MDS3120: Álgebra Lineal Aplicada para Ciencia de Datos

Clase 6: Factorización QR

Instructor: Manuel A. Sánchez, Carlos Sing Long *compilado el 9 de junio de 2024*

Introducción

En esta clase, exploraremos conceptos fundamentales en álgebra lineal que son esenciales en una variedad de aplicaciones. Nos centraremos en tres temas principales: ortogonalización, factorización QR y pseudoinversa.

La ortogonalización es un proceso que implica transformar un conjunto de vectores linealmente independientes en un conjunto ortogonal o ortonormal. Esta técnica es crucial en la resolución de sistemas de ecuaciones lineales, la diagonalización de matrices y la descomposición espectral, entre otros. Discutiremos las ventajas de representar datos en términos de vectores ortogonales.

La factorización QR es una herramienta poderosa que descompone una matriz en el producto de una matriz ortogonal y una matriz triangular superior. Esta factorización encuentra una amplia gama de aplicaciones en la resolución de sistemas de ecuaciones lineales, la aproximación de mínimos cuadrados y la diagonalización de matrices.

Por último, la pseudoinversa es una herramienta fundamental en el análisis de datos y la optimización. Se trata de una generalización de la inversa de una matriz que se utiliza para resolver sistemas de ecuaciones lineales sobredeterminados y calcular soluciones de mínimos cuadrados en problemas de regresión.

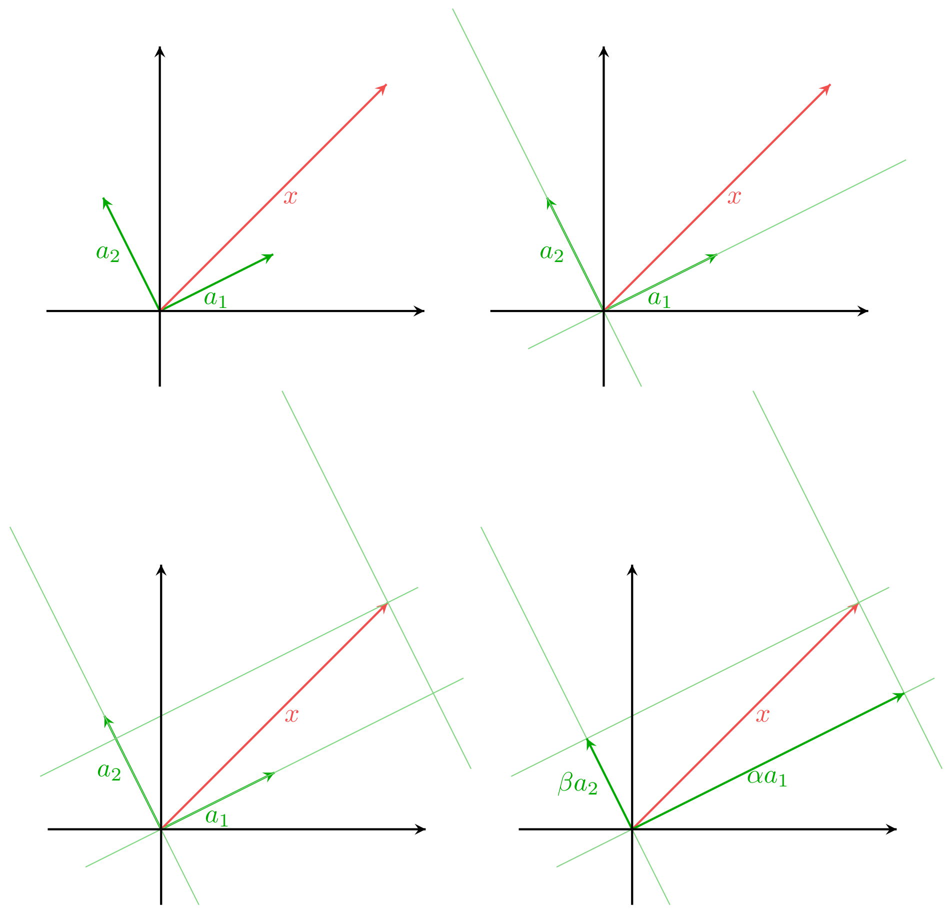
Comenzamos formulando la siguiente pregunta:

¿Cómo calculamos las coordenadas de un vector x en una base $\{a_1, \dots, a_n\}$?

