Clase 20
10/11/22
ALN 2022

Recordar que clase 19 fue charla de Rodrigo sobre Algoritmo QR y convergencia del mismo basado en to Shub-Vasquez.

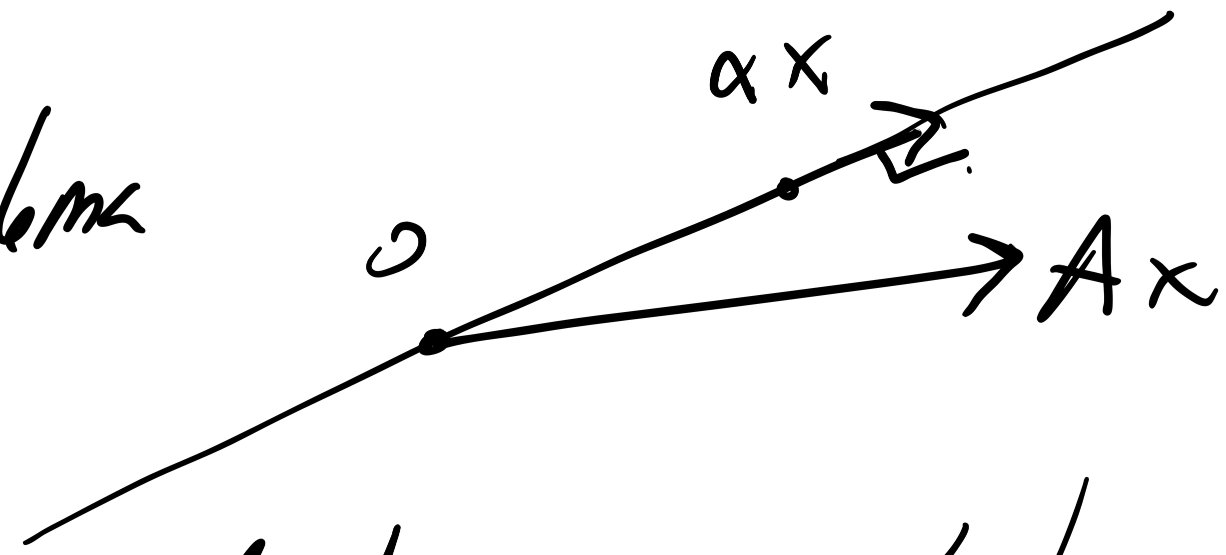
→ Matrices reales simétricas $A \in \mathbb{R}^{m \times m}$ / $A^T = A$

COCIENTE de RAYLEIGH

El C.R. de un vector $x \in \mathbb{R}^m \setminus \{0\}$ (para una matriz A fija) se define como

$$\rho_A(x) = \frac{x^T A x}{x^T x} = \frac{\langle Ax, x \rangle}{\langle x, x \rangle}, \quad x \neq 0.$$

Una forma de motivar ~~su exp~~ el CR es que $\rho_A(x)$ es la solución al problema de minimizar $\| \alpha x - Ax \|$ con $\alpha \in \mathbb{R}$ (que puede verse como un problema de mínimo cuadrados $\| a x - b \|$) ~~mmma~~



Es fácil ver que la solución mínima se satisface para α /

$$\alpha x - Ax \perp \alpha x \quad \text{y por lo } \cancel{\text{tanto}} \quad \langle \alpha x, \alpha x - Ax \rangle = 0$$

sii $\alpha^2 \|x\|^2 = \alpha \langle Ax, x \rangle$.

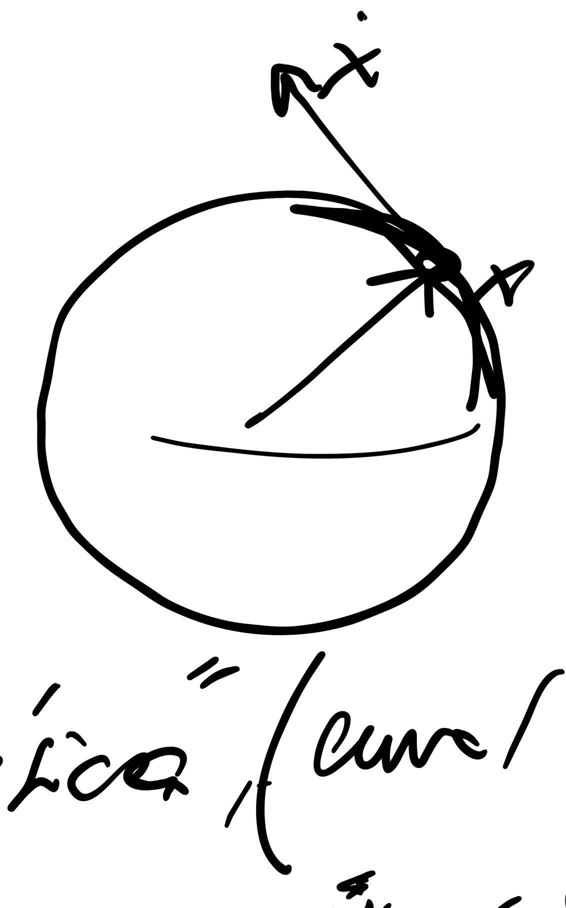
Observar que ρ_A es invariante por cambios de escala:

