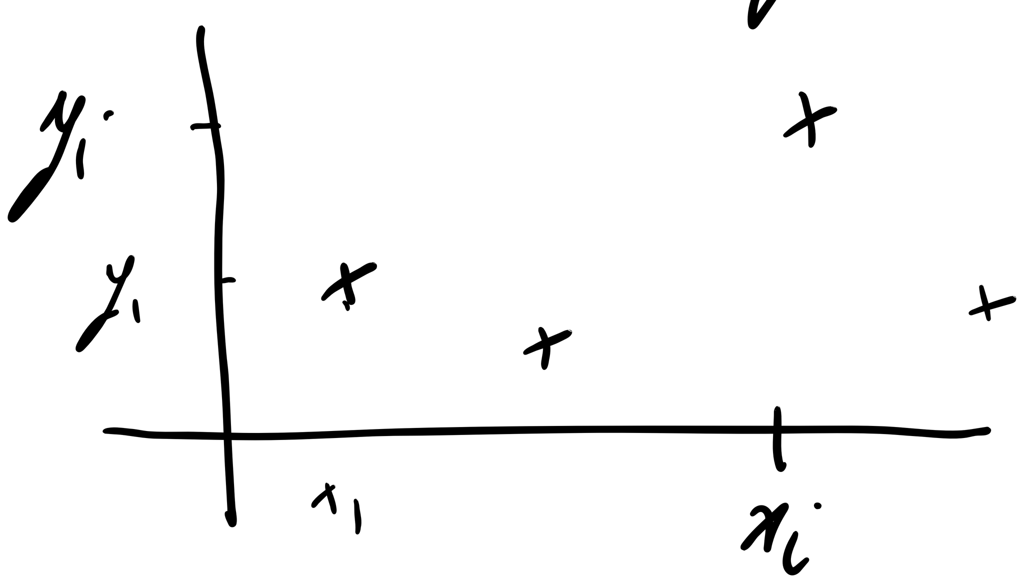
# Mínimos Cuadrados y Pseudo-Inversa (Moore-Penrose)

ALN-2022
Clase 8
20/9/2022

Comencemos con una digresión / motivación sobre interpolación de polinomios.

## Interpolación de Polinomios

Supongamos que tenemos ~~datos~~ m-puntos (≠) $x_1, \dots, x_m \in \mathbb{C}$ y datos $y_1, \dots, y_n \in \mathbb{C}$




Existe un único polinomio de grado $m-1$

$$p(x) = c_0 + c_1 x + \dots + c_{m-1} x^{m-1}$$

tal que interpole los datos: $p(x_i) = y_i \quad (i = 1, \dots, m.)$

Esto resulta de q lo siguiente: los coeficientes de $p$ son solución de

$$\underbrace{\begin{pmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ 1 & x_2 & \cdots & x_2^{m-1} \\ \vdots & & & \\ 1 & x_m & \cdots & x_m^{m-1} \end{pmatrix}}_{V} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

y la matriz de Vandermonde es invertible (asumiendo $x_i \neq x_j$)

Ejercicio: i) probar que $V$ es invertible viendo que si $p(x) = c_0 + \dots + c_{m-1} x^{m-1}$ está en el núcleo entonces $p \equiv 0$.

ii) ~~También~~ Probar que $\det V = \prod_{i>j} (x_i - x_j)$. ~~Idea~~:

Idea: tomar $f(x) = \det \begin{pmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ 1 & x_{m-1} & \cdots & x_{m-1}^{m-1} \\ 1 & x & \cdots & x^{m-1} \end{pmatrix} \longrightarrow$ cambiamos última fila de $V$ ~~por~~ $x$ por $x_m$.

Veamos en la computadora qué sucede con los polinomios que interpolan.

De los ejemplos surge que el polinomio interpolador tiene algún comportamiento no deseado (se discutirá más adelante, pero si no pueden buscar por fenómeno de Runge).

¿Y si aproximamos por polinomios de grados menores?

