Métodos estadísticos y algebraicos

Tratamiento Avanzado de Señal en Comunicaciones

Curso 2009-2010

1: Multiplicaciones matriciales

Si $b = Ax$, entonces b es una combinación lineal de las columnas de A.

- Sea x un vector columna n-dimensional.
- Sea A una matriz $m \times n$ (m filas, n columnas).
- $b = Ax$ es el vector columna m-dimensional

$$b_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m, \qquad (1)$$

- donde b_i denota la i-ésima componente de b,
- a_{ij} denota la componente i, j de A (fila i-ésima, columna j-ésima),
- y x_j denota la componente j-ésima de x.

Mapeado lineal

- El mapeado $x \mapsto Ax$ es lineal, lo que significa que para cualesquiera $x, y \in \mathbb{C}^n$, y cualquier $\alpha \in \mathbb{C}$, se verifica

$$A(x + y) = Ax + Ay,$$

$$A(\alpha x) = \alpha Ax.$$

- La afirmación inversa también es cierta: todo mapa lineal de $\mathbb{C}^n$ sobre $\mathbb{C}^m$ puede expresarse como una multiplicación por una matriz $m \times n$.

Mult. de una matriz por un vector

- Si a_j es la columna j-ésima de la matriz A

$$b = Ax = \sum_{j=1}^{n} x_j a_j. \qquad (2)$$

- Ejemplo

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} ae + bf \\ ce + df \end{bmatrix} = e \begin{bmatrix} a \\ c \end{bmatrix} + f \begin{bmatrix} b \\ d \end{bmatrix}$$

- Solemos pensar en $Ax = b$ como que A actúa sobre x para producir b.
- La fórmula (2) sugiere la interpretación de que x actúa sobre A para producir b.

Multiplicación de dos matrices

- $B = AC$: cada columna de B es una combinación lineal de las columnas de A.
- Si A es una matriz $l \times m$ y C es $m \times n$, entonces B es $l \times n$, con componentes definidas por

$$b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj},$$

donde b_{ij}, a_{ij}, y c_{kj} son las componentes de B, A y C respectivamente.

Multiplicación de dos matrices

- Ejemplo

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \left[e \begin{bmatrix} a \\ c \end{bmatrix} + g \begin{bmatrix} b \\ d \end{bmatrix}, \quad f \begin{bmatrix} a \\ c \end{bmatrix} + h \begin{bmatrix} b \\ d \end{bmatrix} \right]$$

- La ecuación $b = Ax$ se convierte en

$$b_j = (Ac)_j = \sum_{k=1}^{m} c_{kj} a_k. \qquad (3)$$

- Es decir, b_j es una combinación lineal de las columnas a_k con coeficientes c_{kj}.

Alcance de una matriz

El *alcance* de una matriz A:

- es el conjunto de vectores que pueden expresarse como Ax para algún x.
- se denota mediante range(A).
- según (2), es el espacio generado por las columnas de A.
- también se denomina el *espacio de columnas* de A.

Espacio nulo de una matriz

El espacio nulo de $A \in \mathbb{C}^{m \times n}$:

- es el conjunto de vectores x que satisfacen $Ax = 0$, donde 0 es el vector nulo de $\mathbb{C}^m$.
- se denota mediante $\text{null}(A)$.

Las componentes de cada vector $x \in \text{null}(A)$:

- proporcionan los coeficientes de una expansión del vector nulo como una combinación lineal de las columnas de A:
- $0 = x_1 a_1 + \cdots + x_n a_n$.

Rango de una matriz

- El *rango de columnas* de una matriz es la dimensión de su espacio de columnas.
- El rango de filas de una matriz es la dimensión del espacio generado por sus filas.
- Siempre coinciden. Denotaremos esta cantidad como el rango de una matriz.
- Una matriz $\mathrm{m} \times \mathrm{n}$ es de *rango completo* si tiene el mayor rango posible (el mínimo entre m y n).
- Una matriz de rango completo con $\mathrm{m} \geq \mathrm{n}$ debe tener n columnas linealmente independientes.
- Una matriz $A \in \mathbb{C}^{\mathrm{m} \times \mathrm{n}}$ con $\mathrm{m} \geq \mathrm{n}$ tiene rango completo si y sólo si no mapea dos vectores distintos sobre el mismo vector.

Inversa de una matriz

- Una matriz no singular, o invertible, es una matriz cuadrada de rango completo.
- Las m columnas de una matriz no singular A de $m \times m$ forman una base de $\mathbb{C}^{m \times m}$.
- Podemos expresar unívocamente cualquier vector como una combinación lineal de aquellos.
- En particular, los vectores unitarios canónicos (con una componente 1 en la posición j-ésima y ceros en todas las demás) denotados e_j, pueden expandirse como

$$e_j = \sum_{i=1}^{m} z_{ij} a_i. \tag{4}$$

Inversa de una matriz

- Sea Z una matriz de componentes z_{ij}, y denotemos por z_j la columna j-ésima de Z.
- Entonces (4) puede escribirse $e_j = Az_j$.
- Esta ecuación tiene la forma de (3); puede escribirse de forma más concisa como

$$[e_1|\cdots|e_m] = I = AZ,$$

- donde I es la matriz $m \times m$ conocida como *identidad*.
- La matriz Z es la inversa de A.
- Cualquier matriz cuadrada no singular tiene una inversa única, denotada A^{-1}, que satisface $AA^{-1} = A^{-1}A = I$.

Propiedades de matrices invertibles

Para una $A \in \mathbb{C}^{m \times m}$, las siguientes condiciones son equivalentes:

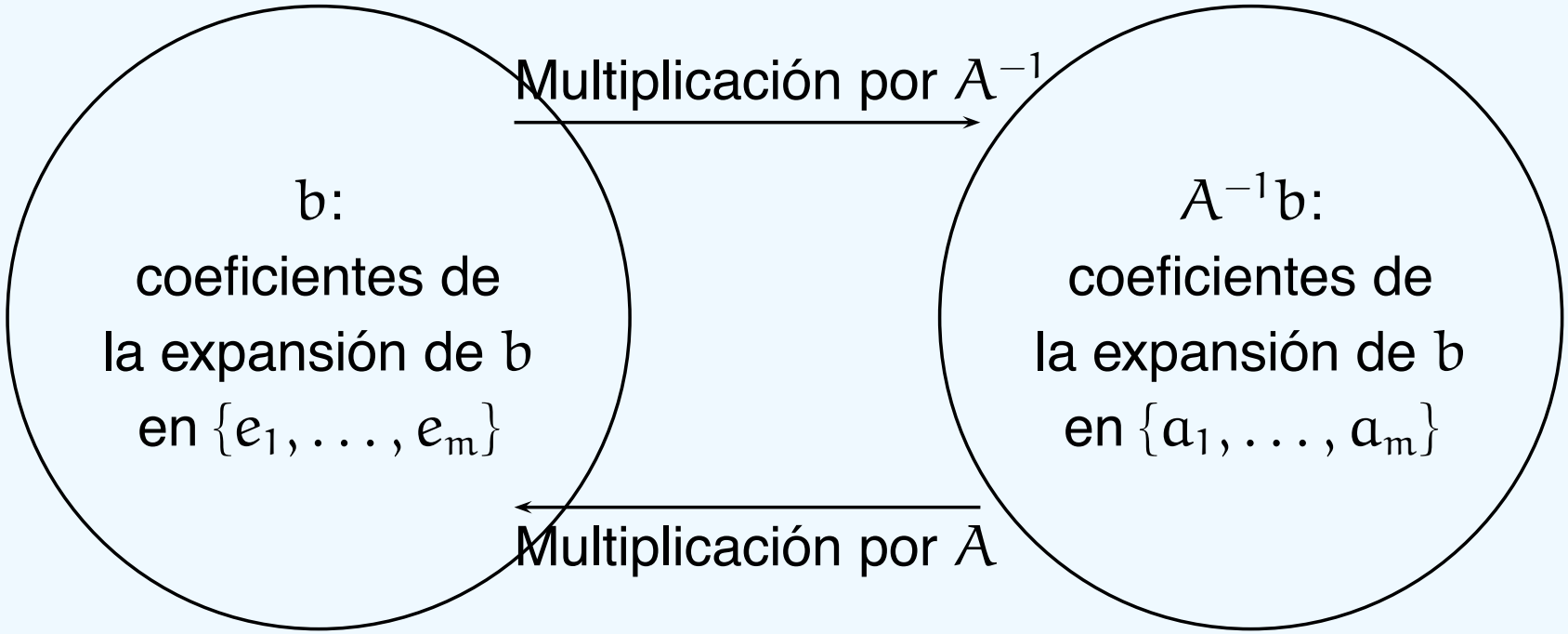
- A tiene una inversa A^{-1},
- $\mathrm{rank}(A) = m$,
- $\mathrm{range}(A) = \mathbb{C}^m$,
- $\mathrm{null}(A) = \{0\}$,
- 0 no es un autovalor de A,
- 0 no es un valor singular de A,
- $\det(A) \neq 0$.

Mult. de una matriz inversa por un vector

- Cuando escribimos el producto $\mathrm{x} = \mathrm{A}^{-1}\mathrm{b}$, es importante no permitir que la notación de la matriz inversa nos oculte qué es lo que está ocurriendo realmente.
- En vez de pensar en x como el resultado de aplicar A^{-1} sobre b, deberíamos entender que x es el único vector que satisface la ecuación $\mathrm{A}\mathrm{x} = \mathrm{b}$.
- Según (2), esto significa que x es el vector de coeficientes de la única expansión lineal de b en la base de las columnas de A; es decir,
- $\mathrm{A}^{-1}\mathrm{b}$ es el vector de coeficientes de la expansión de b en la base formada por las columnas de A.

Mult. de una matriz inversa por un vector

- La multiplicación por A^{-1} es una operación de *cambio de base*:



2: vectores y matrices ortogonales

- La mayor parte de los mejores algoritmos actuales de álgebra lineal numérica se basan de una forma u otra en la ortogonalidad.
- En esta sección presentamos sus ingredientes: vectores ortogonales y matrices ortogonales (unitarias).

Adjunto de una matriz

- La *hermítica conjugada*, o *adjunta* de una matriz $m \times n$ A, denotada A^*, es la matriz $n \times m$ cuya componente i, j es el complejo conjugado de la componente j, i de A:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \longrightarrow A^* = \begin{bmatrix} a_{11}^* & a_{21}^* & a_{31}^* \\ a_{12}^* & a_{22}^* & a_{32}^* \end{bmatrix}.$$

- Si $A = A^*$, A es *hermítica*.
- Una matriz hermítica debe ser cuadrada.

