Álgebra Lineal: Optimización + Eigenvectores

Clase 7 Curso Propedéutico
2017/06/20

Optimización…

- Por escribir….

$\min/\max f(x)$

$\text{sa} \quad g(x)=0$

La solución debe cumplir

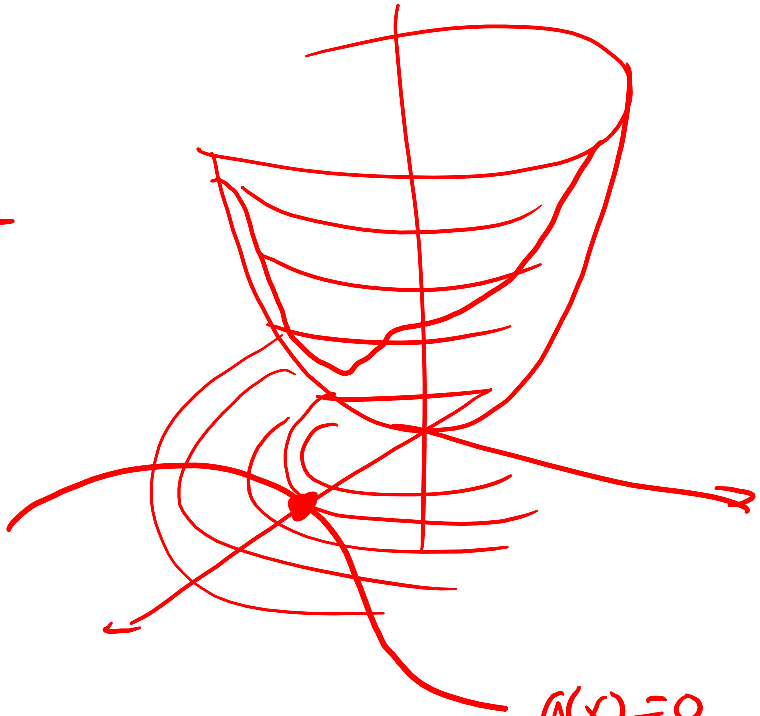
$\nabla f(x) = \lambda \nabla g(x)$

$g(x)=0$

o bien

$\mathcal{L}(x,\lambda) = f(x) - \lambda g(x)$

$\nabla \mathcal{L}(x,\lambda)=0$



Derivación de funciones que usan matrices

2 casos especiales de derivación con matrices y vectores

1) $f(x) := b^T x = x^T b$ $\quad x, b \in \mathbb{R}^n \quad f: \mathbb{R}^n \to \mathbb{R}$

$$Df(x) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b \qquad f(x) = x_1 b_1 + \dots + x_n b_n$$

2) $f(x) = x^T A x$ con A simétrica $\quad A \in \mathbb{R}_{n\times n}$

$x \in \mathbb{R}_n$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} x_j \qquad A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

$$= a_{11}x_1^2 + \dots + a_{nn}x_n^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \dots$$

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i \neq j} a_{ij}x_i x_j$$

simetría

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_i x_j$$

$$= \begin{aligned} & a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 + \dots + a_{2n}x_2x_n \\ & \vdots \\ & + a_{n1}x_nx_1 + \dots + a_{nn}x_n^2 \end{aligned}$$

$$\frac{\partial}{\partial x_k}(x^TAx) = a_{k1}x_1 + \dots + a_{k,k-1}x_{k-1} + 2a_{kk}x_k + \dots + a_{kn}x_n$$
$$+ a_{1k}x_k + \dots + a_{k-1,k}x \quad + a_{k+1,k}x_{k+1} + \dots + a_{nk}x_n$$

$$= 2\sum_{i=1}^{n} a_{ki}x_i = 2x^TA_k$$

entonces $$\nabla(x^TAx) = 2\begin{bmatrix} x^TA_1 \\ \vdots \\ x^TA_n \end{bmatrix} = 2Ax$$

Aplicación: Puntos críticos de una función cuadrática $f(x) = x^TAx + b^Tx + c$

$$\nabla f(x) = 0 \Rightarrow 2Ax + b = 0 \Rightarrow Ax = \frac{-b}{2} \Rightarrow x = -\frac{1}{2}A^{-1}b$$

