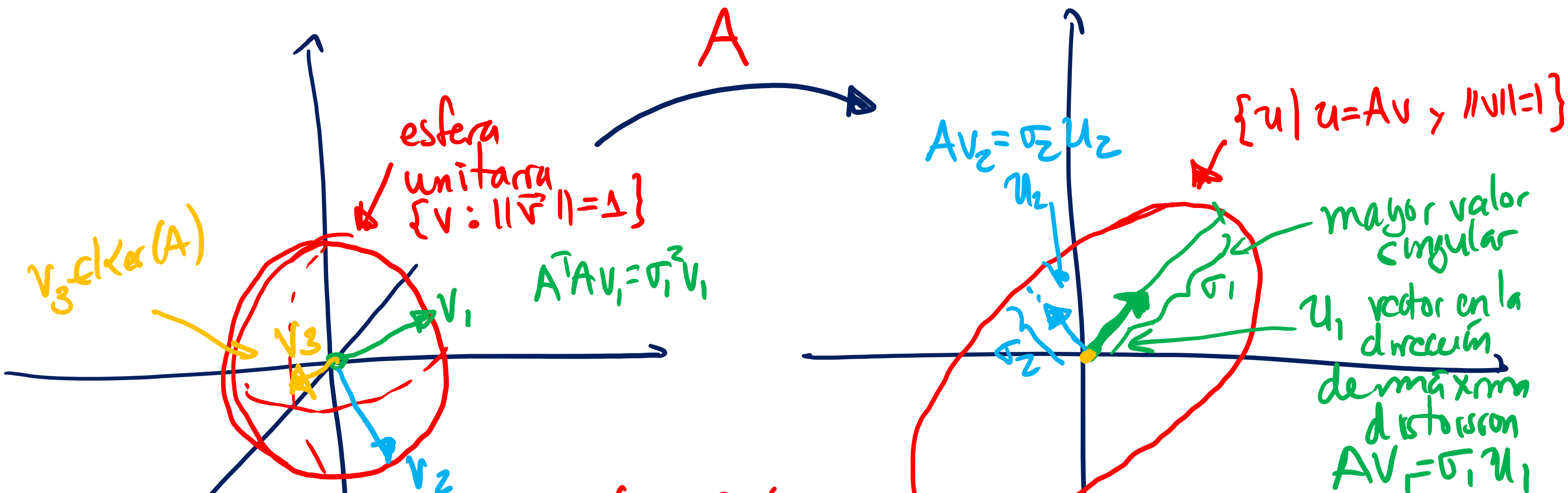
Recap

Clase 8 Curso Propedéutico
2017/06/21

Recap SVD

$$A = U \Sigma V^T$$

U ortogonal
V ortogonal
Σ diagonal acompletada con ceros en dimensiones sobrantes



$$v_i^T v_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$AV = A\left[v^1 | v^2 | v^3\right] = \left[u^1 | u^2\right]\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$$

$$= \left[\sigma_1 u^1 | \sigma_2 u^2 | 0\right]$$

Solución de sistemas de ecuaciones con la SVD

$(AB)^{-1} = B^{-1}A^{-1}$

$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$

- Supongamos que queremos resolver un sistema de ecuaciones

$$Ax = b$$

$\tilde{A}^{-1} = V\tilde{\Sigma}^{-1}U^T$

Si $A = U\Sigma V^{\top}$ entonces $\tilde{A}^{-1}Ax = \tilde{A}^{-1}b$

$$x = \boxed{V\tilde{\Sigma}^{-1}U^T b}$$

Solución al sistema de ecuaciones "si $\sigma_j = 0$ poner 0"

$\tilde{A}^{-1}A = V\tilde{\Sigma}^{-1}U^T U\Sigma V^T = V\tilde{\Sigma}^{-1}\Sigma V^T = VV^T = I$ suponiendo invertible

Es solución.

Tengo Invertir $\Sigma = \begin{bmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & \sigma_m & 0 \end{bmatrix}$ $\quad \tilde{\Sigma}^{-1} = \begin{bmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_m \\ \hline 0 & \cdots & 0 \end{bmatrix}$

ojo: si A es invertible $\tilde{\Sigma}^{-1}$ es la inversa de Σ

Nota!!!! Observemos que este método siempre da una solución incluso si la matriz no es invertible!!!!!!!