Transpuesta de una matriz

- Para una A real, el adjunto simplemente intercambia las filas y columnas de A.
- En este caso, el adjunto se conoce también como la *transpuesta*, y se denota por A^T.
- Si una matriz real es hermítica, es decir, $A = A^T$, se dice también que es *simétrica*.

Producto interno

- El *producto interno* de dos vectores columna $x, y \in \mathbb{C}^m$ es el producto del adjunto de x por y:

$$x^* y = \sum_{i=1}^{m} x_i^* y_i.$$

- La longitud euclídea de x puede escribirse como $||x||$, y puede definirse como la raíz cuadrada del producto interno de x consigo mismo:

$$||x|| = \sqrt{x^* x} = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2}.$$

Producto interno

- El coseno del ángulo α entre x e y puede expresarse también en términos del producto interno:

$$\cos\alpha = \frac{x^* y}{\|x\|\,\|y\|}.$$

- El producto interno es *bilineal*, que significa que es lineal en cada vector por separado:

$$(x_1 + x_2)^* y = x_1^* y + x_2^* y,$$
$$x^*(y_1 + y_2) = x^* y_1 + x^* y_2,$$
$$(\alpha x)^*(\beta y) = \alpha^* \beta x^* y.$$

Producto interno

- Utilizaremos frecuentemente la propiedad de que para cualesquiera matrices o vectores A y B de dimensiones compatibles,

$$(AB)^* = B^* A^*. \tag{5}$$

- Este resultado es análogo a la fórmula igualmente importante para productos de matrices cuadradas invertibles,

$$(AB)^{-1} = B^{-1} A^{-1}.$$

- La notación A^{-*} es una forma abreviada de escribir $(A^*)^{-1}$ o $(A^{-1})^*$, que son iguales, como puede demostrarse aplicando (5) con $B = A^{-1}$.

Vectores ortogonales

- Dos vectores x e y son ortogonales si $x^* y = 0$.
- Si x e y son reales: ángulo recto en $\mathbb{R}^m$.
- Dos conjuntos X e Y son ortogonales si todo $x \in X$ es ortogonal a todo $y \in Y$.
- Un conjunto de vectores no nulos S es *ortogonal* si sus elementos son ortogonales dos a dos; es decir, si para todo $x, y \in S$, $x \neq y \rightarrow x^* y = 0$.
- Un conjunto de vectores es *ortonormal* si es ortogonal y, además, todo $x \in S$ tiene $||x|| = 1$.
- Los vectores en un conjunto ortogonal S son linealmente independientes.
- Si un conjunto ortogonal $S \subseteq \mathbb{C}^m$ contiene m vectores, entonces constituye una base de $\mathbb{C}^m$.

Componentes de un vector

- El producto interno puede utilizarse para descomponer vectores arbitrarios en componentes ortogonales.
- Supongamos que $\{q_1, \ldots, q_n\}$ es un conjunto ortonormal, y sea v un vector arbitrario.
- La cantidad $q_j^* v$ es un escalar.
- Utilizando estos escalares como coordenadas de una expansión, el vector

$$r = v - (q_1^* v) q_1 - (q_2^* v) q_2 - \cdots - (q_n^* v) q_n$$

es ortogonal a $\{q_1, \ldots, q_n\}$.

Componentes de un vector

- Podemos comprobar esta afirmación calculando

$$q_i^* r = q_i^* v - (q_1^* v)(q_i^* q_1) - \cdots - (q_n^* v)(q_i^* q_n).$$

- Esta suma colapsa, ya que $q_i^* q_j = 0$ para $i \neq j$:

$$q_i^* r = q_i^* v - (q_i^* v)(q_i^* q_i) = 0.$$

- Por lo tanto, v puede descomponerse en $n+1$ componentes ortogonales:

$$v = r + \sum_{i=1}^{n} (q_i^* v) q_i = r + \sum_{i=1}^{n} (q_i q_i^*) v. \qquad (6)$$

Componentes de un vector

- En esta descomposición, r es la parte de v ortogonal al conjunto de vectores $\{q_1, \ldots, q_n\}$, o equivalentemente, al subespacio generado por este conjunto de vectores;
- y $(q_i^* v) q_i$ es la parte de v en la dirección de q_i.
- Si $\{q_i\}$ constituye una base de $\mathbb{C}^m$, entonces n debe ser igual a m, y r debe ser el vector nulo, de tal forma que v esté completamente descompuesto en m componentes ortogonales según las direcciones de q_i:

$$v = \sum_{i=1}^{m} (q_i^* v) q_i = \sum_{i=1}^{m} (q_i q_i^*) v. \qquad (7)$$

Componentes de un vector

- Tanto en (6) como en (7), hemos escrito la expresión de dos formas distintas, una con $(q_i^* \nu) q_i$ y otra con $(q_i q_i^*) \nu$.
- Estas expresiones son iguales, pero tienen interpretaciones distintas.
- En el primer caso, vemos ν como una suma de coeficientes $(q_i^* \nu)$ veces los vectores q_i.
- En la segunda, vemos ν como una suma de componentes ortogonales de ν sobre las direcciones de q_i.
- La operación de la proyección i-ésima se lleva a cabo por la especialísima matriz de rango uno $q_i q_i^*$.

Matrices unitarias

- Una matriz cuadrada $Q \in \mathbb{C}^{m \times m}$ es unitaria (*ortogonal* en el caso real) si $Q^* = Q^{-1}$; es decir, si $Q^* Q = I$.
- En términos de las columnas de Q,

$$\left[\begin{array}{c} q_1^* \\ \hline q_2^* \\ \hline \vdots \\ \hline q_m^* \end{array}\right] \left[\begin{array}{c|c|c|c} & & & \\ q_1 & q_2 & \cdots & q_m \\ & & & \end{array}\right] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

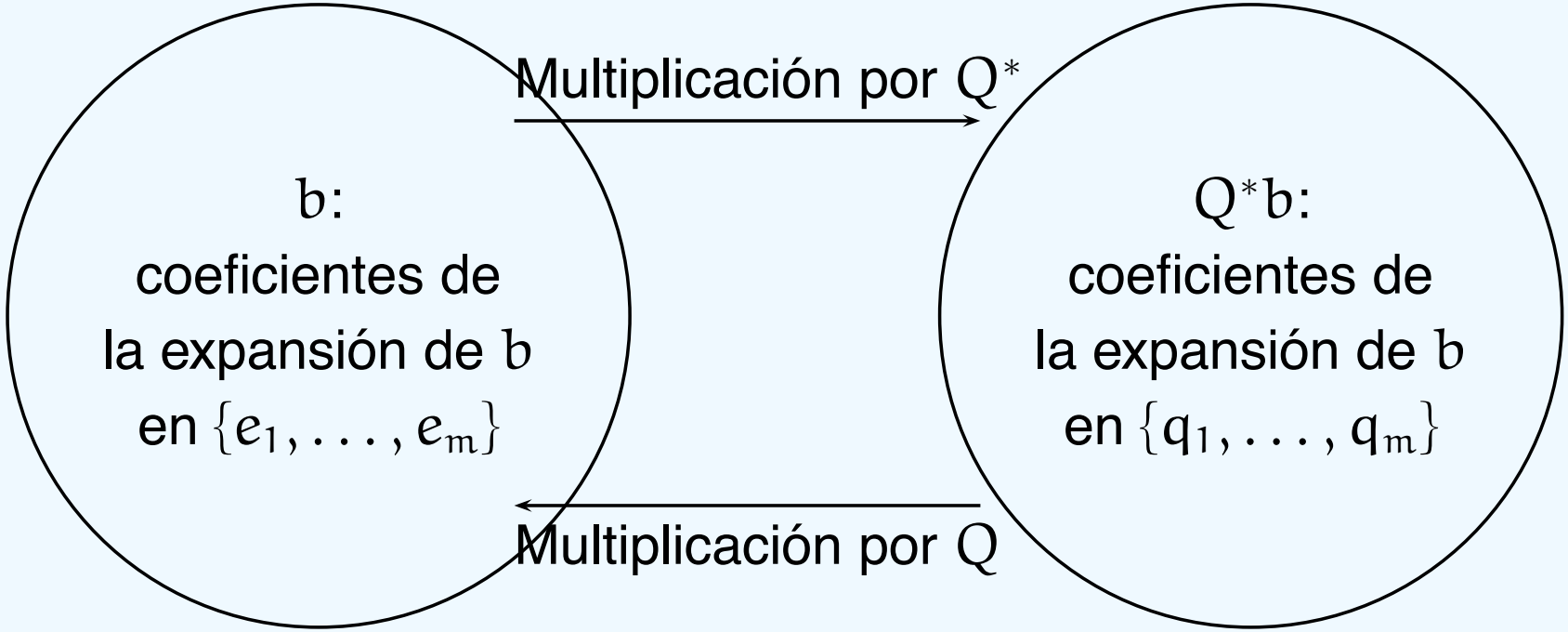
- En otras palabras, $q_i^* q_j = \delta_{ij}$, y las columnas de la matriz unitaria Q forman una base ortonormal de $\mathbb{C}^m$. El símbolo δ_{ij} es la *delta de Kronecker*.

Multiplicación por una matriz unitaria

- Antes comentamos la interpretación de los productos matriz-vector Ax y $A^{-1}b$.
- Si A es una matriz unitaria Q, estos productos se convierten en Qx y $Q^{*}b$, y por supuesto las mismas interpretaciones siguen siendo válidas.
- Como antes, Qx es la combinación lineal de columnas de Q con coeficientes x.
- Análogamente, $Q^{*}b$ es el vector de coeficientes de la expansión de b en la base de las columnas de Q.

Multiplicación por una matriz unitaria

- Esquemáticamente, la situación es la siguiente:



Multiplicación por una matriz unitaria

- La multiplicación por una matriz unitaria o su adjunto conserva la estructura geométrica en el sentido euclídeo (los productos internos).
- Es decir, a partir de (5), para una Q unitaria, $(Qx)^*(Qy) = x^*y$.
- La invarianza de los productos internos significa que los ángulos entre los vectores se mantienen, y también sus longitudes:

$$||Qx|| = ||x||.$$

- En el caso real, la multiplicación por una matriz ortogonal corresponde a una rotación rígida (si el determinante de Q vale uno) o una reflexión (si $\det(Q) = -1$) del espacio vectorial.