$$x = \beta_1 a_1 + \beta_2 a_2 + ... + \beta_n a_n, \quad \beta_i?$$

Existe un tipo de base para la cual es simple determinar las coordenadas de un vector cualquiera

Ejemplo 6.1. Ilustración gráfica de representación en base ortogonal



Dado un vector x y una base $\{a_1, a_2\}$, nos interesa representar x en términos de la base. Graficamente necesitamos encontrar en los espacios generados, las lineas extendidas, de los vectores a_1 y a_2 los vectores cuya suma nos de el vector x. Esto es, buscamos escalares α y β tales que $\alpha a_1 + \beta a_2 = x$. Encontrar esta combinación resulta una tarea sencilla si contamos con una base ortogonal.

Tema 1: Ortogonalización

Discutiremos a continuación el concepto de ortonogonalización de vectores. Comenzamos definiendo un conjunto de vectores ortogonales y ortonormales. Luego mostramos un algoritmo que, dado un conjunto de vectores linealmente independientes, nos permite calcular un conjunto de vectores ortonormales, lo cuales generan el mismo espacio vectorial.

Vectores ortogonales

Definición 6.1. Una colección de vectores $a_1, ..., a_k$ es ortogonal o mutuamente ortogonal si

$$a_i \cdot a_j = a_i^\top a_j = 0, \quad \text{para } 1 \leq i, j \leq k,\ i \neq j.$$

Si además tenemos que $\|a_i\|_2^2 = a_i^\top a_i = 1$, entonces los vectores se dicen ortonormales, es decir

$$a_i \cdot a_j = a_i^\top a_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Ejemplo 6.2. Ejemplos: de vectores ortogonales

- Los vectores canónicos unitarios son ortonormales

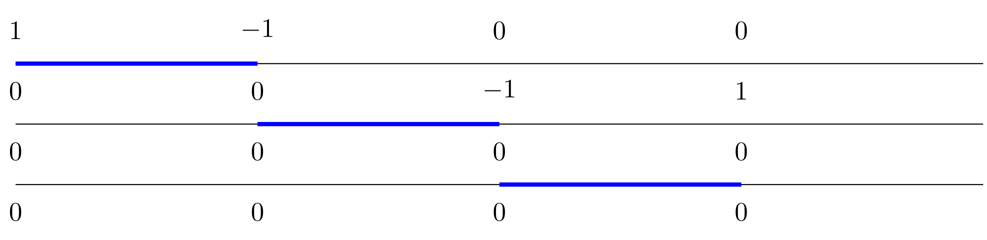
$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Los siguientes vectores son ortonormales

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

- Haar wavelets

$$W = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix}$$



Comentario 6.1. Los vectores ortonormales tienen las siguientes propiedades:

- Si tenemos un conjunto de vectores **ortonormales** entonces tenemos que estos vectores son inmediatamente **linealmente independientes**.

En efecto, si $a_1, ..., a_k$ son ortonormales y existen constantes $\beta_1, ..., \beta_k$ tales que

$$\beta_1 a_1 + ... + \beta_k a_k = 0 \quad \Longrightarrow \quad 0 = a_i^\top(\beta_1 a_1 + ... + \beta_k a_k) = \beta_i \|a_i\|^2 \quad \Longrightarrow a_i = 0.$$

- Si $a_1, ..., a_k$ son vectores **ortonormales** y $x = \beta_1 a_1 + ... + \beta_k a_k$, entonces los **coeficientes** β_i están dados por $\beta_i = a_i^\top x$, $1 \leq i \leq k$.
- Un conjunto de n-vectores ortonormales $a_1, ...a_n$ forman una base del espacio generado por los vectores y se dicen **base ortonormal**.

Ejemplo 6.3. Base ortonormal.

Sea el vector $x \in \mathbb{R}^3$, dado por $\quad x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, y los vectores

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Observe que, $\{a_1, a_2, a_3\}$ forman una base ortonormal. Entonces, describimos x en términos de la base ortonormal por

$$x = (x^\top a_1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (x^\top a_2) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (x^\top a_3) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

$$x = (-3) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + (\frac{3}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + (-\frac{1}{\sqrt{2}}) \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

Observe lo simple que es calcular los coeficientes para escribir x en términos de la base $\{a_1, a_2, a_3\}$.

Dada la importancia de una base ortogonal nos preguntamos si, dado un conjunto de vectores y el espacio generado por estos, es posible encontrar una base ortogonal que genere el mismo espacio. La respuesta es afirmativa y a este proceso lo conocemos como proceso de ortogonalización.