Normas matriciales y optimización

$$\|A\| = \max_{\|u\|=1} \|Au\|$$

Problema equivalente $\quad \max \ \|Au\|^2 \quad$ s.a. $\ \|u\|^2 - 1 = 0$

$u^T u = u^T I_n u$

identidad

$\|Au^*\|^2 = u^{*T}(A^T A u^*) = \lambda u^{*T} u^* = \underline{\lambda}$

$\Rightarrow \|A\| = \sqrt{\lambda_{max}}$

$= u^T (A^T A) u$

Solución

$$L(x,\lambda) = u^T (A^T A) u - \lambda (u^T u - 1)$$

$$\nabla L(x,\lambda) = 0 \iff$$

$$\boxed{\cancel{2} A^T A u = \cancel{2} \lambda u}$$

$(A^T A) u^* = \lambda u^*$ ← solucionar sistema

$A^T A$ es simétrica

$(A^T A)^T = (A^T)(A^T)^T$

$= A^T A$

pues $(AB)^T = B^T A^T$

Conclusión: u^* es eigenvector (?) de $A^T A$

Dos temas nuevos: formas cuadráticas y optimización

Eigenvectores

Dada $A \in R_{n\times n}$ ← ojo: cuadrada

Eigenvector y eigenvector asociado: $Av = \lambda v$

Diagonalización

Si R^n tiene una base de eigenvectores de A entonces estos elementos formar un **Sistema de coordenadas.**

es decir, cualquier elemento $v \in \mathbb{R}^n$ se puede escribir de forma única $\vec{v} = \sum \alpha_i \vec{w}_i$ donde $\vec{w}_1, \dots, \vec{w}_n$ es la base de eigenvectores

solo por ser base

~~La matriz A es **diagonal** en es ese Sistema de coordenadas.~~

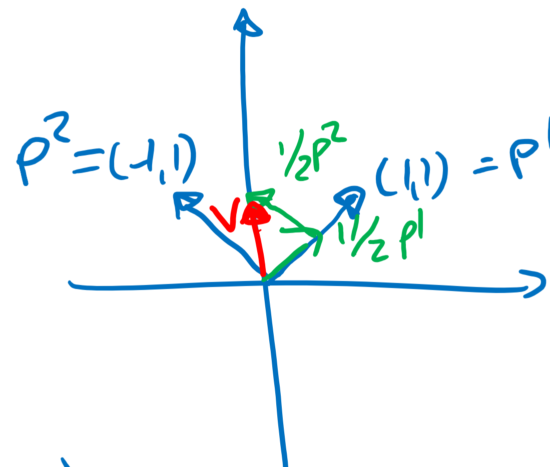
Además la base debe cumplir

$AW = WD$ con $W = [w_1 | \dots | w_n]$ $D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

$\Rightarrow A = WDW^{-1}$

- **Observación**: De manera general, si P es una matriz invertible, sus columnas dan una base de R^n y $P^{-1}v$ nos dice como escribir a v en esas nuevas coordenadas.

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = [p^1 \mid p^2]$$

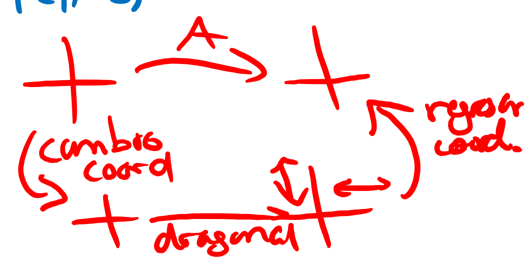


$$v = (0,1) = 0\vec{e}_1 + 1\vec{e}_2$$
$$= 0(1,0) + 1(0,1)$$

<u>Inversa:</u> $P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ $\quad PP^{-1} = P^{-1}P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$P^{-1}v = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \qquad (1/2 \;\; 1/2)_{[p^1,p^2]} = (0,1)_{\{e_1,e_2\}}$$

Conclusión: Si $A = WDW^{-1}$ con D diagonal A es



Eigenvectores como problema de optimización!