3: normas

- Las nociones fundamentales de tamaño y distancia en un espacio vectorial están asociadas al concepto de normas.
- Son los criterios con las que mediremos aproximaciones y convergencias a lo largo de todo el curso.

Normas vectoriales

- Una *norma* es una función $||\cdot|| : \mathbb{C}^m \rightarrow \mathbb{R}$ que asigna una longitud real a cada vector.
- Para concordar con una noción razonable de longitud, una norma debe satisfacer las siguientes tres condiciones para todos los vectores x e y y para todos los escalares $\alpha \in \mathbb{C}$,

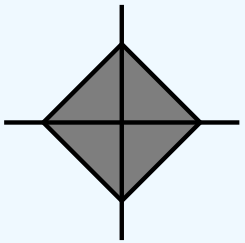
$$||x|| \geq 0, \text{ y } ||x|| = 0 \text{ sólo si } x = 0,$$
$$||x + y|| \leq ||x|| + ||y||,$$
$$||\alpha x|| = |\alpha| ||x||.$$

Normas vectoriales

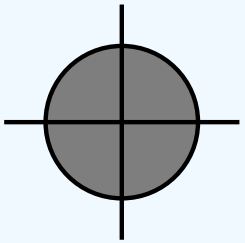
- A continuación definimos la clase más importante de normas vectoriales, las normas-p.
- La bola cerrada unitaria $\{x \in \mathbb{C}^m : ||x|| = 1\}$ correspondiente a cada norma se ilustra mediante la correspondiente figura a la derecha de cada ecuación para el caso $m = 2$.

Normas-p

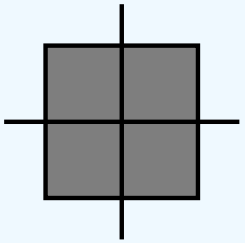
$$||x||_1 = \sum_{i=1}^{m} |x_i|,$$



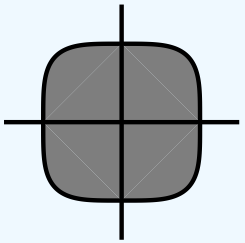
$$||x||_2 = \left(\sum_{i=1}^{m} |x_i|^2\right)^{1/2} = \sqrt{x^* x},$$



$$||x||_\infty = \max_{1 \le i \le m} |x_i|,$$



$$||x||_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p}, \quad (1 \le p < \infty).$$



Normas-p

- La norma-2 es la función longitud Euclídea; su bola unitaria es esférica.
- La norma-1 se utiliza por las líneas aéreas para definir el tamaño máximo permitido para el equipaje de mano.
- La plaza Sergel en Estocolmo, Suecia, tiene la forma de la bola unitaria en la norma-4.
- El poeta danés Piet Hein popularizó esta "superelipse" como una forma agradable para algunos objetos tales como mesas de conferencias.

Normas-p ponderadas

- Después de las normas-p, las normas más útiles son las normas-p ponderadas, donde a cada una de las coordenadas de un espacio vectorial se le da su propio peso.
- En general, dada cualquier norma $||\cdot||$, una norma ponderada puede escribirse como

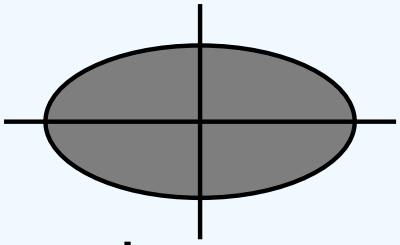
$$||x||_W = ||Wx||,$$

donde W es la matriz diagonal en la que la componente i-ésima de la diagonal es el peso $w_i \neq 0$.

Normas-p ponderadas

- Por ejemplo, una norma-2 ponderada $\|\cdot\|_W$ sobre $\mathbb{C}^m$ viene dada por:

$$\|x\|_W = \left(\sum_{i=1}^{m} |w_i x_i|^2\right)^{1/2},$$



- También puede generalizarse la idea de normas ponderadas permitiendo que W sea una matriz no singular arbitraria, no necesariamente diagonal.
- Las normas más importantes en este curso son la norma-2 no ponderada y su norma matricial inducida.

N. matriciales inducidas por n. vectoriales

- Una matriz $\mathrm{m} \times \mathrm{n}$ puede verse como un vector en un espacio mn-dimensional: cada una de las mn componentes de la matriz es una coordenada independiente.
- Toda norma mn-dimensional puede utilizarse para medir el "tamaño" de dicha matriz.
- Sin embargo, son más útiles las *normas matriciales inducidas*,
- que se defininen en términos del comportamiento de una matriz considerada como un operador entre espacios normados dominio y alcance.

N. matriciales inducidas por n. vectoriales

- Dadas las normas vectoriales $||\cdot||_{(n)}$ y $||\cdot||_{(m)}$ sobre el dominio y alcance de $A \in \mathbb{C}^{m \times n}$,
- la norma matricial inducida $||A||_{(m,n)}$ es el menor número C para el que se verifica la siguiente desigualdad para todo $x \in \mathbb{C}^m$:

$$||Ax||_{(m)} \leq C||x||_{(n)}.$$

- Es decir, $||A||_{(m,n)}$ es el supremo de los cocientes $||Ax||_{(m)}/||x||_{(n)}$ sobre todos los vectores $x \in \mathbb{C}^m$ (el máximo factor por el que A puede "estirar" un vector x).
- Decimos que $||\cdot||_{(m,n)}$ es la norma matricial inducida por $||\cdot||_{(m)}$ y $||\cdot||_{(n)}$.

N. matriciales inducidas por n. vectoriales

- Debido a la condición $||\alpha x|| = |\alpha|||x||$, la acción de A viene determinada por su acción sobre los vectores unitarios.
- Por lo tanto, la norma matricial puede definirse equivalentemente en términos de las imágenes de los vectores unitarios bajo A:

$$||A||_{(m,n)} = \sup_{x \in \mathbb{C}^n, x \neq 0} \frac{||Ax||_{(m)}}{||x||_{(n)}} = \sup_{x \in \mathbb{C}^n, ||x||_{(n)}=1} ||Ax||_{(m)}.$$

- Esta forma de la definición puede resultar conveniente para visualizar normas matriciales inducidas, como en los dibujos anteriores.

Ejemplo

- La matriz

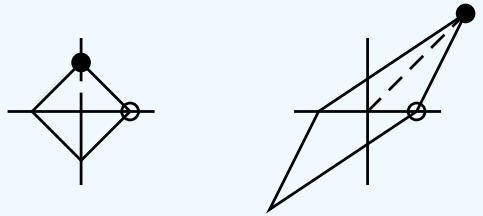
$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

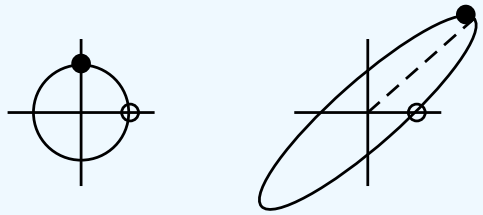
  mapea $\mathbb{C}^2$ sobre $\mathbb{C}^2$.

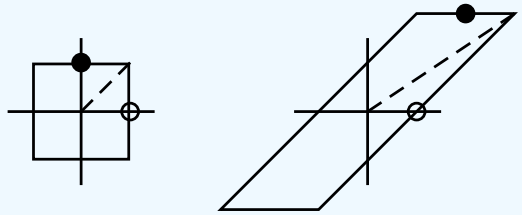
- También mapea $\mathbb{R}^2$ sobre $\mathbb{R}^2$, lo que es más conveniente si queremos hacer dibujos y también es suficiente para determinar normas-p matriciales, ya que los coeficientes de A son reales.

Ejemplo (continuación)

A la izquierda, las bolas unitarias de $\mathbb{R}^2$ con respecto a $||\cdot||_1$, $||\cdot||_2$, y $||\cdot||_\infty$. A la derecha, sus imágenes a través de la matriz A. Las linéas discontinuas marcan los vectores que más son amplificados por

Norma 1:  $||A||_1 = 4,$

Norma 2 :  $||A||_2 \approx 2{,}9208,$

Norma ∞:  $||A||_\infty = 3.$

A en cada norma.

Normas matriciales genéricas

- Como hemos indicado antes, las normas matriciales no tienen porqué ser inducidas por normas vectoriales.
- En general, una norma matricial debe satisfacer simplemente las tres condiciones sobre normas vectoriales aplicadas en el espacio vectorial mn-dimensional de matrices

$$||A|| \geq 0, \text{ y } ||A|| = 0 \text{ sólo si } A = 0,$$
$$||A + B|| \leq ||A|| + ||B||,$$
$$||\alpha A|| = |\alpha| ||A||.$$

Norma de Frobenius

- La norma matricial más importante que no está inducida por una norma vectorial es la *norma de Hilbert-Schmidt* o *norma de Frobenius*,

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

- Es la misma que la norma-2 de la matriz cuando se considera como un vector mn-dimensional.
- La fórmula para la norma de Frobenius puede escribirse también en términos de las filas o columnas individuales.

Norma de Frobenius

- Por ejemplo, si a_j es la columna j-ésima de A,

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^2 \right)^{1/2}.$$

- Esta identidad, así como el resultado análogo basado en las filas en lugar de las columnas, puede expresarse de forma compacta

$$\|A\|_F = \sqrt{\mathrm{tr}(A^* A)} = \sqrt{\mathrm{tr}(A A^*)},$$

donde $\mathrm{tr}(B)$ denota la *traza* de B; es decir, la suma de las componentes de su diagonal.

Invarianza bajo multiplicación unitaria

- La norma-2 matricial, al igual que la norma-2 vectorial, es invariante bajo multiplicaciones por matrices unitarias.
- La misma propiedad se verifica también para la norma de Frobenius.
- Para cualquier $A \in \mathbb{C}^{m \times n}$ y $Q \in \mathbb{C}^{m \times m}$ unitaria,

$$\|QA\|_2 = \|A\|_2, \qquad \|QA\|_F = \|A\|_F.$$

- También es cierto si Q se generaliza a una matriz rectangular con columnas ortonormales, es decir, $Q \in \mathbb{C}^{p \times m}$, con $p > m$.
- Existen identidades análogas si multiplicamos por la derecha por matrices unitarias o por matrices rectangulares con filas ortonormales.

4: Descomposición en valores singulares

- La descomposición en valores singulares (SVD) es una factorización matricial cuyo cálculo es un paso de muchos algoritmos.
- Igualmente importante es la utilización conceptual de la SVD.
- Muchos problemas de álgebra lineal pueden responderse mejor si primero nos hacemos la pregunta: ¿qué ocurre si realizamos la SVD?

