Solving large scale eigenvalue problems

Lecture 6, March 28, 2018: Simple vector iterations

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

The power method (aka. vector iteration) is the simplest method to compute a single eigenvector of a matrix.

- Simple vector iteration (power method)
- Inverse vector iteration
- Rayleigh quotient iteration (RQI)

Simple vector iteration

Let $A \in \mathbb{R}^{n\times n}$.
Starting with arbitrary initial vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$ we form the vector sequence $\left\{\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ defined by

$$\boxed{\mathbf{x}^{(k)} := A\mathbf{x}^{(k-1)}, \qquad k = 1, 2, \ldots} \qquad (*)$$

Clearly,

$$\mathbf{x}^{(k)} := A^k \, \mathbf{x}^{(0)}.$$

We will show that the $\mathbf{x}^{(k)}$ 'converge' to 'the' eigenvector associated with the eigenvalue of largest magnitude.

Algorithm: Simple vector iteration

1: Choose a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$ with $\|\mathbf{x}^{(0)}\| = 1$.
2: $k = 0$.
3: **repeat**
4: $\quad k := k + 1;$
5: $\quad \mathbf{y}^{(k)} := A\,\mathbf{x}^{(k-1)};$
6: $\quad \mu_k := \|\mathbf{y}^{(k)}\|;$
7: $\quad \mathbf{x}^{(k)} := \mathbf{y}^{(k)}/\mu_k;$
8: **until** a convergence criterion is satisfied

Vectors $\mathbf{x}^{(k)}$ have all norm (length) one. $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ is a sequence on the unit sphere of $\mathbb{R}^n$.

Here, the maximum norm is polular as well: $\|\mathbf{y}\|_\infty = \max_i |\mathbf{y}_i|$.

Important note

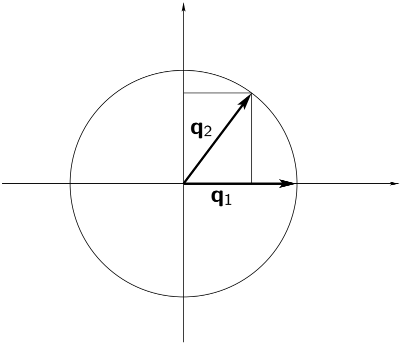
- Let $A = USU^*$ be the Schur decomposition of A. Then,

$$U^*\mathbf{x}^{(k)} := SU^*\mathbf{x}^{(k-1)} \qquad \text{and} \qquad U^*\mathbf{x}^{(k)} := S^k U^*\mathbf{x}^{(0)}.$$

- U unitary: $\|\mathbf{x}^{(k)}\| \quad \Longrightarrow \quad \|U^*\mathbf{x}^{(k)}\| = 1$ for all k.
- If sequence $\left\{\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ converges to $\mathbf{x}_*$ then sequence $\left\{\mathbf{y}^{(k)} = U^*\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ converges to $\mathbf{y}_* = U^*\mathbf{x}_*$.
- So, for convergence analysis: can assume w.l.o.g. that A is upper triangular.
- If we assumed that A is symmetric then for a convergence analysis we could restrict ourselves to *diagonal* matrices.
- Note that some performance issues are excluded here.

Intermezzo: Angles between vectors

Let $\mathbf{q}_1$ and $\mathbf{q}_2$ be *unit* vectors.
Angle between vectors $\mathbf{q}_1$ and $\mathbf{q}_2$:



Intermezzo: Angles between vectors (cont.)

The length of the orthogonal projection of $\mathbf{q}_2$ on span$\{\mathbf{q}_1\}$ is:

$$c := \|\mathbf{q}_1\mathbf{q}_1{}^*\mathbf{q}_2\| = |\mathbf{q}_1{}^*\mathbf{q}_2| \leq 1.$$

The length of the orthogonal projection of $\mathbf{q}_2$ on span$\{\mathbf{q}_1\}^\perp$ is

$$s := \|(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1{}^*)\mathbf{q}_2\|. \qquad (+)$$

As $\mathbf{q}_1\mathbf{q}_1^*$ is an orthogonal projection, by Pythagoras' formula:

$$1 = \|\mathbf{q}_2\|^2 = \|\mathbf{q}_1\mathbf{q}_1{}^*\mathbf{q}_2\|^2 + \|(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1{}^*)\mathbf{q}_2\|^2 = s^2 + c^2.$$