Tomemos $n \leq m$ y ~~buscamos~~ consideremos ~~p~~ polinomios

$$p(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$$

y donde buscamos $p$ tal que $\circledast = \sum_{i=1}^{m} |p(x_i) - y_i|^2$ sea lo más chico posible.

Observa que lo esperable es que $\circledast$ no sea $0$ dado que tenemos $n$-incógnitas (a saber $c_0, \ldots, c_{n-1}$) y $m$-ecuaciones $p(x_i) = y_i$, por lo que nuestro interés se centra en minimizar la diferencia. (en este caso la diferencia cuadrática.)

Veamos algunos casos particulares de esto.

$\boxed{n=0}$ Si $p$ es de grado cero, estamos buscando la constante que mejor se ajusta a los datos: $c_0 / \sum_{i=1}^{m} |c_0 - y_i|^2$ es mínimo.

La solución, muy sencilla es el promedio o baricentro $c_0 = \frac{1}{m} \sum_{i=1}^{m} y_i$.

$\boxed{n=1}$ Regresión Lineal: es la ecuación lineal $c_0 + c_1 x$ que "mejor" se adapta a los datos, i.e. que minimice:

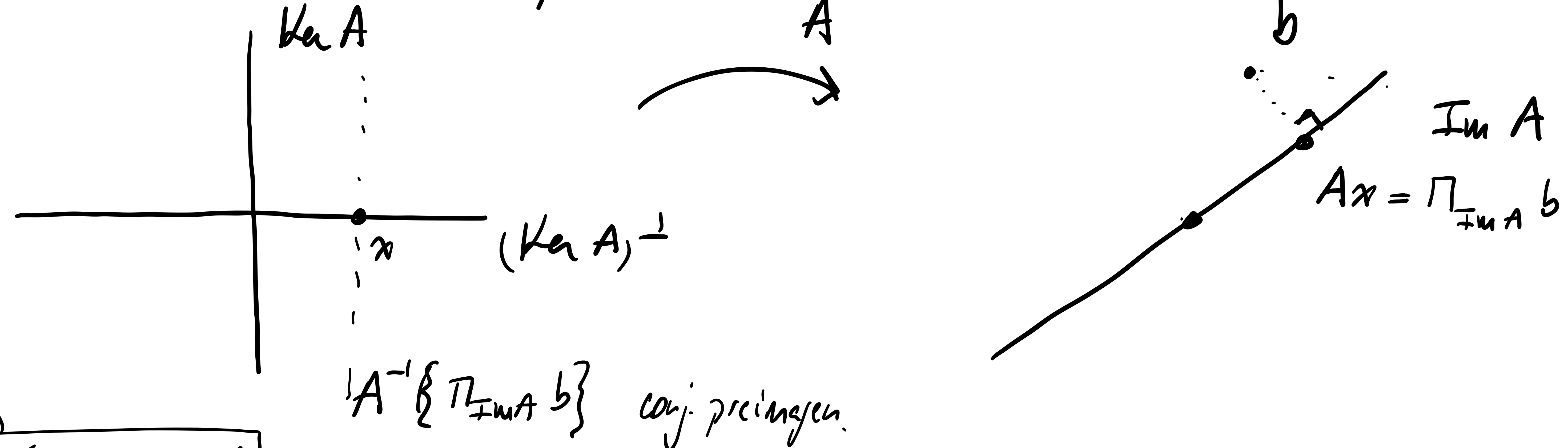
$$\sum_{i=1}^{m} |c_0 + c_1 x_i - y_i|^2 .$$

Todos los problemas anteriores son casos particulares de la siguiente <u>formulación</u> (llamada ~~...~~ <u>problema de mínimos cuadrados</u>).

$$\boxed{\text{Dado } A\in\mathbb{C}^{m\times n},\ m\geq n,\ b\in\mathbb{C}^{m},\ \text{encontrar } x\in\mathbb{C}^{n} \text{ que minimice } \|Ax-b\|_2 \quad \text{(PMC)}}$$

(~~...~~ Vale la pena mencionar que en aplicaciones "modernas" tipo compressed sensing hay interés en resolver el problema anterior para otras normas)

Veamos como resolver este problema.