Proceso de ortogonalización de Gram-Schmidt

El siguiente algoritmo presenta el procedimiento para calcular la base ortonal mencionada anteriormente.

Algoritmo 1: Gram - Schmidt (clásico)

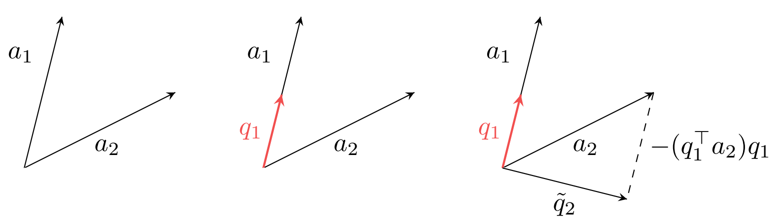
Data: Vectores de $\mathbb{R}^n$, $a_1, ..., a_k$.
Result: Vectores ortonormales de $\mathbb{R}^n$, $q_1, ..., q_\ell$, $\ell \leq k$.
for $i = 1 : k$ **do**
  $\tilde{q}_i = a_i - (q_1^\top a_i)q_1 - \dots (q_{i-1}^\top a_i)q_{i-1}$; `// Ortogonalización`
  if $\tilde{q}_i = 0$ **then**
    break ; `// dependencia lineal`
  else
    $q_i = \tilde{q}_i / \|\tilde{q}_i\|_2$; `// normalización`

Ejemplo 6.4. Aplicación de proceso de ortogonalización de Gram-Schmidt.
Considere los siguientes vectores $a_1 = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}$, $a_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \begin{bmatrix} 0{,}2425 \\ 0{,}9701 \end{bmatrix}, \quad q_2 = \begin{bmatrix} 0{,}9701 \\ -0{,}2425 \end{bmatrix}.$$

El procedimiento lo Ilustramos gráficamente a continuación



Comenzamos con los vectores a_1, a_2. Luego calculamos el primer vector de nuestra base ortonormal q_1, que es a_1 normalizado. Luego calculamos el vector $\tilde{q}_2$ como el vector a_2 menos la proyección de este sobre q_1. Finalmente normalizamos el últimos vector y obtenemos q_2.

Ejemplo 6.5. Ortogonalización Consideramos un segundo ejemplo con vectores de 3 dimensiones. Sean los vectores $a_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, $a_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, $a_3 = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$.
El proceso de ortogonalización de Gram-Schmidt entrega los vectores ortonormales

$$q_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad q_2 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \quad q_3 = \sqrt{14} \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

Observaciones

El proceso de ortogonalización de Gram-Schmidt es una herramienta poderosa que podemos seguir profundizando. A continuación presentamo algunas observaciones respecto a este:

1. Hasta ahora hemos usado Gram-Schmidt comenzando de una colección de vectores linealmente independientes. ¿Que sucede si no tenemos certeza de que los vectores son linealmente independientes? La respuesta esta en el algoritmo. Observamos que si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt se completa entonces los vectores son linealmente independientes. En caso contrario, Si dado un conjunto de vectores $a_1, ..., a_k$ el proceso de Gram-Schmidt **no se completa**, o termina prematuramente en la iteración j, entonces el vector a_j es una combinación lineal de los vectores $q_1, ..., q_{j-1}$ (o también de $a_1, ..., a_{j-1}$).
2. **Algoritmo de Gram-Schmidt Modificado**. En implementaciones una version modificada del algoritmo, pero equivalente matemáticamente, se considera debido a que posee propiedades de errores de redondeo superiores. Otros dos algortimos, Reflexiones de Householder y Rotaciones de Givens, presentan ventajas respecto a la ortogonalidad de los vectores, la cual es sensible a errores de redondeo en los algoritmos de Gram-Schmidt, y son consideradas los algoritmos estándar para resolver el problema de ortogonalización de vectores.

Finalizamos esta sección definiendo el concepto de matriz ortogonal.

Definición 6.2. Definición de matriz ortogonal Una matrix $A \in \mathbb{R}^{m\times n}$ se dice **ortogonal** si satisface que

$$A^\top A = I \in \mathbb{R}^{m\times m}.$$

Observe que las columnas de matrices ortogonales forman una base ortonormal.