From $(+)$:

$$\begin{aligned} s^2 &= \|(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1{}^*)\mathbf{q}_2\|^2 \\ &= \mathbf{q}_2{}^*(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1{}^*)\mathbf{q}_2 \\ &= \mathbf{q}_2{}^*\mathbf{q}_2 - (\mathbf{q}_2{}^*\mathbf{q}_1)(\mathbf{q}_1{}^*\mathbf{q}_2) \\ &= 1 - c^2 \end{aligned}$$

Intermezzo: Angles between vectors (cont.)

So, there is a number, say, ϑ, $0 \le \vartheta \le \frac{\pi}{2}$, such that $c = \cos\vartheta$ and $s = \sin\vartheta$. This uniquely determined number ϑ is the angle between the vectors $\mathbf{q}_1$ and $\mathbf{q}_2$:

$$\vartheta = \angle(\mathbf{q}_1, \mathbf{q}_2).$$

The generalization to arbitrary vectors is straightforward.

Definition

The **angle** θ between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\vartheta = \angle(\mathbf{x}, \mathbf{y}) = \arcsin\left(\left\|\left(I - \frac{\mathbf{x}\mathbf{x}^*}{\|\mathbf{x}\|^2}\right)\frac{\mathbf{y}}{\|\mathbf{y}\|}\right\|\right) = \arccos\left(\frac{|\mathbf{x}^*\mathbf{y}|}{\|\mathbf{x}\|\|\mathbf{y}\|}\right).$$

Convergence analysis

Assume that

$$S = \begin{bmatrix} \lambda_1 & \mathbf{s}_1^* \\ \mathbf{0} & S_2 \end{bmatrix}, \qquad (S_2 \text{ upper triangular}) \tag{1}$$

has eigenvalues

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|.$$

Eigenvector of S corresponding to largest eigenvalue λ_1 is $\mathbf{e}_1$.

We will show that the iterates $\mathbf{x}^{(k)}$ converge to $\mathbf{e}_1$.

More precisely, we will show that $\angle(\mathbf{x}^{(k)}, \mathbf{e}_1) \longrightarrow 0$ as $k \to \infty$.

Convergence analysis (cont.)

Let

$$\mathbf{x}^{(k)} = \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{pmatrix} =: \begin{pmatrix} x_1^{(k)} \\ \mathbf{x}_2^{(k)} \end{pmatrix}$$

with $\|\mathbf{x}^{(k)}\| = 1$. Then,

$$\sin\vartheta^{(k)} := \sin(\angle(\mathbf{x}^{(k)}, \mathbf{e}_1)) = \sqrt{\sum_{i=2}^{n} |x_i^{(k)}|^2} = \sqrt{\frac{\sum_{i=2}^{n} |x_i^{(k)}|^2}{\sum_{i=1}^{n} |x_i^{(k)}|^2}}.$$

The last expression is for non-normalized vectors $\mathbf{x}^{(k)}$, cf. $(*)$.

Convergence analysis (cont.)

First, we simplify the form of S in (1), by eliminating $\mathbf{s}_1^*$,

$$\begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{s}_1^* \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{s}_1^* \\ \mathbf{0} & S_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0}^* \\ \mathbf{0} & S_2 \end{bmatrix}.$$

The vector $\mathbf{t}$ that realizes this transformation has to satisfy

$$\lambda_1 \mathbf{t}^* + \mathbf{s}_1^* - \mathbf{t}^* S_2 = \mathbf{0}^* \quad \Longleftrightarrow \quad \mathbf{s}_1^* = \mathbf{t}^*(S_2 - \lambda_1 I).$$

This equation has a solution if and only if $\lambda_1 \notin \sigma(S_2)$ which is the case by assumption.

Remark: $[1, -\mathbf{t}^*]$ is left eigenvector of S associated with λ_1.

Convergence analysis (cont.)