Una observación geométrica

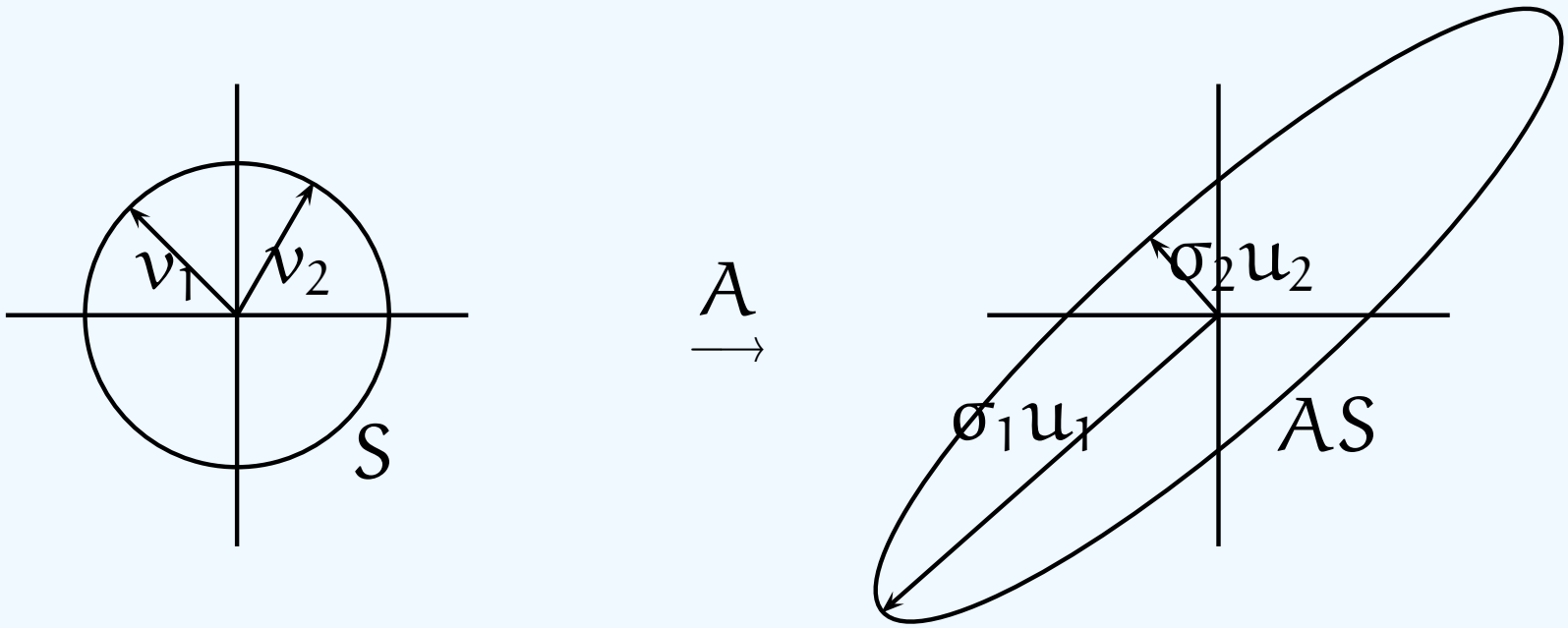
- La SVD está motivada por el siguiente hecho geométrico:
- La imagen de la esfera unitaria por cualquier matrix $\mathrm{m} \times \mathrm{n}$ es una hiperelipse.
- La SVD se puede aplicar tanto a matrices reales como complejas.
- Sin embargo, al describir la interpretación geométrica, supondremos que la matriz es real.

Hiperelipses

- Una hiperelipse es la generalización m-dimensional de una elipse.
- Podemos definir una hiperelipse en $\mathbb{R}^m$ como la superficie que se obtiene al estirar la esfera unitaria de $\mathbb{R}^m$ por factores $\sigma_1, \ldots, \sigma_m$ (que pueden ser cero) según direcciones ortogonales $u_1, \ldots, u_m \in \mathbb{R}^m$.
- Por conveniencia, consideraremos que los u_i son vectores unitarios; es decir, $||u_i||_2 = 1$. Los vectores $\{\sigma_i u_i\}$ son los *semiejes principales* de la hiperelipse, con longitudes $\sigma_1, \ldots, \sigma_m$. Si A tiene rango r, exactamente r de las longitudes σ_i serán no nulas, y en particular, si $m \geq n$, como mucho n de ellos serán no nulos.

Interpretación geométrica

- Por esfera unitaria entendemos la esfera Euclídea convencional en el espacio n-dimensional, es decir, la esfera unitaria según la norma-2; que denotaremos como S.
- Entonces AS, la imagen de S bajo el mapeado A es una hiperelipse.



Definiciones

- Sea S la esfera unitaria en $\mathbb{R}^n$, y tomemos cualquier $A \in \mathbb{R}^{m \times n}$ con $m \geq n$.
- Por simplicidad, supondremos por el momento que A tiene rango completo n.
- La imagen AS es una hiperelipse en $\mathbb{R}^m$.
- Definiremos ahora algunas propiedades de A en términos de la forma de AS.

Definiciones

- Los n *valores singulares* de A son las longitudes de los n semiejes principales de AS, denotados $\sigma_1, \sigma_2, \ldots, \sigma_n$ (numerados en orden descendente, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$).
- Los n *vectores singulares por la izquierda* de A son los vectores unitarios $\{u_1, u_2, \ldots, u_n\}$ orientados según las direcciones de los semiejes principales de AS, numerados según sus valores singulares correspondientes. El vector $\sigma_i u_i$ es el i-ésimo mayor semieje principal de AS.
- Los n *vectores singulares por la derecha* de A son los vectores unitarios $\{v_1, v_2, \ldots, v_n\} \in S$ que son las preimágenes de los semiejes principales de AS, numerados de forma que $Av_j = \sigma_j u_j$.

SVD reducida

- La relación entre los vectores singulares por la derecha $\{v_j\}$ y los vectores singulares por la izquierda $\{u_j\}$ es

$$Av_j = \sigma_j u_j, \quad 1 \leq j \leq n.$$

- Esta colección de ecuaciones vectoriales puede expresarse como una ecuación matricial,

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \left[\begin{array}{c|c|c|c} v_1 & v_2 & \cdots & v_n \end{array}\right] = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \cdots & u_n \end{array}\right]$$

SVD reducida

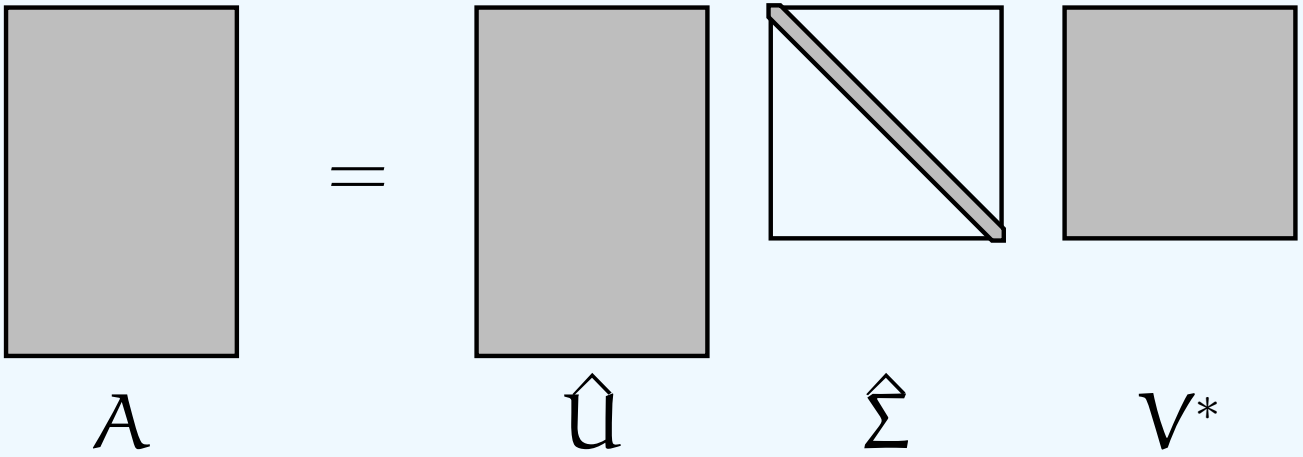
- De forma más compacta, $A V = \hat{U} \hat{\Sigma}$.
- En esta ecuación matricial, $\hat{\Sigma}$ es una matriz diagonal $n \times n$ con componentes reales positivas (A tenía rango completo n),
- $\hat{U}$ es una matriz $m \times n$ con columnas ortonormales,
- y V es una matriz $n \times n$ con columnas ortonormales.
- Por lo tanto V es unitaria, y podemos multiplicar por la derecha por su inversa V^* para obtener

$$A = \hat{U} \hat{\Sigma} V^*. \tag{8}$$

SVD reducida

- Esta factorización de A se conoce como la *descomposición en valores singulares reducida*, o *SVD reducida*, de A.
- Esquemáticamente, tiene la siguiente representación:

SVD reducida ($m \geq n$)



SVD completa

- En la mayor parte de las aplicaciones, la SVD se utiliza tal y como la acabamos de describir.
- Sin embargo, no es de esta forma como se formula la SVD habitualmente en la bibliografía.
- Hemos introducido el término “reducida” y los acentos sobre $\mathbf{U}$ y $\mathbf{\Sigma}$ para distinguir la factorización (8) de la más estándar SVD “completa”.

SVD completa

- La idea es la siguiente. Las columnas de $\hat{\mathrm{U}}$ son n vectores ortonormales en el espacio m-dimensional $\mathbb{C}^{\mathrm{m}}$.
- A no ser que $\mathrm{m} = \mathrm{n}$, no forman una base de $\mathbb{C}^{\mathrm{m}}$, ni $\hat{\mathrm{U}}$ es una matriz unitaria.
- Sin embargo, si añadimos $\mathrm{m} - \mathrm{n}$ columnas ortonormales adicionales, $\hat{\mathrm{U}}$ puede extenderse a una matriz unitaria.
- Haremos esto de forma arbitraria, y denotaremos al resultado U.

