# Solución del problema de mínimos cuadrados por medio de una descomposición QR (un tema del curso “Álgebra Lineal Numérica”)

Egor Maximenko

Instituto Politécnico Nacional
Escuela Superior de Física y Matemáticas
México

18 de mayo de 2021

**Objetivo:**

resolver el problema de mínimos cuadrados
usando una factorización QR delgada.

Nos restringimos al caso $A \in \mathbb{R}^{n \times m}$, $\mathrm{r}(A) = m$.

**Prerrequisitos:**

- algoritmos para construir una factorización QR;
- algoritmos para resolver sistemas triangulares;
- proyección ortogonal sobre un subespacio generado por una lista ortonormal.

# Problema de mínimos cuadrados

Están dados $A \in \mathbb{R}^{n\times m}$ y $b \in \mathbb{R}^n$.

Consideramos la función $f\colon \mathbb{R}^m \to \mathbb{R}$,

$$f(x) := \|Ax - b\|.$$

En otras palabras,

$$f(x) = \sqrt{\sum_{j=1}^{n} \left( \sum_{k=1}^{m} A_{j,k} x_k \right)^2}.$$

Encontrar un punto $s$ en $\mathbb{R}^n$, donde $f$ alcanza su mínimo, esto es,

$$\forall x \in \mathbb{R}^n \qquad f(x) \geq f(s).$$

# Observación

Consideramos

$$f(x) := \|Ax - b\|, \qquad g(x) := \|Ax - b\|^2.$$

Como la función $t \mapsto t^2$ es estrictamente creciente en $[0, +\infty)$, las funciones $f$ y $g$ tienen los mismos puntos de mínimo.

# El sentido geométrico: la imagen de una matriz

Sea $A \in \mathbb{R}^{n\times m}$. Denotemos por $a_1, \ldots, a_m$ las columnas de $A$:

$$A = \left[a_1, \ldots, a_m\right].$$

Denotemos por $S$ la imagen de $A$:

$$\begin{aligned}
S &:= A(\mathbb{R}^m) = \{Ax \colon x \in \mathbb{R}^m\} = \{y \in \mathbb{R}^n \colon \quad \exists x \in \mathbb{R}^m \quad y = Ax\} \\
&= \left\{y \in \mathbb{R}^n \colon \quad \exists x_1, \ldots, x_m \in \mathbb{R}^m \quad y = \sum_{k=1}^{m} x_k a_k\right\} \\
&= \ell(a_1, \ldots, a_m).
\end{aligned}$$

Por eso $S$ también se conoce como el espacio columna de $A$.

# El sentido geométrico del problema de mínimos cuadrados

Sean $A = [a_1, \ldots, a_m] \in \mathbb{R}^{n\times m}$ y $b \in \mathbb{R}^n$. Denotemos por $S$ la imagen de $A$:

$$S := A(\mathbb{R}^m) = \ell(a_1, \ldots, a_m).$$

Consideramos $f\colon \mathbb{R}^m \to \mathbb{R}$,

$$f(x) = \|Ax - b\|.$$

Entonces

$$\inf_{x\in\mathbb{R}^m} g(x) = \inf_{x\in\mathbb{R}^m} \|Ax - b\| = \inf_{u\in S} \|u - b\| = d(b, S).$$

# El sentido geométrico del problema de mínimos cuadrados

Sean $A = [a_1, \ldots, a_m] \in \mathbb{R}^{n\times m}$ y $b \in \mathbb{R}^n$. Denotemos por $S$ la imagen de $A$:

$$S := A(\mathbb{R}^m) = \ell(a_1, \ldots, a_m).$$

Consideramos $f\colon \mathbb{R}^m \to \mathbb{R}$,

$$f(x) = \|Ax - b\|.$$

Entonces

$$\inf_{x\in\mathbb{R}^m} g(x) = \inf_{x\in\mathbb{R}^m} \|Ax - b\| = \inf_{u\in S} \|u - b\| = d(b, S).$$

Encontrar un punto mínimo de $g$ equivale a encontrar un $S$ un punto más cercano a $b$.

# Criterio de la inyectividad de una matriz en términos de su rango

## Proposición

*Sea* $A \in \mathbb{R}^{n \times m}$*. Pongamos*

$$\mathrm{r}(A) := \dim(A(\mathbb{R}^m)).$$

*Entonces las siguientes condiciones son equivalentes:*

(a) $\mathrm{r}(A) = m$;

(b) *las columnas de* $A$ *son linealmente independientes;*

(c) *la transformación lineal* $x \mapsto Ax$ *es inyectiva.*

# La proyección ortogonal es el punto más cercano (repaso)

## Proposición

*Sea* $S$ *un subespacio de* $\mathbb{R}^n$, *sea* $b \in \mathbb{R}^n$ *y sean* $u \in S$, $w \in S^\perp$ *tales que* $u + w = b$.