~~Ejercicio~~

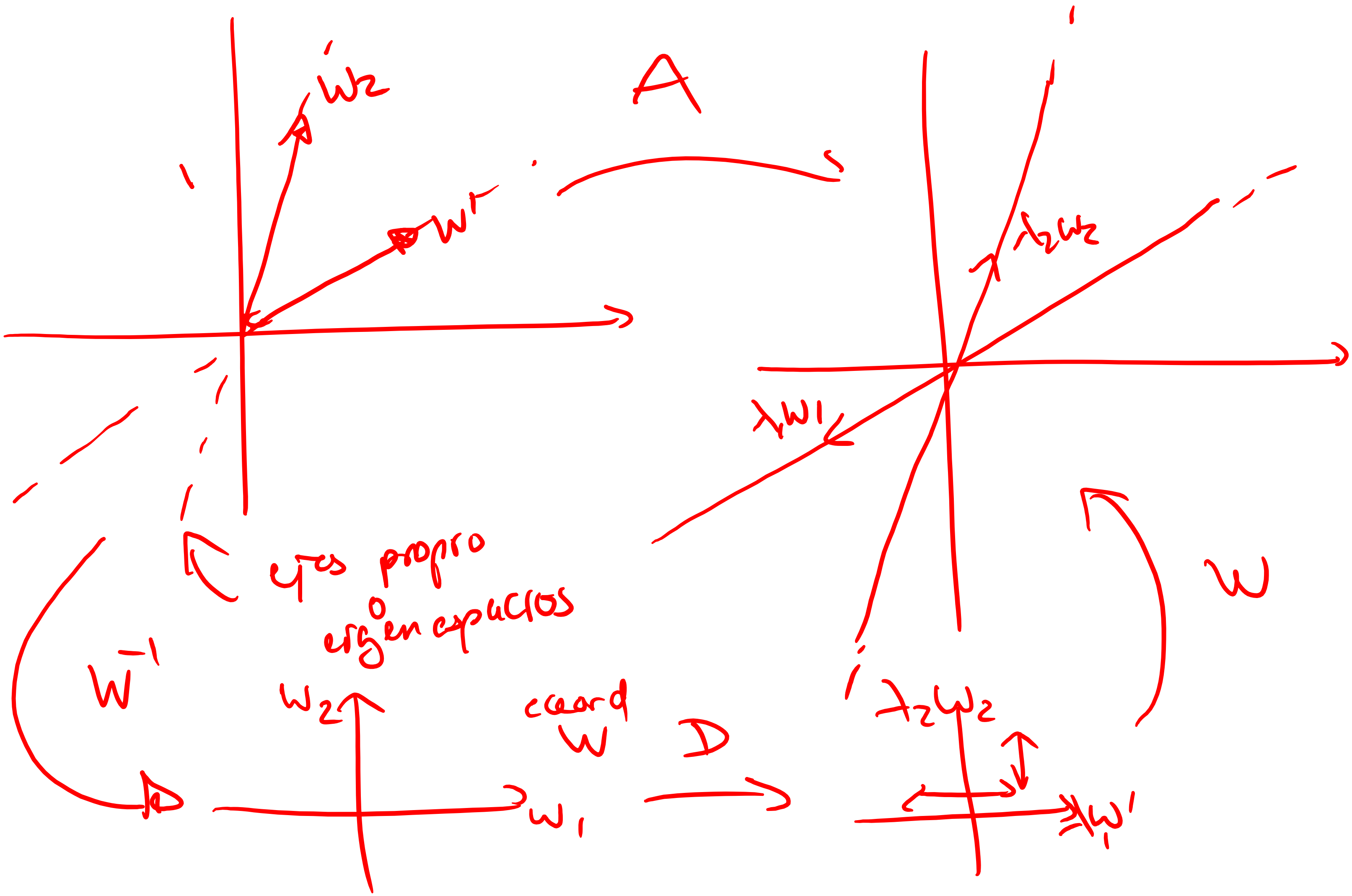
Un eigenvector es la solución a un problema de optimización

$$\max_{\|w\|=1} w^T A w$$

Si A es simétrica.

A matriz diagonalizable

$A = W D W^{-1}$



Regresando al determinante

Recordar ① $\det(AB) = \det(A)\det(B)$ ② $\det(W^{-1}) = \dfrac{1}{\det(W)}$

Si $A = WDW^{-1}$ $\quad \det(A) = \det(W)\det(D)\det(W^{-1})$

$$= \frac{\cancel{\det(W)}}{\cancel{\det(W)}}\det(D)$$

$$= \det(D)$$

$$= \det\left(\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}\right)$$

$$= \prod \lambda_i$$

El determinante de una matriz es el producto de sus eigenvalores

Obs práctica: Todas las matrices que van a ver en su vida son diagonalizables. De hecho, si permiten números complejos, todas las que existen lo son.

Algoritmos para cálculos de Eigenvectores

El método de la potencia devuelve el eigenvalores asociado al eigenvalor más grande (en valor absoluto).