SVD completa

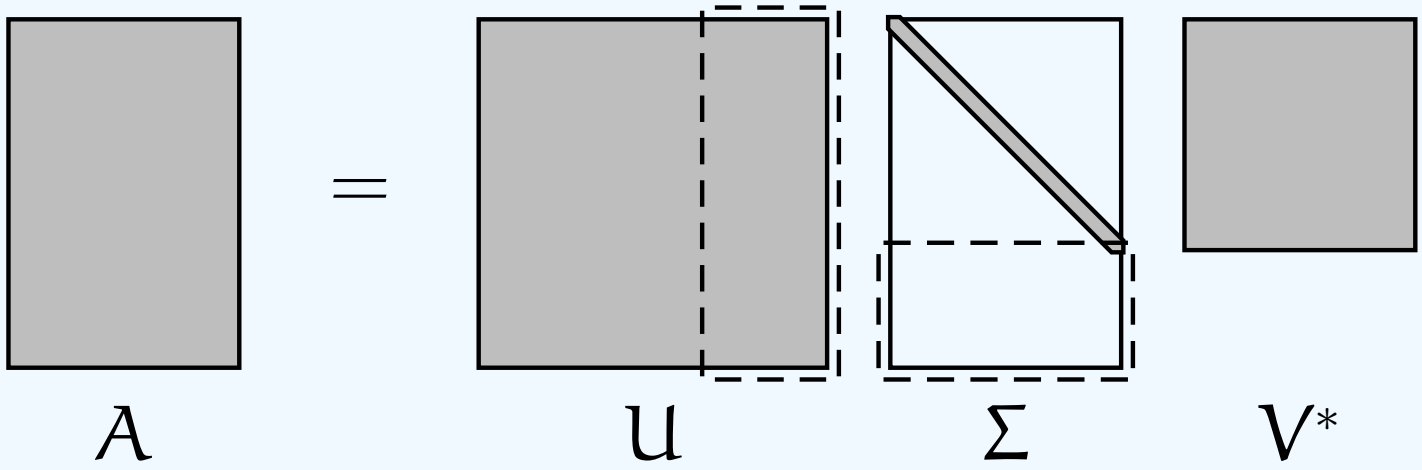
- Si $\hat{U}$ se reemplaza por U en (8), entonces $\hat{\Sigma}$ debe también cambiar.
- Para que el producto se mantenga inalterado, las últimas $m - n$ columnas de U deberían multiplicarse por cero.
- Por lo tanto, sea Σ la matriz $m \times n$ que consiste en $\hat{\Sigma}$ en el bloque $n \times n$ superior y $m - n$ filas de ceros en la parte inferior.
- Ahora tendremos una nueva factorización, la *SVD completa* de A:

$$A = U\Sigma V^*. \tag{9}$$

SVD completa

- Aquí U es $\mathrm{m} \times \mathrm{m}$ y unitaria, V es $\mathrm{n} \times \mathrm{n}$ y unitaria, y Σ es $\mathrm{m} \times \mathrm{n}$ y diagonal con componentes reales positivas.
- Esquemáticamente:

SVD completa ($\mathrm{m} \geq \mathrm{n}$)



- Las líneas discontinuas indican las columnas "silenciosas" de U y las filas de Σ que se descartan al pasar de (9) a (8).

SVD completa

- Si A es deficiente en rango, la factorización (9) también es válida.
- Ahora sólo r en lugar de n de los vectores singulares por la izquierda de A están determinados por la geometría de la hiperelipse.
- Para construir la matriz unitaria U, introducimos $m - r$ en lugar de sólo $m - n$ columnas ortonormales arbitrarias adicionales.
- La matriz V necesitará también $n - r$ columnas ortonormales arbitrarias para extender las r columnas determinadas por la geometría.
- La matriz Σ tendrá ahora r componentes diagonales positivas, siendo cero las restantes $n - r$ componentes.

SVD reducida

- Análogamente, la SVD reducida (8) también tiene sentido para matriz A de rango menor que completo.
- Se puede tomar $\hat{\mathrm{U}}$ de $\mathrm{m} \times \mathrm{n}$, con $\hat{\Sigma}$ de dimensiones $\mathrm{n} \times \mathrm{n}$ con algunos ceros en la diagonal,
- o comprimir aún más la representación de modo que $\hat{\mathrm{U}}$ es $\mathrm{m} \times \mathrm{r}$ y $\hat{\Sigma}$ es $\mathrm{r} \times \mathrm{r}$ con componentes estrictamente positivas en la diagonal.

Definición formal

- Sean m y n arbitrarios, no impondremos $m \geq n$.
- Dada $A \in \mathbb{C}^{m \times n}$, no necesariamente de rango completo,
- una *descomposición en valores singulares* (SVD) de A es una factorización

$$A = U \Sigma V^*, \tag{10}$$

- donde

$U \in \mathbb{C}^{m \times m}$ es unitaria,

$V \in \mathbb{C}^{n \times n}$ es unitaria,

$\Sigma \in \mathbb{R}^{m \times n}$ es diagonal.

Definición formal

- Además, se supone que las componentes diagonales σ_j de Σ
  - son no negativas y ordenadas no decrecientemente
  - $(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0)$, donde $p = \text{mín}(m, n)$.
- Observemos que
  - la matriz diagonal Σ tiene la misma forma que A,
  - pero que U y V son siempre matrices unitarias cuadradas.

Definición formal

Podemos comprobar ahora que la imagen de la esfera unitaria en $\mathbb{R}^n$ bajo el mapeado $A = U\Sigma V^*$ debe ser una hiperelipse en $\mathbb{R}^m$, ya que:

1. El mapeado unitario V^* conserva la esfera,
2. la matriz diagonal Σ estira la esfera en una hiperelipse alineada con la base canónica,
3. y el mapeado final unitario U rota o refleja la hiperelipse sin cambiar su forma.

- Por lo tanto, si toda matriz tiene una SVD, la imagen de la esfera unitaria bajo cualquier mapeado lineal es una hiperelipse, como afirmamos antes.

Existencia y unicidad

- Toda matriz $A \in \mathbb{C}^{m \times n}$ tiene una descomposición en valores singulares (10).
- Aún más, los valores singulares $\{\sigma_j\}$ están determinados de forma única, y, si A es cuadrada y los σ_j son distintos, los vectores singulares por la izquierda $\{u_j\}$ y por la derecha $\{v_j\}$ están determinados de forma única salvo por signos complejos (es decir, factores escalares complejos de módulo 1).

Un cambio de base

- La SVD nos permite decir que toda matriz es diagonal: basta con que utilicemos las bases apropiadas para los espacios dominio y rango.
- Cualquier $\mathbf{b} \in \mathbb{C}^m$ puede expandirse en la base de los vectores singulares por la izquierda de A (columnas de $\mathbf{U}$),
- y cualquier $\mathbf{x} \in \mathbb{C}^n$ puede expandirse en la base de los vectores singulares por la derecha de A (columnas de $\mathbf{V}$).
- Los vectores coordenados para estas expansiones son

$$\mathbf{b}' = \mathbf{U}^* \mathbf{b}, \qquad \mathbf{x}' = \mathbf{V}^* \mathbf{x}.$$

Un cambio de base

- Seqún (9), la relación $b = Ax$ puede espresarse en términos de b' y x':

$$b = Ax \iff U^* b = U^* Ax = U^* U \Sigma V^* x$$
$$\iff b' = \Sigma x'.$$

- Cuando quiera que $b = Ax$, tenemos $b' = \Sigma x'$. Por tanto A se reduce a la matriz diagonal Σ cuando el rango se expresa en la base de las columnas de U y el dominio se expresa en la base de las columnas de V.

SVD vs. descomposición en autovalores

- La diagonalización de una matriz mediante la expresión en términos de una nueva base también subyace el estudio de autovalores.
- Una matriz cuadrada no deficiente A puede expresarse como una matriz diagonal de autovalores Λ, si el rango y el dominio se representan en una base de vectores propios.
- Si las columnas de una matriz $X \in \mathbb{C}^{m \times m}$ contienen vectores propios independientes de $A \in \mathbb{C}^{m \times m}$, la *descomposición en autovalores* de A es

$$A = X \Lambda X^{-1}, \tag{11}$$

- donde Λ es una matriz diagonal $m \times m$ cuyas componentes son los autovalores de A.

SVD vs. descomposición en autovalores

- Esto implica que si definimos, para $b, x \in \mathbb{C}^m$ que verifican $b = Ax$,

$$b' = X^{-1}b, \quad x' = X^{-1}x,$$

- entonces los recién expandidos vectores b' y x' satisfacen $b' = \Lambda x'$.
- En las aplicaciones, los autovalores tienden a ser relevantes en problemas relacionados con el comportamiento de formas iteradas de A, tales como potencias matriciales A^k o exponenciales e^{tA}, mientras que los vectores singulares tienden a ser relevantes en problemas relacionados con el comportamiento de la propia A, o de su inversa.

SVD vs. descomposición en autovalores

Existen diferencias fundamentales entre la SVD y la descomposición en autovalores:

- La SVD utiliza dos bases distintas (los conjuntos de vectores singulares izquierdos y derechos), mientras que la descomposición en autovalores utuliza sólo una (los vectores propios).
- La SVD utiliza bases ortonormales, mientras que la descomposición en autovalores utiliza una base que generalmente no es ortogonal.
- No todas las matrices (ni tan siquiera las cuadradas) tienen una descomposición en autovalores, pero todas las matrices (incluso las rectangulares) tienen una descomposición en valores singulares.

Propiedades matriciales utilizando la SVD

En los resultados siguientes utilizaremos la siguiente nomenclatura:

- Supondremos que A tiene dimensiones $\mathrm{m} \times \mathrm{n}$.
- p el mínimo entre m y n
- $\mathrm{r} \leq \mathrm{p}$ es el número de valores singulares no nulos de A
- $\langle \mathrm{x}, \mathrm{y}, \ldots, \mathrm{z} \rangle$ es el espacio generado por los vectores $\mathrm{x}, \mathrm{y}, \ldots, \mathrm{z}$.

Propiedades matriciales utilizando la SVD

1. El rango de A es r, el número de valores singulares no nulos.
2. $\text{range}(A) = \langle u_1, \ldots, u_r \rangle$, y $\text{null}(A) = \langle v_{r+1}, \ldots, v_n \rangle$.
3. $\|A\|_2 = \sigma_1$, y $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$
4. Los valores singulares no nulos de A son las raíces cuadradas de los autovalores no nulos de A^*A o AA^*. (Estas matrices tienen los mismos autovalores no nulos).
5. Si $A = A^*$, entonces los valores singulares de A son los valores absolutos de los autovalores de A.
6. Para $A \in \mathbb{C}^{m \times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.

Aproximaciones de rango inferior

- Pero ¿qué es la SVD?
- Otra aproximación a una explicación es considerar cómo podríamos representar una matriz A como una suma de r matrices de rango unidad.

