Algorithms for Finding Eigenvalues-Eigenvectors

Y.Cem Subakan

Roadmap

- In Lecture 26, we investigate triangulation to find the Schur decomposition of a matrix. This works for finding eigenvalues of general matrices
- In Lecture 27, we talk about finding eigenvectors of a real Hermitian matrix. Power iterations
-

Lecture 26
Reduction to Hessenberg or Tridiagonal Form

NLA Reading Group Spring'13

By Cem Subakan

Reduction to Hessenberg?

Eigenvalue-Revealing Factorizations

In the preceding pages we have described three examples of *eigenvalue-revealing factorizations*, factorizations of a matrix that reduce it to a form in which the eigenvalues are explicitly displayed. We can summarize these as follows.

A diagonalization $A = X\Lambda X^{-1}$ exists if and only if A is nondefective.

A unitary diagonalization $A = Q\Lambda Q^*$ exists if and only if A is normal.

A unitary triangularization (Schur factorization) $A = QTQ^*$ always exists.

- In Lecture 24, we saw that the upper triangular matrix T in Schur factorization gives us eigenvalues in itsdiagonal
  - To find T:

$$\underbrace{Q_j^* \cdots Q_2^* Q_1^*}_{Q^*} A \underbrace{Q_1 Q_2 \cdots Q_j}_{Q}$$

converges to an upper-triangular matrix T as $j \to \infty$.

Two Phases until we get iT

$$
\underset{A \neq A^*}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}}
\xrightarrow{\text{Phase 1}}
\underset{H}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix}}
\xrightarrow{\text{Phase 2}}
\underset{T}{\begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ & & & & \times \end{bmatrix}}.
$$

-The first phase is to find Hessenberg matrices.
-In the second phase, a sequence of reduction to Hessenberg matrices converge to T.

How to do iT?

- Basic idea is to find a series of similarity transforms so that we converge to T.
-

$$\underbrace{Q_j^* \cdots Q_2^* Q_1^*}_{Q^*} \, A \, \underbrace{Q_1 Q_2 \cdots Q_j}_{Q}$$

converges to an upper-triangular matrix T as $j \to \infty$.

Greed is not good for iT

$$\underset{A}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{Q_1^* \cdot} \underset{Q_1^* A}{\begin{bmatrix} \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ 0 & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ 0 & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ 0 & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ 0 & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}} .$$

We need to complete the similarity transform:

$$\underset{Q_1^* A}{\begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\cdot Q_1} \underset{Q_1^* A Q_1}{\begin{bmatrix} \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}} .$$

Be less ambiTious

$$
\underset{A}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix}}
\xrightarrow{Q_1^* \cdot}
\underset{Q_1^* A}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}
\xrightarrow{\cdot Q_1}
\underset{Q_1^* A Q_1}{\begin{bmatrix} \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}.
$$

$$
\underset{Q_1^* A Q_1}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & \times & \times & \times & \times \end{bmatrix}}
\xrightarrow{Q_2^* \cdot}
\underset{Q_2^* Q_1^* A Q_1}{\begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \mathbf{0} & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}
\xrightarrow{\cdot Q_2}
\underset{Q_2^* Q_1^* A Q_1 Q_2}{\begin{bmatrix} \times & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ \times & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & \times & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \\ & & \boldsymbol{\times} & \boldsymbol{\times} & \boldsymbol{\times} \end{bmatrix}}.
$$

We have iT!

After repeating this process $m-2$ times, we have a product in Hessenberg form, as desired:

$$\underbrace{Q_{m-2}^* \cdots Q_2^* Q_1^*}_{Q^*} A \underbrace{Q_1 Q_2 \cdots Q_{m-2}}_{Q} = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix} = H.$$

Lecture 27

Rayleigh Quotient and Inverse Iterations

NLA Reading Group Spring’13

By Y. Cem Subakan

Real Symmetric Matrices

- Now, we restrict ourselves to symmetric, real matrices.
- It implies that
  1) All eigenvalues are real
  2) We have complete set of eigenvectors
  3) Eigenvectors are orthogonal to each other

Rayleigh Quotient

The *Rayleigh quotient* of a vector $x \in \mathbb{R}^m$ is the scalar

$$r(x) = \frac{x^T A x}{x^T x}.$$

Notice that if x is an eigenvector, then $r(x) = \lambda$ is the corresponding eigenvalue. One way to motivate this formula is to ask: given x, what scalar α "acts most like an eigenvalue" for x in the sense of minimizing $\|Ax - \alpha x\|_2$?