$\rho_A(\theta x) = \rho_A(x)$ por lo que tiene sentido definir

$$\rho_A : S^{m-1} \longrightarrow \mathbb{R}.$$

Proposición: Sean $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ los val. propios de A simétrica.

Entonces $\lambda_1 = \max_{x \in S^{n-1}} \rho_A(x)$, $\lambda_n = \min_{x\in S^{n-1}} \rho_A(x)$. Además los puntos críticos de ρ_A en S^{n-1} son los vectores propios.



Dem: Una forma sencilla de ver esto es observar que dado $x \in S^{n-1}$ y $\dot{x} \in x^{\perp}$ ($\|\dot{x}\|=1$) entonces podemos considerar la función $t \overset{f}{\longmapsto} \rho_A(\underbrace{\cos t\, x + \operatorname{sen} t\, \dot{x}}_{\gamma(t)})$ $\longrightarrow$ "geodésica" (curva $\gamma(0)=x$, $\gamma(t)\in S^{n-1}$, $\dot{\gamma}(0)=\dot{x}$)

Luego se tiene $D\rho_A(x)\dot{x} = \frac{d}{dt} f(t)\Big|_{t=0} = \frac{d}{dt}\rho_A(\gamma(t))\Big|_{t=0}$

que no es otra cosa que $\cancel{\frac{d}{dt}}\rho_A$

$$\frac{d}{dt} f(t)\Big|_{t=0} = \frac{d}{dt}\Big|_{t=0} \langle A(\cos t\, x + \operatorname{sen} t\, \dot{x}), \cos t\, x + \operatorname{sen} t\, \dot{x}\rangle$$

$$= \frac{d}{dt}\Big|_{t=0} \cos^2 t \langle Ax, x\rangle + \cos t \operatorname{sen} t \langle Ax, \dot{x}\rangle + \operatorname{sen}^2 t \langle A\dot{x}, \dot{x}\rangle$$

$(*)$ $$= \Big[-2\cos t \operatorname{sen} t \langle Ax, x\rangle + (\cos^2 t - \operatorname{sen}^2 t)\langle Ax, \dot{x}\rangle + 2\operatorname{sen} t \cos t \langle A\dot{x}, \dot{x}\rangle\Big]_{t=0}$$

$$= \langle Ax, \dot{x}\rangle.$$

Por lo tanto, $D\rho_A(x)\dot{x} = 0$ sii $\dot{x} \perp Ax$ $\forall \dot{x} \in x^{\perp}$,

por lo que $Ax = \alpha x$ para algún $\alpha \in \mathbb{R}$, i.e.

x es crítico sii es vector propio de ρ_A. $\square$

Si queremos podemos clasificar los puntos críticos considerando el "Hessiano", en nuestro caso se puede hacer estudiando $\frac{d^2}{dt^2}\Big|_{t=0} \rho_A(\gamma_{\dot x}(t))$ variando $\dot x \in x^{\perp}$.

De (*) resulta $$\frac{d^2}{dt^2}\Big|_{t=0} \rho_A(\gamma_{\dot x}(t)) = -2\langle Ax, x\rangle + 2\langle A\dot x, \dot x\rangle$$

i.e $$Hess_x(\dot x, \dot x) = 2\left[\langle A\dot x, \dot x\rangle - \langle Ax, x\rangle\right].$$

Siendo x vector propio, y $x^{\perp}$ invariante resulta en en $x = v_1$ $Hess_x \leq 0$ (máx) y $x = v_m$ $Hess_x > 0$ (mínimo) y el resto es punto silla.

Un punto interesante del cociente de Rayleigh es que ~~...~~ si x es cercano a vector propio v_j entonces (utilizando Taylor de orden 1 a la función $f(t) = \rho_A(\gamma_{\dot x}(t))$) podemos ver

$$\left| \rho_A(x) - \underbrace{\rho_A(v_j)}_{\lambda_j} \right| = O(\|x - v_j\|^2).$$

Por lo tanto si tenemos una buena aproximación de un vector propio, ρ_A nos da una muy buena aproximación del valor propio.