$\Sigma\tilde{\Sigma}^{-1} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 0 \ \ddots \ 0 \end{bmatrix}$

Este método de solución de ecuaciones tiene muchas ventajas

- Estabilidad numérica
- Menos sensibilidad a errores de rendondeo

Algunos conceptos de cálculo numérico

http://ta.twi.tudelft.nl/users/vuik/wi211/disasters.html

Número de condición

- Los errores numéricos surgen por errores de redondeo. La computadora solo puede representar un número con una cantidad finita de elementos (bits).

- El **número de condición** mide cuánto cambia la solución de un sistema de ecuaciones si el vector independiente tiene un error de redondeo e

$$\kappa(A) = \boxed{\frac{\sigma_{max}}{\sigma_{min}}}$$

tienen b y $\tilde{b} = b + \epsilon$

Comparemos las soluciones de $Ax_1 = b$ y $Ax_2 = \tilde{b}$

$x_1 - x_2 = A^{-1}b - A^{-1}\tilde{b} = A^{-1}\epsilon$ ← error en la solución

Pregunta ¿qué tan grande puede ser el cambio en la solución relativo al tamaño del error ϵ?

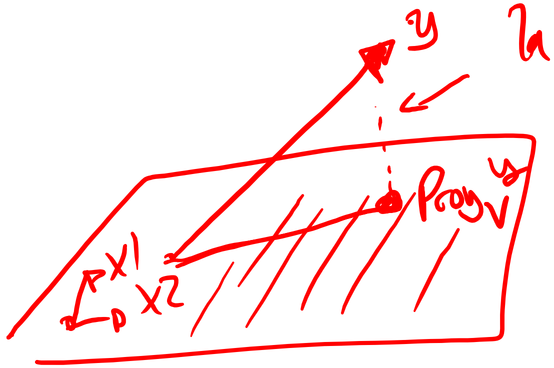
$$\kappa(A) = \max_{b, \epsilon \neq 0} \left\{ \underbrace{\frac{\|A^{-1}\epsilon\|}{\|A^{-1}b\|}}_{\text{error en solución relativo a tamaño solución}} \Big/ \underbrace{\frac{\|\epsilon\|}{\|b\|}}_{\text{tamaño error relativo a input } b} \right\}$$

$$= \max_{\epsilon} \left(\frac{\|A^{-1}\epsilon\|}{\|\epsilon\|} \right) \max_{b} \left(\frac{\|b\|}{\|A^{-1}b\|} \right)$$

$$= \max_{\epsilon} \left(\frac{\|A^{-1}\epsilon\|}{\|\epsilon\|} \right) \max_{u} \left(\frac{\|Au\|}{\|u\|} \right) = \boxed{\|A^{-1}\| \|A\|}$$

σ_{min} σ_{max}

Proyecciones en subespacios



la proyección minimiza el error $\|y - Proy_V y\|$

$$V = span(X^1 | \dots | X^n)$$

$$X = [X^1 | \dots | X^n]$$

$$Proy_V y = \beta_1 X^1 + \dots + \beta_n X^n = X\beta$$

Problema de Proyección

$$\min_{\beta} \|y - X\beta\| \sim \min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \sim \min_{\beta} \underbrace{\frac{1}{2}(y - X\beta)^T(y - X\beta)}_{f(\beta)}$$

Derivar e igualar ~~a cero~~

$$\nabla f(\beta) = \frac{1}{2}\nabla\left(y^T y + \beta^T X^T X \beta - 2\beta^T X^T y \right)$$

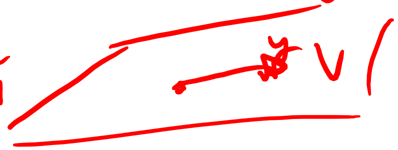
$$= \frac{1}{2}\left(2X^T X\beta - 2X^T y \right)$$

$$\nabla f(\beta^*) = 0 \iff \beta^* = \underline{(X^T X)^{-1} X^T y}.$$

$$\therefore\ Proy_V y = X\beta^* = X(X^T X)^{-1} X^T y = \hat{H} Y$$

Definición abstracta

P se llama matriz de proyección en un espacio V si $Py = y \ \forall y \in V$



- Pueden checar que $X(X^TX)^{-1}X^T = \hat{H}$ es la proyección en span$(X^1, \dots, X^n)$.