Maximizer of RQ

$$\nabla r(x) = \frac{2}{x^T x}(Ax - r(x)x).$$

From this formula we see that at an eigenvector x of A, the gradient of $r(x)$ is the zero vector. Conversely, if $\nabla r(x) = 0$ with $x \neq 0$, then x is an eigenvector and $r(x)$ is the corresponding eigenvalue.

Maximizer of RQ

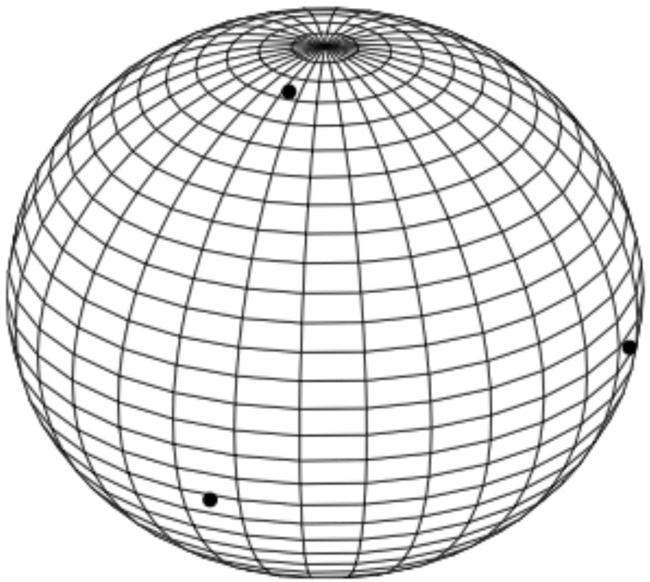


Figure 27.1. *The Rayleigh quotient* $r(x)$ *is a continuous function on the unit sphere* $\|x\| = 1$ *in* $\mathbb{R}^m$, *and the stationary points of* $r(x)$ *are the normalized eigenvectors of* A. *In this example with* $m = 3$, *there are three orthogonal stationary points (as well as their antipodes).*

Goodness of RQ as an eigenvalue estimate

Let q_J be one of the eigenvectors of A. From the fact that $\nabla r(q_J) = 0$, together with the smoothness of the function $r(x)$ (everywhere except at the origin $x = 0$), we derive an important consequence:

$$r(x) - r(q_J) = O(\|x - q_J\|^2) \quad \text{as } x \to q_J. \tag{27.3}$$

Thus the Rayleigh quotient is a *quadratically accurate* estimate of an eigenvalue. Herein lies its power.

Goodness of RQ as an eigenvalue estimate

A more explicit way to derive (27.3) is to expand x as a linear combination of the eigenvectors $q_1, \ldots, q_m$ of A. If $x = \sum_{j=1}^{m} a_j q_j$, then $r(x) = \sum_{j=1}^{m} a_j^2 \lambda_j / \sum_{j=1}^{m} a_j^2$. Thus $r(x)$ is a weighted mean of the eigenvalues of A, with the weights equal to the squares of the coordinates of x in the eigenvector basis. Because of this squaring of the coordinates, it is not hard to see that if $|a_j/a_J| \leq \epsilon$ for all $j \neq J$, then $r(x) - r(q_J) = O(\epsilon^2)$.

Power Iterations

Algorithm 27.1. Power Iteration

$v^{(0)} =$ some vector with $\|v^{(0)}\| = 1$

for $k = 1, 2, \ldots$

$w = Av^{(k-1)}$	apply A
$v^{(k)} = w/\|w\|$	normalize
$\lambda^{(k)} = (v^{(k)})^T Av^{(k)}$	Rayleigh quotient

Analysis of Power Iterations

We can analyze power iteration easily. Write $v^{(0)}$ as a linear combination of the orthonormal eigenvectors q_i:

$$v^{(0)} = a_1 q_1 + a_2 q_2 + \cdots + a_m q_m.$$

Since $v^{(k)}$ is a multiple of $A^k v^{(0)}$, we have for some constants c_k

$$\begin{aligned} v^{(k)} &= c_k A^k v^{(0)} \\ &= c_k (a_1 \lambda_1^k q_1 + a_2 \lambda_2^k q_2 + \cdots + a_m \lambda_m^k q_m) \\ &= c_k \lambda_1^k \left(a_1 q_1 + a_2 (\lambda_2/\lambda_1)^k q_2 + \cdots + a_m (\lambda_m/\lambda_1)^k q_m \right). \end{aligned} \qquad (27.4)$$

