{aligned}
\min_{\pi\in\mathbb{P}_{j-1}} \max_{\lambda_2\le\lambda\le\lambda_m} \frac{|\pi(\lambda)|}{|\pi(\lambda_1)|} &= \frac{\max\limits_{\lambda_2\le\lambda\le\lambda_m} T_{j-1}(\lambda;[\lambda_2,\lambda_m])}{T_{j-1}(\lambda_1;[\lambda_2,\lambda_m])} \\
&= \frac{1}{T_{j-1}(\lambda_1;[\lambda_2,\lambda_m])} \\
&= \frac{1}{T_{j-1}(1+2\gamma)}, \qquad \gamma = \frac{\lambda_2-\lambda_1}{\lambda_m-\lambda_2}.
\end{aligned}
$$

$T_{j-1}(1+2\gamma)$ is the value of the Chebyshev polynomial corresponding to the 'normal' interval $[-1, 1]$.

Error bounds by Saad (cont.)

The point $1 + 2\gamma$ is obtained if the affine transformation

$$[\lambda_2, \lambda_m] \ni \lambda \longrightarrow \frac{2\lambda - \lambda_2 - \lambda_m}{\lambda_1 - \lambda_2} \in [-1,\ 1]$$

is applied to λ_1.

Error bounds by Saad (cont.)

Thus we have proved the first part of the following

Theorem 4 (Saad [3])

Let $\vartheta_1^{(j)}, \ldots, \vartheta_j^{(j)}$ be the Ritz values of A in $\mathcal{K}^j(\mathbf{x})$ and let $(\lambda_\ell, \mathbf{u}_\ell)$, $\ell = 1, \ldots, m$, be the eigenpairs of A (in $\mathcal{K}^m(\mathbf{x})$). Then

$$0 \le \vartheta_1^{(j)} - \lambda_1 \le (\lambda_m - \lambda_1)\left[\tan\varphi \frac{1}{T_{j-1}(1+2\gamma)}\right]^2, \quad \gamma = \frac{\lambda_2 - \lambda_1}{\lambda_m - \lambda_2},$$

and

$$\tan\angle(\mathbf{u}_1, \text{projection of } \mathbf{u}_1 \text{ on } \mathcal{K}^j) \le \tan\varphi \cdot \frac{1}{T_{j-1}(1+2\gamma)}.$$

Error bounds by Saad (cont.)

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

$$\mathbf{x} = \mathbf{u}_1 \cos\varphi + \mathbf{h}\sin\varphi.$$

Then

$$\mathbf{s} = \pi(A)\mathbf{x} = \pi(\lambda_1)\mathbf{u}_1 \cos\varphi + \pi(A)\mathbf{h}\sin\varphi$$

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

$$\tan\angle(\mathbf{s}, \mathbf{u}_1) = \frac{\sin\varphi\ \|\pi(A)\mathbf{h}\|}{\cos\varphi\ |\pi(\lambda_1)|}.$$

The rest is similar as above. □

Error bounds by Saad (cont.)

Theorem 4 can be rewritten to give error bounds for $\lambda_m - \vartheta_j^{(j)}$ but also for the interior eigenvalues, see lecture notes.
For the largest eigenvalue we have

$$0 \leq \lambda_m - \vartheta_j^{(j)} \leq (\lambda_m - \lambda_1) \tan^2 \varphi_m \frac{1}{T_{j-1}(1 + 2\gamma_m)^2}, \qquad (7)$$

with

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

From more general results one sees that the eigenvalues at the beginning and at the end of the spectrum are approximated the quickest.

Error bounds by Saad (cont.)