~~Recap de todo el curso~~

Regresamos a SVD

dado $A = U\Sigma V^T$ $\qquad \Sigma^+ = \tilde{\Sigma}^{-1}$ notación.

la pseudo inversa de A es la matriz

$$A^+ = \tilde{A}^{-1} = U\tilde{\Sigma}^{-1}V^T = V\Sigma^+U^T$$

Si queremos resolver $Ax = b \Rightarrow x = A^{-1}b$ si A es invertible

pero si no es invertible $x = A^+b$ es la "mejor aproximación a una inversa".

Tarea: 1) Checar que A^+ proyecta en span$(A^1 \dots A^b)$ 2) Encontrar relación entre A^+ y $\hat{H}$

¿Qué hemos visto?

1. Vectores y proyecciones
2. Matrices/Transformaciones Lineales
3. Kernel e imagen (espacio columna) de una transformación/relación con sistemas de ecuaciones.
4. Determinante como distorsión del volumen.
5. Matrices inversas y relación con el determinante
6. Matrices Ortogonales
7. Eigenvectores
8. Optimización con y sin restricciones
9. Funciones cuadráticas
10. Descenso gradiente
11. Descomposición SVD
12. Relación entre SVD, optimización y normas matriciales
13. SVD, Eigenvectores y Optimización.

1 Vector = dirección + magnitud / flecha / lista ordenada / $v \in \mathbb{R}^n$

2 matriz = $\underbrace{\text{"arrays" o "listas de listas"}}_{\text{parte computacional}}$ / transformaciones: sus columnas son el efecto sobre vectores canónicos

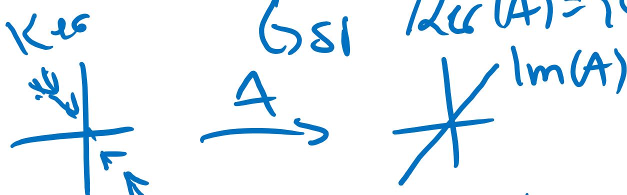
/ $A \in \mathbb{R}_{m \times n}$ $\quad T: \mathbb{R}^n \to \mathbb{R}^m \quad T(x) = Ax$

3 Kernel e imagen:
- subespacios
- $Ker(A) = \{v \mid Av = 0\}$
- $Im(A) = \{w \mid$ existe v tal que $Av = w\}$

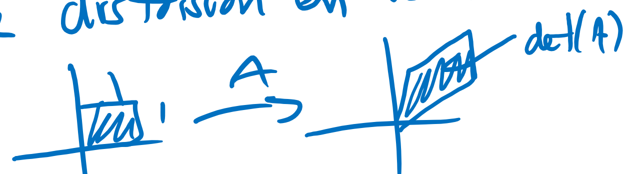
- $\underbrace{dim(Ker)}_{\text{nulidad}} + \underbrace{dim(Im)}_{\text{rango}} = n$

Dado un sistema de eq. lin. $Ax = b$
- tiene solución $\iff b \in Im(A) \iff \exists X: b = \sum x_i A^i$
- si hay una solución dice cuántas hay

↳ si $Ker(A) = \{0\}$ solo hay una



4 Determinante: mide distorsión en volumen $\quad det(A) = \dfrac{Vol(A(I))}{Vol(I)}$



- Para matrices cuadradas
- ($det(A) = \prod \lambda_i$ producto de los eigenvalores de A)

5
- A invertible $\iff det(A) \neq 0$

6) Ortogonales: transpuesta e inversa son la misma

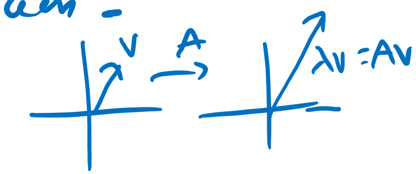
$$A^TA = I \qquad \langle A^i, A^j \rangle = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i \neq j \end{cases}$$

