Clase 17. Matrices Ortogonales y el Proceso de Gram-Schmidt.

Curso ‘Linear Algebra’ del MIT.

Resumen

La clase pasada terminamos estudiando preliminarmente los **vectores ortonormales**. En esta ocasión nos concentraremos en ellos y en las **matrices ortogonales**, viendo sus características y cómo obtenerlas cuando, inicialmente, dichos objetos no son perpendiculares. Para esto último, usaremos el **Proceso de Gram-Schmidt**.

1. Vectores Ortonormales.

La clase anterior señalamos que los **vectores ortonormales** son aquellos que se caracterizan por ser:

- Ortogonales entre sí.
- Vectores Unitarios (i.e, de magnitud igual a 1).

Formalmente se indica que dos (o más) vectores $\vec{q_i}$ y $\vec{q_j}$ son **ortonormales** si:

$$\vec{q_i}^T \cdot \vec{q_j} = \begin{cases} 0 \text{ cuando } i \neq j \\ 1 \text{ cuando } i = j \end{cases}$$

La primera condición es la de ortogonalidad. Recordemos que dos vectores de igual dimensionalidad, pero de componentes distintos, son perpendiculares si en su producto punto:

$$\vec{q_i}^T \cdot \vec{q_j} = 0 \text{ (con } i \neq j)$$

Mientras que la segunda condición hace referencia a que sean unitarios y también se puede explicar a partir de su producto punto. Como ya hemos estudiado, el producto punto entre un vector y sí mismo coincide con ser su magnitud al cuadrado.

$$\vec{q}_i^T \cdot \vec{q}_j = ||\vec{q}||^2 \text{ (con } i = j\text{)}$$

Si $||\vec{q}|| = 1$ (i.e, si $\vec{q}$ es unitario), entonces:

$$\vec{q}_i^T \cdot \vec{q}_j = 1 \text{ (con } i = j\text{)}$$

Los **vectores unitarios** también reciben el nombre de **vectores normalizados** y son aquellos de **magnitud** 1 que **van en la misma dirección de otro vector**, los cuales se obtienen multiplicando a este último por el recíproco de su magnitud.

Por ejemplo, sea $\vec{a} \neq \vec{0}$. Su versión normalizada $\hat{a}$ se calcula como:

$$\hat{a} = \frac{1}{||\vec{a}||} \cdot \vec{a}$$

La **norma** (o norma vectorial) es otra forma de referirnos a la magnitud de un vector.

De este modo, ahora podemos entender por qué los vectores **ortonormales** se llaman así:

- Orto $\rightarrow$ Perpendiculares (u ortogonales).
- Normal $\rightarrow$ Unitarios.

Un conjunto de vectores ortonormales, al ser ortogonales, siempre son **linealmente independientes** entre sí. A continuación se demuestra esta afirmación.

Demostración. Sean $\vec{v}_1,\ \vec{v}_2,\ \cdots,\ \vec{v}_n$ vectores ortogonales distinto de $\vec{0}$ y $c_1,\ c_2,\ \cdots,\ c_n$ valores escalares. Estos vectores son linealmente independientes si:

$$\sum_{i=1}^{n} c_i \vec{v}_i = \vec{0} \iff c_1,\ c_2,\ \cdots,\ c_n = 0$$

Calculemos el producto punto en la igualdad de arriba con la transpuesta del vector ortogonal $\vec{v}_i \neq \vec{0}$ para cada $i = 1,\ 2,\ \cdots,\ n$:

$$\begin{aligned} \vec{v}_i^T \cdot \sum_{i=1}^{n} c_i \vec{v}_i &= \vec{v}_i^T \cdot \vec{0} \\ \vec{v}_i^T \cdot (c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n) &= 0 \\ c_1(\vec{v}_i^T \vec{v}_1) + c_2(\vec{v}_i^T \vec{v}_2) + \cdots + c_n(\vec{v}_i^T \vec{v}_n) &= 0 \end{aligned}$$

Cada vez que $i \neq j$, con $j = 1,\ 2,\ \cdots,\ n$, $\vec{v}_i^T \cdot \vec{v}_j = 0$. Por lo tanto, cuando $i = j$:

$$c_i||\vec{v}_i||^2 = 0$$

Como $\vec{v}_i \neq \vec{0}$, entonces $||\vec{v}_i|| \neq 0$, por lo que la única posibilidad de que $c_i||\vec{v}_i||^2 = 0$, es que $c_i = 0$.