Convergence quality

Theorem 27.1. *Suppose* $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m| \geq 0$ *and* $q_1^T v^{(0)} \neq 0$. *Then the iterates of Algorithm 27.1 satisfy*

$$\|v^{(k)} - (\pm q_1)\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right), \qquad |\lambda^{(k)} - \lambda_1| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right) \tag{27.5}$$

as $k \to \infty$. *The* $\pm$ *sign means that at each step* k, *one or the other choice of sign is to be taken, and then the indicated bound holds.*

Proof. The first equation follows from (27.4), since $a_1 = q_1^T v^{(0)} \neq 0$ by assumption. The second follows from this and (27.3). If $\lambda_1 > 0$, then the $\pm$ signs are all $+$ or all $-$, whereas if $\lambda_1 < 0$, they alternate. $\square$

Downsides of Power Iterations

On its own, power iteration is of limited use, for several reasons. First, it can find only the eigenvector corresponding to the largest eigenvalue. Second, the convergence is linear, reducing the error only by a constant factor $\approx |\lambda_2/\lambda_1|$ at each iteration. Finally, the quality of this factor depends on having a largest eigenvalue that is significantly larger than the others. If the largest two eigenvalues are close in magnitude, the convergence will be very slow.

Fortunately, there is a way to amplify the differences between eigenvalues.

1) Can only find eigenvector corresponding to the largest eigenvalue
2) Convergence is linear:

$$\|v^{(k)} - (\pm q_1)\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right),$$

3) Quality of convergence is dependent on spectral gap.

Remedy for close eigenvalues: Inverse Iterations

For any $\mu \in \mathbb{R}$ that is not an eigenvalue of A, the eigenvectors of $(A - \mu I)^{-1}$ are the same as the eigenvectors of A, and the corresponding eigenvalues are $\{(\lambda_j - \mu)^{-1}\}$, where $\{\lambda_j\}$ are the eigenvalues of A. This suggests an idea. Suppose μ is close to an eigenvalue λ_J of A. Then $(\lambda_J - \mu)^{-1}$ may be much larger than $(\lambda_j - \mu)^{-1}$ for all $j \neq J$. Thus, if we apply power iteration to $(A - \mu I)^{-1}$, the process will converge rapidly to q_J. This idea is called *inverse iteration*.