*Entonces* $u$ *es el punto de* $S$ *más cercano a* $b$:

$$\forall y \in S \setminus \{u\} \qquad \|y - b\| > \|u - b\|.$$

# La proyección ortogonal es el punto más cercano (repaso)

## Proposición

*Sea $S$ un subespacio de $\mathbb{R}^n$, sea $b \in \mathbb{R}^n$ y sean $u \in S$, $w \in S^\perp$ tales que $u + w = b$.*

*Entonces $u$ es el punto de $S$ más cercano a $b$:*

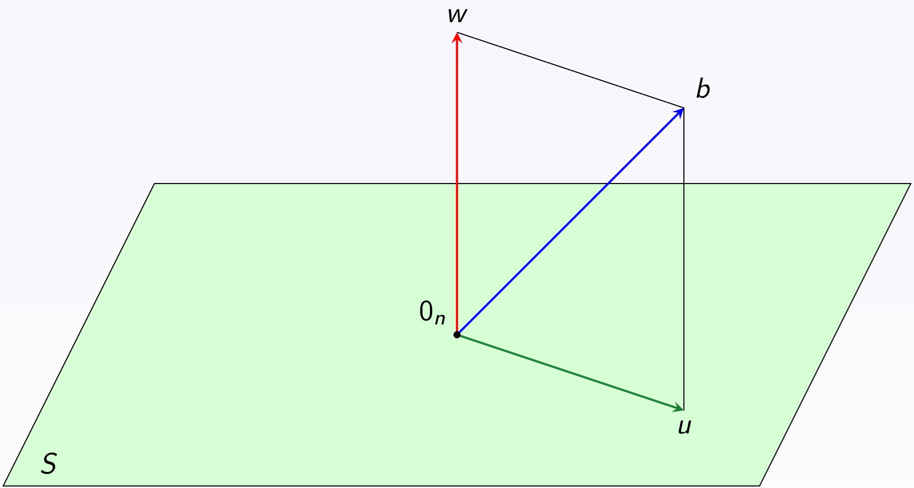
$$\forall y \in S \setminus \{u\} \qquad \|y - b\| > \|u - b\|.$$

**Demostración.**

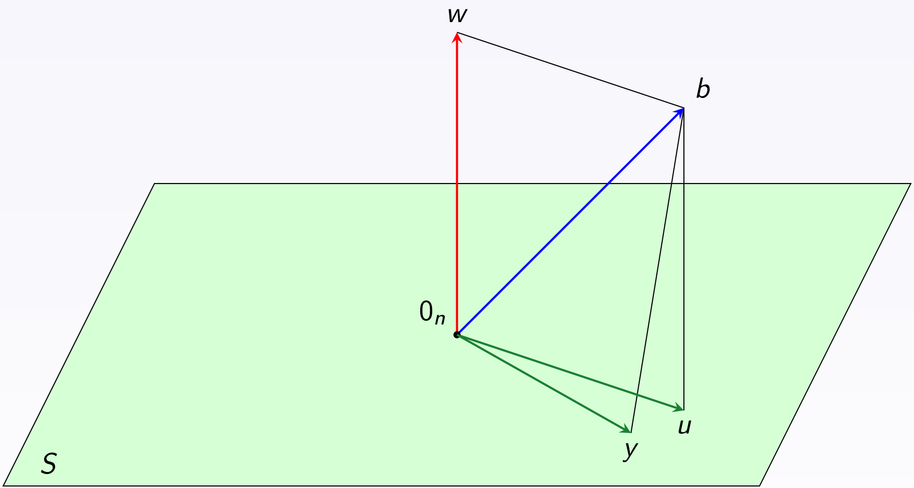
$$\|y - b\|^2 = \|y - u + u - b\|^2 = \|(y - u) + (u - b)\|^2$$

Notamos que $y - u \in S$, $u - b = -w \in S^\perp$ y aplicamos el teorema de Pitágoras.