En consecuencia, se demuestra que los vectores ortogonales $\vec{v}_1,\ \vec{v}_2,\ \cdots,\ \vec{v}_n$ **son linealmente independientes** porque su combinación lineal puede ser igual a $\vec{0}$ solo cuando $c_1,\ c_2,\ \cdots, c_n = 0$.

En ese sentido, como los **vectores ortonormales** son ortogonales, entonces también son **linealmente independientes** y el conjunto de ellos forma una **base** (*basis*).

2. Matrices Ortogonales.

Con una base ortonormal podemos formar una matriz que suele denotarse como Q, donde los primeros son los vectores columna de esta última.

La principal característica de una matriz Q, es que satisface la siguiente igualdad:

$$Q^TQ = I$$

Demostración. Sean $\vec{q}_1,\ \vec{q}_2,\ \cdots,\ \vec{q}_m$ vectores columna ortonormales de Q en $\mathbb{R}^n$. Entonces:

$$\begin{aligned} Q^TQ &= I \\ \begin{bmatrix} \vec{q}_1^T \\ \vec{q}_2^T \\ \vdots \\ \vec{q}_m^T \end{bmatrix} \cdot \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_m \end{bmatrix} &= I \\ \begin{bmatrix} (\vec{q}_1^T \cdot \vec{q}_1) & (\vec{q}_1^T \cdot \vec{q}_2) & \cdots & (\vec{q}_1^T \cdot \vec{q}_m) \\ (\vec{q}_2^T \cdot \vec{q}_1) & (\vec{q}_2^T \cdot \vec{q}_2) & \cdots & (\vec{q}_2^T \cdot \vec{q}_m) \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{q}_m^T \cdot \vec{q}_1) & (\vec{q}_m^T \cdot \vec{q}_2) & \cdots & (\vec{q}_m^T \cdot \vec{q}_m) \end{bmatrix} &= I \end{aligned}$$

Recordemos que:

$$\vec{q}_i^T \cdot \vec{q}_j = \begin{cases} 0 \text{ cuando } i \neq j \\ 1 \text{ cuando } i = j \end{cases}$$

donde $\vec{q}_i$ y $\vec{q}_j$ son ortonormales. Por lo tanto:

$$\begin{bmatrix} (\vec{q}_1^T \cdot \vec{q}_1) & (\vec{q}_1^T \cdot \vec{q}_2) & \cdots & (\vec{q}_1^T \cdot \vec{q}_m) \\ (\vec{q}_2^T \cdot \vec{q}_1) & (\vec{q}_2^T \cdot \vec{q}_2) & \cdots & (\vec{q}_2^T \cdot \vec{q}_m) \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{q}_m^T \cdot \vec{q}_1) & (\vec{q}_m^T \cdot \vec{q}_2) & \cdots & (\vec{q}_m^T \cdot \vec{q}_m) \end{bmatrix} = I$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

La diagonal de Q^TQ serán solo 1s y el resto de sus entradas, ceros. Así, $Q^TQ = I$ (Q.E.D).

Ahora bien, la igualdad $Q^TQ = I$ se cumple solo si en Q de $n \times m$, $n \geq m$. En cambio, si $n \leq m$, solo se satisface $QQ^T = I$. No obstante, cuando $n = m$, ocurre que:

$$Q^TQ = QQ^T = I$$

La igualdad de arriba es el **criterio de invertibilidad** que estudiamos en la Clase 3: Si una matriz A es **cuadrada e invertible**, se satisface que:

$$A^{-1}A = AA^{-1} = I$$

En otras palabras, cuando Q es cuadrada (i.e, $n = m$), se cumple que $Q^T = Q^{-1}$ y, por consiguiente, que:

$$Q^{-1}Q = QQ^{-1} = I$$

En ese momento, Q recibe el nombre de **Matriz Ortogonal**, la cual se caracteriza por:

- Tener vectores columna **ortonormales**.
- Ser cuadrada.
- Ser invertible, donde $Q^{-1} = Q^T$.

