



**UC** | Chile



**UC** | Chile

## **Algebra Lineal Aplicada para Ciencia de Datos**



# Clase 10. Problemas cuadráticos: Solución de mínima norma



- 1 Motivación
- 2 Formulación del problema
- 3 Resolución con argumentos geométricos
- 4 El multiplicador de Lagrange
- 5 Las ecuaciones de punto silla
- 6 El caso general



**UC** | Chile

**Motivación**

# Motivación



En problemas modernos de análisis de datos frecuentemente tenemos datos

$$(x_{1,1}, \dots, x_{1,n}, y_1), \quad \dots, \quad (x_{m,1}, \dots, x_{m,n}, y_m)$$

donde  $x_{i,1}, \dots, x_{i,n}$  es la  $i$ -ésima observación de  $n$  **predictores** y  $y_i$  es la  $i$ -ésima **respuesta**

Se tienen  $n$  predictores y  $m$  observaciones

# Motivación



Al igual que en la clase previa, podemos modelar la relación entre **predictores** y **respuestas** usando el **modelo afín**

$$y_i \approx \beta_0 + \beta_1 x_{i,1} + \dots + \beta_n x_{i,n} \quad \text{con} \quad i \in \{1, \dots, m\}.$$

para factores  $\beta_0, \beta_1, \dots, \beta_n$  por determinar

# Motivación



En este caso el **residuo** es

$$r_i = y_i - \beta_0 - \beta_1 x_{i,1} - \dots - \beta_n x_{i,n} \quad \text{para } i \in \{1, \dots, m\}$$

Matricialmente la **suma de los residuos al cuadrado** es

$$\sum_{i=1}^m r_i^2 = \|r\|_2^2 = \|y - X\beta\|_2^2$$

donde  $r, y \in \mathbb{R}^m$ ,  $\beta \in \mathbb{R}^{n+1}$  y  $X \in \mathbb{R}^{m \times n}$  están definidos de manera análoga a la clase previa

## Motivación

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i,1} + \dots + \beta_n x_{i,n} \\ &\downarrow \qquad \qquad \qquad \downarrow \\ &= \begin{cases} 1 & \text{si el gen se expresa en el individuo } i \\ 0 & \text{si no} \end{cases} \\ &\downarrow \\ &= \begin{cases} 1 & \text{si el individuo } i \text{ tiene la condición} \\ 0 & \text{si no} \end{cases} \end{aligned}$$

En **genética** habitualmente hay un **gran número de predictores** pero es **costoso** obtener muchas observaciones

# Motivación



Esto quiere decir que se tienen **más predictores que observaciones** y el sistema

$$y = X\beta$$

esta **subdeterminado**: tenemos **más incógnitas que ecuaciones**

Si el sistema es **consistente** entonces tiene una **infinidad de soluciones**

# Motivación



Esto quiere decir que existe una **infinidad** de vectores  $\beta$  para los cuales

$$\|y - X\beta\|_2^2 = 0$$

Por lo tanto, el problema es **seleccionar** una de estas soluciones a través de un **criterio sistemático**



**UC** | Chile

## Formulación del problema

# Formulación del problema



Consideramos un vector  $y \in \mathbb{R}^m$  y una matriz  $A \in \mathbb{R}^{m \times n}$  con

$$m \leq n$$

La matriz es **ancha**: tiene **a lo más** tantas filas como columnas.

Supondremos además que es de **rango completo**: para una matriz ancha, esto quiere decir que

$$\text{im}(A) = \mathbb{R}^m \quad \Rightarrow \quad \text{nul}(A^\top) = \{0\}$$