(Pitágoras)

Geométricamente vemos que $x$ es solución si $r=b-Ax \perp \mathrm{Im}\,A$ (i.e. $Ax$ es el punto más cercano en $\mathrm{Im}\,A$ a $\underline{\underline{b}}$.)

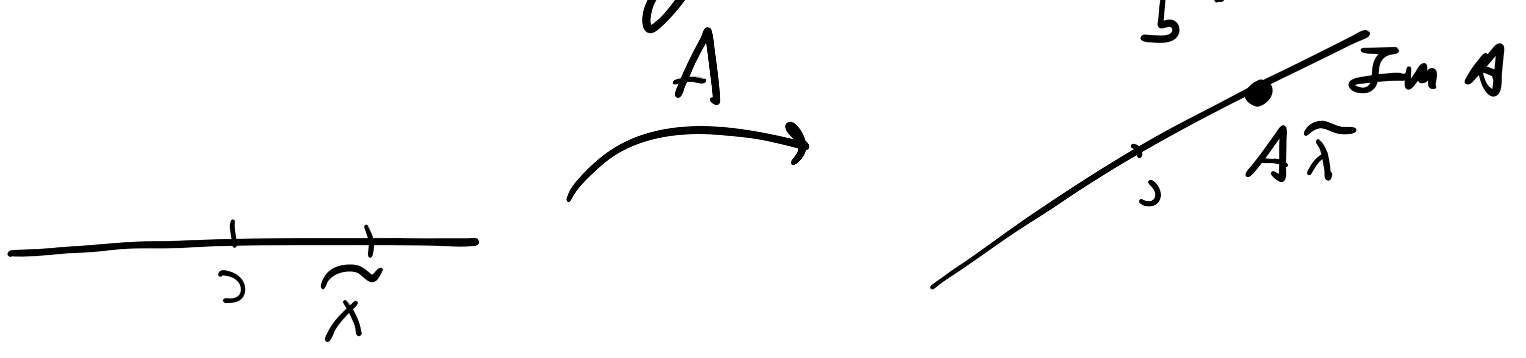
Por lo tanto $A^*(b-Ax)=0$ i.e. $\qquad A^*Ax=A^*b \quad (I)$

o equivalentemente $\Pi_{\mathrm{Im}A}\, b = Ax$

siendo $\Pi_{\mathrm{Im}A}$ la proy. ortogonal sobre $\mathrm{Im}\,A$.

Del dibujo además tenemos que si $\mathrm{Ker}\,A$ es no trivial hay todo un subespacio afín, a saber, $\tilde{x}+\mathrm{Ker}\,A$, solución de (PMC), siendo $\tilde{x}$ cualquier solución de $A\tilde{x}=\Pi_{\mathrm{Im}A}\, b$.

Si $A$ tiene <u>rango máximo</u> entonces $\mathrm{Ker}\,A=\{0\}$ y por lo tanto $\exists!\ \tilde{x}$ solución, $\boxed{\tilde{x}=(A^*A)^{-1}A^*\, b}$



# Inversa Generalizada (Moore-Penrose)

La transformación resultante, cuando $A$ tiene rango máximo, se denota

$A^\dagger : \mathbb{C}^m \to \mathbb{C}^n$. $\quad A^\dagger = (A^*A)^{-1}A^*$ y se llama Inversa Moore-Penrose.

Más en general, si $L: V \to W$ es una transformación lineal entre espacios con productos internos podemos considerar la descomposición de $L$ en los siguientes subespacios:

$$L: \underbrace{\operatorname{Ker} L \oplus (\operatorname{Ker} L)^\perp}_{V} \longrightarrow \underbrace{\operatorname{Im} L \oplus (\operatorname{Im} L)^\perp}_{W}$$

donde si asumimos

$\dim V = n$, $\dim W = m$, $\operatorname{rango}(A) = r$ entonces $\dim \operatorname{Ker} L = n-r$

$\dim(\operatorname{Ker} L)^\perp = r$ $\quad \dim \operatorname{Im} L = r$, $\dim(\operatorname{Im} L)^\perp = m-r$