If the Lanczos algorithmus is applied with $(A-\sigma I)^{-1}$ then we form Krylov spaces $\mathcal{K}^j(\mathbf{x},(A-\sigma I)^{-1})$. Here the eigenvalues are $\frac{1}{\hat\lambda_1} \geq \frac{1}{\hat\lambda_2} \geq \cdots \geq \frac{1}{\hat\lambda_m}$, $\hat\lambda_i = \lambda_i - \sigma$.
Eq. (7) then becomes

$$0 \leq \frac{1}{\hat\lambda_1} - \frac{1}{\hat\vartheta_m^{(j)}} \leq \left(\frac{1}{\hat\lambda_1} - \frac{1}{\hat\lambda_m}\right) \frac{\tan^2\varphi_1}{T_{j-1}(1+2\hat\gamma_1)^2}, \quad \hat\gamma_1 = \frac{\frac{1}{\hat\lambda_1} - \frac{1}{\hat\lambda_2}}{\frac{1}{\hat\lambda_2} - \frac{1}{\hat\lambda_m}}.$$

One can show that

$$1 + 2\hat\gamma_1 = 2(1+\hat\gamma_1) - 1 \geq 2\frac{\hat\lambda_2}{\hat\lambda_1} - 1 > 1.$$

Error bounds by Saad (cont.)

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

$$\frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\vartheta}_1^{(j)}} \leq c_1 \left(\frac{1}{T_{j-1}(2\frac{\hat{\lambda}_2}{\hat{\lambda}_1} - 1)} \right)^2 \qquad (8)$$

With the simple inverse vector iteration we had

$$\frac{1}{\hat{\lambda}_1} - \frac{1}{\hat{\lambda}_1^{(j)}} \leq c_2 \left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2} \right)^{2(j-1)} \qquad (9)$$

Error bounds by Saad (cont.)

$\frac{\hat{\lambda}_2}{\hat{\lambda}_1}$	$j = 5$	$j = 10$	$j = 15$	$j = 20$	$j = 25$
2.0	$\frac{3.00e-06}{3.91e-03}$	$\frac{6.63e-14}{3.81e-06}$	$\frac{1.46e-21}{3.72e-09}$	$\frac{3.24e-29}{3.63e-12}$	$\frac{7.17e-37}{3.55e-15}$
1.1	$\frac{2.71e-02}{4.66e-01}$	$\frac{5.45e-05}{1.79e-01}$	$\frac{1.08e-07}{6.93e-02}$	$\frac{2.14e-10}{2.67e-02}$	$\frac{4.24e-13}{1.03e-02}$
1.01	$\frac{5.60e-01}{9.23e-01}$	$\frac{1.04e-01}{8.36e-01}$	$\frac{1.48e-02}{7.56e-01}$	$\frac{2.02e-03}{6.85e-01}$	$\frac{2.75e-04}{6.20e-01}$

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

An orthogonal basis for $\mathcal{K}_m$

Problem: The matrix

$$K_m(A, \mathbf{x}) := \left[\begin{array}{c|c|c|c} \mathbf{x} & A\mathbf{x} & \cdots & A^{m-1}\mathbf{x} \end{array} \right]$$

becomes more and more ill-conditioned as m increases.
(Remember vector iteration for computing largest eigenvalue.)

Solution: We have to find a well-conditioned basis of $\mathcal{K}_m$.

Arnoldi & Lanczos algorithms

Task: For $j = 1, 2, \dots, m$, compute orthonormal bases $\{\mathbf{v}_1, \dots, \mathbf{v}_j\}$ for the Krylov spaces

$$\mathcal{K}_j = \text{span}\left\{\mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \dots, A^{j-1}\mathbf{x}\right\}.$$

The algorithms that do this are

- Lanczos algorithm for A symmetric/Hermitian.
- Arnoldi algorithm for A nonsymmetric.

Difficulty: Because of ill-conditioning, do not want to *explicitly* form $\mathbf{x}, A\mathbf{x}, \dots, A^j\mathbf{x}$.

Arnoldi & Lanczos algorithms (cont.)

Instead of using $A^j\mathbf{x}$ we proceed with $A\mathbf{v}_j$.
(Notice that $A\mathbf{v}_i \in \mathcal{K}_{i+1} \subset \mathcal{K}_j$ for all $i < j$.)