o bien, que sus **filas** son **linealmente independientes**

## Formulación del problema

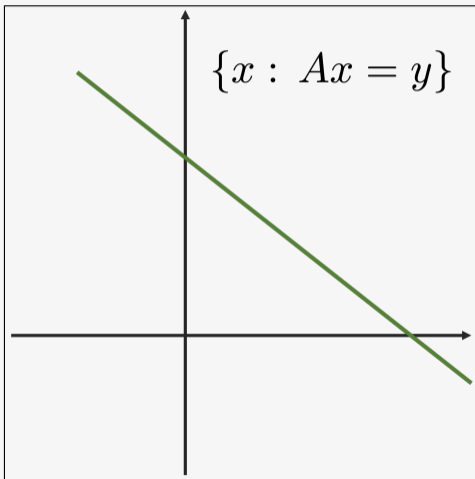


El sistema

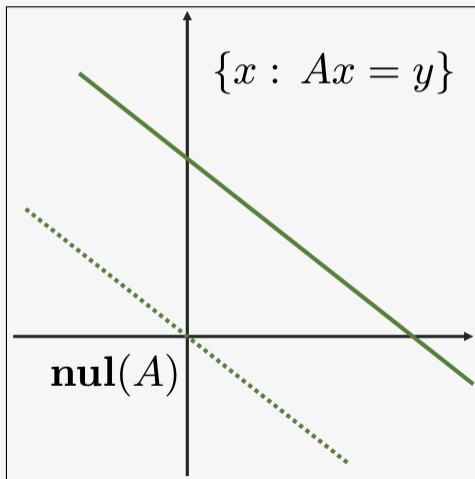
$$y = Ax$$

está **subdeterminado** pero es **consistente** ya que  $A$  es de rango completo: tiene una **cantidad infinita** de soluciones

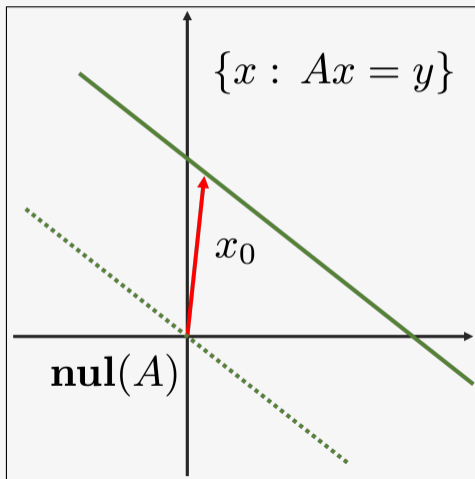
## Formulación del problema. Ejemplo



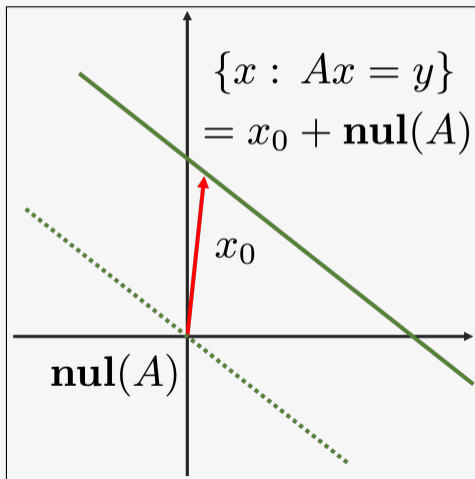
## Formulación del problema. Ejemplo



## Formulación del problema. Ejemplo



## Formulación del problema. Ejemplo



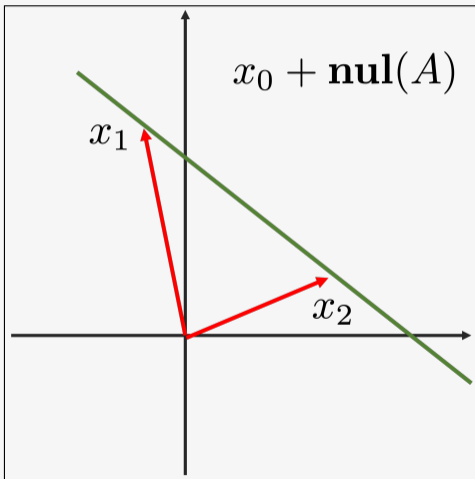
# Formulación del problema



Buscamos la solución de **mínima norma Euclidea**

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} \quad \|x\|_2^2 \quad \text{sujeto a} \quad Ax = y$$

## Formulación del problema. Ejemplo



## Formulación del problema. Ejemplo



Para

$$y = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad y \quad A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

el sistema

$$y = Ax$$

es **consistente** pero **subdeterminado**

## Formulación del problema. Ejemplo

Cualquier vector de la forma

$$x(\alpha) = \underbrace{\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}}_{x_0} + \alpha \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1/3 \end{bmatrix}}_{\text{nul}(A)} = \begin{bmatrix} \alpha \\ -1 + \alpha \\ 1 + \alpha/3 \end{bmatrix}$$

para algún escalar  $\alpha$  es solución de la ecuación

## Formulación del problema. Ejemplo



Por lo tanto, el vector de mínima norma Euclidea es aquel para el que

$$\begin{aligned}\|x(\alpha)\|_2^2 &= \alpha^2 + (-1 + \alpha)^2 + (1 + \alpha/3)^2 \\ &= \frac{19}{9}\alpha^2 - \frac{4}{3}\alpha + 2\end{aligned}$$

sea mínimo

# Formulación del problema. Ejemplo





**UC** | Chile

## **Resolución con argumentos geométricos**

## Resolución con argumentos geométricos

$$\boxed{\begin{array}{l} \text{minimizar} \quad \|x\|_2^2 \\ x \in \mathbb{R}^n \end{array} \quad \text{sujeto a} \quad Ax = y}$$