So, we have

$$\mathbf{x}^{(k)} = \begin{pmatrix} x_1^{(k)} \\ \mathbf{x}_2^{(k)} \end{pmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{s}_1^* \\ \mathbf{0} & S_2 \end{bmatrix} \mathbf{x}^{(k-1)} = \cdots = \begin{bmatrix} \lambda_1 & \mathbf{s}_1^* \\ \mathbf{0} & S_2 \end{bmatrix}^k \mathbf{x}^{(0)}$$

$$= \begin{bmatrix} 1 & \mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{0}^* \\ \mathbf{0} & S_2 \end{bmatrix}^k \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \begin{pmatrix} x_1^{(0)} \\ \mathbf{x}_2^{(0)} \end{pmatrix}.$$

We define

$$\mathbf{y}^{(k)} := \frac{1}{\lambda_1^k} \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \mathbf{x}^{(k)} \tag{2}$$

Convergence analysis (cont.)

$$\begin{aligned}
\mathbf{y}^{(k)} &= \frac{1}{\lambda_1^k} \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} S\,\mathbf{x}^{(k-1)} \\
&= \left(\frac{1}{\lambda_1} \begin{bmatrix} \lambda_1 & \mathbf{0}^* \\ \mathbf{0} & S_2 \end{bmatrix} \right) \frac{1}{\lambda_1^{k-1}} \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \mathbf{x}^{(k-1)} \\
&= \begin{bmatrix} 1 & \mathbf{0}^* \\ \mathbf{0} & \frac{1}{\lambda_1} S_2 \end{bmatrix} \begin{pmatrix} y_1^{(k-1)} \\ \mathbf{y}_2^{(k-1)} \end{pmatrix} = \begin{bmatrix} 1 & \mathbf{0}^* \\ \mathbf{0} & \frac{1}{\lambda_1} S_2 \end{bmatrix} \mathbf{y}^{(k-1)}.
\end{aligned}$$

Let us assume that $y_1^{(0)} = 1$. Then, $y_1^{(k)} = 1$ for all k.

Need to show that $\mathbf{y}_2^{(k)}$ goes to zero as $k \to \infty$, and how fast.

Convergence analysis (cont.)

$$\mathbf{y}_2^{(k)} = \frac{1}{\lambda_1} S_2 \mathbf{y}_2^{(k-1)}$$

$$\frac{1}{\lambda_1} S_2 = \begin{bmatrix} \mu_2 & * & \cdots & * \\ & \mu_3 & \cdots & * \\ & & \ddots & \vdots \\ & & & \mu_n \end{bmatrix}, \qquad |\mu_k| = \frac{|\lambda_k|}{|\lambda_1|} < 1.$$

Convergence analysis (cont.)

Theorem

Let $|||\cdot|||$ *be any matrix norm. Then*

$$\lim_{k\to\infty} |||M^k|||^{1/k} = \rho(M) = \max_i |\lambda_i(M)|. \qquad (3)$$

Proof.

See Horn-Johnson, *Matrix Analysis*, 1985, pp.297-299. □

So, for *any* $\varepsilon > 0$ there is an integer $K(\varepsilon)$ such that

$$|||M^k|||^{1/k} \leq \rho(M) + \varepsilon, \qquad \text{for all } k > K(\varepsilon). \qquad (4)$$

Convergence analysis (cont.)

So, for *any* $\varepsilon > 0$ there is an integer $K(\varepsilon)$ such that

$$|||M^k|||^{1/k} \le \rho(M) + \varepsilon, \qquad \text{for all } k > K(\varepsilon). \tag{4}$$

In our case:

$$\rho\left(\frac{1}{\lambda_1} S_2\right) = |\mu_2| < 1.$$

Can choose ε such that $|\mu_2| + \varepsilon < 1$. For any such ε we have

$$\sin(\angle(\mathbf{y}^{(k)}, \mathbf{e}_1)) = \frac{\|\mathbf{y}_2^{(k)}\|}{\|\mathbf{y}^{(k)}\|} = \frac{\|\mathbf{y}_2^{(k)}\|}{\sqrt{1 + \|\mathbf{y}_2^{(k)}\|}}$$

$$\le \|\mathbf{y}_2^{(k)}\| \le \|\frac{1}{\lambda_1} S\|^k \|\mathbf{y}_2^{(0)}\| \le (|\mu_2| + \varepsilon)^k \|\mathbf{y}_2^{(0)}\|.$$

Convergence analysis (cont.)