Luego la restricción de $L$ a $(\operatorname{Ker} L)^\perp$ es un isomorfismo entre $(\operatorname{Ker} L)^\perp$ y $\operatorname{Im} L$:

$$L|_{(\operatorname{Ker} L)^\perp} : (\operatorname{Ker} L)^\perp \longrightarrow \operatorname{Im} L .$$

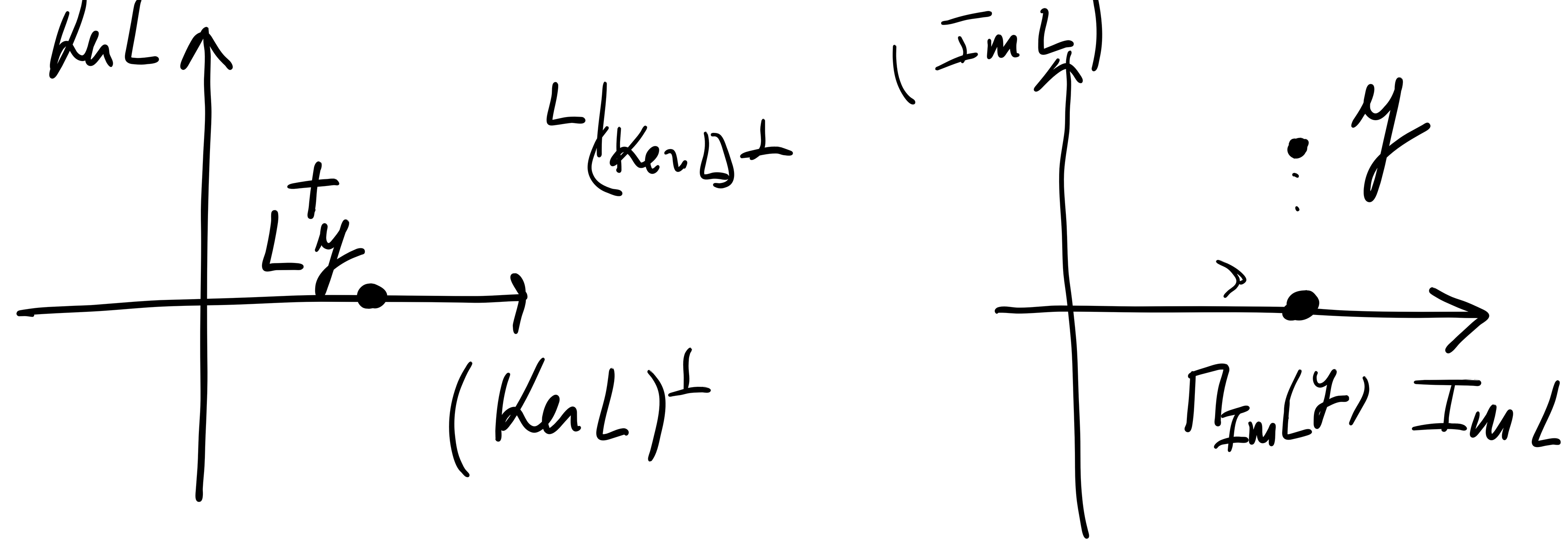
(esto resulta porque ambos espacios tienen igual dimensión, y $L|_{\operatorname{Ker}^\perp}$ inyectiva).

La inversa Moore-Penrose se define como el mapa lineal $L^\dagger : W \to V$ definido por

$$L^\dagger = \iota_{(\operatorname{Ker} L)^\perp} \cdot \left( L|_{(\operatorname{Ker} L)^\perp} \right)^{-1} \cdot \Pi_{\operatorname{Im} L}$$