Método de la Potencia

$v_0 =$ (casi) cualquier semilla

For $k = 1$ **...,** convergencia

  Definir $\lambda_{k+1} = v_k^\top A v_k$

  Definir $v_{k+1} = \frac{Av_k}{||Av_k||}$

v_k converge al eigenvector más grande y λ_k el eigenvalor más grande

$V_k \propto A^K V_0$

¿Por qué sirve el método de la potencia?

Si $w_1, \ldots, w_n$ es una ***base*** de eigenvectores de $A = PDP^\top$ entonces v_k se puede escribir como

$$v_k = \alpha_1 w_1 + \cdots + \alpha_n w_n$$

donde $\alpha = P^{-1} v_k$ es las coordenadas con la base de eigenvectores. Entonces:

Intuición

$$v_k = \frac{A v_{k-1}}{\|A v_{k-1}\|} \propto \cdots A^k v_0 = A^k \alpha_1 w_1 + \dots + A^k \alpha_n w_n$$

$$= \alpha_1 \lambda_1^k w_1 + \dots + \alpha_n \lambda_n^k w_n$$

$$\frac{A^k v_0}{\|A^k v_0\|} = \frac{\cancel{\alpha_1 \lambda_1^k} w_1 + \cancel{\dots + \alpha_n \lambda_n^k w_n}^{\to 0}}{\|\cancel{\alpha_1 \lambda_1^k} w_1 + \cancel{\dots + \alpha_n \lambda_n^k w_n}\|}$$

Algoritmos para cálculos de Eigenvectores

Algoritmo QR (el más usado)

Utiliza un método conocido como **descomposición QR** (lo vamos a ver más adelante) que escribe

$A = QR$ con Q ortogonal y R triangular superior.

ALGORITMO QR
$A_0 = A$
For $k = 1$ **...,** convergencia
  Factorizar $A_k = Q_k R_k$
  Definir $A_{k+1} = R_k Q_k$
La matriz Q_k tiene los eigenvectores y R_k los eigenvalores en la diagonal.

Cuando se demuestra la convergencia de este algoritmo (algo que no haremos en este curso…) se ve que detrás hay un mecanismo muy similar al método de la potencia. Normalmente este método se usa en conjunto con otros métodos que aceleran su convergencia como las reflexiones de Householder

Matrices simétricas y formas cuadráticas

- Una de las "aplicaciones" de los eigenvectores es que permite estudiar de manera muy sencilla estudiar las formas cuadráticas.

Una forma cuadrática es una **función** $f\colon R^n \to R$ de la forma

$$f(x) = f(x_1, \dots, x_n) = x^\top A x + b^\top x + c$$

Con A una matriz simétrica

- Son funciones "polinomiales" de orden dos.
- Ej: $A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathrm{c} = -1$

- Antes de estudiar la teoría de formas cuadráticas vamos a ver algunos problemas bonitos que surgen con formas cuadráticas (algunos los resolveremos aquí).

Ejemplos:

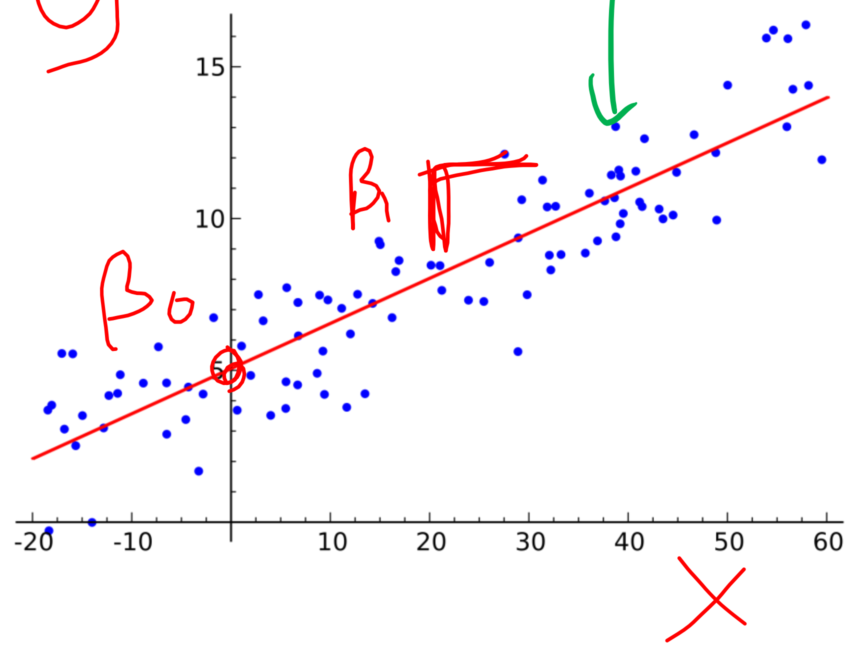
- *Regresión Lineal:* Queremos explicar una variable y con una variable x