Analogous result for $\mathbf{x}^{(k)}$:

$$\begin{pmatrix} x_1^{(k)} \\ \mathbf{x}_2^{(k)} \end{pmatrix} = \lambda_1^k \begin{bmatrix} 1 & \mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \begin{pmatrix} y_1^{(k)} \\ \mathbf{y}_2^{(k)} \end{pmatrix},$$

So,

$$|x_1^{(k)}| = \lambda_1^k |y_1^{(k)} + \mathbf{t}^* \mathbf{y}_2^{(k)}| = \lambda_1^k |1 + \mathbf{t}^* \mathbf{y}_2^{(k)}|.$$

Since $\|\mathbf{y}_2^{(k)}\| \le (|\mu_2| + \varepsilon)^k \|\mathbf{y}_2^{(0)}\|$, there is a $\tilde{K} \ge K(\varepsilon)$ such that

$$|\mathbf{t}^* \mathbf{y}_2^{(k)}| < \frac{1}{2}, \qquad \forall\, k > \tilde{K}$$

and

$$|1 + \mathbf{t}^* \mathbf{y}_2^{(k)}| > \frac{1}{2}, \qquad \forall\, k > \tilde{K}.$$

Convergence analysis (cont.)

$$\begin{aligned}
\sin(\angle(\mathbf{x}^{(k)}, \mathbf{e}_1)) = \frac{\|\mathbf{x}_2^{(k)}\|}{\|\mathbf{x}^{(k)}\|} &= \frac{\|\mathbf{y}_2^{(k)}\|}{\sqrt{|y_1^{(k)} + \mathbf{t}^*\mathbf{y}_2^{(k)}|^2 + \|\mathbf{y}_2^{(k)}\|^2}} \\
&< \frac{\|\mathbf{y}_2^{(k)}\|}{\sqrt{\frac{1}{4}|y_1^{(k)}|^2 + \frac{1}{4}\|\mathbf{y}_2^{(k)}\|^2}} \\
&\leq 2\|\mathbf{y}_2^{(k)}\| \\
&\leq 2(|\mu_2| + \varepsilon)^k \|\mathbf{y}_2^{(0)}\|.
\end{aligned}$$

Convergence analysis (cont.)

We assumed $y_1^{(0)} = 1$ or, more generally, $y_1^{(0)} \neq 0$.

From (2) we have

$$\mathbf{y}^{(0)} := \begin{bmatrix} 1 & -\mathbf{t}^* \\ \mathbf{0} & I \end{bmatrix} \mathbf{x}^{(0)}.$$

Thus,

$$y_1^{(0)} = [1, -\mathbf{t}^*]\mathbf{x}^{(0)}.$$

Therefore,

$$y_1^{(0)} \neq 0 \quad \Longleftrightarrow \quad [1, -\mathbf{t}^*]\mathbf{x}^{(0)} \neq 0.$$

Remember: $[1, -\mathbf{t}^*]$ is the left eigenvector of S associated with λ_1.

Convergence analysis (cont.)

Since we can choose ε arbitrarily small, we have proved

Theorem

Let the eigenvalues of $A \in \mathbb{R}^{n\times n}$ be arranged such that $|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_n|$. Let $\mathbf{u}_1$ and $\mathbf{v}_1$ be right and left eigenvectors of A corresponding to λ_1, respectively. Then, the vector sequence generated by simple vector iteration converges to $\mathbf{u}_1$ in the sense that

$$\sin\vartheta^{(k)} = \sin(\angle(\mathbf{x}^{(k)},\mathbf{u}_1)) \le c \cdot \left|\frac{\lambda_2}{\lambda_1}\right|^k \qquad (5)$$

provided that $\mathbf{v}_1^*\mathbf{x}^{(0)} \neq 0$.