2.1 Ejemplos de Matrices Ortogonales.

Una **matriz permutación** cuadrada es un ejemplo de una matriz ortogonal. Como recordaremos, consiste de ceros y un 1 en cada columna, la cual se usa para reordenar las entradas de otra matriz al multiplicarla. A continuación tenemos un ejemplo:

Sea P la siguiente matriz permutación:

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Los vectores columna de P son ortogonales entre sí y todos de magnitud igual a 1. Ahora multipliquémosle su transpuesta.

$$P^T P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Otro ejemplo de una matriz ortogonal es la **Matriz de Rotación** R, la cual es utilizada para hacer rotar un objeto en un espacio euclidiano:

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Las columnas de R son ortogonales y unitarias. Además, cuando es multiplicada por su transpuesta, obtenemos la matriz identidad:

$$R^T R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

En ocasiones nos encontraremos con matrices de columnas ortogonales, pero que no son unitarias. Un ejemplo son las **Matrices Hadamard** H. Éstas son cuadradas, con entradas de valores 1 y -1 y sus columnas son perpendiculares entre sí. A continuación tenemos una de 4×4:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Veamos que todas las columnas de H son de magnitud 2. Por lo tanto, podemos normalizarlas muliplicando a esta matriz por $1/2$ (i.e, el recíproco de sus normas), pasando a ser **ortonormales** y, en consecuencia, $H = Q$.

$$\frac{1}{2}\cdot H=\frac{1}{2}\cdot\begin{bmatrix}1&1&1&1\\1&-1&1&-1\\1&1&-1&-1\\1&-1&-1&1\end{bmatrix}=\begin{bmatrix}0.5&0.5&0.5&0.5\\0.5&-0.5&0.5&-0.5\\0.5&0.5&-0.5&-0.5\\0.5&-0.5&-0.5&0.5\end{bmatrix}=Q$$

Al ser cuadrada, se cumplirá que $Q^TQ=I$.

2.2 Aplicación de las Matrices Ortogonales.

Las matrices ortogonales son útiles, por ejemplo, para **buscar una matriz de proyección**.

Como recordaremos de la Clase 15, la matriz de proyección P para buscar a un vector $\vec{p}\in C(A)$ que se proyecte sobre $\vec{b}\notin C(A)$, se obtiene a partir de la siguiente fórmula:

$$P=A(A^TA)^{-1}A^T$$

conllevando a que:

$$\vec{p}=P\vec{b}$$

En ciertas ocasiones podría ser más preferible trabajar con una matriz Q de vectores columna ortonormales y, por consiguiente, buscar a $\vec{p}\in C(Q)$ proyectado sobre $\vec{b}$:

$$P=Q(Q^TQ)^{-1}Q^T$$

Como las columnas de Q son ortonormales, entonces son linealmente independientes y podemos dar por asegurado al menos que $Q^TQ=I$:

$$P=Q(I)^{-1}Q^T=QIQ^T=QQ^T$$

En ese sentido, si Q es cuadrada, entonces:

$$P=I$$

Esta última igualdad hace sentido, puesto que si Q es cuadrada y sus columnas linealmente independientes[1], $\vec{b}$ estará sí o sí en $C(Q)$, de manera que:

$$\vec{p}=I\vec{b}=\vec{b}$$

[1]Y, por consiguiente, de rango completo.