Recordar que el método de la potencia aproxima el vector propio dominante, pero la velocidad de aproximación es lineal:

$$d_{\mathbb{P}}\left(A^k(v), v_1\right) \simeq \left|\frac{\lambda_2}{\lambda_1}\right|^k.$$

Y además solo serviría para el vector propio dominante. Pero una simple observación puede convertir el método de la potencia en un algoritmo muy útil.

Si $\mu \in \mathbb{R}$ es cercano a λ_{j_0} (val. propio de A), entonces $A-\mu I_m$ tiene a v_j como vector propio y $\lambda_j-\mu$ como valor propio. Y análogamente $(A-\mu I_m)^{-1}$ tiene mismos vectores propios y los valores propios son $(\lambda_j-\mu)^{-1}$.

Por lo tanto si $\mu \simeq \lambda_{j_0} \Rightarrow \lambda_{j_0}-\mu \simeq 0$ y $(\lambda_{j_0}-\mu)^{-1}$ es gigante. En otras palabras si μ es muy cercano a un valor propio ($\min_i |\mu-\lambda_i| = |\mu-\lambda_{j_0}|$) $\Rightarrow$ v_{j_0} es el vector propio dominante de $(A-\mu I_m)^{-1}$ y por lo tanto el método de la potencia sirve para aproximarlo.

Una idea se viene a la mente, si tenemos una buena estimativa del valor propio λ_{j_0}, entonces tendremos una buena estimativa del vector propio, y utilizando el cociente de Rayleigh tendremos una mejor estimativa del valor propio.

Este análisis sugiere un algoritmo conocido con

<u>Iteración de Rayleigh Quotient</u> (de cociente de Ray.)

(1) Elegir $v^{(0)} \in \mathbb{R}^m$

For $k = 0, 1, 2, \ldots$

(2) Si $(A - \rho_A(v^{(k)}) I_m)$ singular, resolver $(\cdot = 0)$ para encontrar vector propio

De lo contrario Resolver $\left(A - \rho_A(v^{(k)}) I_m\right) w = v^{(k)}$

(3) Normalizar $v^{(k+1)} := \frac{w}{\|w\|}$

<u>Teo</u> (Kahan-Ostrowski) Si $\phi_k = d_{\mathbb{P}}(v^{(k)}, v_{j_0})$ → vector propio más próximo.

Entonces $v^{(k)} \longrightarrow v_{j_0}$ vector propio

$$\lim_{k} \frac{\phi_{k+1}}{\phi_k^{3}} \leq 1. \qquad \text{(Caso simétrico)}$$

Por mucho tiempo se pensó que este algoritmo siempre convergía en el caso de matrices no simétricas. Pero:

<u>Teo</u>: (Batterson-Smillie) $\forall n \geq 3$ $\exists$ un abierto de matrices donde el algoritmo falla (en conjunto positivo de $v^{(0)}$ iniciales el algoritmo no converge.)

Desde 1960 el algoritmo más utilizado hasta 1990 es el siguiente

"Practical" QR Algoritmo (Francis)

(1) $(Q^{(0)})^T A^{(0)} Q^{(0)} = A$, → $A^{(0)}$ Hessenberg de A (tridiagonal

(2) for $k = 1, 2, \dots$

elegimos $\mu^{(k)}$ → por ejemplo $\mu^{(k)}$ $A^{(k-1)}_{mm}$ o valor propio de bloque $\begin{pmatrix} A^{(k-1)}_{m-1\,m-1} & A^{(k)}_{m-1\,m} \\ A^{(k-1)}_{m\,m-1} & A^{(k)}_{mm} \end{pmatrix}$ más cercano a $A^{(k)}_{mm}$

$$Q^{(k)} R^{(k)} = A^{(k-1)} - \mu^{(k)} I_m$$

$$A^{(k)} = R^{(k)} Q^{(k)} + \mu^{(k)} I_m$$

(damos vuelta Q y R y shifteamos)

(3) Si $A^{(k)}_{j,j+1}$ cercano a cero, ⟶ $\begin{pmatrix} * & * & 0 & - \\ * & * & * & 0 \\ 0 & 0 & * & * \\ & & * & * \end{pmatrix}$

redefinimos $A^{(k)} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$

y aplicamos algoritmo anterior a cada factor por separado

¿Qué podemos decir de la convergencia?

Conjetura Shub: el algoritmo falla en un abierto de condiciones iniciales.