Remarks

- μ_k in the algorithm converges to $|\lambda_1|$. The sign of λ_1 can be found by comparing single components of $\mathbf{y}^{(k)}$ and $\mathbf{x}^{(k-1)}$.
- If $\mathbf{v}_1^*\mathbf{x}^{(0)} = 0$ then the vector iteration converges to an eigenvector corresponding to the second largest eigenvalue. Rounding errors usually prevent this: after a long initial phase the $\mathbf{x}^{(k)}$ turn to $\mathbf{u}_1$.
- Convergence of vector iteration is faster the smaller $|\lambda_2|/|\lambda_1|$. Convergence may be very slow (cf. QR algorithm).
- In case that $\lambda_1 \neq \lambda_2$ but $|\lambda_1| = |\lambda_2|$ there may be no convergence at all. An example is

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{x}^{(0)} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Symmetric case

The symmetric case is treated very similarly to the non-symmetric case. However, we can approximate the eigenvalue by the Rayleigh quotient.

$$\lambda^{(k)} := \mathbf{x}^{(k)^*} A \mathbf{x}^{(k)}, \qquad \|\mathbf{x}^{(k)}\| = 1.$$

Note that we form $A\mathbf{x}^{(k)}$ during the iteration anyway.

Simple vector iteration for Hermitian matrices

Choose a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$ with $\|\mathbf{x}^{(0)}\| = 1$.
$\mathbf{y}^{(0)} := A\mathbf{x}^{(0)}$.
$\lambda^{(0)} := \mathbf{y}^{(0)^*}\mathbf{x}^{(0)}$.
$k := 0$.
while $\|\mathbf{y}^{(k)} - \lambda^{(k)}\mathbf{x}^{(k)}\| > \mathrm{tol}$ **do**
    $k := k + 1$;
    $\mathbf{x}^{(k)} := \mathbf{y}_{k-1}/\|\mathbf{y}_{k-1}\|$;
    $\mathbf{y}^{(k)} := A\mathbf{x}^{(k)}$;
    $\lambda^{(k)} := \mathbf{y}^{(k)^*}\mathbf{x}^{(k)}$;
end while

Simple symmetric vector iteration: theory

Theorem

Let $A = A^*$ with spectral decomposition

$$A = U\Lambda U^*, \qquad U = [\mathbf{u}_1, \ldots, \mathbf{u}_n], \quad \Lambda = diag(\lambda_1, \ldots, \lambda_n).$$

Then, the simple vector iteration computes sequences $\left\{\lambda^{(k)}\right\}_{k=0}^{\infty}$ and $\left\{\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ that converge linearly towards the largest eigenvalue λ_1 of A and the corresponding eigenvector $\mathbf{u}_1$ provided that $\mathbf{u}_1^*\mathbf{x}^{(0)} \neq 0$. The convergence rates are given by

$$\sin\vartheta^{(k)} \le \left|\frac{\lambda_2}{\lambda_1}\right|^k \sin\vartheta^{(0)}, \qquad |\lambda_1 - \lambda^{(k)}| \le (\lambda_1 - \lambda_n) \left|\frac{\lambda_2}{\lambda_1}\right|^{2k} \sin^2\vartheta^{(0)}.$$

where $\vartheta^{(k)} = \angle(\mathbf{x}^{(k)}, \mathbf{u}_1)$ and $\lambda^{(k)} = \rho(\mathbf{x}^{(k)})$.

Inverse vector iteration

- Convergence of simple vector iteration is potentially very slow
- Polynomial in A has the same eigenvectors as A.
- May try to find a polynomial that enhances the eigenvalue that we are looking for. Not successful in the most critical case when the wanted eigenvalue is very close to unwanted.
- **Shift-and-invert** spectral transformation is the way to go.
- Transform the matrix by the rational function

$$f(\lambda) = 1/(\lambda - \sigma)$$

  where σ is so-called **shift** close to the desired eigenvalue.
- Inverse vector iteration: Simple vector iteration with matrix $(A - \sigma I)^{-1}$

Inverse vector iteration

Choose starting vector $\mathbf{x}_0 \in \mathbb{R}^n$ and a shift σ.
Compute the LU factorization of $A - \sigma I$: $LU = P(A - \sigma I)$
$\mathbf{y}^{(0)} := U^{-1} L^{-1} P \mathbf{x}^{(0)}$.
$\mu^{(0)} = \mathbf{y}^{(0)^*} \mathbf{x}^{(0)}, \quad \lambda^{(0)} := \sigma + 1/\mu^{(0)}. \quad k := 0.$
while $\|\mathbf{x}^{(k)} - \mathbf{y}^{(k)}/\mu^{(k)}\| > \mathrm{tol}\|\mathbf{y}^{(k)}\|$ **do**
  $k := k + 1.$
  $\mathbf{x}^{(k)} := \mathbf{y}_{k-1}/\|\mathbf{y}_{k-1}\|.$
  $\mathbf{y}^{(k)} := U^{-1} L^{-1} P \mathbf{x}^{(k)}.$
  $\mu^{(k)} := \mathbf{y}^{(k)^*} \mathbf{x}^{(k)}, \quad \lambda^{(k)} := \sigma + 1/\mu^{(k)}.$
end while