β α

$$\min_{\beta_0,\beta_1} \sum_i |y_i - x_i\beta_1 - \beta_0|^2$$

$\beta_0 + \beta_1 X$

y

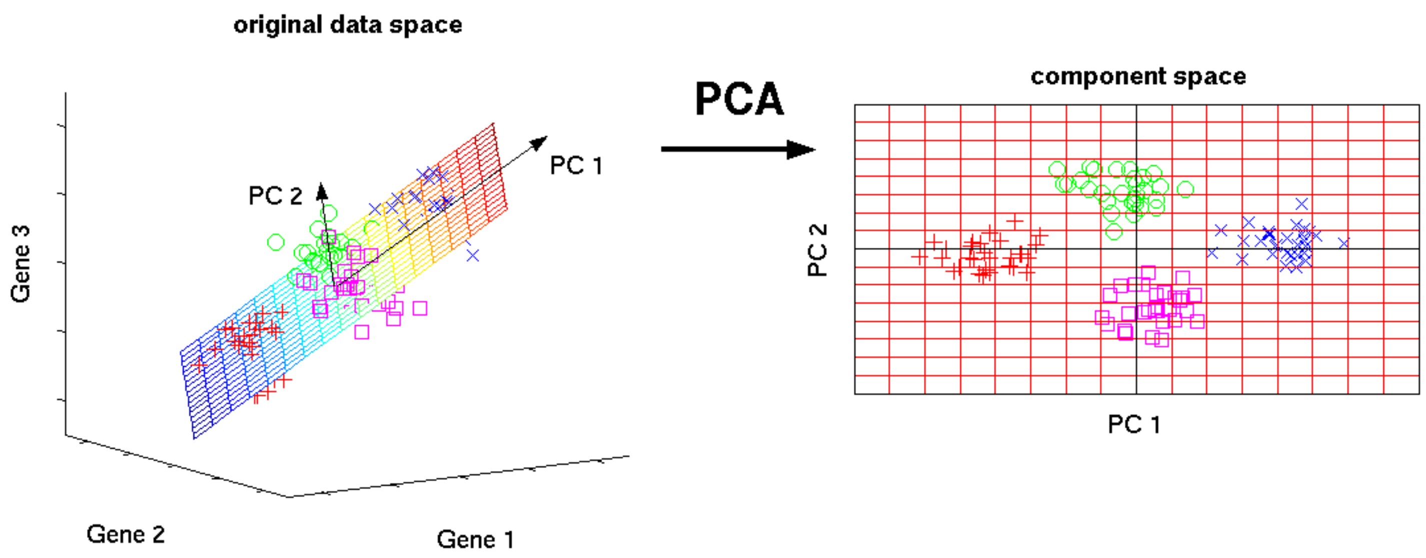


- *Reducción de dimensionalidad (PCA) y clasificación lineal*

Encontrar una combinación lineal de vectores $X^1, \ldots, X^p$ que "mejor" represente a las variables.

