Note Title

Perturbation theory for the eigenvalue/e-vector problem.

I) An important (& often used) framework for e-value computation is to produce a sequence of similarity transformations $\{X_k\}$ such that $X_k^{-1} A X_k$ are progressively "more diagonal" i.e. the sequence $\{X_k^{-1} A X_k\}$ converges to an almost diagonal matrix.

(i) In this context, a basic question is - what information do the diagonal entries of A give about the eigenvalues of A?

The answer is provided by Gerschgorin's theorem.

(refer to the reading assignment).

(ii) How are the eigenvalues of A affected by small perturbations in A?

Bauer-Fike theorem: Suppose E is small in norm & suppose μ is an eigenvalue of $A+E$.

If $X^{-1} A X = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then

$A \quad A+E$

$\lambda \quad \lambda + \delta\lambda = \mu.$

$|\delta\lambda|$

$$\min_{\lambda \in \lambda(A)} |\lambda - \mu| \leq K_p(X) \cdot \|E\|_p$$

($\|\cdot\|_p$ is any p-norm.)

(iii) Recall that for any $n \times n$ matrix A, we have the Schur decomposition $Q^T A Q = T$, where Q is unitary & T is upper Δ^r with e-values of A on the diagonal.

$\therefore$ We may write $Q^T A Q = T = D + N$

where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$

& N is strictly upper Δ^r

When A is normal, $Q^T A Q = D$ & in this case $N=0$.

The foll. result describes the effect of perturbations in A on the e-values of A in the language of Schur decomposition:

<u>Theorem</u>: If μ is an e-value of $A+E$ & p is the smallest positive integer such that $|N|^p = 0$, then

$$\min_{\lambda \in \lambda(A)} |\lambda - \mu| \leq \max\{\theta, \theta^{1/p}\}$$ where $\theta = \|E\|_2 \cdot \sum_{k=0}^{p-1} \|N\|_2^k$

An upshot of the above theorem is that if A is normal, then $N=0 \Rightarrow \theta = \|E\|_2 \cdot \|N\|_2^0 = \|E\|_2$, so extreme eigenvalue sensitivity does not occur.

On the other hand, the theorem also suggests that eigenvalues of non-normal matrices may be sensitive to perturbations. However, non-normality does not necessarily imply eigenvalue sensitivity. A matrix can have a mix of well-conditioned & ill-conditioned e-values.

So for non-normal matrices, it is useful to consider sensitivity of individual eigenvalues rather than of the spectrum as a whole.

<u>Example</u>: $A = \begin{bmatrix} 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ \vdots & & \ddots & 1 \\ \varepsilon & \cdots & \cdots & 0 \end{bmatrix}_{n \times n}$, char. poly. of A is

$$\lambda^n - \varepsilon$$

$\therefore$ The e-values of A are

$$\lambda = \sqrt[n]{\varepsilon}.$$

$$\frac{d\lambda}{d\varepsilon} = \frac{\varepsilon^{\frac{1}{n}-1}}{n} \quad ; \quad \text{if } n \geq 2, \text{ then for } \varepsilon \to 0$$

$$\frac{d\lambda}{d\varepsilon} = \infty .$$

$\therefore$ the e-value λ will be very badly conditioned.

This reflects the ill-conditioning that occurs for eigenvalues with multiplicity > 1.

(iv) Let λ be a simple e-value of A with $Ax = \lambda x$ & $y^T A = \lambda y$.

It can be shown that $s(\lambda) = \frac{1}{|y^T x|}$ determines the conditioning of λ.

(v) Conditioning of e-vector computation is dependent on the multiplicity of its e-value. For simple e-values, the computation is likely to be well conditioned, hence stable. For e-vectors of repeated e-values the problem is more complicated.