Por lo tanto, para Q de $n \times m$ podemos resumir que:

$$P = \begin{cases} I \text{ si } n = m \\ QQ^T \text{ si } n > m \end{cases}$$

Las propiedades de P, $P^2 = P$ y $P^T = P$, también se aplican en QQ^T:

$$(QQ^T)^2 = (QQ^T) \cdot (QQ^T) = Q(Q^TQ)Q^T = QIQ^T = QQ^T$$

$$(QQ^T)^T = (Q^T)^T Q^T = QQ^T$$

También se hace más fácil resolver las **ecuaciones normales del método de mínimos cuadrados** usando matrices ortogonales.

En el método de mínimos cuadrados, como estudiamos en la Clase 16, las ecuaciones normales se obtienen como:

$$A^T A\hat{x} = A^T \vec{b}$$

donde $\hat{x}$ es el vector de las mejores soluciones del sistema $A\vec{x} = \vec{b}$ cuando no tiene soluciones.

Si, en vez de trabajar con una matriz A, lo hacemos con una Q de vectores columna ortonormales, entonces:

$$\begin{aligned} Q^T Q\hat{x} &= Q^T \vec{b} \\ I\hat{x} &= Q^T \vec{b} \\ \hat{x} &= Q^T \vec{b} \end{aligned}$$

Por lo tanto, sean $\hat{x}_i$ los componentes del vector $\hat{x}$, para $i = 1,\ 2,\ \cdots,\ n$; y $\vec{q}_j$ los vectores columnas ortogonales de Q, donde $j = 1,\ 2,\ \cdots,\ m$, los primeros corresponderán a:

$$\hat{x}_i = \vec{q}_i^{\,T} \cdot \vec{b}$$

3. Proceso de Gram-Schmidt.

Cuando tenemos un conjunto de **vectores linealmente independientes**, es posible **ortonormalizarlos** a través de un algoritmo llamado **Proceso de Gram-Schmidt**.

Digamos que el conjunto $X = \{\vec{x}_1,\ \vec{x}_2,\ \cdots,\ \vec{x}_m\}$ son vectores linealmente independientes; los de $U = \{\vec{u}_1,\ \vec{u}_2,\ \cdots,\ \vec{u}_m\}$ son ortogonales; y los de $W = \{\vec{w}_1,\ \vec{w}_2,\ \cdots,\ \vec{w}_m\}$, ortonormales.

La tarea del proceso de Gram-Schmidt es convertir a los vectores de X en los de U y luego en los de W, pero la condición es que los primeros sean linealmente independientes.

Paso 1. El primer paso del proceso de Gram-Schmidt, es establecer que

$$\vec{u}_1 = \vec{x}_1$$

y luego normalizamos a $\vec{u}_1$ para obtener a la versión ortonormalizada de $\vec{x}_1$, $\vec{w}_1$.

$$\vec{w}_1 = \frac{1}{||\vec{u}_1||} \cdot \vec{u}_1$$

Paso 2. Ahora debemos buscar a $\vec{w}_2$ a partir de $\vec{x}_2$, tal que $\vec{w}_1 \perp \vec{w}_2$. Para ello, definimos a $\vec{u}_2$ como la resta entre $\vec{x}_2$ y la proyección de este último sobre $\vec{u}_1$.

$$\vec{u}_2 = \vec{x}_2 - \left(\vec{u}_1 \cdot \frac{\vec{u}_1^T \vec{x}_2}{\vec{u}_1^T \vec{u}_1} \right)$$

El vector $\vec{u}_2$ es, básicamente, el vector error que estudiamos en la Clase 15.

Ahora, veamos que la proyección de $\vec{x}_1$ sobre $\vec{u}_1$ la podemos simplificar de la siguiente manera:

$$\vec{u}_1 \cdot \frac{\vec{u}_1^T \vec{x}_2}{\vec{u}_1^T \vec{u}_1} = \frac{\vec{u}_1 \cdot \vec{u}_1^T}{||\vec{u}_1||^2} \cdot \vec{x}_2 = \frac{\vec{u}_1}{||\vec{u}_1||} \cdot \frac{\vec{u}_1^T}{||\vec{u}_1||} \cdot \vec{x}_2$$

