July 4, 2014

Inverse iteration

Idea 1: Use A^{-1} to compute the smallest eigenvalue.

(Note: $\Lambda(A^{-1}) = \{ 1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n \}$.)

Thus $v^{(0)} = c_1 q_1 + c_2 q_2 + \ldots + c_n q_n$

$A^{-1} v^{(0)} = c_1 1/\lambda_1 q_1 + \ldots + c_n 1/\lambda_n q_n$

$\vdots$

$A^{-k} v^{(0)} = c_1 (1/\lambda_1)^k q_1 + \ldots + c_n (1/\lambda_n)^k q_n$

$= (1/\lambda_n)^k [c_1 (\lambda_n/\lambda_1)^k q_1 + \ldots + c_{n-1} (\lambda_n/\lambda_{n-1})^k q_{n-1} + c_n q_n]$

$\therefore$ $A^{-k} v^{(0)} \sim c_n (1/\lambda_n)^k q_n$ for large k

Idea 2: Shifting.

Consider $B = A - \mu I$, μ is not an eigenvalue of A. Then B has the same eigenvectors of A and its eigenvalues are $\{ \lambda_j - \mu \}$, $\lambda_j \in \Lambda(A)$.

If μ is close to λ_J, $\lambda_J - \mu$ would be the smallest eigenvalue of B.

We can apply idea 1 to compute $\lambda_J - \mu$.

Example

$$A = \begin{bmatrix} 21 & 7 & -1 \\ 5 & 7 & 7 \\ 4 & -4 & 20 \end{bmatrix}, \quad \Lambda(A) = \{8, 16, 24\}, \quad \mu = 15$$

$v^{(0)} = (1, 1, 1)^T$

$w = (A-\mu I)^{-1} v^{(0)} = (0.032, 0.16, 0.30)^T$

$v^{(1)} = w/||w|| = (0.093, 0.46, 0.88)^T$

$\lambda^{(1)} = r(v^{(1)}) = 19.2000$

$w = (A-\mu I)^{-1} v^{(1)} = (-0.33, 0.40, 0.76)^T$

$v^{(2)} = w/||w|| = (-0.36, 0.44, 0.83)^T$

$\lambda^{(2)} = r(v^{(2)}) = 15.9749$

$w = (A-\mu I)^{-1} v^{(2)} = (-0.39, 0.40, 0.79)^T$

$v^{(3)} = w/||w|| = (-0.40, 0.41, 0.82)^T$

$\lambda^{(3)} = r(v^{(3)}) = 16.0290$

:

:

$q_2 = (-0.4082, 0.4082, 0.8165)^T, \quad \lambda_2 = 16.$

Algorithm

$v^{(0)}$ = initial guess, $||v^{(0)}|| = 1$

for $k = 1, 2, \ldots$

Solve $(A - \mu I)\, w = v^{(k-1)}$ (i.e. $w = (A - \mu I)^{-1} v^{(k-1)}$)

$v^{(k)} = w / ||w||$

$\lambda^{(k)} = (v^{(k)})^T A\, v^{(k)}$

end

Notes

1) Like power iteration, inverse iteration has linear convergence.

2) Unlike power iteration, we can choose which eigenvector to compute by choosing μ close to the corresponding λ_J.

3) Theorem: Suppose λ_J is closest to μ and λ_L is the second closest, i.e. $|\mu - \lambda_J| < |\mu - \lambda_L| \leq |\mu - \lambda_j|$ $j \neq J$. Also suppose $q_J^T v^{(0)} \neq 0$. Then

$$\left\| v^{(k)} - (\pm q_J) \right\| = O\left(\left| \frac{\mu - \lambda_J}{\mu - \lambda_L} \right|^k \right), \quad \left| \lambda^{(k)} - \lambda_J \right| = O\left(\left| \frac{\mu - \lambda_J}{\mu - \lambda_L} \right|^{2k} \right)$$

as $k \to \infty$.

Rayleigh quotient iteration

- Rayleigh quotient gives an eigenvalue estimate from an eigenvector estimate.

- Inverse iteration gives an eigenvector estimate from an eigenvalue estimate.

Idea: combine the two.

Algorithm

$v^{(0)}$ = initial guess with $||v^{(0)}|| = 1$

$\lambda^{(0)} = (v^{(0)})^T A v^{(0)} = r(v^{(0)})$

for k = 1, 2, . . .

Solve $(A - \lambda^{(k-1)} I) w = v^{(k-1)}$

$v^{(k)} = w / ||w||$

$\lambda^{(k)} = (v^{(k)})^T A v^{(k)}$

end

Theorem: RQI converges for most starting vector $v^{(0)}$. The convergence is cubic:

$$\left\| v^{(k+1)} - (\pm q_J) \right\| = O\left(\left\| v^{(k)} - (\pm q_J) \right\|^3 \right), \quad \left| \lambda^{(k+1)} - \lambda_J \right| = O\left(\left| \lambda^{(k)} - \lambda_J \right|^3 \right)$$

Example

$$A = \begin{bmatrix} 21 & 7 & -1 \\ 5 & 7 & 7 \\ 4 & -4 & 20 \end{bmatrix}, \qquad v^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Apply Rayleigh Quotient iteration:

$\lambda^{(0)} = 22$

$\lambda^{(1)} = 24.0812$

$\lambda^{(2)} = 24.0013$

$\lambda^{(3)} = 24.00000017$

Complexity

- Each step of power iteration involves $Av^{(k-1)} \rightarrow O(n^2)$ flops.
- Each step of inverse iteration solves $(A-\mu I)w = v^{(k-1)} \rightarrow O(n^3)$ flops. One can pre-compute and store L, U factors of $A - \mu I$. Thus each step $\rightarrow O(n^2)$ flops for forward and back solves.
- The matrix $A - \lambda^{(k-1)}I$ changes in each step of RQI $\rightarrow O(n^3)$ flops in general.
- If A is tridiagonal, all 3 methods $\rightarrow O(n)$ flops per iteration.

