Descomposición SVD

Luciano A. Perez

10 de septiembre de 2015

1. Introducción

1.1. La Descomposición

Sean $n, m \in \mathbb{N}$, y sea $A \in \mathbb{C}^{m \times n}$; una *descomposición en valores singulares* (o *SVD*) de A es una factorización

$$A = U\Sigma V^* \tag{1}$$

donde:

$$\begin{cases} U \in \mathbb{C}^{m\times m} & \text{es unitaria} \quad (UU^* = U^*U = I) \\ V \in \mathbb{C}^{n\times n} & \text{es unitaria} \\ \Sigma \in \mathbb{R}^{m\times n} & \text{es diagonal} \quad (\Sigma_{ij} = 0, i \neq j) \end{cases}$$

los elementos diagonales de Σ, $\sigma_i = \Sigma_{ii}$ son además no negativos y están en orden decreciente, i.e. $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p$ (donde $p = min\{m, n\}$). Estos σ_i son reales y se llaman *valores singulares* de la matriz A.

Nótese que no estamos pidiendo que $m \geq n$ ni tampoco que A sea de rango completo (o sea puede ser $r(A) < p$). En la descomposción, U y V son siempre cuadradas, mientras que Σ tiene la forma de A.

Las columnas de U, pensadas como vectores de $\mathbb{C}^m$ se conocen como *vectores singulares por izquierda* de A, y las columnas de V como vectores singulares por derecha.

En principio no parece trivial que toda matriz tenga una factorización de este tipo (Sec.2). Lo que sí es fácil ver es, *dada* la existencia de esta descomposición quiénes son U, V y Σ.

Para ello notemos que si $A = U\Sigma V^*$, entonces $A^* = V\Sigma^* U^*$ y haciendo el producto AA^*:

$$AA^* = U\Sigma \overbrace{V^*V}^{I} \Sigma^* U^* = U\Sigma\Sigma^* U^*$$

de donde se obtiene, dado que U es unitaria:

$$U^*AA^*U = \Sigma\Sigma^* \tag{2}$$

Pero como Σ es una matriz real de $m \times n$ diagonal, entonces el producto $\Sigma\Sigma^*$ es una matriz diagonal de $m \times m$ cuyas entradas diagonales son σ_i^2 en el mismo orden que los de Σ:

$$\begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_m & \cdots & 0 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots \\ 0 & \sigma_2 & 0 & \cdots \\ \vdots & \cdots & \ddots & \cdots \\ \vdots & \cdots & \ddots & \sigma_m \\ \vdots & \cdots & \ddots & \cdots \\ 0 & \cdots & 0 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots \\ 0 & \sigma_2^2 & 0 & \cdots \\ \vdots & \cdots & \ddots & \cdots \\ \vdots & \cdots & \ddots & \sigma_m^2 \end{pmatrix}$$

Entonces hemos encontrado una diagonalización de la matriz AA^*, y los valores singulares de A son las raices cuadradas los valores propios de AA^* (que es hermitiana y semidefinida positiva).

Nótese que las columnas de U son autovectores de AA^*, puesto que si en 2 multiplicamos a izquierda por U tenemos, dado que $UU^* = I$ y que $\Sigma\Sigma^*$ es diagonal:

$$AA^*U = \Sigma\Sigma^*U$$

Es decir, si llamamos u_j a las columnas de U, tenemos $AA^*u_j = \sigma_j^2 u_j$.

Esto nos da un *método* para hallar la descomposición SVD:

1. Dada $A \in \mathbb{C}^{m\times n}$ formamos la matriz AA^*

2. Resolvemos el problema de autovalores $AA^*u = \lambda u$. Esto siempre se puede hacer porque AA^* es hermitiana.

   Como además es semidefinida positiva, los λ_j $(j = 1, ..., m)$ que hallamos son no negativos y podemos definir $\sigma_j = \sqrt{\lambda_j}$.

   Teniendo los σ_j ordenados de manera decreciente armamos, con los correspondientes vectores u_j como columnas, la matriz $U \in \mathbb{C}^{m\times m}$.

3. Ahora para obtener V simplemente resolvemos el sistema que se obtiene de 1 multiplicando a derecha por V, esto es $AV = U\Sigma$.

Una observación evidente que puede hacerse es qué hubiera pasado si desde un principio, en lugar de multiplicar AA^* hubieramos hecho el producto en el orden inverso, es decir A^*A.

Tenemos $A^*A = V\Sigma^*\Sigma V^*$ o sea $V^*A^*AV = \Sigma^*\Sigma$. De manera que ahora el rol de V es como el de U anteriormente y podríamos enunciar “otro” método para hallar la descomposición SVD de la matriz A.