- Como transformaciones son rotaciones y reflexiones
- $\det(A) = \pm 1$

7) Eigenvectores y diagonalización:

- $Av = \lambda v$ ← eigen vector

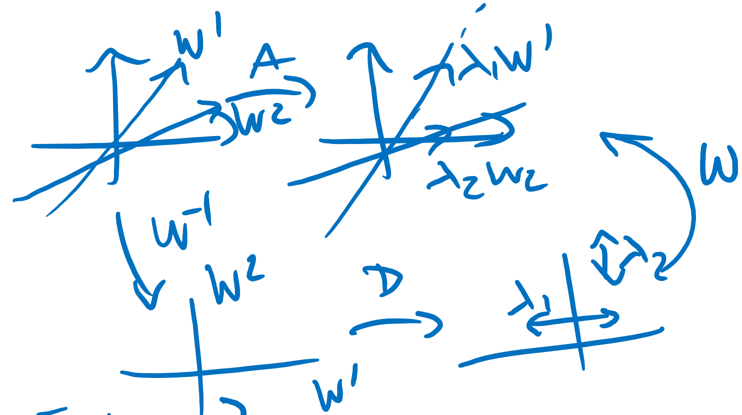
  ↑ cte eigenvalor



Diagonalización

- A es diagonalizable si existe W ortogonal y D diagonal tal que

$$A = W D W^{-1}$$
$$(AW = WD)$$
$$W = [w^1 | \dots | w^n]$$
$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

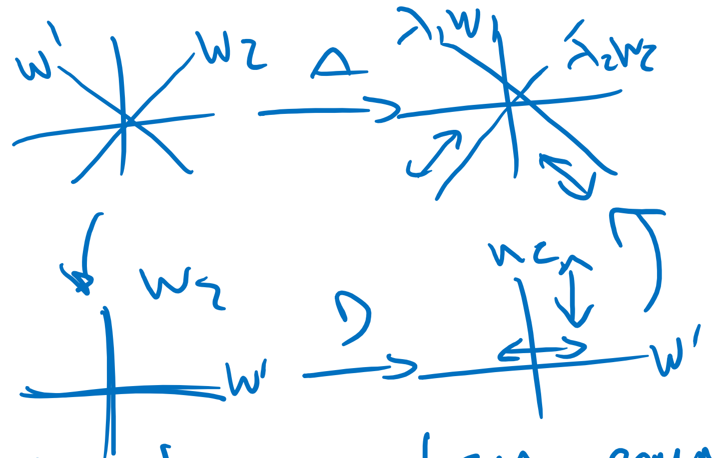


multiplicar por W^{-1} es cambiar a coordenadas $[w^1 | \dots | w^n]$

- Diagonalizar equivale a encontrar base de eigenvectores

- Caso A simétrica
  - siempre es diagonalizable
  - la base de eigenvectores es ortonormal

$A = W D W^T =$ rotar/reflejar → reescalar → desrotar/desreflejar



(caso)

Lección: "Todas las matrices equivalen a una matriz diagonal salvo rotaciones"

8, 9,) Repaso pero nos ayudó a entender normas matriciales entre otras cosas

10) Descenso gradiente:
- ∇f apu~~nta al~~ máximo ascenso
- $-\nabla f$ al máximo descenso

- El método favorito de optimización en big data:
- Moviéndos $x_k = x_{k-1} - \alpha \nabla f(x_{k-1})$ llegamos a un mínimo

Final) SVD :

$A = U\Sigma V^T$

- El resultado más padre de álgebra
- Dice cómo aproximar matrices reduciendo complejidad $A = \sum \sigma_i u_i v_i^T$
- Matrices de cualquier tamaño
- Se construye definiendo el efecto de A en una esfera
- Describe A como rotación → reescalamiento → otra rotación
- $\sigma_1 = \|A\| = \max \frac{\|Ax\|}{\|x\|}$
- $\sigma_1, \sigma_2, \ldots$ son los tamaños de ejes del elipsoide imagen de la esfera por A
- siempre existe
- permite solucionar sistemas de ecuaciones establemente
- define una pseudoinversa
- etc
- ⋮ U son los eigenvec de AA^T y V de A^TA