Theorem

A SPD with spectral decomposition $A = U\Lambda U^*$. *Let* $\lambda_1', \ldots, \lambda_n'$ *be a renumeration of the eigenvalues of A such that*

$$\frac{1}{|\lambda_1' - \sigma|} > \frac{1}{|\lambda_2' - \sigma|} \geq \cdots \geq \frac{1}{|\lambda_n' - \sigma|}$$

If $\mathbf{u}_1'^{\,*}\mathbf{x}^{(0)} \neq 0$, *then inverse vector iteration constructs sequences* $\left\{\lambda^{(k)}\right\}_{k=0}^{\infty}$ *and* $\left\{\mathbf{x}^{(k)}\right\}_{k=0}^{\infty}$ *that converge linearly towards that eigenvalue* λ_1' *closest to the shift* σ *and to the corresponding eigenvector* $\mathbf{u}_1'$. *The bounds*

$$\sin\vartheta^{(k)} \leq \left|\frac{\lambda_1' - \sigma}{\lambda_2' - \sigma}\right|^k \sin\vartheta^{(0)}, \qquad \lambda^{(k)} - \lambda_1 \leq \delta \left|\frac{\lambda_1' - \sigma}{\lambda_2' - \sigma}\right|^{2k} \sin^2\vartheta^{(0)}.$$

hold with $\vartheta^{(k)} = \angle(\mathbf{x}^{(k)}, \mathbf{u}_1)$ *and* $\delta = \mathrm{spread}(\sigma((A - \sigma I)^{-1}))$.

Discussion of inverse iteration

- Can compute eigenvectors corresponding to any (simple and well separated) eigenvalue if we choose the shift properly
- Very good convergence rates, if shift is close to an eigenvalue.
- However, one may feel uncomfortable solving an 'almost singular' system of equations.
- An analysis using the SVD of $A - \sigma I$ shows that the tiny smallest singular value σ_n blows up the component in direction of $\mathbf{v}_n$. So, the vector $\mathbf{z}$ points in the desired 'most singular' direction.

The generalized eigenvalue problem

Generalized eigenvalue problem $A\mathbf{x} = \lambda B\mathbf{x}$.

Simple vector iteration

$$\mathbf{x}^{(k)} := B^{-1}A\mathbf{x}^{(k-1)}, \qquad k = 1, 2, \ldots$$

Shift-and-invert iteration

$$(A - \sigma B)^{-1}B\mathbf{x} := \mu\mathbf{x}, \qquad \mu = \frac{1}{\lambda - \sigma}.$$

Computing higher eigenvalues

In order to compute higher eigenvalues $\lambda_2, \lambda_3, \ldots,$ we make use of the mutual orthogonality of the eigenvectors of symmetric matrices, or of Schur vectors of nonsymmetric matrices.

We can compute the j-th eigenpair $(\lambda_j, \mathbf{u}_j)$ by inverse iteration, keeping the iterated vectors $\mathbf{x}^{(k)}$ orthogonal to the already known or computed eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$.

Rayleigh quotient iteration

- Assume that matrix A is Hermitian (or symmetric).
- Inverse iteration is effective way to compute eigenpairs, if good approximation of desired eigenvalue is known.
- Approximation is used as shift.
- Rayleigh quotient of eigenvector gives a very good approximation of its eigenvalue.

