Eigenvalues and Eigenvectors

Definition 5.1: Suppose $A \in \mathbb{R}^{n \times n}$.

If $\lambda \in \mathbb{C}$ satisfies

$$A x = \lambda x$$

and $\underline{x} \neq \underline{0}$, then λ is called an eigenvalue of A. x is the associated eigenvector.

~~Fact~~:

Remark: Why do we care about eigenvalues and eigenvectors?

Consider a general dynamical system

$$\frac{dx}{dt} = Ax, \quad A \in \mathbb{R}^{n \times n}.$$

$$x(0) = x_0.$$

A solution can be rewritten in the form:

$$x(t) = x(0) \underbrace{\exp(At)}_{\text{matrix exponential}}.$$

If we can "diagonalize" A, $A = Q D Q^{-1}$,

here Q is a matrix with eigenvectors in the columns, D is a diagonal matrix with the eigenvalues on the diagonal,

then the system of ODEs can be rewritten

$$\frac{dx}{dt} = Q D Q^{-1} x$$

$$\Leftrightarrow \quad \frac{d\, Q^{-1} x}{dt} = D Q^{-1} x.$$

Defining $y = Q^{-1} x$, we obtain a new, decoupled set of ODES:

$$\frac{dy}{dt} = Dy, \quad \underline{\text{or}}:$$

$$\frac{dy_i}{dt} = \lambda_i y_i ,$$

an the eigenvalues tell us which components of y <u>grow</u> or <u>decay</u>.

Gerschgorin Thms

Definition 5.5 $A \in \mathbb{C}^{n\times n}$, $n \geq 2$., $A = (a_{ij})$

D_i = Gerschgorin discs

$$= \{ z \in \mathbb{C} : |z - a_{ii}| \leq R_i \}.$$

with $$R_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|.$$

Thm 5.4 $A \in \mathbb{C}^{n\times n}$, $n \geq 2$. All of the eigenvalues of A are in $D = \bigcup_{i=1} D_i$,

D_i = Gerschgorin disc.

Pf. Let (x, λ) be an eigenpair. by definition

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i, \quad i = 1, \ldots, n.$$

Take x_k to be the component with largest absolute value: $|x_j| \leq |x_k|$, $j = 1, \ldots, n$.

Then

$$|\lambda - a_{kk}||x_k| = |\lambda x_k - a_{kk} x_k|$$

$$= \left| \sum_{j=1}^{n} a_{kj} x_j - a_{kk} x_k \right|$$

$$= \left| \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} x_j \right|$$

$$\leq |x_k| \, R_k .$$

This shows that $|\lambda - a_{kk}| \leq R_k$, i.e. λ is in D_k.

Thm 5.5: (Gerschgorin's 2^{nd} theorem).
let $n \geq 2$.
let $1 \leq p \leq n-1$.

Suppose the Gerschgorin disks of $A \in \mathbb{C}^{n \times n}$ can be divided into two disjoint subsets $\underbrace{D^{(p)}}_{p \text{ disks}}$ and $\underbrace{D^{(q)}}_{n-p \text{ discs}}$.

Then $D^{(p)}$ contains exactly p eigenvalues and $D^{(q)}$ contains exactly $n-p$ eigenvalues.

Proof: Neat idea: define a matrix

$$B(\varepsilon) = (b_{ij}(\varepsilon)), \qquad b_{ij}(\varepsilon) = \begin{cases} a_{ii} & i = j \\ \varepsilon a_{ij} & i \neq j \end{cases}$$

Note that $B(1) = A$. Also, note that the eigenvalues of $B(0)$ are the diagonal elements of A. Now use continuity of the eigenvalues as a function of ε to finish the proof.

<u>Fact</u>: The eigenvalues of a matrix A are roots of the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$.

<u>Fact</u>: There is a one-to-one correspondence between eigenvalue problems and root-finding problems.

i.e. If $p(z) = z^m + a_{m-1} z^{m-1} + \dots + a_1 z + a_0$,

then the roots of p are actually the eigenvalues of the matrix

$$\begin{bmatrix} 0 & & & & -a_0 \\ 1 & \ddots & & & \vdots \\ & \ddots & \ddots & & \vdots \\ & & 1 & 0 & -a_{m-2} \\ & & & 1 & -a_{m-1} \end{bmatrix}$$

The above matrix is called the <u>Companion Matrix</u> for p.

<u>Thm</u> (Abel 1824).
For any $m \geq 5$, there is a polynomial $p(z)$ of degree m with rational coeff's that has a real root with the property that it cannot be written in any expression involving rational numbers, $+$, $-$, $*$, $\div$, and k^{th} roots.

Further issue: determining roots of a characteristic polynomial, or more generally any polynomial, is ill-conditioned.

Example: Consider the 2x2 identity matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

writing down the characteristic polynomial, we get $p(\lambda) = (\lambda-1)^2 = \lambda^2 - 2\lambda + 1$.

If we perturb the constant coeff in this polynomial by 10^{-4}:

$$\tilde{p}(\lambda) = \lambda^2 - 2\lambda + 0.9999,$$

the roots of this perturbed poly are:

$$\tilde{p}(\lambda) = (\lambda - 0.99)(\lambda - 1.01),$$

i.e. they are perturbed by 10^{-2} from the original roots of $p(\lambda)$!!!

$\boxed{\text{Power method}}$ for computing eigenvalues and eigenvectors.

Let's assume $A \in \mathbb{R}^{m \times m}$ is a real, symmetric matrix.

Definition: (Rayleigh Quotient).