$$\max_{\|v\|=1} \|Xv\|^2$$

$$Xv = v_1 X^1 + \ldots + v_p X^p$$

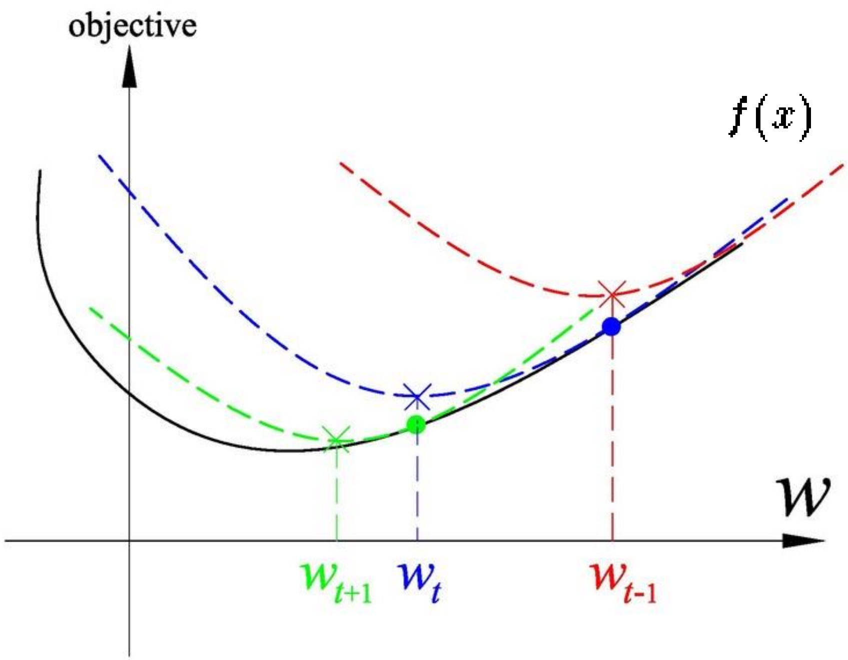


Optimización en General

- Cualquier función puede aproximarse por una función cuadrática. Este es el corazón del famosísimo método de Newton-Raphson que vamos a ver.

$$\min_{x} f(x)$$

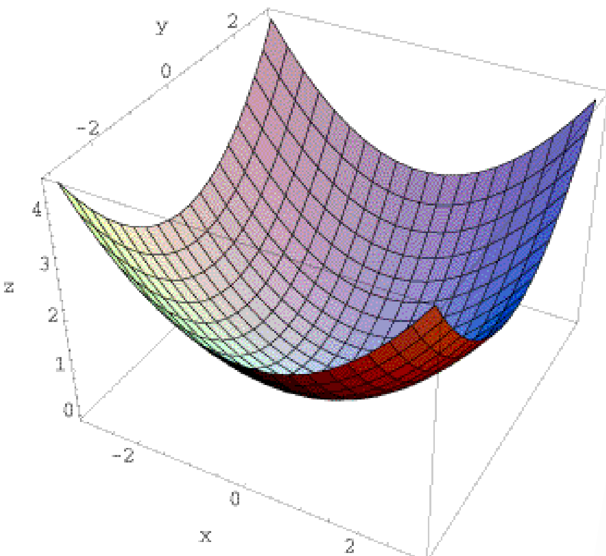
Taylor: $f(x+h) \approx \mathrm{h}^{\top}\nabla^{2}f(x)h + h^{\top}\nabla f(x) + f(\mathrm{x})$

