Orthogonal & QR iteration.

Note Title

(I) Orthogonal / simultaneous / block power iteration / subspace iteration

The idea of simultaneous/orthogonal iteration is to apply the power iteration to several vectors at once.

Instead of starting with one vector x_0, suppose we start with n linearly independent (& orthogonal) vectors $x_1^{(0)}, \ldots, x_n^{(0)}$.

Let $A \in \mathbb{C}^{m\times m}$. Under suitable assumptions, it can be shown that $\langle A^k x_1^{(0)}, \ldots, A^k x_n^{(0)} \rangle$ converges to the e-vectors of A corresponding to the n largest e-values of A (in abs. value).

In matrix notation, let Z_0 denote the initial matrix

$$Z_0 = [x_1^{(0)} | \ldots | x_n^{(0)}]_{m\times n}.$$

Define Z_k to be the result after k applications of A, i.e. $Z_k = A^k Z_0 = [x_1^{(k)} | \ldots | x_n^{(k)}]_{m\times n}$.

Since our interest is in the column space of Z_k, we will extract a well-behaved basis for the span of the columns of Z_k (denote by span Z_k).

This is done by computing the reduced QR factorization of Z_k as $Z_k = Q_k R_k$, where $\dim Q_k = m\times n$, $\dim R_k = n\times n$.

Then as $k\to\infty$, the successive columns of Q_k are expected to converge to the required e-vectors.

Algorithm: Let Z_0 be a $m \times n$ orthogonal matrix

$i = 1$ to convergence

$$Z_k = A Q_{k-1} \qquad (Q_0 = Z_0)$$
$$Q_k R_k = Z_k \qquad (\text{QR factorization})$$
$$\lambda(Q_k^* A Q_k) = \{\lambda_1^{(k)}, \dots, \lambda_n^{(k)}\} \qquad (\text{sequence of e-value estimates})$$

end

Note that if $n=1$, this is the power method & the sequence $\{Q_k x_1^{(0)}\}$ is precisely the sequence of vectors produced by the power method with starting vector $x_1^{(0)}$

Analysis: We assume A is diagonalizable with
$$A = S \Lambda S^{-1}.$$

We require 2 assumptions:

1. the leading $(n+1)$ e-values are distinct in abs. value i.e. $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > |\lambda_{n+1}| \geq \dots \geq |\lambda_m|$.

2. all the leading principal minors of Z_0 are non-singular. This condition is analogous to requiring that the first component of x_0 be non-zero in the power method. There it means that x_0 has a non-zero component in the direction of e_1. Here the condition means that none of the vectors $\{z_i, \dots, z_j\}$ are orthogonal to the span of $\{e_i, \dots, e_j\}$ $\forall\, 1 \leq j \leq n$, i.e. z_0 is not deficient in certain eigen-directions.

$$\begin{aligned} \text{span}(Q_k) = \text{span}(Z_k) &= \text{span}(A Q_{k-1}) \\ &= \text{span}(A^2 Q_{k-2}) \\ &\;\vdots \\ &= \text{span}(A^k Z_0) \end{aligned}$$

$A = S\Lambda S^{-1}$

$A^k = S\Lambda^k S^{-1}$

$$= \text{span}(S\Lambda^k S^{-1} Z_0)$$

Now $S\Lambda^k S^{-1} Z_0 = S \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_m^k \end{bmatrix} S^{-1} Z_0$

$$= \lambda_n^k \, S \underbrace{\begin{bmatrix} (\lambda_1/\lambda_n)^k & & & & \\ & \ddots & & & \\ & & 1 & & \\ \text{n×n block} & & & \ddots & \\ & & & & (\lambda_m/\lambda_n)^k \end{bmatrix}_{m\times m}}_{} \underset{m\times m}{S^{-1}} \underset{m\times n}{Z_0}$$

$|\lambda_1| > \cdots > |\lambda_n| > |\lambda_{n+1}| \geq \cdots \geq |\lambda_m|$

$\left|\frac{\lambda_i}{\lambda_n}\right| \geq 1$ for $i \leq n$ & $\left|\frac{\lambda_i}{\lambda_n}\right| < 1$ for $i > n$.