$$\|y - b\|^2 = \|y - u\|^2 + \|u - b\|^2 > \|u - b\|^2.$$

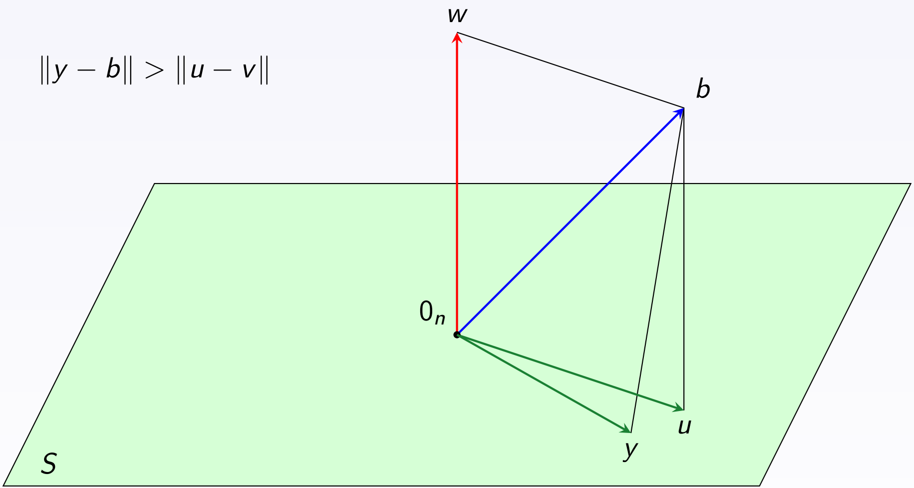


w
b
$0_n$
u
S



$w$
$b$
$0_n$
$u$
$y$
$S$

$$\|y - b\| > \|u - v\|$$



# Receta para la proyección ortogonal al subespacio en términos de una base ortonormal del subespacio (repaso)

## Proposición

*Sea $S$ un subespacio de $\mathbb{R}^n$, sea $b \in \mathbb{R}^n$, y sea $(q_1, \ldots, q_m)$ una base ortonormal de $S$.*

*Pongamos*

$$u := \sum_{k=1}^{m} \langle b, q_k \rangle \, q_k, \qquad w := b - u.$$

*Entonces $u \in S$, $w \in S^\perp$ y $u + w = b$.*

# Receta para la proyección ortogonal al subespacio en términos de una base ortonormal del subespacio (repaso)

**Proposición**

*Sea $S$ un subespacio de $\mathbb{R}^n$, sea $b \in \mathbb{R}^n$, y sea $(q_1, \ldots, q_m)$ una base ortonormal de $S$. Pongamos*

$$u := \sum_{k=1}^{m} \langle b, q_k \rangle \, q_k, \qquad w := b - u.$$

*Entonces $u \in S$, $w \in S^{\perp}$ y $u + w = b$.*

La misma receta en una forma matricial: si $Q = [q_1, \ldots, q_m]$, entonces

$$u =$$

# Receta para la proyección ortogonal al subespacio en términos de una base ortonormal del subespacio (repaso)

**Proposición**

*Sea $S$ un subespacio de $\mathbb{R}^n$, sea $b \in \mathbb{R}^n$, y sea $(q_1, \ldots, q_m)$ una base ortonormal de $S$.*
*Pongamos*

$$u := \sum_{k=1}^{m} \langle b, q_k \rangle \, q_k, \qquad w := b - u.$$

*Entonces $u \in S$, $w \in S^\perp$ y $u + w = b$.*

La misma receta en una forma matricial: si $Q = [q_1, \ldots, q_m]$, entonces

$$u = QQ^\top b.$$

# Descomposición QR delgada (repaso)

Está dada una matriz $A \in \mathbb{R}^{n \times m}$ con $m \leq n$ y $\mathrm{r}(A) = m$.

Se dice que $(Q, R)$ es una descomposición QR delgada de $A$ si

- $Q \in \mathbb{R}^{n \times m}$ y las columnas de $Q$ forman una lista ortonormal, esto es, $Q^\top Q = I_m$,
- $R \in \mathbb{R}^{m \times m}$ y la matriz $R$ es triangular superior,
- $A = QR$.

# Descomposición QR delgada (repaso)

Está dada una matriz $A \in \mathbb{R}^{n\times m}$ con $m \le n$ y $\mathrm{r}(A) = m$.