Reformulamos el problema escogiendo primero una solución  $x_0$  cualquiera a la ecuación

$$Ax_0 = y$$

Esto nos permite escribir

$$Ax = y \quad \Leftrightarrow \quad Ax = Ax_0 \quad \Leftrightarrow \quad A(x_0 - x) = 0.$$

## Resolución con argumentos geométricos



$$\boxed{\text{minimizar}_{x \in \mathbb{R}^n} \|x\|_2^2 \text{ sujeto a } Ax = y}$$

Reformulamos el problema escogiendo una posible solución  $x_0$  a la ecuación

$$Ax_0 = y$$

Esto nos permite escribir

$$Ax = y \quad \Leftrightarrow \quad Ax = Ax_0 \quad \Leftrightarrow \quad A(x_0 - x) = 0.$$

## Resolución con argumentos geométricos

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} \quad \|x\|_2^2 \quad \text{sujeto a} \quad A(x_0 - x) = 0$$

Siempre podemos representar  $x \in \mathbb{R}^n$  como

$$x = x_0 - z \quad \text{con} \quad z = x_0 - x$$

de modo que en vez de buscar un vector  $x \in \mathbb{R}^n$  que resuelva este problema, podemos buscar un vector  $z \in \mathbb{R}^n$  que resuelva este mismo problema

## Resolución con argumentos geométricos

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} \quad \|x\|_2^2 \quad \text{sujeto a} \quad A(x_0 - x) = 0$$

Siempre podemos representar  $x \in \mathbb{R}^n$  como

$$x = x_0 - z \quad \text{con} \quad z = x_0 - x$$

de modo que en vez de buscar un vector  $x \in \mathbb{R}^n$  que resuelva este problema, podemos buscar un vector  $z \in \mathbb{R}^n$  que resuelva este mismo problema

## Resolución con argumentos geométricos



$$\underset{z \in \mathbb{R}^n}{\text{minimizar}} \quad \|x_0 - z\|_2^2 \quad \text{sujeto a} \quad Az = 0$$

Finalmente,

$$Az = 0 \quad \Leftrightarrow \quad z \in \mathbf{nul}(A)$$

## Resolución con argumentos geométricos



$$\underset{z \in \mathbb{R}^n}{\text{minimizar}} \quad \|x_0 - z\|_2^2 \quad \text{sujeto a} \quad Az = 0$$

Finalmente,

$$Az = 0 \quad \Leftrightarrow \quad z \in \mathbf{nul}(A)$$

## Resolución con argumentos geométricos

$$\underset{z \in \mathbb{R}^n}{\text{minimizar}} \quad \|x_0 - z\|_2^2 \quad \text{sujeto a} \quad z \in \mathbf{nul}(A)$$

El vector  $z$  que resuelve este problema es aquel que minimiza la distancia entre  $x_0$  y cualquier vector de  $z \in \mathbf{nul}(A)$

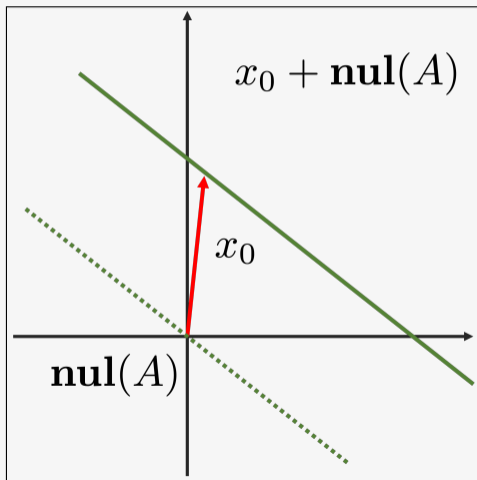
$$z^* = \mathbf{proy}_{\mathbf{nul}(A)}(x_0)$$