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*. \tag{12}$$

- Hay muchas formas de expresar una matriz $m \times n$ A como una suma de matrices de rango unidad.

Aproximaciones de rango inferior

- Por ejemplo, podríamos escribir A como la suma de sus m filas, o sus n columnas, o sus mn componentes.
- Como otro ejemplo, la eliminación de Gauss reduce A a la suma de una matriz completa de rango uno, una matriz de rango uno cuyas primeras fila y columna son cero, una matriz de rango uno cuyas dos primeras filas y columnas son cero, etc.

Aproximaciones de rango inferior

- La fórmula (12), sin embargo, representa una descomposición en matrices de rango uno con una propiedad más profunda:
- *la suma parcial ν-ésima captura tanta energía de $\mathbf{A}$ como es posible*.
- Esta afirmación es cierta definiendo "energía" como la norma-2 o como la norma de Frobenius.
- Podemos expresarlo de forma precisa formulando un problema de la mejor aproximación de una matriz $\mathbf{A}$ mediante matrices de rango inferior.

Aproximaciones de rango inferior

- Para cualquier ν con $0 \leq \nu \leq r$, definimos

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^*; \tag{13}$$

- si $\nu = p = \min\{m, n\}$, definimos $\sigma_{\nu+1} = 0$.
- Entonces

$$\|A - A_\nu\|_2 = \inf_{\substack{B \in \mathbb{C}^{m \times n} \\ \text{rank}(B) \leq \nu}} \|A - B\|_2 = \sigma_{\nu+1}.$$

Interpretación geométrica

- ¿Cuál es la mejor aproximación de un hiperelipsoide por un segmento de línea? Utiliza como segmento de línea el mayor eje.
- ¿Cuál es al mejor aproximación por un elipsoide bidimensional? Toma el elipsoide generado por el mayor y el segundo mayor eje.
- Continuando de esta forma, a cada paso mejoramos la aproximación añadiendo a nuestra aproximación el mayor eje del hiperelipsoide que no hayamos incluido todavía. Tras r pasos, hemos capturado todo A.
- Esta idea tiene ramificaciones en áreas tan dispares como compresión de imágenes y análisis funcional.

Aproximaciones de rango inferior

- Existe un resultado análogo para la norma de Frobenius.
- Para cualquier ν con $0 \leq \nu \leq r$, la matriz A_ν de (13) verifica también

$$\|A - A_\nu\|_F = \inf_{\substack{B \in \mathbb{C}^{m \times n} \\ \mathbf{rank}(B) \leq \nu}} \|A - B\|_F = \sqrt{\sigma_{\nu+1}^2 + \cdots + \sigma_r^2}.$$

5: Proyectores

- Un *proyector* es una matriz cuadrada P que satisface

$$\mathrm{P}^2 = \mathrm{P}. \tag{14}$$

- Una matriz con esta propiedad se denomina también *idempotente*.
- Esta definición incluye tanto a los proyectores ortogonales como a los no ortogonales.
- Para evitar confusiones, podemos utilizar el término *proyector oblicuo* en el caso no ortogonal.

Interpretación geométrica

- El término proyector surge de la noción de que si iluminamos desde la derecha el subespacio range(P), entonces $P\nu$ sería la sombra proyectada por el vector ν.
- Si $\nu \in$ range(P), entonces reside exactamente sobre su propia sombra, y aplicar el proyector resulta en el propio ν.
- Matemáticamente, tenemos $\nu = Px$ para algún x, y

$$P\nu = P^2x = Px = \nu.$$

Interpretación geométrica

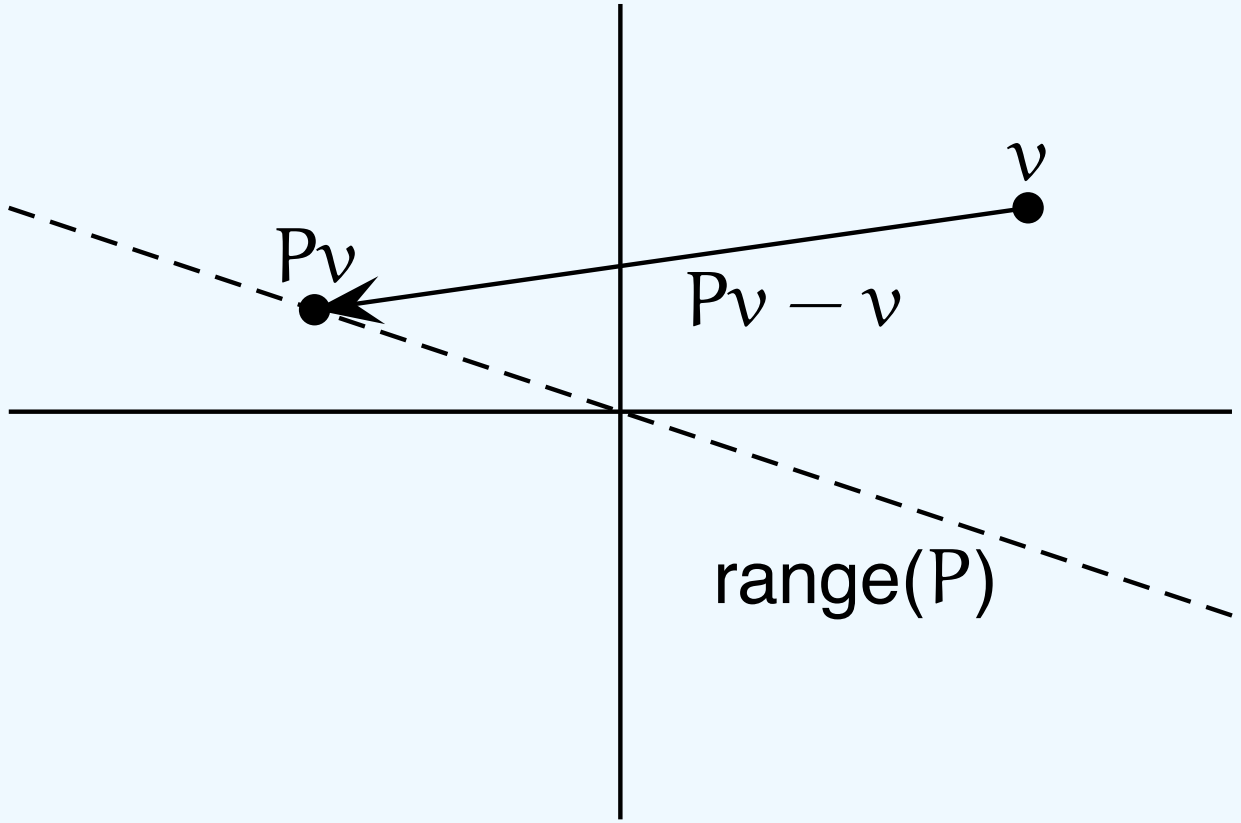
- ¿Desde qué dirección proviene la luz cuando $v \neq Pv$?
- La respuesta depende de v;
- para cualquier v concreto, se puede deducir dibujando la línea desde v a Pv, $Pv - v$.
- Aplicando el proyector a este vector proporciona un resultado nulo:

$$P(Pv - v) = P^2v - Pv = 0.$$

- Lo que significa que $Pv - v \in \text{null}(P)$. Es decir, la dirección de la luz puede ser diferente para cada v, pero siempre está descrita por un vector en $\text{null}(P)$.

Interpretación geométrica

Proyección oblícua:



Proyectores complementarios

- Si P es un proyector, $\mathrm{I} - \mathrm{P}$ es también un proyector, ya que también es idempotente:

$$(\mathrm{I} - \mathrm{P})^2 = \mathrm{I} - 2\mathrm{P} + \mathrm{P}^2 = \mathrm{I} - \mathrm{P}.$$

- La matriz $\mathrm{I} - \mathrm{P}$ se denomina el *proyector complementario* a P.
- ¿Sobre qué espacio proyecta $\mathrm{I} - \mathrm{P}$?
- ¡Exáctamente sobre el espacio nulo de P!
- Sabemos que $\mathrm{range}(\mathrm{I} - \mathrm{P}) \supseteq \mathrm{null}(\mathrm{P})$, ya que si $\mathrm{P}v = 0$, tenemos $(\mathrm{I} - \mathrm{P})v = v$.
- Al contrario, sabemos que $\mathrm{range}(\mathrm{I} - \mathrm{P}) \subseteq \mathrm{null}(\mathrm{P})$, ya que para cualquier v, tenemos $(\mathrm{I} - \mathrm{P})v = v - \mathrm{P}v \in \mathrm{null}(\mathrm{P})$.

Proyectores complementarios

- Por lo tanto, para cualquier proyector P,

$$\text{range}(\mathrm{I}-\mathrm{P}) = \text{null}(\mathrm{P}).$$

- Si escribimos $\mathrm{P} = \mathrm{I} - (\mathrm{I}-\mathrm{P})$ obtenemos el resultado complementario

$$\text{null}(\mathrm{I}-\mathrm{P}) = \text{range}(\mathrm{P}).$$

- Podemos ver también que $\text{null}(\mathrm{I}-\mathrm{P}) \cap \text{null}(\mathrm{P}) = \{0\}$: cualquier vector v en ambos conjuntos satisface $v = v - \mathrm{P}v = (\mathrm{I}-\mathrm{P})v = 0$.
- Otra forma de enunciar este hecho es

$$\text{range}(\mathrm{P}) \cap \text{null}(\mathrm{P}) = \{0\}.$$

Proyectores complementarios

- Estos cálculos muestran que un *proyector separa* $\mathbb{C}^m$ *en dos subespacios*.
- Al contrario, sean S_1 y S_2 dos subespacios de $\mathbb{C}^m$ tales que $S_1 \cap S_2 = \{0\}$ y $S_1 + S_2 = \mathbb{C}^m$, donde $S_1 + S_2$ denota la generación de S_1 y S_2, es decir, el conjunto de vectores $s_1 + s_2$ con $s_1 \in S_1$ y $s_2 \in S_2$.
- Entonces existe un proyector P tal que $\text{range}(P) = S_1$ y $\text{null}(P) = S_2$.
- Decimos que P es el proyector *sobre* S_1 *según* S_2.