Se dice que $(Q, R)$ es una descomposición QR delgada de $A$ si

- $Q \in \mathbb{R}^{n\times m}$ y las columnas de $Q$ forman una lista ortonormal, esto es, $Q^\top Q = I_m$,
- $R \in \mathbb{R}^{m\times m}$ y la matriz $R$ es triangular superior,
- $A = QR$.

$$\underbrace{\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,1} & A_{3,2} & A_{3,3} \\ A_{4,1} & A_{4,2} & A_{4,3} \\ A_{5,1} & A_{5,2} & A_{5,3} \end{bmatrix}}_{\text{columnas lin. indep.}} = \underbrace{\begin{bmatrix} Q_{1,1} & Q_{1,2} & Q_{1,3} \\ Q_{2,1} & Q_{2,2} & Q_{2,3} \\ Q_{3,1} & Q_{3,2} & Q_{3,3} \\ Q_{4,1} & Q_{4,2} & Q_{4,3} \\ Q_{5,1} & Q_{5,2} & Q_{5,3} \end{bmatrix}}_{\text{columnas ortonormales}} \underbrace{\begin{bmatrix} R_{1,1} & R_{1,2} & R_{1,3} \\ 0 & R_{2,2} & R_{2,3} \\ 0 & 0 & R_{3,3} \end{bmatrix}}_{\text{matriz triangular superior}}.$$

# Descomposición QR delgada y los subespacios generados

## Proposición

*Sea* $A \in \mathbb{R}^{n\times m}$ *con* $m \leq n$ *y* $\mathrm{r}(A) = m$*, y sea* $(Q, R)$ *una descomposición QR delgada de* $A$*. Denotemos por* $a_1, \ldots, a_m$ *las columnas de* $A$ *y por* $q_1, \ldots, q_m$ *las columnas de* $Q$*:*

$$A = [a_1, \ldots, a_m], \qquad Q = [q_1, \ldots, q_m].$$

*Entonces*

$$\ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

**Idea de demostración.**

Como $A = QR$, cada columna de $A$ es una combinación lineal de las columnas de $Q$:

$$a_k = A_{*,k} = QR_{*,k} = \sum_{j=1}^{m} Q_{*,j} R_{j,k} = \sum_{j=1}^{m} R_{j,k} q_j \in \ell(q_1, \ldots, q_m).$$

Como $\mathrm{r}(A) = m$ y $A = QR$, necesariamente $\mathrm{r}(R) = m$, y la matriz $R$ es invertible. Luego

$$Q = AR^{-1}.$$

Esta igualdad implica que cada columna de $Q$ es una combinación lineal de las columnas de $A$.

# Solución del problema de mínimos cuadrados en términos de una descomposición QR

**Teorema**

*Sea* $A \in \mathbb{R}^{n\times m}$ *con* $m \leq n$ *y* $\mathrm{r}(A) = m$, *sea* $b \in \mathbb{R}^n$,
*y sea* $(Q, R)$ *una descomposición QR delgada de* $A$*.*

# Solución del problema de mínimos cuadrados en términos de una descomposición QR

## Teorema

*Sea* $A \in \mathbb{R}^{n\times m}$ *con* $m \le n$ *y* $\mathrm{r}(A) = m$, *sea* $b \in \mathbb{R}^n$,
*y sea* $(Q, R)$ *una descomposición QR delgada de* $A$. *Definimos*

$$s := R^{-1} Q^\top b.$$

## Solución del problema de mínimos cuadrados en términos de una descomposición QR

### Teorema

*Sea $A \in \mathbb{R}^{n\times m}$ con $m \le n$ y $\mathrm{r}(A) = m$, sea $b \in \mathbb{R}^n$, y sea $(Q, R)$ una descomposición QR delgada de $A$. Definimos*

$$s := R^{-1}Q^\top b.$$

*Entonces $s$ es el punto mínimo global estricto de la función*

$$f\colon \mathbb{R}^m \to \mathbb{R}, \qquad f(x) = \|Ax - b\|.$$

# Solución del problema de mínimos cuadrados en términos de una descomposición QR

**Teorema**

*Sea* $A \in \mathbb{R}^{n\times m}$ *con* $m \le n$ *y* $\mathrm{r}(A) = m$*, sea* $b \in \mathbb{R}^n$*,*
*y sea* $(Q, R)$ *una descomposición QR delgada de* $A$*. Definimos*

$$s := R^{-1} Q^\top b.$$

*Entonces* $s$ *es el punto mínimo global estricto de la función*

$$f\colon \mathbb{R}^m \to \mathbb{R}, \qquad f(x) = \|Ax - b\|.$$

*En otras palabras, para cada* $x$ *en* $\mathbb{R}^m \setminus \{s\}$ *se tiene*

# Solución del problema de mínimos cuadrados en términos de una descomposición QR

**Teorema**

*Sea* $A \in \mathbb{R}^{n\times m}$ *con* $m \le n$ *y* $\mathrm{r}(A) = m$, *sea* $b \in \mathbb{R}^n$,
*y sea* $(Q, R)$ *una descomposición QR delgada de* $A$. *Definimos*

$$s := R^{-1} Q^\top b.$$

*Entonces* $s$ *es el punto mínimo global estricto de la función*

$$f\colon \mathbb{R}^m \to \mathbb{R}, \qquad f(x) = \|Ax - b\|.$$

*En otras palabras, para cada* $x$ *en* $\mathbb{R}^m \setminus \{s\}$ *se tiene* $\|Ax - b\| > \|As - b\|$.