Recordemos que $(1/||\vec{u}_1||) \cdot \vec{u}_1 = \vec{w}_1$. Por lo tanto:

$$\vec{u}_1 \cdot \frac{\vec{u}_1^T \vec{x}_2}{\vec{u}_1^T \vec{u}_1} = \vec{w}_1 \cdot (\vec{w}_1^T \cdot \vec{x}_2)$$

De este modo, encontramos una forma más sencilla de calcular a $\vec{u}_2$:

$$\vec{u}_2 = \vec{x}_2 - (\vec{w}_1^T \cdot \vec{x}_2) \cdot \vec{w}_1$$

Posteriormente, normalizamos a $\vec{u}_2$.

$$\vec{w}_2 = \frac{1}{||\vec{u}_2||} \cdot \vec{u}_2$$

Paso 3. Continuamos con $\vec{w}_3$, donde $\vec{w}_3 \perp \vec{w}_1$ y $\vec{w}_3 \perp \vec{w}_2$. Primero buscamos a $\vec{u}_3$ como la resta entre $\vec{x}_3$ y las proyecciones de este vector sobre $\vec{u}_1$ y sobre $\vec{u}_3$.

$$\vec{u}_3 = \vec{x}_3 - (\vec{w}_1^T \cdot \vec{x}_3) \cdot \vec{w}_1 - (\vec{w}_2^T \cdot \vec{x}_3) \cdot \vec{w}_2$$

para después normalizarlo, resultando en $\vec{w}_3$:

$$\vec{w}_3 = \frac{1}{||\vec{u}_3||} \cdot \vec{u}_3$$

Así, podemos resumir el proceso de Gram-Schmidt para los m vectores de W como:

$$\begin{aligned}
\vec{u}_1 &= \vec{x}_1, & \vec{w}_1 &= \frac{1}{||\vec{u}_1||} \cdot \vec{u}_1 \\
\vec{u}_2 &= \vec{x}_2 - (\vec{w}_1^T \cdot \vec{x}_2) \cdot \vec{w}_1, & \vec{w}_2 &= \frac{1}{||\vec{u}_2||} \cdot \vec{u}_2 \\
\vec{u}_3 &= \vec{x}_3 - (\vec{w}_1^T \cdot \vec{x}_3) \cdot \vec{w}_1 - (\vec{w}_2^T \cdot \vec{x}_3) \cdot \vec{w}_2, & \vec{w}_3 &= \frac{1}{||\vec{u}_3||} \cdot \vec{u}_3 \\
\vec{u}_4 &= \vec{x}_4 - (\vec{w}_1^T \cdot \vec{x}_4) \cdot \vec{w}_1 - (\vec{w}_2^T \cdot \vec{x}_4) \cdot \vec{w}_2 - (\vec{w}_3^T \cdot \vec{x}_4) \cdot \vec{w}_3, & \vec{w}_4 &= \frac{1}{||\vec{u}_4||} \cdot \vec{u}_4 \\
&\vdots & &\vdots \\
\vec{u}_m &= \vec{x}_m - \sum_{i=1}^{m-1} (\vec{w}_i^T \cdot \vec{x}_m) \cdot \vec{w}_i, & \vec{w}_m &= \frac{1}{||\vec{u}_m||} \cdot \vec{u}_m
\end{aligned}$$

4. Descomposición $A = QR$.

Continuemos con los conjuntos de vectores X, U y W usados en la sección anterior.

Una consecuencia del Proceso de Gram-Schmidt, es que cada vector de X es una combinación lineal de los de W y viceversa. Esto significa[2] que:

$$\text{span}(X) = \text{span}(W)$$

Recordemos que:

$$\vec{w}_1 = \frac{1}{||\vec{u}_1||} \cdot \vec{u}_1$$

Si despejamos $\vec{u}_1$, entonces:

$$\vec{u}_1 = ||\vec{u}_1|| \cdot \vec{w}_1$$

Y como $\vec{u}_1 = \vec{x}_1$:

$$\vec{x}_1 = ||\vec{u}_1|| \cdot \vec{w}_1$$

[2] Anton y Rorres (2013). *Elementary Linear Algebra. Applications Version.* Pp. 200.