QR iteration

Def: If $X \in R^{nxn}$ is nonsingular, then $A \rightarrow X^{-1} A X$ is called a similarity transformation of A.

Def: A and B are similar if $B = X^{-1} A X$ for some nonsingular X.

Theorem: If A, B are similar, then they have the same characteristic polynomial and hence the same eigenvalues.

Pf: $p_B(z) = \det(zI - X^{-1}AX) = \det(X^{-1}(zI - A)X)$

$= \det(X^{-1}) \det(zI - A) \det(X)$

$= \det(zI - A) = p_A(z)$

Idea: Apply a sequence of similarity transformation to A which will converge to a diagonal matrix.

Consider $A^{(k-1)}$. Compute QR factorization of $A^{(k-1)}$.

i.e. $R^{(k)} = (Q^{(k)})^T A^{(k-1)}$ (e.g. Householder transform)

Then $R^{(k)}Q^{(k)} = (Q^{(k)})^T A^{(k-1)} Q^{(k)} \equiv A^{(k)}$

Clearly $A^{(k-1)}$ and $A^{(k)}$ are similar.

Algorithm (QR iteration)

$A^{(0)} = A$

for $k = 1, 2, \ldots$

$Q^{(k)} R^{(k)} = A^{(k-1)}$ (QR factorization of $A^{(k-1)}$)

$A^{(k)} = R^{(k)} Q^{(k)}$

end

How does it work?

Example

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix} = A^{(0)}$$

$A^{(0)} = Q^{(1)} R^{(1)}$

$A^{(1)} = R^{(1)} Q^{(1)}$

$$= \begin{bmatrix} 4.17 & 1.10 & -1.27 \\ 1.10 & 2.00 & 0 \\ -1.27 & 0 & 2.83 \end{bmatrix}$$

$A^{(1)} = Q^{(2)} R^{(2)}$

$A^{(2)} = R^{(2)} Q^{(2)}$

$$= \begin{bmatrix} 5.09 & 0.16 & 0.62 \\ 0.16 & 1.86 & -0.55 \\ 0.62 & -0.55 & 2.05 \end{bmatrix}$$

$A^{(2)} = Q^{(3)} R^{(3)}$

$A^{(3)} = R^{(3)} Q^{(3)}$

$$= \begin{bmatrix} 5.20 & -0.08 & -0.21 \\ -0.08 & 2.18 & 0.50 \\ -0.21 & 0.50 & 1.62 \end{bmatrix}$$

Eigenvalues of A: 5.2143, 2.4608, 1.3249

Note: $A^{(k)} \rightarrow$ diagonal matrix.

Simultaneous iteration/Block power iteration

- Apply power iteration to several vectors at once and maintain linearly independence among the vectors.

- Start with: $v_1^{(0)}, v_2^{(0)}, \ldots, v_p^{(0)}$
  Then $A^k v_1^{(0)}$ converges to q_1 where $|\lambda_1|$ is largest.
  Thus span $\{ A^k v_1^{(0)}, \ldots, A^k v_p^{(0)} \}$ should converge to $\{ q_1, \ldots, q_p \}$ where $\lambda_1, \ldots, \lambda_p$ are the p largest eigenvalues.

- Write $V^{(0)} = [v_1^{(0)} \; v_2^{(0)} \; \ldots \; v_p^{(0)}]$.

  Define $V^{(k)} = A^{(k)} V^{(0)} = [v_1^{(k)} \; v_2^{(k)} \; \ldots \; v_p^{(k)}]$.

- As $k \to \infty$, the vectors $v_1^{(k)}, \ldots, v_p^{(k)}$ all converge to multiples of the same dominant eigenvector q_1.

- Orthogonalize the vectors at each step.

Algorithm

Pick $\hat{Q}^{(0)} \in R^{nxp}$ with orthonormal columns

for k = 1, 2, . . .

$Z^{(k)} = A \hat{Q}^{(k-1)}$	power iteration
$\hat{Q}^{(k)} \hat{R}^{(k)} = Z^{(k)}$	reduced QR factorization

end

Note: The column space of $\hat{Q}^{(k)}$ and $Z^{(k)}$ are the same. They are both equal to that of $A^{(k)} \hat{Q}^{(0)}$.

- Assumption 1: The leading p+1 e.v. are distinct in absolute values:

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_p| > |\lambda_{p+1}| \geq |\lambda_{p+2}| \ldots \geq |\lambda_n|$$

- Assumption 2: All the leading principal minors of $\hat{Q}^T V^{(0)}$ are nonsingular.

Theorem: Suppose the block power iteration is carried out and assumptions 1 & 2 hold. Then as $k \to \infty$,

$$\left\| q_j^{(k)} - (\pm q_j) \right\| = O(c^k) \qquad j = 1, 2, \ldots, p$$

where $c = \max_{1 \leq k \leq p} \left| \frac{\lambda_{k+1}}{\lambda_k} \right| < 1$