Proyectores complementarios

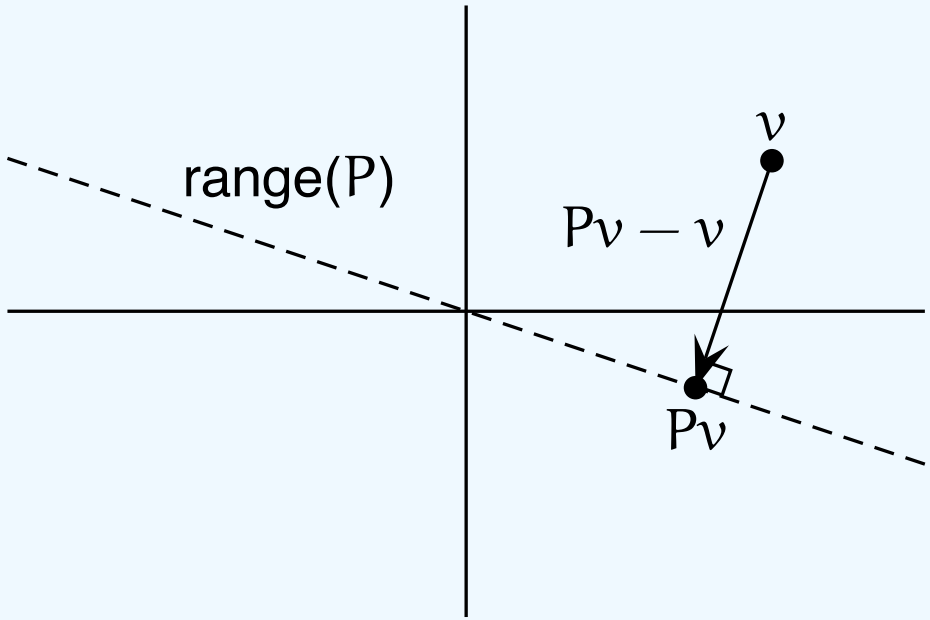
- Este proyector y su complementario pueden considerarse como la única solución del siguiente problema:
- Dado v, encontrar vectores $v_1 \in S_1$ y $v_2 \in S_2$ tales que $v_1 + v_2 = v$.
- La proyección Pv proporciona v_1, y la proyección complementaria $(I - P)v$ proporciona v_2.
- Estos vectores son únicos ya que todas las soluciones deben ser de la forma

$$(Pv + v_3) + ((I - P)v - v_3) = v,$$

- donde es obvio que v_3 debe estar tanto en S_1 como en S_2, es decir, $v_3 = 0$.

Proyectores ortogonales

- Un *proyector ortogonal* es uno que proyecta sobre un subespacio S_1 según un espacio S_2, donde S_1 y S_2 son ortogonales.



- Algebraicamente, Un proyector P es ortogonal si y sólo si $\mathrm{P} = \mathrm{P}^*$ (hermítico).
- Su SVD completa es $\mathrm{P} = \mathrm{Q}\Sigma\mathrm{Q}^*$.

Proyección con una base ortonormal

- Como un proyector ortogonal tiene algunos valores singulares nulos (excepto en el caso trivial $\mathrm{P} = \mathrm{I}$), es natural eliminar las columnas silenciosas de Q en $\mathrm{P} = \mathrm{Q}\Sigma\mathrm{Q}^*$ y utilizar la SVD reducida en lugar de la completa.
- Obtenemos

$$\mathrm{P} = \hat{\mathrm{Q}}\hat{\mathrm{Q}}^*, \tag{15}$$

- donde las columnas de $\hat{\mathrm{Q}}$ son ortonormales.
- En (15), la matriz $\hat{\mathrm{Q}}$ no tiene porque venir de una SVD.

Proyección con una base ortonormal

- Sea $\{q_1, \ldots, q_n\}$ cualquier conjunto de n vectores ortonormales en $\mathbb{C}^m$,
- y sea $\hat{Q}$ la matriz $m \times n$ correspondiente.
- De (6) sabemos que

$$v = r + \sum_{i=1}^{n} (q_i q_i^*) v$$

representa una descomposición de un vector $v \in \mathbb{C}^m$ en una componente en el espacio de columnas de $\hat{Q}$ más una componente en el espacio ortogonal.

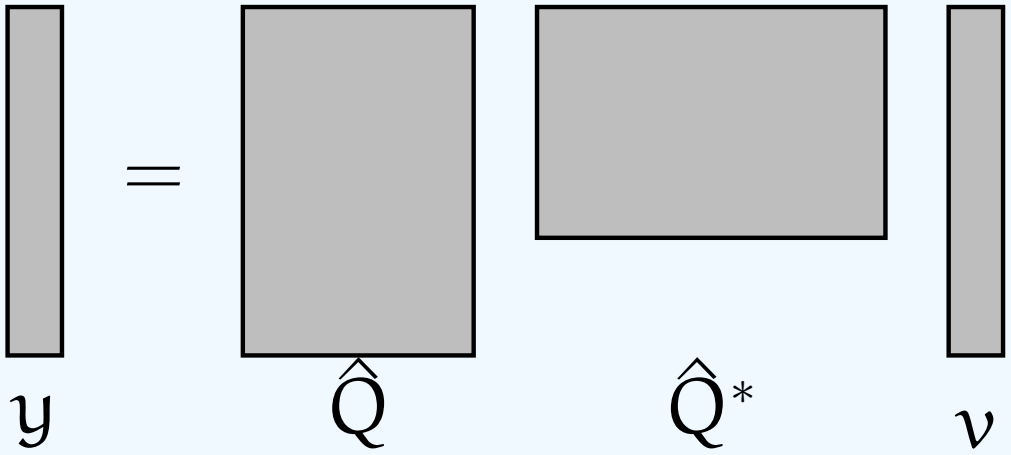
Proyección con una base ortonormal

- Por lo tanto el mapeado

$$v \mapsto \sum_{i=1}^{n} (q_i q_i^*) v \tag{16}$$

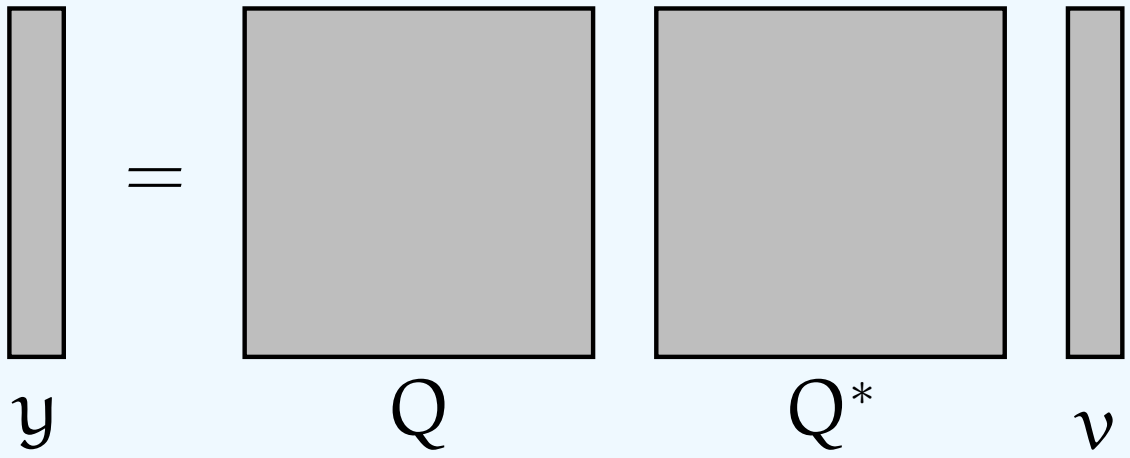
es un proyector ortogonal sobre range$(\hat{Q})$,

- y, en forma matricial, $y = \hat{Q}\hat{Q}^* v$:



Proyección con una base ortonormal

- Por tanto cualquier producto $\hat{\mathrm{Q}}\hat{\mathrm{Q}}^*$ es siempre un proyector sobre el espacio de columnas de $\hat{\mathrm{Q}}$, siempre que sus columnas sean ortonormales.
- Quizá $\hat{\mathrm{Q}}$ se haya obtenido eliminando algunas columnas y filas de una factorización completa $\nu = \mathrm{Q}\mathrm{Q}^*\nu$ de la identidad,



- y quizá no.

Proyección con una base ortonormal

- El complementario de un proyector ortogonal es también un proyector ortogonal.
- Demostración: $\mathrm{I} - \hat{\mathrm{Q}}\hat{\mathrm{Q}}^*$ es hermítica.
- El complementario proyecta sobre el espacio ortogonal a range$(\hat{\mathrm{Q}})$.
- Un caso especial importante de proyectores ortogonales el proyector ortogonal de rango uno que aísla la componente en un dirección única q, que puede escribirse

$$\mathrm{P}_\mathrm{q} = \mathrm{q}\mathrm{q}^*. \tag{17}$$

Proyección con una base ortonormal

- Éstas son las piezas a partir de las cuales pueden construirse proyectores de rango superior, como en (16).
- Sus complementarios son los proyectores ortogonales de rango $m-1$ que eliminan las componentes en la dirección de q:

$$P_{\perp q} = I - pp^*. \tag{18}$$

- En (17) y (18) q es un vector unitario.
- Para vectores no nulos arbitrarios a,

$$P_a = \frac{aa^*}{a^*a}, \qquad P_{\perp a} = I - \frac{aa^*}{a^*a}$$

Proyección con una base arbitraria

- También puede construirse un proyector ortogonal sobre un subespacio de $\mathbb{C}^m$ comenzando con una base arbitraria, no necesariamente ortogonal.
- Supongamos que el subespacio es generado por los vectores linealmente independientes $\{a_1, \ldots, a_n\}$, y sea A la matriz $m \times n$ cuya columna j-ésima es a_j.
- Al pasar de v a su proyección ortogonal $y \in \text{range}(A)$, la diferencia $y - v$ debe ser ortogonal a $\text{range}(A)$.
- Esto es equivalente a la afirmación de que y debe satisfacer $a_j^*(y - v) = 0$ para todo j.

Proyección con una base arbitraria

- Como $y \in \text{range}(A)$, podemos hacer $y = Ax$ y escribir esta condición como $a_j^*(Ax - v) = 0$ para todo j, o equivalentemente, $A^*(Ax - v) = 0$ o $A^*Ax = A^*v$.
- Como A es de rango completo, A^*A es no singular.
- Por lo tanto

$$x = (A^*A)^{-1}A^*v.$$

- Por último, la proyección de v, $y = Ax$, es $y = A(A^*A)^{-1}A^*v$.