Del mismo modo, como $\vec{w}_2 = (1/||\vec{u}_2||) \cdot \vec{u}_2$, entonces:

$$\vec{u}_2 = ||\vec{u}_2|| \cdot \vec{w}_2$$

Si reemplazamos a $\vec{u}_2$ en la ecuación $\vec{u}_2 = \vec{x}_2 - (\vec{w}_1^T \cdot \vec{x}_2) \cdot \vec{w}_1$ y despejamos $\vec{x}_2$ en ella, entonces:

$$\vec{x}_2 = (\vec{w}_1^T \cdot \vec{x}_2) \cdot \vec{w}_1 + ||\vec{u}_2|| \cdot \vec{w}_2$$

En otras palabras:

$$\begin{aligned}
\vec{x}_1 &= ||\vec{u}_1|| \cdot \vec{w}_1 \\
\vec{x}_2 &= (\vec{w}_1^T \cdot \vec{x}_2) \cdot \vec{w}_1 + ||\vec{u}_2|| \cdot \vec{w}_2 \\
\vec{x}_3 &= (\vec{w}_1^T \cdot \vec{x}_3) \cdot \vec{w}_1 + (\vec{w}_2^T \cdot \vec{x}_3) \cdot \vec{w}_2 + ||\vec{u}_3|| \cdot \vec{w}_3 \\
&\vdots \\
\vec{x}_m &= \sum_{i=1}^{m-1} \left[(\vec{w}_i^T \cdot \vec{x}_m) \cdot \vec{w}_i \right] + ||\vec{u}_m|| \cdot \vec{w}_m
\end{aligned}$$

donde $(\vec{w}_i^T \cdot \vec{x}_m)$ y $||\vec{u}_m||$ son **escalares**.

Veamos esta idea en matrices. Digamos que A es una matriz de $n \times m$ con columnas linealmente independientes, o de rango$(A) = m$, y las ortonormalizamos usando el proceso de Gram-Schmidt, obteniendo a Q también de $n \times m$ y de rango$(Q) = m$.

Sea $A = [\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_m]$, sabemos que coincide que span$(\vec{a}_1, \ \vec{a}_2, \ \cdots, \ \vec{a}_m) = C(A)$. De igual modo, si $Q = [\vec{q}_1 \ \vec{q}_2 \ \cdots \ \vec{q}_m]$, entonces span$(\vec{q}_1, \ \vec{q}_2, \ \cdots, \ \vec{q}_m) = C(Q)$. Por lo tanto, a partir de lo que estudiamos anteriormente:

$$C(A) = C(Q)$$

Es decir, cada vector columna de A es una combinación lineal de los Q.

Supongamos que U es una matriz que consiste de las columnas ortogonalizadas de A. Que $C(A) = C(Q)$ significa que:

$$\begin{aligned}
\vec{a}_1 &= ||\vec{u}_1|| \cdot \vec{q}_1 \\
\vec{a}_2 &= (\vec{q}_1^T \cdot \vec{a}_2) \cdot \vec{q}_1 + ||\vec{u}_2|| \cdot \vec{q}_2 \\
&\vdots \\
\vec{a}_m &= (\vec{q}_1^T \cdot \vec{a}_m) \cdot \vec{q}_1 + (\vec{q}_2^T \cdot \vec{a}_m) \cdot \vec{q}_2 + \cdots + (\vec{q}_{m-1}^T \cdot \vec{a}_m) \cdot \vec{q}_{m-1} + ||\vec{u}_m|| \cdot \vec{q}_m
\end{aligned}$$