de modo que

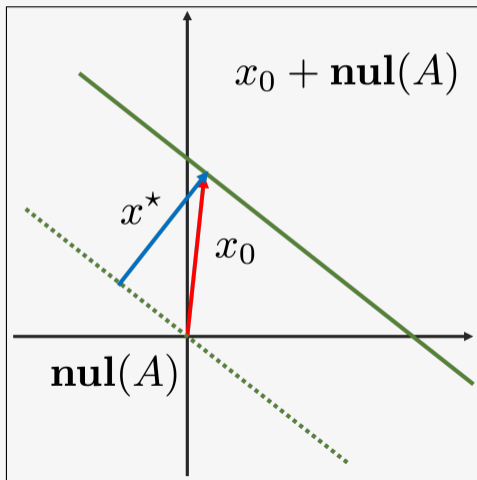
$$x^* = x_0 - z^* = x_0 - \mathbf{proy}_{\mathbf{nul}(A)}(x_0)$$

y la solución es el **complemento ortogonal** de  $x_0$  sobre  $\mathbf{nul}(A)$

## Resolución con argumentos geométricos. Ejemplo



## Resolución con argumentos geométricos. Ejemplo



## Resolución con argumentos geométricos



De la propiedad fundamental de la proyección ortogonal deducimos que

$$z \cdot x^* = 0 \quad \text{para cualquier } z \in \mathbf{nul}(A)$$

Sin embargo, en este caso no tenemos una colección de generadores **natural** de  $\mathbf{nul}(A)$

Como consecuencia, no obtenemos un sistema de ecuaciones como las ecuaciones normales

Esto nos lleva a introducir un **multiplicador de Lagrange**



**UC** | Chile

# El multiplicador de Lagrange

# El multiplicador de Lagrange



La propiedad fundamental de la proyección ortogonal se puede escribir como

$$x^{\star} \in \mathbf{nul}(A)^{\perp}$$

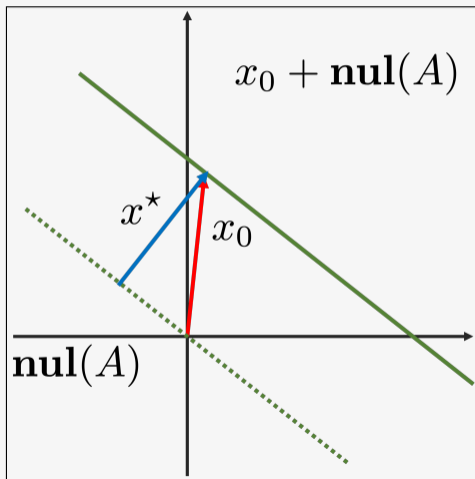
Ya que

$$\mathbf{nul}(A)^{\perp} = \mathbf{im}(A^{\top})$$

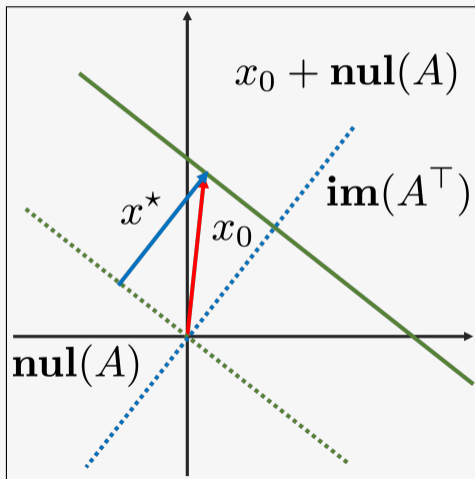
esta propiedad es equivalente a

$$x^{\star} \in \mathbf{im}(A^{\top})$$

## El multiplicador de Lagrange. Ejemplo



## El multiplicador de Lagrange. Ejemplo



# El multiplicador de Lagrange



Por lo tanto, existe un **multiplicador de Lagrange**  $\lambda^* \in \mathbb{R}^m$  tal que

$$x^* = -A^\top \lambda^* \quad \Rightarrow \quad x^* + A^\top \lambda^* = 0$$

y la solución  $x^*$  debe satisfacer las ecuaciones

$$x^* + A^\top \lambda^* = 0$$

$$Ax^* = y$$

**¿Podemos resolver estas ecuaciones?**



**UC** | Chile

# Las ecuaciones de punto silla

# Las ecuaciones de punto silla



Las ecuaciones

$$\begin{aligned}x^{\star} + A^{\top} \lambda^{\star} &= 0 \\ Ax^{\star} &= y\end{aligned}$$

se pueden escribir matricialmente como

$$\begin{bmatrix} I & A^{\top} \\ A & 0 \end{bmatrix} \begin{bmatrix} x^{\star} \\ \lambda^{\star} \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