Tema 2: Factorización QR

El proceso de ortogonalización de de vectores presentado en la sección anterior esconde una herramienta poderosa de factorización de matrices. Esta es la factorización QR la cual discutimos a continuación.

Factorización QR

Dado un conjunto de vectores linealmente independientes $\{a_1, ..., a_n\}$, con $a_i \in \mathbb{R}^m$ para $1 \leq i \leq n$, el proceso de Gram-Schmidt nos entrega los vectores ortonormales $\{q_1, ..., q_n\}$. Esto además nos permite escribir cada vector a_i de la siguiente forma

$$a_i = (q_1^\top a_i)q_1 + ... + (q_{i-1}^\top a_i)_{i-1} + +(q_i^\top a_i)q_i$$

Si definimos los coeficiente de una matriz $R \in \mathbb{R}^{n\times n}$ por $R_{ji} = (q_j^\top a_i)$, entonces podemos reescribir la represetación de a_i en términos de esta matriz como

$$a_i = R_{1i}q_1 + ... + R_{i-1i}q_{i-1} + R_{ii}q_i$$

Por lo tanto, si ahora tenemos una matriz $A \in \mathbb{R}^{m\times n}$ con columnas linealmente independientes, podemos usar el proceso de ortogonalización de Gram-Schmidt para las columnas de A y obtenemos la siguiente factorización para A

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} R = QR$$

donde $R \in \mathbb{R}^{n\times n}$ es triangular superior con elementos en la diagonal no cero y $Q \in \mathbb{R}^{m\times n}$ es ortogonal.

Teorema 6.1. *Factorización QR.*
Sea $A \in \mathbb{R}^{m\times n}$, $m \geq n$ una matriz con rango completo n. Entonces, existe $Q \in \mathbb{R}^{m\times n}$, ortogonal y $R \in \mathbb{R}^{n\times n}$ triangular superior, tales que:

$$A = QR \qquad \textit{factorización reducida}$$

Una factorización QR completa es una matriz ortogonal $\tilde{Q} \in \mathbb{R}^{m\times m}$ y una matriz $\tilde{R} \in \mathbb{R}^{m\times n}$, con ceros bajo la diagonal

$$A = \tilde{Q}\tilde{R}$$

Ejemplo 6.6. Factorización QR.
Se la matriz

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -2 \\ -1 & 2 & 3 \end{bmatrix}$$

Observe que los vectores columnas de esta matriz fueron usados anteriormente en un ejemplo del proceso de ortogonalización. Entonces, la factorización QR de A es

$$A = \underbrace{\begin{bmatrix} 1/\sqrt{3} & 4/\sqrt{42} & 2/\sqrt{14} \\ 1/\sqrt{3} & 1/\sqrt{42} & -3/\sqrt{14} \\ -1/\sqrt{3} & 5/\sqrt{42} & -1/\sqrt{14} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{3} & -1/\sqrt{3} & -\sqrt{3} \\ 0 & \sqrt{14}/\sqrt{3} & \sqrt{21}/\sqrt{2} \\ 0 & 0 & \sqrt{7}/\sqrt{2} \end{bmatrix}}_{R}$$

donde observamos que la matriz Q tiene como vectores columnas los vectores resultantes del proceso de ortogonalización y la matriz R es triangular superior.

Resolución de sistemas de ecuaciones lineales con QR.

Veremos a continuación que la factorización QR es una herramienta poderosa en el ámbito de la resolución de sistemas lineales. Al descomponer una matriz en el producto de una matriz ortogonal y una matriz triangular superior, la factorización QR facilita la resolución eficiente y estable de sistemas de ecuaciones lineales. Esta técnica es particularmente útil en problemas de regresión lineal, optimización y procesamiento de señales, donde la estabilidad numérica y la eficiencia computacional son fundamentales.

Considere primero el caso de un sistema lineal cuadrado con solución única, es decir $A \in \mathbb{R}^{n\times n}$ es una matriz no singular y sea el vector $b \in \mathbb{R}^n$. Entonces estamos interesados en encontrar un vector

$x \in \mathbb{R}^n$ tal que:

$$Ax = b.$$

La observación fundamental es que, dada la factorización $A = QR$, podemos reescribir el problema como

$$(QR)x = b$$

Luego, dado que Q es una matriz ortogonal, podemos multiplicar por la izquierda ambos lados de la ecuación por $Q^\top$, obteniendo

$$Rx = Q^\top b.$$

Este nuevo sistema lineal, es un sistema triangular superior, el cual puede ser resuelto utilizando sustitución regresiva tal como discutimos anteriormente. El procedimiento es resumido en el siguiente algoritmo.