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

Note that if x is an eigenvector of A, then $r(x) = \lambda$ is its corresponding eigenvalue.

Via taking derivatives, we can see that $\nabla r(x) = \frac{2}{x^T x} (Ax - r(x)x)$.

So, if x is an eigenvector of A, then $\nabla r(x) = 0$. And, if $\nabla r(x) = 0$ with $x \neq 0$, then x is an eigenvector of A, with eigenvalue $r(x)$.

Remark: let q_J be an eigenvector of A. We know that $\nabla r(q_J) = 0$.

By Taylor expansion, we have:

$$r(x) - r(q_J) = \nabla r(q_J)(x - q_J)$$
$$+ (x - q_J)^T \nabla^2 r(\xi)(x - q_J)$$

for some ξ on the line between x and q_J.

This shows that

$$|r(x) - r(q_J)| \leq C \, \| x - q_J \|^2,$$

since $\nabla r(q_J) = 0$.

The idea is if we have a good estimate for an ~~eigenvalue~~ eigenvector, we can get a quadratically accurate estimate for the corresponding eigenvalue.

Power Method

let $v^{(0)}$ with $\|v^{(0)}\| = 1$ be some initial guess for an eigenvector of A. The power method, or power iteration, is defined as:

for $k = 1, 2, \ldots$

$$w = A v^{(k-1)} \qquad \text{apply } A$$
$$v^{(k)} = w / \|w\| \qquad \text{Normalize}$$
$$\lambda^{(k)} = (v^{(k)})^T A v^{(k)} \qquad \text{Rayleigh Quotient.}$$

The claim is that $\lambda^{(k)}$ converges to an eigenvalue of A and $v^{(k)}$ converges to the corresponding eigenvector, as $k \to \infty$. This iteration should converge to the <u>largest eigenvalue in absolute value.</u>

<u>Thm:</u> Suppose $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m| \geq 0$

and $q_1^T v^{(0)} \neq 0$. Then the power iteration above creates iterates satisfying

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

Pf: Take the 1st iterate, $v^{(0)}$, and expand in the basis of eigenvectors of A:

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

$w = Av^{(k-1)}$ is actually a multiple of $A^k v^{(0)}$, up to normalizing constants c_k, so:

$$\begin{aligned} w &= c_k A^k v^{(0)} \\ &= c_k \left(a_1 \lambda_1^k q_1 + a_2 \lambda_2^k q_2 + \cdots + a_m \lambda_m^k q_m \right) \\ &= c_k a_1 \lambda_1^k \Big(q_1 + \underbrace{\tfrac{a_2}{a_1}\left(\tfrac{\lambda_2}{\lambda_1}\right)^k q_2 + \cdots + \tfrac{a_m}{a_1}\left(\tfrac{\lambda_m}{\lambda_1}\right)^k q_m}_{z_k} \Big). \\ &= c_k a_1 \lambda_1^k \left(q_1 + z_k \right). \end{aligned}$$

Now, we normalize:

Why? Because we don't know the sign of c_k. So:

$$v^{(k)} = \frac{w}{\|w\|} = (\pm)\left(\frac{q_1 + z_k}{\|q_1 + z_k\|} \right)$$

$$v^{(k)} - (\pm) q_1 = \frac{q_1 + z_k}{\|q_1 + z_k\|} - q_1 \frac{\|q_1 + z_k\|}{\|q_1 + z_k\|}$$

$$= \frac{\left(1 - \|q_1 + z_k\|\right) q_1}{\|q_1 + z_k\|} + \frac{z_k}{\|q_1 + z_k\|}.$$

So: $\| v^{(k)} - q_1 \|$

$$\leq \frac{\left| 1 - \|q_1 + z_k\| \right| \|q_1\|}{\|q_1 + z_k\|} + \frac{\|z_k\|}{\|q_1 + z_k\|}$$

$$= \left(\left| 1 - \|q_1 + z_k\| \right| + \|z_k\| \right) \frac{1}{\|q_1 + z_k\|}$$

$$= \left(\left| \|q_1\| - \|q_1 + z_k\| \right| + \|z_k\| \right) \frac{1}{\|q_1 + z_k\|}$$

Back half of Δ-inequality $\leq \dfrac{2\|z_k\|}{\|q_1 + z_k\|}$)

i.e. $\| v^{(k)} - q_1 \| \to 0$ as $\|z_k\| \to 0$.

Note that $\|z_k\| \to 0$ with a rate that depends on the ratio $\left(\frac{\lambda_2}{\lambda_1}\right)^k$.

The rate of convergence for the eigenvalues approximations $\lambda^{(k)}$ is given by the properties of the Rayleigh quotient.

Inverse Iteration:

What about finding other eigenvalues and eigenvectors?

Idea: "Shift" the matrix A so that we change the dominant eigenvalue:

Take $\mu \in \mathbb{R}$, μ not an eigenvalue of A.

Let $\{\lambda_j\}$ be the eigenvalues of A.

Then $\{\lambda_j - \mu\}$ are the eigenvalues of

$$(A - \mu I).$$

and $\{(\lambda_j - \mu)^{-1}\}$ are the eigenvalues of

$$(A - \mu I)^{-1}.$$

If $\mu \approx \lambda_j$, then $(\lambda_j - \mu)^{-1}$ is really big, and probably the largest eigenvalue of $(A-\mu I)^{-1}$ in magnitude.

$\rightsquigarrow$ Apply power method to $(A-\mu I)^{-1}$.

Inverse Iteration Algorithm:

$v^{(0)}$, $\|v^{(0)}\| = 1$.

for $k = 1, 2, \ldots$

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

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

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