Estas son las **ecuaciones de punto silla**

Cuando  $A$  es de rango completo, la solución **existe** y es **única**

## Las ecuaciones de punto silla. Ejemplo

Para

$$y = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad y \quad A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

la ecuación de punto silla es

$$\begin{bmatrix} 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 3 \end{bmatrix}$$

## Las ecuaciones de punto silla



La primera ecuación implica que

$$x^* = -A^T \lambda^*$$

y aplicando esto en la segunda obtenemos

$$y = Ax^* = A(-A^T \lambda^*) = -AA^T \lambda^*$$

Ya que  $A$  es de rango completo la matriz  $AA^T$  es **no singular**

## Las ecuaciones de punto silla



Si existe  $\lambda \in \mathbb{R}^m$  tal que

$$AA^T\lambda = 0 \quad \Rightarrow \quad A^T\lambda \in \mathbf{nul}(A)$$

entonces, ya que,

$$\mathbf{im}(A^T) = \mathbf{nul}(A)^\perp$$

concluimos que

$$A^T\lambda \in \mathbf{nul}(A) \cap \mathbf{nul}(A)^\perp \quad \Rightarrow \quad A^T\lambda = 0$$

y, ya que  $\mathbf{nul}(A) = \{0\}$ , concluimos que  $\lambda = 0$

## Las ecuaciones de punto silla



Por lo tanto

$$y = Ax^* = A(-A^\top \lambda^*) = -AA^\top \lambda^* \quad \Rightarrow \quad \lambda^* = -(AA^\top)^{-1}y$$

y

$$x^* = -A^\top \lambda^* = A^\top (AA^\top)^{-1}y$$

## Las ecuaciones de punto silla. Ejemplo

Para

$$y = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad y \quad A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

se tiene

$$AA^T = \begin{bmatrix} 2 & -1 \\ -1 & 10 \end{bmatrix} \Rightarrow (AA^T)^{-1} = \begin{bmatrix} 10/19 & 1/19 \\ 1/19 & 2/19 \end{bmatrix}$$

y

$$\lambda^* = - \begin{bmatrix} 13/19 \\ 7/19 \end{bmatrix} \Rightarrow x^* = \begin{bmatrix} 6/19 \\ -13/19 \\ 21/19 \end{bmatrix}.$$

## Las ecuaciones de punto silla



En la práctica usamos la **factorización QR** para encontrar  $x^*$

Cuando  $A \in \mathbb{R}^{m \times n}$  es **ancha** y de **rango completo** entonces

$$A^\top = QR \quad \text{con} \quad Q \in \mathbb{R}^{n \times m} \quad \text{y} \quad Q^\top Q = I \quad \text{y} \quad R \in \mathbb{R}^{m \times m} \quad \text{invertible}$$

Por lo tanto

$$AA^\top \lambda^* = (QR)^\top (QR) \lambda^* = R^\top Q^\top QR \lambda^* = R^\top R \lambda^* = -y$$

## Las ecuaciones de punto silla



Por lo tanto, de manera similar a las ecuaciones normales, podemos resolver

$$R^{\top} \eta^{\star} = -y$$

$$R \lambda^{\star} = \eta^{\star}$$

usando **sustitución progresiva** y **sustitución regresiva** respectivamente

Luego, obtenemos  $x^{\star}$  calculando

$$x^{\star} = -A^{\top} \lambda^{\star}$$



**UC** | Chile

**El caso general**

## El caso general



**¿Qué ocurre cuando  $A$  es ancha, pero no es de rango completo?**

En este caso, podría ocurrir que

$$y \notin \mathbf{im}(A)$$

y este caso lo estudiaremos en clases futuras

Cuando

$$y \in \mathbf{im}(A)$$

entonces las ecuaciones de punto silla siguen siendo válidas

## El caso general



En este caso el multiplicador de Lagrange **no es único** ya que el sistema

$$AA^T \lambda^* = -y$$

es **consistente** pero tiene una **cantidad infinita** de soluciones

Sin embargo, la solución óptima  $x^*$  es **única**



**UC** | Chile