Algoritmo 2: Resolución de sistemas de ecuaciones lineales con factorización QR

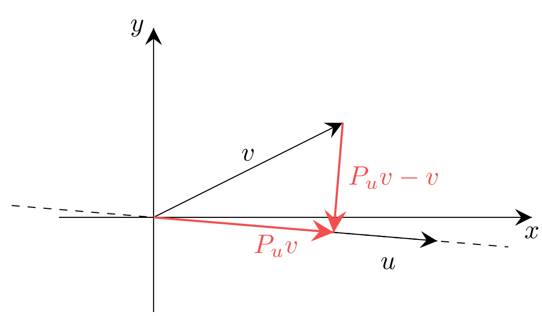
Data: $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$
Result: $x \in \mathbb{R}^n$.
Calcular factorización $A = QR$;
Calcular $Q^\top b$;
Resolver sistema triangular $Rx = Q^\top b$;

Matriz de proyección

Otra aplicación de la factorización es su uso para calcular proyecciones de vectores en subespacios vectoriales. Comencemos definiendo el concepcto de proyección.

Definición 6.3. Una **proyección** es una matriz cuadrada P que satisface $P^2 = P$. Además, decimos que una proyección es **ortogonal** si $P^\top = P$.

Ejemplo 6.7. Ilustramos graficamente el concepto de proyección considerando dos vectores v, $u \in \mathbb{R}^2$. La matriz de $P_u = uu^\top/\|u\|^2$ es una matriz de proyección, $P_u^2 = P_u$. Representamos esto graficamente en la siguiente figura. Se observa que $P_u v$ es un vector que pertenece al espacio generado por el vector u.



Observación: Si Q es una matriz ortogonal alta, entonces $P = QQ^\top$ es una proyeccoón ortogonal. En efecto,

$$P^2 = (QQ^\top)(QQ^\top) = QQ^\top = P$$
$$P^\top = (QQ^\top)^\top = QQ^\top = P$$

Así, dada una matriz ortogonal, la matriz $P = QQ^\top$ proyecta vectores en espacio generado por las columnas de Q.

6.1. Tema 3: Pseudoinversa

A continuación discutiremos una herramienta que nos permitirá escribir soluciones de sistemas de ecuaciones lineales no necesariamente cuadrados, esta es la pseudoinversa.

Matriz de Gram

Definición 6.4. Sea $A \in \mathbb{R}^{m\times n}$. Se define la **matriz de Gram** asociada como la matriz cuadrada $A^\top A$.

Comentario 6.2. Note que, A tiene columnas linealmente independientes si y sólo sí su matriz de Gram es invertible. En efecto,

$$\begin{aligned}(A^\top A)x = 0 \quad &\longrightarrow \quad 0 = x^\top (A^\top A)x = x^\top A^\top Ax = \|Ax\|^2\\ &\longrightarrow \quad Ax = 0\\ &\longrightarrow \quad x = 0\end{aligned}$$

Para el recíproco observe que si existe $x \neq 0$ tal que $Ax = 0$, entonces $A^\top Ax = 0$, lo que implica que la matriz de Gram no es invertible.

Definición 6.5. Definición de Pseudoinversa
Sea $A \in \mathbb{R}^{m\times n}$.

- Si A tiene **columnas linealmente independientes** entonces la matriz $(A^\top A)^{-1}A^\top$ es una inversa por la izquierda de A (matrices altas o cuadradas). La **peudoinversa** se define entonces
$$A^\dagger = (A^\top A)^{-1}A^\top$$
- Si A tiene **filas linealmente independientes** entonces la matriz $A^\top(AA^\top)^{-1}$ es una inversa por la derecha de A (matrices anchas o cuadradas.) La **peudoinversa** se define entonces
$$A^\dagger = A^\top(AA^\top)^{-1}$$