Proyección con una base arbitraria

- Por lo tanto el proyector ortogonal sobre range(A) puede expresarse mediante la fórmula

$$P = A(A^*A)^{-1}A^*.$$

- Observemos que esta es una generalización multidimensional del proyector P_a.
- En el caso ortonormal $A = \hat{Q}$, el término entre paréntesis colapsa a la identidad y recuperamos (15).

6: Factorización QR

- Existe una idea algorítmica en álgebra lineal numérica más importante que todas las demás:
- la factorización QR.

Factorización QR reducida

- En muchas aplicaciones, estamos interesados en los espacios de columnas de una matriz A.
- Observemos el prural: son los espacios *sucesivos* generados por las columnas $a_1, a_2, \ldots$ de A:

$$\langle a_1 \rangle \subseteq \langle a_1, a_2 \rangle \subseteq \langle a_1, a_2, a_3 \rangle \subseteq \ldots$$

- $\langle a_1 \rangle$ es el espacio unidimensional generado por a_1, $\langle a_1, a_2 \rangle$ es el espacio bidimensional generado por a_1 y a_2, y así sucesivamente.
- La idea de la factorización QR es la construcción de una secuencia de vectores ortonormales $q_1, q_2, \ldots$ que generen estos espacios sucesivos.

Factorización QR reducida

- Supongamos de momento que $A \in \mathbb{C}^{m \times n}$ ($m \geq n$) tiene rango completo n.
- Queremos que la secuencia $q_1, q_2, \ldots$ tenga la propiedad

$$\langle q_1, q_2, \ldots, q_j \rangle = \langle a_1, a_2, \ldots, a_j \rangle, \quad j = 1, \ldots, n.$$

- Equivalentemente (suponiendo $r_{kk} \neq 0$),

$$\left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_n \\ & & & \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ q_1 & q_2 & \cdots & q_n \\ & & & \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix},$$

Factorización QR reducida

- De forma expandida,

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned} \tag{19}$$

- En forma matricial tenemos

$$A = \hat{Q}\hat{R},$$

donde $\hat{Q}$ es $m \times n$ con columnas ortonormales y $\hat{R}$ es $n \times n$ y triangular superior.

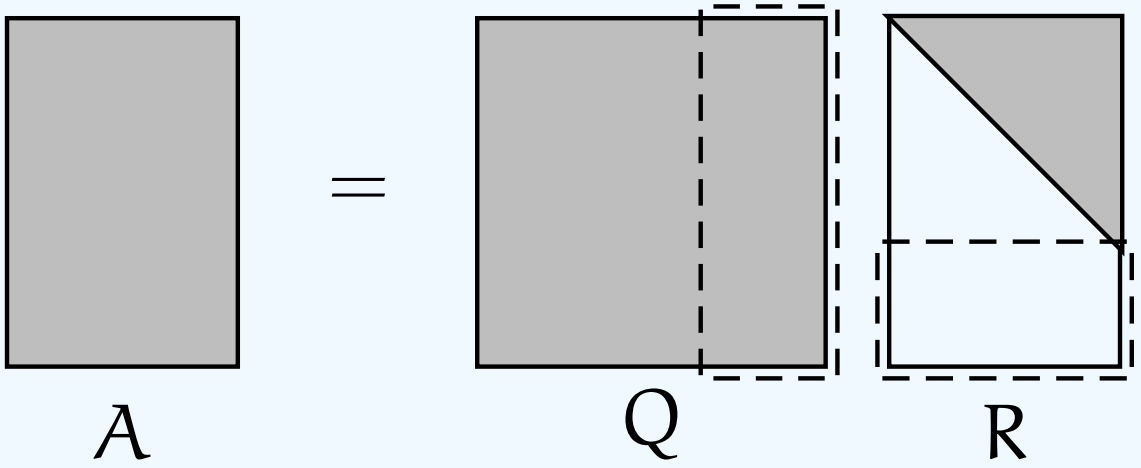
- Ésta es una *factorización QR reducida de* A.

Factorización QR completa

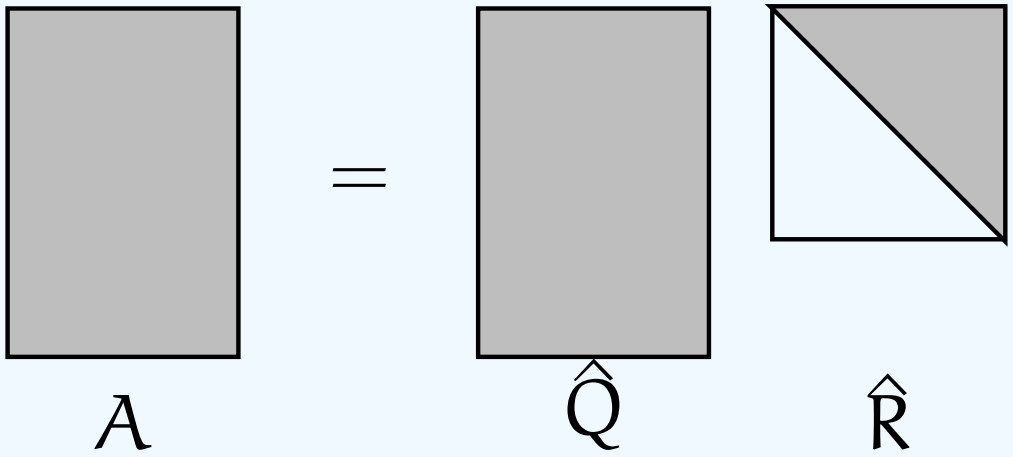
- Una *factorización QR completa* de $A \in \mathbb{C}^{m \times n}$ $(m \geq n)$ añade $m - n$ columnas ortonormales adicionales a $\hat{Q}$, convirtiéndola en una matriz unitaria $m \times m$ Q.
- Se añaden filas de ceros a $\hat{R}$ de forma que se convierte en una matriz $m \times n$ R, todavía triangular superior.
- Todas las matrices tienen factorizaciones QR, y bajo restricciones apropiadas, son únicas. Establecemos primero el resultado sobre existencia.

Comparación entre factorizaciones

Factorización QR completa ($m \geq n$)



Factorización QR reducida ($m \geq n$)



Comparación entre factorizaciones

- En la factorización QR completa, Q es $\mathrm{m} \times \mathrm{m}$, R es $\mathrm{m} \times \mathrm{n}$, y las últimas $\mathrm{m} - \mathrm{n}$ columnas de Q están multiplicadas por ceros en R (encerrados en líneas a trazos).
- En la factorización QR reducida, las columnas y filas silenciadas son eliminadas. Ahora $\hat{\mathrm{Q}}$ es $\mathrm{m} \times \mathrm{n}$, $\hat{\mathrm{R}}$ es $\mathrm{n} \times \mathrm{n}$, y ninguna de las filas de $\hat{\mathrm{R}}$ son necesariamente cero.
- En la factorización QR completa, las columnas q_j para $\mathrm{j} > \mathrm{n}$ son ortogonales a range(A). Suponiendo que A es de rango completo n, constituyen una base ortogonal de range$(A)^\perp$ o, equivalentemente, de null(A).

Ortogonalización de Gram-Schmidt

- Las ecuaciones (19) sugieren un método para calcular la factorización QR reducida:
- Dados $a_1, a_2, \ldots$, podemos construir los vectores $q_1, q_2, \ldots$ y las componentes r_{ij} mediante un proceso de ortogonalización sucesiva.
- Ésta es una idea antigua, conocida como *ortogonalización de Gram-Schmidt*.

Ortogonalización de Gram-Schmidt

- El proceso funciona de la siguiente forma.
- En el paso j-ésimo, deseamos encontrar un vector unitario $q_j \in \langle a_1, \ldots, a_j \rangle$ que sea ortogonal a $q_1, \ldots, q_{j-1}$.
- El vector

$$v_j = a_j - (q_1^* q_j) q_1 - \cdots - (q_{j-1}^* q_j) q_{j-1} \quad (20)$$

  es del tipo buscado, salvo que todavía no está normalizado.
- Si dividimos por $||v_j||_2$, el resultado es un vector apropiado q_j.

Ortogonalización de Gram-Schmidt

- Con esto en mente, reescribamos (19) de la forma

$$
\begin{aligned}
q_1 &= \frac{a_1}{r_{11}}, \\
q_2 &= \frac{a_2 - r_{12} q_1}{r_{22}}, \\
q_3 &= \frac{a_3 - r_{13} q_1 - r_{23} q_2}{r_{33}}, \\
&\vdots \\
q_n &= \frac{a_n - \sum_{i=1}^{n-1} r_{in} q_i}{r_{nn}}.
\end{aligned}
\tag{21}
$$

Ortogonalización de Gram-Schmidt

- De (20) se desprende que una definición apropiada para los coeficientes r_{ij} en los numeradores de (21) es

$$r_{ij} = q_i^* a_j, \quad (i \neq j). \tag{22}$$

- Los coeficientes r_{jj} en lo denominadores se escogen para normalizar:

$$|r_{jj}| = \left\| a_j - \sum_{i=1}^{j-1} r_{ij} q_i \right\|_2 . \tag{23}$$

- Si escogemos $r_{jj} > 0$, $\hat{R}$ tendrá componentes positivas en la diagonal.

Ortogonalización de Gram-Schmidt

- Éste es el algoritmo clásico de Gram-Schmidt.

for $j = 1$ **to** n
  $v_j = a_j$
  for $i = 1$ **to** $j - 1$
    $r_{ij} = q_i^* a_j$
    $v_j = v_j - r_{ij} q_i$
  $r_{jj} = \|v_j\|_2$
  $q_j = v_j / r_{jj}$

- Es correcto matemáticamente, pero inestable numéricamente.
- Lo que se puede corregir mediante la *iteración de Gram-Schmidt modificada*.

Solución de $Ax = b$ mediante QR

- Supongamos que queremos resolver en x la ecuación $Ax = b$, donde $A \in \mathbb{C}^{m \times m}$ es no singular.
- Si $A = QR$ es una factorización QR, entonces podemos escribir $QRx = b$, o

$$Rx = Q^* b.$$

- El miembro de la derecha de está ecuación es sencillo de calcular, si Q es conocido, y el sistema de ecuaciones lineales implícitas en el miembro de la izquierda también es sencillo de resolver ya que es triangular.

Solución de $Ax = b$ mediante QR

- Esto sugiere el siguiente método para calcular la solución de $Ax = b$:
  1. Calcula una factorización QR $A = QR$.
  2. Calcula $y = Q^* b$.
  3. Resuelve $Rx = y$ para x.
- Éste es un método excelente para resolver $Ax = b$, pero requiere el doble de operaciones que la eliminación de Gauss.