proy. ortogonal

mapa inclusión de $(\operatorname{Ker} L)^\perp$ en $V$

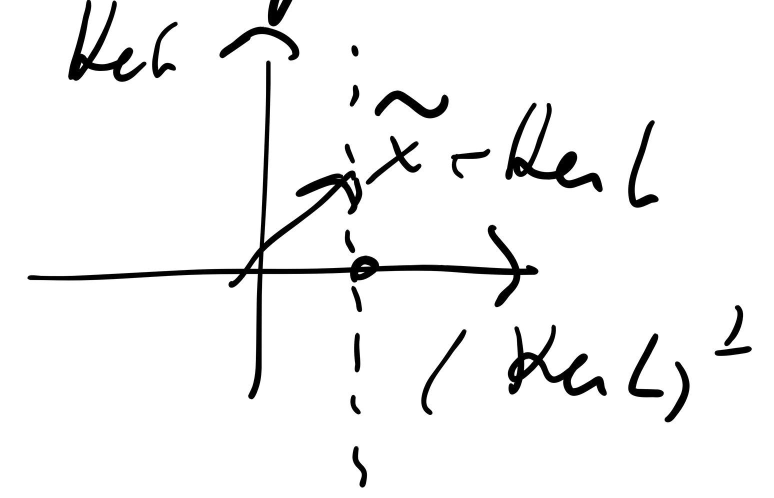



O sea $L^\dagger y$ da como resultado $x \in (\operatorname{Ker} L)^\perp$ tal que $Lx = \Pi_{\operatorname{Im} L}(y)$.

Una forma equivalente de definirlo es como la solución del problema.

$$A^{\dagger}y := \underset{x\in V:\ Ax=\Pi_{\mathrm{Im}\,L}(y)}{\arg\min} \|x\|$$

Esto resulta de que $L^{-1}(\Pi_{\mathrm{Im}\,L}(y))$ es un subespacio afín tipo $\tilde{x}+\ker L$.




Propiedades: $L^{\dagger}$ satisface:

1. $L^{\dagger}\circ L = \Pi_{(\ker L)^{\perp}}$ (proyección ortogonal en $(\ker L)^{\perp}$).
2. $L\circ L^{\dagger} = \Pi_{\mathrm{Im}\,L}$

(Dem: quedan como ejercicio).

Corolario: 1. Si $L$ inyectiva $\Longrightarrow$ $L^{\dagger}\circ L = Id_V$
2. Si $L$ sobreyectiva $\Longrightarrow$ $L\circ L^{\dagger} = Id_W$
3. Si $L$ isomorfismo $\Longrightarrow$ $L^{\dagger} = L^{-1}$.

Proposición: Sea $A\in\mathbb{C}^{m\times n}$, $(m\geq n)$ de rango máximo.

$$\Longrightarrow \quad A^{\dagger} = (A^{*}A)^{-1}A^{*}$$

Dem: En página 41 (abajo) concluimos que $\exists!\ \tilde{x}$ solución del problema de mínimos cuadrados $\tilde{x} = \underset{x\in\mathbb{C}^n}{\arg\min}\|Ax-y\|_2$ y que tiene la expresión $\tilde{x} = (A^{*}A)^{-1}A^{*}y$.

Además $A\tilde{x} = \Pi_{\mathrm{Im}\,A}\,y$ y $\ker A = \{0\} \Longrightarrow \tilde{x} = A^{\dagger}y$. $\square$

Ejercicio: Si $A\in\mathbb{C}^{m\times n}$ $m\geq n$, de rango máximo, y la SVD $A = UDV^{*}$

$$\Longrightarrow \quad A^{\dagger} = VD^{-1}U^{*}.$$

# Algoritmos para Min. Cuadrados

Consideramos el caso de $A$ de rango máximo ($A \in \mathbb{C}^{m\times n}$, $m \geq n$, $rk(A) = n$).

La ecuación normal es $\boxed{A^*Ax = A^*b}$, donde lo que buscamos es la solución $x$ (que sabemos que es única porque $(A^*A)$ invertible).

## *Alg. vía Cholesky:

El método clásico de resolver el PMC es vía la factorización de Cholesky que veremos más adelante donde toda matriz $P$ definida positiva puede ser escrita $P = R^*R$ con $R$ triangular superior.