So $\Lambda^k S^{-1} Z_0 = \begin{bmatrix} [V_k]_{n\times n} \\ \hline [W_k] \end{bmatrix}_{m\times n}$ & let $S = [\underbrace{s_1 | \cdots | s_n}_{S_n} | \underbrace{\cdots | s_m}_{S_{m-n}}]$

here $W_k \to 0$ ($\&$ $V_k \not\to 0$) as $k \to \infty$

$$\therefore S\Lambda^k S^{-1} Z_0 = \lambda_n^k S \begin{bmatrix} [V_k] \\ \hline [W_k] \end{bmatrix}$$

$$= \lambda_n^k \left[S_n V_k + S_{m-n} W_k \right]$$

Finally, $\text{span}(Q_k) = \text{span}(A^k Z_0)$

$$= \text{span}(S_n V_k + S_{n-m} W_k)$$

$$\longrightarrow \text{span}(S_n V_k) \quad \text{as } k \to \infty$$

= span of first n e-vectors of A.

Note that, we showed convergence of spans, but not column-by-column convergence — this is true but harder to prove.

Fact: first column of $Q_k \longrightarrow$ first e-vector of A

$$
\begin{array}{ccc}
\langle q_1^{(k)} \rangle & \longrightarrow & s_1 \\
\cap & & \\
\langle q_1^{(k)}, q_2^{(k)} \rangle & \longrightarrow & \langle s_1, s_2 \rangle \\
\cap \vdots & & \\
\langle q_1^{(k)}, \dots, q_n^{(k)} \rangle & \longrightarrow & \langle s_1, \dots, s_n \rangle.
\end{array}
$$

QR iteration

	$m \times m$	$n < m$	
The	Power $\rightarrow$	Orthogonal $\rightarrow$	QR.
	highest.	first n e-values.	all e-values
	$\lvert\lambda_1\rvert > \lvert\lambda_2\rvert$	$\lvert\lambda_1\rvert > \dots > \lvert\lambda_n\rvert > \lvert\lambda_{n+1}\rvert$	$\lvert\lambda_1\rvert > \dots > \lvert\lambda_m\rvert$.

As before, lets assume that $|\lambda_1| > \dots > |\lambda_m|$. ✩

Let $Z_0 = [x_1^{(0)} | \dots | x_m^{(0)}]$ be the initial matrix consisting of m orthogonal columns.

✩ Suppose all principal submatrices of Z_0 are non-singular. Then from the analysis done for orthogonal iteration,

the matrix $T_k = Q_k^* A Q_k$ converges to the upper Δ^r form as $k \to \infty$.

The algorithm for QR iteration can be obtained by considering how T_k is obtained from T_{k-1}.

Algorithm: $A \in \mathbb{C}^{m\times m}$, Z_0 unitary in $\mathbb{C}^{m\times m}$.
$T_0 = Z_0^* A Z_0$.

$k = 1$ to convergence

$$T_{k-1} = Q_k R_k \quad \text{(QR factorization)}$$

$$T_k = R_k Q_k$$

end

On the one hand, $T_{k-1} = Q_{k-1}^* A Q_{k-1} = Q_{k-1}^* (A Q_{k-1})$

$$= Q_{k-1}^* (Z_k)$$

$$= Q_{k-1}^* (Q_k R_k)$$

$$= \underline{(Q_{k-1}^* Q_k)} R_k$$

& on the other hand,

$$T_k = Q_k^* A Q_k = (\underbrace{Q_k^* A Q_{k-1}}) (\underbrace{Q_{k-1}^* Q_k})$$

$AQ_{k-1} = Z_k = Q_k R_k$

so $R_k = Q_k^* A Q_{k-1}$

$$= \underline{R_k (Q_{k-1}^* Q_k)}$$

Thus T_k is determined by computing the QR factorization of T_{k-1} & then multiplying the factors together in reverse order.

The algorithm produces a sequence of matrices $T_1, T_2, \ldots, T_k, \ldots$ which converges to the upper Δ^r Schur form of A.