Lemma

Let $\mathbf{q}$ be any nonzero vector. The number ρ that minimizes $\|A\mathbf{q} - \rho\mathbf{q}\|$ is the Rayleigh quotient

$$\rho = \frac{\mathbf{q}^* A\mathbf{q}}{\mathbf{q}^*\mathbf{q}}.$$

Rayleigh quotient iteration (RQI)

```
Choose a starting vector y_0 ∈ ℝ^n, ‖y_0‖ = 1, and tolerance ε.
for k = 1, 2, ... do
    ρ_k := y_{k-1}^* A y_{k-1}.
    Solve (A − ρ_k I) z_k = y_{k-1}   for z_k.
    σ_k = ‖z_k‖.
    y_k := z_k / σ_k.
    if σ_k > 10/ε then
        return y_k
    end if
end for
```

Convergence of Rayleigh quotient iteration

- Rayleigh quotient iteration usually converges, however not always towards the desired solution.
- Let's assume that $\mathbf{y}_k \xrightarrow[k\to\infty]{} \mathbf{x}$ with $A\mathbf{x} = \lambda\mathbf{x}$.
- Let $\|\mathbf{x}\| = \|\mathbf{y}_k\| = 1$ for all k and $\varphi_k = \angle(\mathbf{x}, \mathbf{y}_k)$.
  Assumption implies $\{\varphi_k\} \xrightarrow[k\to\infty]{} 0$.

Can write

$$\mathbf{y}_k = \mathbf{x}\cos\varphi_k + \mathbf{u}_k \sin\varphi_k, \qquad \|\mathbf{x}\| = \|\mathbf{y}_k\| = \|\mathbf{u}_k\| = 1.$$

Rayleigh quotient

$$\rho_k = \rho(\mathbf{y}_k) = \frac{\mathbf{y}_k{}^* A \mathbf{y}_k}{\mathbf{y}_k{}^* \mathbf{y}_k} = \mathbf{y}_k{}^* A \mathbf{y}_k$$

Convergence of Rayleigh quotient iteration (cont.)

$$\begin{aligned}\lambda - \rho_k &= \lambda - \cos^2\varphi_k \underbrace{\mathbf{x}^* A\mathbf{x}}_{\lambda} - \cos\varphi_k \sin\varphi_k \underbrace{\mathbf{x}^* A\mathbf{u}_k}_{0} - \sin^2\varphi_k \mathbf{u}_k^* A\mathbf{u}_k \\ &= \lambda(1-\cos^2\varphi_k) - \sin^2\varphi_k \rho(\mathbf{u}_k) \\ &= (\lambda - \rho(\mathbf{u}_k))\sin^2\varphi_k.\end{aligned}$$

We now have

Theorem (Cubic convergence of Rayleigh quotient iteration)

With the above assumption we have $\lim_{k\to\infty}\left|\frac{\varphi_{k+1}}{\varphi_k^3}\right| \leq 1.$

Convergence of Rayleigh quotient iteration (cont.)

Proof:

$$\begin{aligned}\mathbf{z}_{k+1} = (A - \rho_k I)^{-1}\mathbf{y}_k &= \mathbf{x}\cos\varphi_k/(\lambda - \rho_k) + (A - \rho_k I)^{-1}\mathbf{u}_k \sin\varphi_k \\ &= \mathbf{x}\underbrace{\frac{\cos\varphi_k}{\lambda - \rho_k}}_{\|\mathbf{z_{k+1}}\|\cos\varphi_{k+1}} + \mathbf{u}_{k+1}\underbrace{\sin\varphi_k\|(A - \rho_k I)^{-1}\mathbf{u}_k\|}_{\|\mathbf{z_{k+1}}\|\sin\varphi_{k+1}},\end{aligned}$$

where

$$\mathbf{u}_{k+1} := (A - \rho_k I)^{-1}\mathbf{u}_k/\|(A - \rho_k I)^{-1}\mathbf{u}_k\|$$

Convergence of Rayleigh quotient iteration (cont.)

Thus,

$$\begin{aligned}
\tan\varphi_{k+1} &= \frac{\sin\varphi_{k+1}}{\cos\varphi_{k+1}} \\
&= \sin\varphi_k \left\|(A-\rho_k I)^{-1}\mathbf{u}_k\right\| \frac{\lambda-\rho_k}{\cos\varphi_k} \\
&= (\lambda-\rho_k) \left\|(A-\rho_k I)^{-1}\mathbf{u}_k\right\| \tan\varphi_k \\
&= (\lambda-\rho(\mathbf{u}_k)) \left\|(A-\rho_k I)^{-1}\mathbf{u}_k\right\| \sin^2\varphi_k \tan\varphi_k .
\end{aligned}$$

So,

$$(A-\rho_k I)^{-1}\mathbf{u}_k = (A-\rho_k I)^{-1}\left(\sum_{\lambda_i\neq\lambda}\beta_i\mathbf{x}_i\right) = \sum_{\lambda_i\neq\lambda}\frac{\beta_i}{\lambda_i-\rho_k}\mathbf{x}_i$$

Convergence of Rayleigh quotient iteration (cont.)