Las combinaciones lineales de la derecha de cada vector columna $\vec{a}$ son equivalentes a la multiplicación entre una matriz y las componentes de un vector:

$$
\begin{aligned}
\vec{a}_1 &= \begin{bmatrix} \vec{q}_1 \end{bmatrix} \cdot \begin{bmatrix} ||\vec{u}_1|| \end{bmatrix} \\
\vec{a}_2 &= \begin{bmatrix} \vec{q}_1 & \vec{q}_2 \end{bmatrix} \cdot \begin{bmatrix} \vec{q}_1^T \cdot \vec{a}_2 \\ ||\vec{u}_2|| \end{bmatrix} \\
&\vdots \\
\vec{a}_m &= \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_{m-1} & \vec{q}_m \end{bmatrix} \cdot \begin{bmatrix} \vec{q}_1^T \cdot \vec{a}_m \\ \vec{q}_2^T \cdot \vec{a}_m \\ \vdots \\ \vec{q}_{m-1}^T \cdot \vec{a}_m \\ ||\vec{u}_m|| \end{bmatrix}
\end{aligned}
$$

Los vectores del lado derecho los denotaremos como $\vec{r}_i$, para cada $i = 1,\ 2,\ \cdots,\ m$:

$$
\vec{r}_1 = \begin{bmatrix} ||\vec{u}_1|| \end{bmatrix} \quad \vec{r}_2 = \begin{bmatrix} \vec{q}_1^T \cdot \vec{a}_2 \\ ||\vec{u}_2|| \end{bmatrix} \quad \cdots \quad \vec{r}_m = \begin{bmatrix} \vec{q}_1^T \cdot \vec{a}_m \\ \vec{q}_2^T \cdot \vec{a}_m \\ \vdots \\ \vec{q}_{m-1}^T \cdot \vec{a}_m \\ ||\vec{u}_m|| \end{bmatrix}
$$

Reemplazando en las ecuaciones de arriba:

$$
\begin{aligned}
\vec{a}_1 &= \begin{bmatrix} \vec{q}_1 \end{bmatrix} \cdot \vec{r}_1 \\
\vec{a}_2 &= \begin{bmatrix} \vec{q}_1 & \vec{q}_2 \end{bmatrix} \cdot \vec{r}_2 \\
&\vdots \\
\vec{a}_m &= \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_{m-1} & \vec{q}_m \end{bmatrix} \cdot \vec{r}_m
\end{aligned}
$$

Los vectores $\vec{r}_i$ podemos considerarlos como las columnas de una matriz muy relevante que se denota como R:

$$
R = \begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \cdots & \vec{r}_m \end{bmatrix}
$$

La matriz R es relevante porque las combinaciones lineales del último conjunto de ecuaciones de arriba corresponden a la multiplicación entre Q y las columnas de R, que es lo mismo a calcular el producto QR y con ellas obtenemos a las columnas de A. Es decir, podemos factorizar a esta última matriz como:

$$A = QR$$

y a esta ecuación se la conoce como **Descomposición** $A = QR$.

En general, si tenemos una **matriz** A **de columnas linealmente independientes** (i.e, de rango columna completo), es posible factorizarla como el producto entre Q, una matriz con las columnas ortonormalizadas de A; y R, cuyas columnas contienen las constantes de las combinaciones lineales de Q.

4.1 Explicación de la Matriz R en $A = QR$.

La matriz R se caracteriza por ser **cuadrada**, **invertible** y **triangular superior**, donde:

$$R = \begin{bmatrix} ||\vec{u}_1|| & \vec{q}_1^T \cdot \vec{a}_2 & \cdots & \vec{q}_1^T \cdot \vec{a}_m \\ 0 & ||\vec{u}_2|| & \cdots & \vec{q}_2^T \cdot \vec{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & ||\vec{u}_m|| \end{bmatrix}$$

La diagonal principal de la matriz R está compuesta de las magnitudes de los vectores ortogonales $\vec{u}_i$, para $i = 1,\ 2,\ \cdots\ ,\ m$ y la parte inferior a ella solo consiste de **ceros**.

Ahora, recordemos que las columnas de Q son linealmente independientes al ser ortogonales, por lo tanto se cumple $Q^TQ = I$. En ese sentido, podemos obtener a la matriz R multiplicando a la ecuación $A = QR$ por Q^T:

$$\begin{aligned} A &= QR \\ Q^TA &= Q^TQR \\ Q^TA &= IR \\ Q^TA &= R \end{aligned}$$

Veamos que R será **cuadrada** porque Q y A son de igual dimensión. Si Q y A son de $n \times m$, entonces Q^T es de $m \times n$. Por lo tanto, $Q^TA = R$ será de $n \times n$.

