Note Title

Eigenvalue methods for symmetric matrices.

(I) Rayleigh quotient iteration

Recall: Let $A : V \to V$ be a linear map (dim $V = m$)

① The Rayleigh quotient R_A is defined as a function

$$R_A : V \setminus \{0\} \longrightarrow \mathbb{R}$$
$$v \longmapsto \frac{v^T A v}{v^T v}$$

② If x is an e-vector corr. to an e-value λ of A, then $R_A(x) = \lambda$.

The problem of estimating λ given a vector x can be formulated as a least squares problem -

"find a scalar λ which minimizes $\| Ax - \lambda x \|_2$".

This is a system of equations, with variable λ and x being a $m \times 1$ vector & LHS of $Ax = \lambda x$ being the known quantity.

$AX = b$ ← usual setting

$A x = \lambda x$ ← present setting

In usual setting X_{LS} is given by set of normal eqns. $A^T A x = A^T b$.

In the present setting, the system of normal eqns. is

$$x^T x \lambda = x^T (A x).$$

i.e. $$\lambda = \frac{x^T A x}{x^T x} = R_A(x).$$

So $R_A(x)$ is a reasonable estimate of the e-value if x is an approximate e-vector.

The above ideas can be made precise by investigating the local behaviour of R_A in a neighbourhood of x, using derivatives -

considering x as a variable in $\mathbb{R}^m$ & $R_A : \mathbb{R}^m \to \mathbb{R}$,

(check) $\nabla R_A(x) = \frac{2}{x^T x}\left(A(x) - R_A(x) \cdot x \right)$ { using matrix derivatives, details on page 203, T-B.

So x is an e-vector with e-value λ $\Leftrightarrow$ $\nabla R_A(x) = 0$

Geometrically speaking, the e-vectors of A are the stationary points of R_A & e-values of A are the values of R_A at those stationary points.

It can be shown that if q is an e-vector of A, then $R_A(x) - R_A(q) = \Theta\left(\|x - q\|^2 \right)$ as $x \to q$.

$x^{(0)} \downarrow x^{(1)} \downarrow \vdots\ x^{(k)} \downarrow x^{(k+1)}$ $\quad \lambda^{(0)} \downarrow \lambda^{(1)} \vdots\ \lambda^{(k)} \downarrow \lambda^{(k+1)}$

This implies that the Rayleigh quotient is a quadratically accurate estimate of an e-value; i.e. $R_A(x)$ approaches $R_A(q)$ quadratically as fast as x approaches q.

This constitutes the strength of this method.

What is a good starting vector x_0?

Theorem: The Rayleigh quotient iteration converges to e-value/e-vector pair for all except a set of measure zero (i.e., a negligible set) of starting vectors $v^{(0)}$. When it converges, the convergence is cubic in the following sense -

if λ is an e-value of A & v is a vector sufficiently close to e-vector x, then

$$\| v^{(k+1)} - x \| = \theta\left(\| v^{(k)} - x \|^3 \right)$$

$$\& \quad \| \lambda^{(k+1)} - \lambda \| = \theta\left(\| \lambda^{(k)} - \lambda \|^3 \right)$$

Idea :
- start with a vector $v^{(0)}$
- Calculate $R_A(v^{(0)})$, this is the first approximation to $\lambda^{(0)}$.
- apply inverse iteration to $\lambda^{(0)}$ to get $v^{(1)}$, then $\lambda^{(1)} = R_A(v^{(1)})$ & so on.

$$\left\{ \begin{matrix} \text{approx.} \\ \text{e-vector} \end{matrix} \right\} \underset{\text{inverse iteration}}{\overset{R_A}{\rightleftarrows}} \left\{ \begin{matrix} \text{approx.} \\ \text{e-value} \end{matrix} \right\}$$

Algorithm : choose $v^{(0)}$ with $\| v^{(0)} \| = 1$.

$$\lambda^{(0)} = v^{(0)T} A v^{(0)} \qquad \text{i.e. } R_A(v^{(0)})$$

for $k = 1$ to convergence.

(inverse iteration) ... solve $(A - \lambda^{(k-1)} I) w = v^{(k-1)}$ for w

(approx. e-vector) ... $v^{(k)} = w / \|w\|_2$

(approx. e-value) ... $\lambda^{(k)} = R_A(v^{(k)}) = v^{(k)T} A v^{(k)}$.

end.

(II) Jacobi's method.

$$\underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{A} \xrightarrow{P^{-1}AP} \begin{bmatrix} * & & \\ & * & \\ & & * \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} * & \approx 0 & \approx 0 \\ & * & \approx 0 \\ \approx 0 & * & \\ \approx 0 & \approx 0 & * \end{bmatrix}$$

The basic idea is to try & reduce the magnitude of off-diagonal elements, so that eventually they become small enough & can be declared to be zero.

The idea is formulated as follows - consider the Frobenius norm of the off-diagonal elements -

$$\mathrm{off}(A) = \sqrt{\sum_{i=1}^{n}\sum_{\substack{i\neq j \\ j=1}}^{n} a_{ij}^2}$$

The similarity transformations are applied in such a way that off(A) is systematically reduced. We use Jacobi's rotation matrices -

$$J(p,q,\theta) = \begin{array}{c} \\ \\ p \\ \\ q \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & p & & q & \end{array} \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ \cdots & c & \cdots & s & \cdots \\ & & \ddots & & \\ \cdots & -s & \cdots & c & \cdots \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \end{array}$$

The basic step involves choosing an index pair (p,q) & computing the corresponding sine-cosine pair (c,s) such that

$$\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}^T \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix}$$

$(b_{pq} = b_{qp} \underset{\text{want}}{=} 0)$ $\qquad$ $(a_{pq} = a_{qp})$

the LHS matrix is diagonal.

Observe: (1) The matrix $B = J^T A J$ agree with A in all entries except rows & columns p & q.

(2) The Frobenius norm is preserved by orthogonal transf., so we have:

$$a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 = b_{pp}^2 + b_{qq}^2 + 2\underbrace{b_{pq}^2}_{=0}$$

$$= b_{pp}^2 + b_{qq}^2 \qquad (\star)$$

As a result:

$$\text{off}(B)^2 = \|B\|_F^2 - \sum_{i=1}^{n} b_{ii}^2$$

$$= \|A\|_F^2 - \sum_{\substack{i \neq p,q \\ i=1}}^{n} b_{ii}^2 - b_{pp}^2 - b_{qq}^2 + (a_{pp}^2 + a_{qq}^2) - (a_{pp}^2 + a_{qq}^2)$$

$$= \|A\|_F^2 - \left(\sum_{\substack{i \neq p,q \\ i=1}}^{n} a_{ii}^2 + a_{pp}^2 + a_{qq}^2\right)$$

$$+ \underbrace{(a_{pp}^2 + a_{qq}^2 - b_{pp}^2 - b_{qq}^2)}_{\overset{\star}{=} -2a_{pq}^2}$$

$$= \underbrace{\|A\|_F^2 - \sum_{i=1}^{n} a_{ii}^2}\ - 2a_{pq}^2$$

$$\therefore \underbrace{\text{off}(B)^2} = \underbrace{\text{off}(A)^2} - \underbrace{2a_{pq}^2}_{>0}.$$