Inverse Iterations

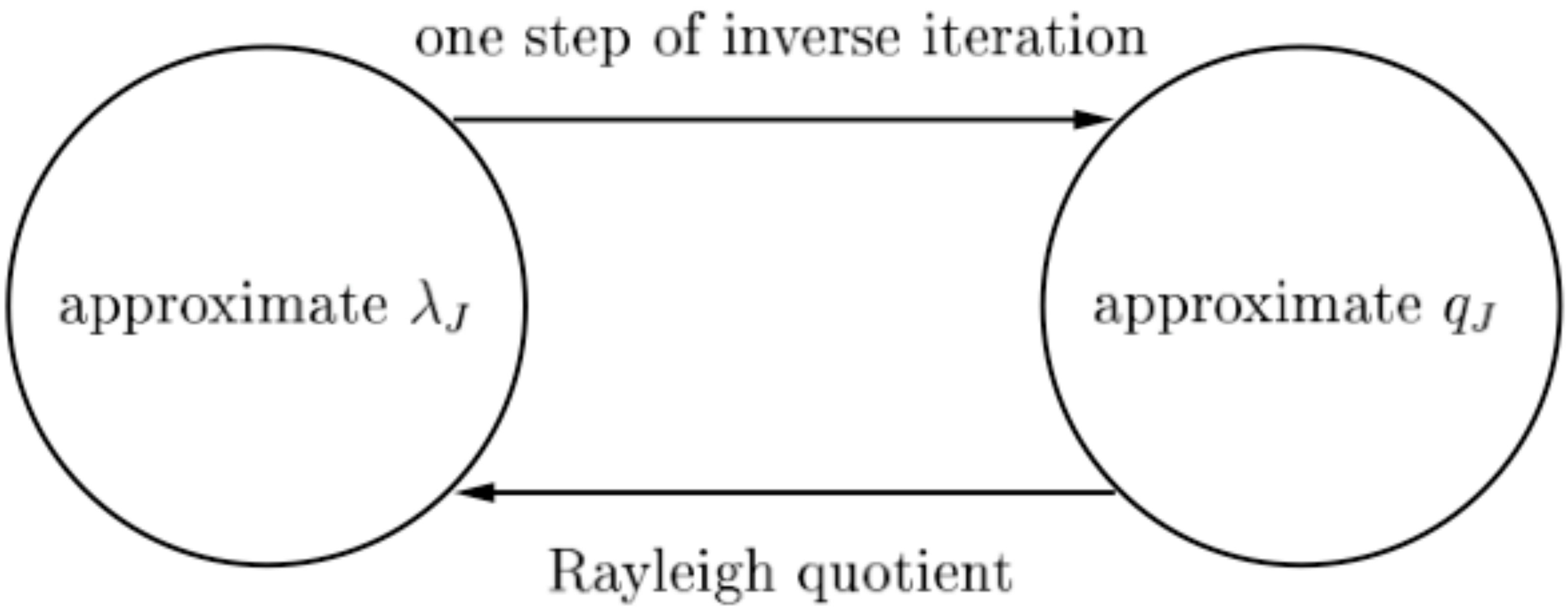
Algorithm 27.2. Inverse Iteration

$v^{(0)}$ = some vector with $\|v^{(0)}\| = 1$

for $k = 1, 2, \ldots$

Solve $(A - \mu I)w = v^{(k-1)}$ for w	apply $(A - \mu I)^{-1}$
$v^{(k)} = w/\|w\|$	normalize
$\lambda^{(k)} = (v^{(k)})^T A v^{(k)}$	Rayleigh quotient

Rayleigh Quotient Iterations



Rayleigh Quotient Iterations

Algorithm 27.3. Rayleigh Quotient Iteration

$v^{(0)} =$ some vector with $\|v^{(0)}\| = 1$

$\lambda^{(0)} = (v^{(0)})^T A v^{(0)} =$ corresponding Rayleigh quotient

for $k = 1, 2, \ldots$

Solve $(A - \lambda^{(k-1)} I)w = v^{(k-1)}$ for w	apply $(A - \lambda^{(k-1)} I)^{-1}$
$v^{(k)} = w/\|w\|$	normalize
$\lambda^{(k)} = (v^{(k)})^T A v^{(k)}$	Rayleigh quotient

Rayleigh Quotient Iterations, Cubic Convergence

Theorem 27.3. *Rayleigh quotient iteration converges to an eigenvalue/eigenvector pair for all except a set of measure zero of starting vectors* $v^{(0)}$. *When it converges, the convergence is ultimately cubic in the sense that if* λ_J *is an eigenvalue of* A *and* $v^{(0)}$ *is sufficiently close to the eigenvector* q_J, *then*

$$\|v^{(k+1)} - (\pm q_J)\| = O(\|v^{(k)} - (\pm q_J)\|^3) \tag{27.6}$$

and

$$|\lambda^{(k+1)} - \lambda_J| = O(|\lambda^{(k)} - \lambda_J|^3) \tag{27.7}$$

as $k \to \infty$. *The* $\pm$ *signs are not necessarily the same on the two sides of* (27.6).

Exercises

27.2. Again let $A \in \mathbb{C}^{m\times m}$ be arbitrary. The set of all Rayleigh quotients of A, corresponding to all nonzero vectors $x \in \mathbb{C}^m$, is known as the *field of values* or *numerical range* of A, a subset of the complex plane denoted by $W(A)$.

(a) Show that $W(A)$ contains the convex hull of the eigenvalues of A.

(b) Show that if A is normal, then $W(A)$ is equal to the convex hull of the eigenvalues of A.

Exercises

27.4. Every real symmetric square matrix can be orthogonally diagonalized, and the developments of this lecture are invariant under orthogonal changes of coordinates. Thus it would have been sufficient to carry out each derivation of this lecture under the assumption that A is a diagonal matrix with entries ordered by decreasing absolute value. Making this assumption, describe the form taken by (27.4), (27.5), and Algorithm 27.3.