Por otra parte, debido a que las columnas de A y de Q son linealmente independientes y a que Q^TA conlleva a una matriz cuadrada, la matriz R siempre será **invertible**. Si lo evaluamos por su rango, será de **rango completo**.

Además, la multiplicación Q^TA nos ayuda a entender por qué las entradas que están por debajo de **la diagonal de** R **son iguales a cero**.

$$Q^T A = R$$

$$\begin{bmatrix} \vec{q}_1^T \\ \vec{q}_2^T \\ \vdots \\ \vec{q}_m^T \end{bmatrix} \cdot \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_m \end{bmatrix} = R$$

$$\begin{bmatrix} \vec{q}_1^T \cdot \vec{a}_1 & \vec{q}_1^T \cdot \vec{a}_2 & \cdots & \vec{q}_1^T \cdot \vec{a}_m \\ \vec{q}_2^T \cdot \vec{a}_1 & \vec{q}_2^T \cdot \vec{a}_2 & \cdots & \vec{q}_2^T \cdot \vec{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \vec{q}_m^T \cdot \vec{a}_1 & \vec{q}_m^T \cdot \vec{a}_2 & \cdots & \vec{q}_m^T \cdot \vec{a}_m \end{bmatrix}$$

Todos los productos punto que están debajo de la diagonal principal de R son entre **vectores ortogonales**. Por ejemplo, $\vec{q}_2$ es la versión ortonormal de $\vec{u}_2$ la que, a su vez, es ortogonal a $\vec{a}_1$. Aquello explica por qué $\vec{q}_2^T \cdot \vec{a}_1 = 0$ y por el mismo motivo los demás productos puntos que están en esa zona, son iguales a ese valor.

4.2 Aplicación en el Método de Mínimos Cuadrados.

La descomposición $A = QR$ también podemos usarla para resolver las ecuaciones normales del método de mínimos cuadrados:

$$A^T A\hat{x} = A^T \vec{b}$$

Recordemos que el método de mínimos cuadrados trata con el problema de que en A de $n \times m$, $n > m$. Por lo tanto existe la posibilidad de que A sea de rango$(A) = m$. De ser el caso, podemos factorizarla como $A = QR$ y podemos reemplazar el producto QR en las ecuaciones normales:

$$\begin{aligned} (QR)^T \cdot QR\hat{x} &= (QR)^T \vec{b} \\ (R^T Q^T QR)\hat{x} &= R^T Q^T \vec{b} \\ R^T I R \; \hat{x} &= R^T Q^T \vec{b} \\ R^T R\hat{x} &= R^T Q^T \vec{b} \end{aligned}$$

Las ecuaciones normales podemos trabajarlas de esa manera o es posible multiplicar esta ecuación por $(R^T)^{-1}$, ya que al ser R cuadrada e invertible, entonces su transpuesta también tiene una inversa, donde $(R^T)^{-1} = (R^{-1})^T$.

$$\begin{aligned}((R^T)^{-1}R^T)R\hat{x} &= ((R^T)^{-1}R^T)Q^T\vec{b} \\ IR\hat{x} &= IQ^T\vec{b} \\ R\hat{x} &= Q^T\vec{b}\end{aligned}$$

No obstante, podemos ir más allá multiplicando a esta última por R^{-1}:

$$\begin{aligned}R^{-1}R\hat{x} &= R^{-1}Q^T\vec{b} \\ I\hat{x} &= R^{-1}Q^T\vec{b} \\ \hat{x} &= R^{-1}Q^T\vec{b}\end{aligned}$$

La principal ventaja de trabajar con $A = QR$, es que toma menos tiempo resolver las ecuaciones normales con R al ser triangular superior que al trabajar solo con $A^TA\hat{x} = A^T\vec{b}$.