Taking norms,

$$\|(A-\rho_k I)^{-1}\mathbf{u}_k\|^2 = \sum_{\lambda_i\neq\lambda} \frac{\beta_i^2}{|\lambda_i-\rho_k|^2} \geq \frac{1}{\min_{\lambda_i\neq\lambda}|\lambda_i-\rho_k|^2} \underbrace{\sum_{\lambda_i\neq\lambda}\beta_i^2}_{\|\mathbf{u}_k\|^2=1}$$

Gap between λ and rest of A's spectrum: $\gamma := \min_{\lambda_i\neq\lambda}|\lambda_i-\lambda|$.

Assumption $\Longrightarrow \exists k_0 \in \mathbb{N}$ s.t. $|\lambda-\rho_k| < \frac{\gamma}{2}$, $\forall k > k_0$. Thus,

$$|\lambda_i-\rho_k| > \frac{\gamma}{2} \qquad \text{for all } \lambda_i \neq \lambda,$$

and

$$\|(A-\rho_k I)^{-1}\mathbf{u}_k\| \leq \frac{1}{\min_{\lambda_i\neq\lambda}|\lambda_i-\rho_k|} \leq \frac{2}{\gamma}, \qquad k > k_0.$$

Convergence of Rayleigh quotient iteration (cont.)

$$
\begin{aligned}
|\tan\varphi_{k+1}| &= \left|\frac{\sin\varphi_{k+1}}{\cos\varphi_{k+1}}\right| \\
&= |\lambda - \rho(\mathbf{u}_k)|\,\|(A-\rho_k I)^{-1}\mathbf{u}_k\|\,|\sin^2\varphi_k||\tan\varphi_k| \\
&\leq \frac{2}{\gamma}|\lambda - \rho(\mathbf{u}_k)|\,|\sin^2\varphi_k||\tan\varphi_k|
\end{aligned}
$$

$\tan\varphi_k \approx \sin\varphi_k \approx \varphi_k$ if $\varphi_k \ll 1 \implies$ cubic convergence rate.

Note: The sequence $\{\mathbf{u}_k\}$ may converge to an eigenvector of A, or, if A has two different eigenvalues that are in equal distance to λ, jump back and forth between the corresponding two eigenvectors.

Remarks on RQI

1. We did not prove global convergence.
   RQI converges 'almost always'. But it is not clear *a priori*, towards which eigenpair.
2. Alternative: first apply inverse vector iteration and switch to Rayleigh quotient iteration as soon as the iterate is close enough to the solution.
3. Rayleigh quotient iteration is expensive. In every iteration step another system of equations has to be solved, i.e., *in every iteration step a matrix has to be factorized*.

   RQI is usually applied only to tridiagonal matrices.

A numerical example

The following MATLAB script demonstrates the power of Rayleigh quotient iteration. It expects as input a matrix `A`, an initial vector `x` of length one.

```
%        Initializations
k = 0; rho = 0; ynorm = 0;

while abs(rho)*ynorm < 1e+15,
   k = k + 1; if k>20, break, end
   rho = x'*A*x;
   y = (A - rho*eye(size(A)))\x;
   ynorm = norm(y);
   x = y/ynorm;
end
```

A numerical example (cont.)

We invoke this routine with the 1D Poisson matrix

```
e=ones(9,1); T=spdiags([-e,2*e,-e],[-1:1],9,9);
```

and the initial vector $\mathbf{x} = [-4, -3, \ldots, 3, 4]^T$.

k	rho	ynorm
1	0.6666666666666666	3.1717e+00
2	0.4155307724080958	2.9314e+01
3	0.3820048793104663	2.5728e+04
4	0.3819660112501632	1.7207e+13
5	0.3819660112501051	2.6854e+16

The cubic convergence is evident.

References

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

[2] D. B. Szyld, *Criteria for combining inverse and Rayleigh quotient iteration*, SIAM J. Numer. Anal., 25 (1988), pp. 1369–1375.