En nuestro caso tenemos que $A^*A = R^*R$ (de hecho este caso particular de $P = A^*A$ se puede usar descomp. QR)

Luego tenemos de la ecuación normal. $R^*Rx = A^*b$, y por tanto consideramos el siguiente algoritmo:

1. Realizar $A^*A$ y $A^*b$ $\qquad \sim m \cdot n^2$ (por simetría)
2. Cholesky fact. $A^*A = R^*R$ $\qquad \sim \frac{n^3}{3}$ (lo veremos)
3. Resolver sistema triangular $R^*y = A^*b$ (para $y$) $\qquad \sim$ orden menor
4. " " " $Rx = y$ (para $x$) $\qquad \sim$ " "

Complejidad total: $\sim m \cdot n^2 + \frac{1}{3} \cdot n^3$.

## * Vía fact. QR

Tenemos que $A = \hat{Q}\hat{R}$ luego de ecuación normal $\hat{R}^* \overbrace{\hat{Q}^*\hat{Q}}^{Id_m} \hat{R}x = \hat{R}^*\hat{Q}^*b$ de donde resulta ($\hat{Q}^*\hat{Q} = Id_m$) $\hat{R}x = \hat{Q}^*b$.

Luego el algoritmo sería:

1. Computar fact. QR: $A = \hat{Q} \cdot \hat{R}$ $\quad \sim 2mn^2 - \frac{2}{3}n^3$
2. $\hat{Q}^* b$ $\quad \sim$ costo menor
3. Resolver sist. triangular $\hat{R}x = \hat{Q}^* b$ (para $x$) $\quad$ " "

Complejidad: $2mn^2 - \frac{2}{3}n^3$.

* Vía SVD

Supongamos que por alguna razón conocemos la SVD de $A$:

$A = \hat{U} \hat{\Sigma} V^*$, luego tenemos que $\Pi_{\operatorname{Im} A} = \hat{U}\hat{U}^*$ (el proyector asociado) y luego tenemos:

$$Ax = \Pi_{\operatorname{Im} A} b, \text{ i.e. } \hat{U}\hat{\Sigma}V^* x = \hat{U}\hat{U}^* b$$

y por lo tanto $\underline{\underline{\hat{\Sigma} V^* x = \hat{U}^* b}}$ y obtenemos el algoritmo:

1. Computar (o conocer) SVD: $A = \hat{U}\hat{\Sigma}V^*$
2. " $\hat{U}^* b$
3. Resolver sist. diagonal $\hat{\Sigma} y = \hat{U}^* b$
4. Output (Computar) $x = Vy$

Observar que QR reduce al cálculo de un sistema triangular y la SVD a un sistema diagonal. Pero ¿cuál es la complejidad del SVD? Eso es un tema de investigación actual. De hecho no se conoce la complejidad "media" de ningún algoritmo para resolverlo.

# Mínimos Cuadrados como problema de Optimización

(Veamos el caso real)

~~Una forma~~ Otra forma de buscar una solución al problema de mínimos cuadrados es la siguiente.

Consideremos la función $f: \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \|Ax-b\|^2$.

Luego el (PMC) es encontrar $x \in \mathbb{R}^n$ que minimice $f$. Para eso observar que $f$ es $C^\infty$ positiva y no acotada superiormente (siempre que $A \neq 0$)
(En particular tiene un mínimo.)

Diferenciando obtenemos $Df(x)\mathring{x} = \lim_{t \to 0} \frac{1}{t}\left(f(x+t\mathring{x}) - f(x)\right) \overset{(*)}{=} 2\langle Ax-b, A\mathring{x}\rangle$

traspuesta

luego despejando $\mathring{x}$ obtenemos $Df(x)\mathring{x} = 2\langle A^*(Ax-b), \mathring{x}\rangle$

Por lo tanto,

$$\nabla f(x) = 2A^T(Ax-b).$$

y el mínimo se alcanza donde $\nabla f(x) = 0$ i.e. $A^T(Ax-b) = 0$

que es exactamente la ecuación normal

$$A^TAx = A^Tb.$$

Al igual que antes, si $A$ tiene rango máximo resulta $A^TA$ invertible y por lo tanto $x = (A^TA)^{-1}A^Tb$.

(*)