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) =$$

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$.

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$. Representamos $u$ como la imagen de un vector bajo $A$:

$$u = QQ^\top b = QRR^{-1}Q^\top b =$$

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$. Representamos $u$ como la imagen de un vector bajo $A$:

$$u = QQ^\top b = QRR^{-1}Q^\top b = A(R^{-1}Q^\top b) = As.$$

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$. Representamos $u$ como la imagen de un vector bajo $A$:

$$u = QQ^\top b = QRR^{-1}Q^\top b = A(R^{-1}Q^\top b) = As.$$

Sea $x \in \mathbb{R}^m \setminus \{s\}$. Como las columnas de $A$ son linealmente independientes, tenemos

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$. Representamos $u$ como la imagen de un vector bajo $A$:

$$u = QQ^\top b = QRR^{-1}Q^\top b = A(R^{-1}Q^\top b) = As.$$

Sea $x \in \mathbb{R}^m \setminus \{s\}$. Como las columnas de $A$ son linealmente independientes, tenemos $Ax \neq u$.

# Demostración

Denotamos por $S$ el subespacio generado por las columnas de $A$:

$$S := \ell(a_1, \ldots, a_m) = \ell(q_1, \ldots, q_m).$$

Sea $u := QQ^\top b$. Como ya sabemos, $u$ es el punto de $S$ más cercano al punto $b$. Representamos $u$ como la imagen de un vector bajo $A$:

$$u = QQ^\top b = QRR^{-1}Q^\top b = A(R^{-1}Q^\top b) = As.$$

Sea $x \in \mathbb{R}^m \setminus \{s\}$. Como las columnas de $A$ son linealmente independientes, tenemos $Ax \neq u$. Luego

$$\|Ax - b\| > \|u - b\| = \|As - b\|.$$

# Comentario al teorema

En particular, el teorema implica que si $A$ en $\mathbb{R}^{n\times m}$ y $\mathrm{r}(A) = m$, entonces el problema de mínimos cuadrados tiene una única solución.

Si $\mathrm{r}(A) < m$, entonces se puede demostrar que el problema de mínimos cuadrados tiene varias soluciones diferentes (el mínimo se alcanza en varios puntos).

# Sugerencia: como pasar de una descompoción QR completa a una descomposición QR delgada

Supongamos que está dada una descomposión QR completa de $A$:

$$A = \widetilde{Q}\widetilde{R}, \qquad \widetilde{Q} \in \mathbb{R}^{n\times n}, \qquad \widetilde{R} \in \mathbb{R}^{n\times m}.$$

Entonces es fácil construir una descomposición QR delgada de $A$:

$$Q = \widetilde{Q}[\,:\,,1:m], \qquad R = \widetilde{R}[1:m,\,:\,].$$

# Programación: solución del problema de mínimos cuadrados

$$s = R^{-1} Q^\top b.$$

Sea $u = Q^\top b$. Se recomienda calcular $s$ como la solución del sistema triangular $Rs = u$.

Evitar la construcción de la matriz $R^{-1}$.

# Programación: solución del problema de mínimos cuadrados

$$s = R^{-1} Q^\top b.$$

Sea $u = Q^\top b$. Se recomienda calcular $s$ como la solución del sistema triangular $Rs = u$. Evitar la construcción de la matriz $R^{-1}$.

```
def least_squares(A, b):
    (Q, R) = my_favorite_qr_thin_algorithm(A)
    u = ???
    s = solve_ut(???, ???)
    return s
```

# Programación: solución del problema de mínimos cuadrados

$$s = R^{-1} Q^\top b.$$

Sea $u = Q^\top b$. Se recomienda calcular $s$ como la solución del sistema triangular $Rs = u$. Evitar la construcción de la matriz $R^{-1}$.

```
def least_squares(A, b):
    (Q, R) = my_favorite_qr_thin_algorithm(A)
    u = ???
    s = solve_ut(???, ???)
    return s
```

**Ejercicio.** Calcular el número total de multiplicaciones de números reales en este algoritmo.

# Ejercicio: solución de un sistema de ecuaciones lineales por medio de una descomposición QR

Supongamos que $A \in \mathbb{R}^{n\times n}$, $\mathrm{r}(A) = n$, $b \in \mathbb{R}^n$.

Resolver el sistema $Ax = b$ usando una descomposición $QR$ de la matriz $A$.