Pseudoinvera por factorización QR

Suponga que $A \in \mathbb{R}^{m\times n}$, y sus columnas son linealmente independientes, y sea su factorización $A = QR$. Entonces:

$$A^\top A = (QR)^\top (QR) = R^\top R$$

Por lo tanto, la pseudoinversa queda:

$$A^\dagger = (A^\top A)^{-1} A = (R^\top R)^{-1}(QR)^\top = R^{-1}Q^\top$$

Resolviendo sistemas sobredeterminados

Una propiedad práctica de la pseudoinversa es que nos permite representar soluciones de sistemas sobredeterminados. Para observar esto, considere el sistema sobredeterminado $Ax = b$, con $A \in \mathbb{R}^{m\times n}$, con $m > n$ y $b \in \mathbb{R}^m$. Además suponga que A tiene n columnas linealmente independientes. Entonces, en este caso la pseudoinversa de $A^\dagger$ es una inversa por la izquierda de A, por lo tanto

$$x = A^\dagger b.$$

Ilustremos esto con un ejemplo

Ejemplo 6.8. Considere el siguiente sistema lineal sobredeterminado:

$$Ax = b \quad \iff \quad \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

Seguimos la idea anterior para encontrar la solución. Utilicemos la factorización QR de A para encontrar la pseudoinversa. La factorización QR de A es

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -0{,}5883 & 0{,}4576 \\ 0{,}7845 & 0{,}5230 \\ 0{,}1961 & -0{,}7191 \end{bmatrix} \begin{bmatrix} 5{,}0990 & 7{,}2563 \\ 0 & 0{,}5883 \end{bmatrix}$$

La solución usando la pseudoinversa es (no necesariamente solución del sistema lineal)

$$x = A^\dagger b = R^{-1}Q^\top = \begin{bmatrix} -1{,}2222 & -1{,}1111 & 1{,}7778 \\ 0{,}7778 & 0{,}8889 & -1{,}2222 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Conclusión

En esta clase, hemos introducido el concepto de factorización QR de una matriz y explorado sus aplicaciones en la resolución de sistemas lineales, tanto cuadrados como sobredeterminados. Notamos que además de su utilidad en la resolución directa de sistemas lineales, la factorización QR encuentra aplicaciones en problemas de mínimos cuadrados, ajuste de curvas, estimación de parámetros y análisis de datos.

Pregunta Formativa

Pregunta #1: ¿Cuál es el propósito principal del proceso de Gram-Schmidt en la factorización QR de una matriz?

1. Reducir la matriz original a una forma triangular inferior.
2. Ortogonalizar las columnas de la matriz original.
3. Descomponer la matriz en una matriz ortogonal y una matriz triangular superior.
4. Encontrar la matriz inversa de la matriz original

Feedback Pregunta #1

1. **Respuesta incorrecta:** Esta opción describe la factorización LU, donde la matriz se descompone en una matriz triangular inferior y una matriz triangular superior.
2. **Respuesta correcta:** El proceso de Gram-Schmidt se utiliza para transformar un conjunto de vectores linealmente independientes en un conjunto ortogonal. Esto es fundamental en la factorización QR para obtener una matriz ortogonal que pueda ser utilizada para resolver sistemas de ecuaciones lineales de manera más eficiente
3. **Respuesta incorrecta:** Esta opción describe el resultado final de la factorización QR, pero no el proceso específico de Gram-Schmidt.
4. **Respuesta incorrecta:** Esta opción no está relacionada con el proceso de Gram-Schmidt o la factorización QR.

Pregunta #2: ¿Cuál de las siguientes afirmaciones describe correctamente el concepto de pseudoinversa de una matriz?

1. La pseudoinversa de una matriz es la inversa de la matriz original.
2. La pseudoinversa de una matriz se utiliza para calcular la descomposición QR de la matriz.
3. La pseudoinversa de una matriz se puede utilizar para representar soluciones sistemas de ecuaciones lineales no necesariamente cuadrados.
4. La pseudoinversa de una matriz solo existe cuando la matriz es cuadrada.

Feedback Pregunta #2

1. **Respuesta incorrecta:** Esta es cierto solo en los casos donde la inversa exista.

2. **Respuesta incorrecta:** Esto no representa la definicion de la pseudoinversa.
3. **Respuesta correcta:** Esta opción describe una aplicación de la pseudoinversa en el contexto de sistemas lineales.
4. **Respuesta correcta:** La pseudoinversa puede calcularse para matrices no necesariamente cuadradas.

6.2. Referencias bibliográficas

Referencias

[1] Boyd, S., & Vandenberghe, L. (2018). *Introduction to applied linear algebra: vectors, matrices, and least squares.* Cambridge university press.

[2] Strang, G. (2019). *Linear algebra and learning from data.* Wellesley-Cambridge Press.