Orthogonalize $A\mathbf{v}_j$ against $\mathbf{v}_1, \ldots, \mathbf{v}_j$ by the Gram–Schmidt:

$$\mathbf{w}_j = A\mathbf{v}_j - \sum_{i=1}^{j} \mathbf{v}_i h_{ij}.$$

$\mathbf{w}_j$ points in the desired new direction (unless it is $\mathbf{0}$). Therefore,

$$\mathbf{v}_{j+1} = \mathbf{w}_j / \|\mathbf{w}_j\|.$$

Arnoldi algo to compute orthonormal basis of Krylov space

Let $A \in \mathbb{R}^{n \times n}$. This algorithm computes orthonormal basis for $\mathcal{K}^j(\mathbf{x})$.
$\mathbf{v}_1 = \mathbf{x}/\|\mathbf{x}\|_2$;
for $j = 1, \ldots$ **do**
  $\mathbf{r} := A\mathbf{v}_j$;
  for $i = 1, \ldots, j$ **do** {Gram-Schmidt orthogonalization}
    $h_{ij} := \mathbf{v}_i^* \mathbf{r}, \quad \mathbf{r} := \mathbf{r} - \mathbf{v}_i h_{ij}$;
  end for
  $h_{j+1,j} := \|\mathbf{r}\|$;
  if $h_{j+1,j} = 0$ **then** {Found an invariant subspace}
    return $(\mathbf{v}_1, \ldots, \mathbf{v}_j, H \in \mathbb{R}^{j \times j})$
  end if
  $\mathbf{v}_{j+1} = \mathbf{r}/h_{j+1,j}$;
end for
return $(\mathbf{v}_1, \ldots, \mathbf{v}_{j+1}, H \in \mathbb{R}^{j+1 \times j})$

Arnoldi relation

The Arnoldi algorithm returns if $h_{m+1,m} = 0$, i.e., if it has found an invariant subspace. The vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ then form an invariant subspace of A,

$$AV_m = V_m H_m, \qquad V_m = [\mathbf{v}_1, \ldots, \mathbf{v}_m].$$

The eigenvalues of H_m are eigenvalues of A as well and the Ritz vectors are eigenvectors of A.

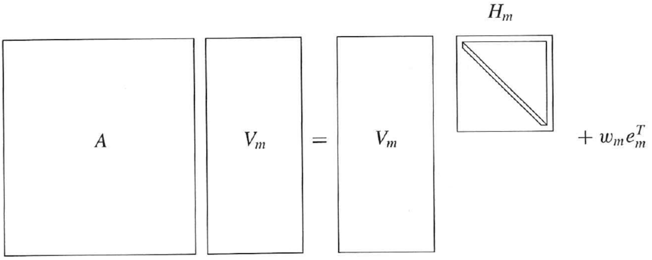
This algorithm costs j matrix-vector multiplications, $n^2/2 + \mathcal{O}(n)$ inner products, and the same number of _axpy's.

In general, we cannot afford to store the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ because of limited memory space. So, we stop prematurely

Arnoldi relation (cont.)

Define $V_m := [\mathbf{v}_1, \ldots, \mathbf{v}_m]$. Then we get the Arnoldi relation

$$AV_m = V_m H_m + \mathbf{w}_m \mathbf{e}_m^T = V_{m+1} \bar{H}_m.$$



Picture from Saad: *Iterative Methods for Sparse Linear Systems*:

Arnoldi relation (cont.)

Here,

$$\bar{H}_m = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1,m} \\ h_{21} & h_{22} & \cdots & h_{2,m} \\ & h_{3,2} & \cdots & h_{3,m} \\ & & \ddots & \vdots \\ & & & h_{m+1,m} \end{bmatrix}$$

The square matrix $H_m \in \mathbb{R}^{m \times m}$ is obtained from $\bar{H}_m \in \mathbb{R}^{(m+1) \times m}$ by deleting the last row.

Notice that

$$H_m = V_m^T A V_m.$$

If A is symmetric $\Longrightarrow H_m \equiv T_m$ is symmetric and thus tridiagonal!

The Lanczos relation is $AV_m = V_m T_m + \mathbf{w}_m \mathbf{e}_m^T = V_{m+1} \bar{T}_m$.

References

[1] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th edition. The Johns Hopkins University Press, Baltimore, 2012.

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

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