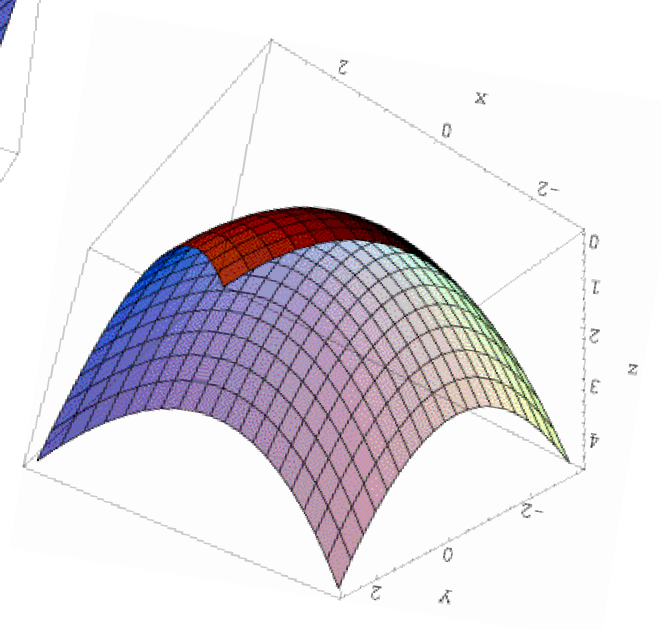


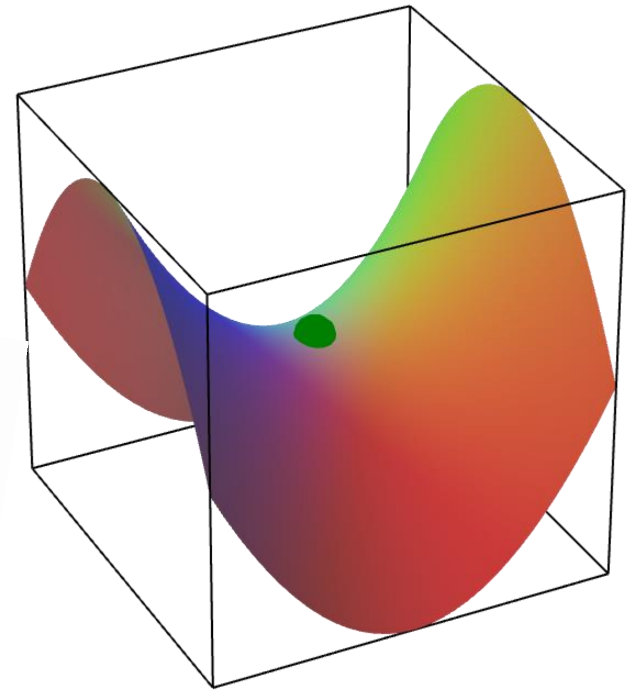
$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 \\ &\quad + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \frac{f''''(x_0)}{4!}(x-x_0)^4 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n. \end{aligned}$$

Formas cuadráticas

- Los eigenvalores de la matriz A determinan la forma cuadrática.

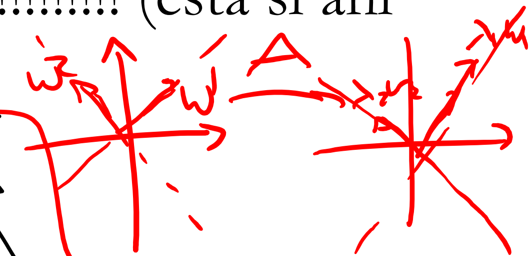






$A = WDW^{-1}$ $\quad W^{-1} = W^T \Rightarrow A = WDW^T$

- Las matrices simétricas cumplen lo siguiente que verlo con detalle escapa el objetivo de este curso
  - Siempre son diagonalizables con eigenvalores números reales (Teorema Espectral)
  - Los eigenvectores SON ortogonales!!!!!!!!!!!!!!!!!!!!! (esta si ahí va la prueba)….



$v_1 \quad v_2$ eigenvectors

$\lambda_1 \neq \lambda_2 \Rightarrow v_1^T v_2 = 0$

$$v_1^T (A v_2) = \lambda_2 v_1^T v_2$$

$$v_1^T A v_2 = v_1^T A^T v_2$$
$$= (A v_1)^T v_2 = \lambda_1 v_1^T \cdot v_2$$

SVD

"Diferencias con eigenvalores
- Eigenvalores es solo para matrices cuadradas
- En eigenvalores $U = V$; U es ortogonal solo si
- "A" es simétrica
- ¡¡SVD siempre existe en números reales y eigenvalores siempre en complejos!!

- Recordatorio Descomposición SVD

Cualquier $A = U \Sigma V^T$ $A \in R^{m \times n}$

con U ortogonal $U \in R^{m \times m}$

V ortogonal $V \in R^{n \times n}$

Σ cuasidiagonal $\Sigma \in R^{m \times n}$

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_m & \end{array}\right] \quad \text{ó} \quad \Sigma = \left[\begin{array}{ccc} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ \hline & 0 & \end{array}\right]$$

la diagonal de Σ se llaman valores singulares

$\sigma_1 \geq \sigma_2 \geq \dots \geq \dots \sigma_{\min(m,n)} \geq 0$

y las columnas de U y V son vectores singulares

Relación entre SVD y Eigenvectores

$A = U \Sigma V^T$

$A^T A = (U \Sigma V^T)^T (U \Sigma V^T) = V \Sigma^T \underbrace{U^T U}_{= I_n \text{ pues } U \text{ ortogonal}} \Sigma V^T$

$= V \Sigma^T \Sigma V^T$

pero

$$\Sigma^T \Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \end{bmatrix}$$

es diagonal

$\Rightarrow$ $A^T A = V D V^T$ con D diagonal (Diagonalización)

$\therefore$ Los vectores singulares V son los eigenvectores de $A^T A$ con eigenvalores los cuadrados de los valores singulares

Similarmente

$AA^T = U \Sigma V^T V \Sigma^T U^T = U D U^T$

$\therefore$ U son eigenvectores de AA^T

Resumen

Descomposición SVD

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

$\Rightarrow$ U son eigenvectores de AA^T

V son " " A^TA

y $\sigma_1^2, \sigma_2^2, \ldots$ son eigenvalores asociados a U y V.

$A^\top A$ y $AA^\top$ siempre tienen eigenvalores no negativos.

Recordar que $A^\top A$ y $AA^\top$ definen formas cuadráticas pues son matrices simétricas.

El valor del eigenvalores son justamente los valor de la máxima distorsión:

$A^\top Av = \lambda v$ implica $||Av||^2/||v||^2 = \lambda$
Similar para $AA^\top u = \lambda u$.

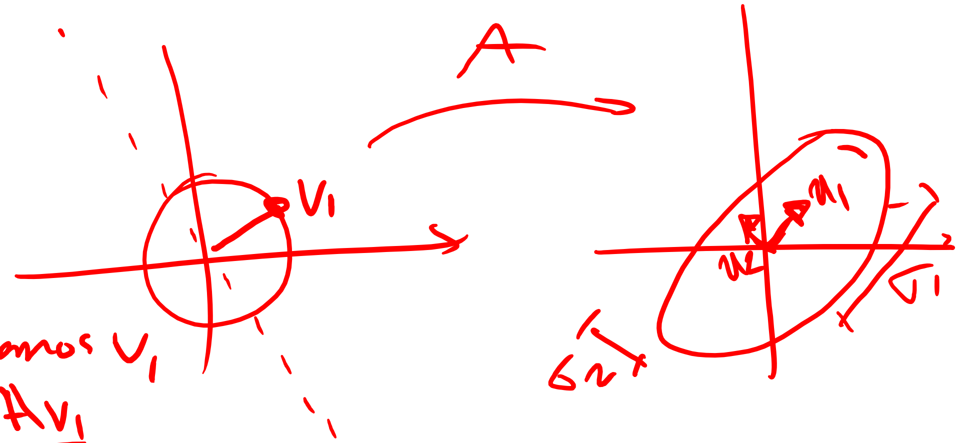
- Las matrices de la descomposición SVD son precisamente los “ejes de máxima distorsión” de la matriz.
- Eso explica porque resumen el efecto de la transformación lineal

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

Derivación constructiva sin eigenvectores

(1) Dado A, resolvemos

$$\sigma_1 = \|A\| = \max_{\text{s.a } \|v\|=1} \|Av\|$$



Al vector solución lo llamamos v_1

y denotamos $u_1 = \dfrac{Av_1}{\|Av_1\|}$

La relación entre u_1 y v_1 es $Av_1 = \sigma_1 u_1$

Obs: De la solución que resolvimos antes, sabemos v_1 es eigenvector de de A^TA.

(2) Hay que resolver un nuevo problema de optimización, calcular

$$\sigma_2 = \max_{\text{s.a}} \|Av\| \quad \|v\|=1 \ \& \ v^Tv_1 = 0$$

llamamos v_2 al vector solución y

$u_2 = \dfrac{Av_2}{\|Av_2\|} \Rightarrow Av_2 = \sigma_2 u_2$

Por construcción

$v_1^T v_2 = 0$ $\qquad Av_1 = \sigma_1 u_1$

$\|v_1\| = \|v_2\| = 1$ $\qquad Av_2 = \sigma_2 u_2$

v_2 eigenvector de A^TA

y

$$u_1^T u_2 = \frac{1}{\sigma_1 \sigma_2} v_1^T A^T A v_2 = \frac{\lambda_2}{\sigma_1\sigma_2} \cancel{v_1^T v_2}^{0} = 0$$

entonces $u_1^T u_2 = 0$

③ Continuando así, encontramos $v_1, \ldots, v_{\min(m,n)}$

y $u_1, \ldots, u_{\min(m,n)}$ y $\sigma_1, \ldots, \sigma_{\min(m,n)}$

tales que $\quad Av_i = \sigma_i u_i \quad$ con $\quad u_i^T u_i = 1 \quad u_i^T u_j = 0$

$v_i^T v_i = 1 \quad v_i^T v_j = 0$

La parte tricky es completar U o V con elementos del kernel de A o A^T según $m > n$ o $m < n$, pero la idea ya está clara.

con eso tenemos $\quad V = [v^1 | v^2 | \ldots] \quad U = [u^1 | \ldots] \quad \Sigma = [\sigma_1 \ldots \sigma \,|\, 0]$

con $V^TV = I \quad U^TU = I$ y $AV = U\Sigma$ ✓