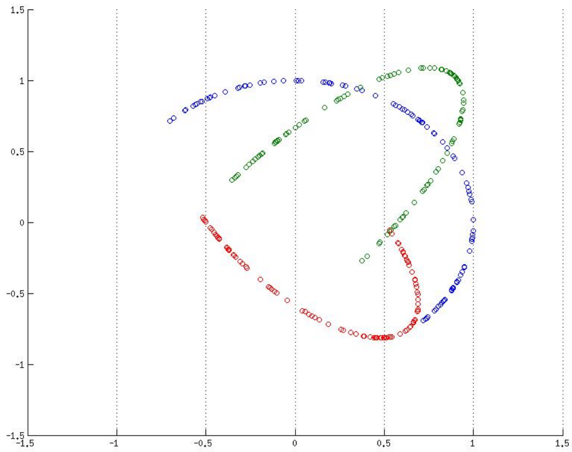

Figura 1: Efecto de matrices arbitrarias sobre la circunferencia unitaria

Ahora bien, si lo hacemos de esta forma, los valores singulares que vamos a obtener son las raices de los autovalores de A^*A. Es fácil ver que los autovalores *distintos de cero* de A^*A y de AA^* son los mismos. La diferencia que puede haber entre el espectro de A^*A y de AA^* es que si una de las dos es de dimensión más grande, entonces vamos a tener al cero con mayor multiplicidad.

Así que para definir Σ no hay ambigüedad: podemos buscar los autovalores de A^*A o de AA^* y con ellos armamos una matriz diagonal del mismo tamaño de A.

Por practicidad, conviene elegir A^*A o AA^* según cual sea de menor tamaño, así el polinomio característico es de menor grado.

Naturalmente, como las matrices U y V se armaron a partir de autovectores, no gozan de la misma unicidad que Σ. Lo que puede probarse (Sec.2) es que, si A es cuadrada y los σ_j son distintos, entonces los vectores singulares están determinados a menos de un signo (o sea un factor $e^{i\theta}$). Si la matriz A es real resulta claro que podemos elegir a U y V reales también.

1.2. Motivación Geométrica

La idea geométrica de que debe existir una descomposición de este tipo para cualquier matriz parte de la observación de que una matriz cualquiera transforma circunferencias en elipses. En la Fig.1, por ejemplo, tomamos puntos de la circunferencia unitaria y los transformamos por dos matrices elegidas al azar (de norma no muy grande ni chica para que la escala del gráfico permita visualizar el efecto), comprobando efectivamente este hecho intuitivo.

No es trivial en principio probar este resultado intuitivo (o *experimental* si se quiere) y justamente la idea que nos lleva a la SVD es aislar las direcciones de los ejes principales de las elipses que obtenemos por una matriz A. Esta idea se generaliza a dimensiones m, n arbitrarias. Más detalles de esta idea en Trefethen y Bau [2].

2. Existencia y Unicidad

Teorema 2.0.1 (Descomposición SVD)**.** Toda $A \in \mathbb{C}^{m\times n}$ tiene una descomposición en valores singulares del tipo 1. Los valores singulares están determinados de manera única. Si A es cuadrada y los σ_j son distintos, las columnas de U y V son únicas a menos del signo (i.e. un factor $e^{i\theta}$).

Demostración. Para probar la existencia empezamos por el valor singular más grande y hacemos inducción. Sea $\sigma_1 = \|A\|_2$ y sean $v_1 \in \mathbb{C}^n$ con $\|v_1\|_2 = 1$, $u_1 \in \mathbb{C}^n$, $\|u_1\|_2 = \sigma_1$ con $Av_1 = u_1$ (existen porque la norma se realiza).

Extendemos a bases ortonormales $\{v_j\}$ de $\mathbb{C}^n$ y $\{u_j\}$ de $\mathbb{C}^m$ y con dichos vectores como columnas armamos las matrices U_1 y V_1. Evidentemente resulta:

$$U_1^*AV_1 = S = \begin{pmatrix} \sigma_1 & w^* \\ 0 & B \end{pmatrix}$$

donde el 0 es un vector columna de $m-1$ ceros, w^* un vector fila de dimensión $n-1$, y $B \in \mathbb{C}^{m-1\times n-1}$. Consideremos la acción de S sobre el vector columna $\left(\begin{smallmatrix} \sigma_1 \\ w \end{smallmatrix}\right)$ y tomemos norma:

$$\left\| \begin{pmatrix} \sigma_1 & w^* \\ 0 & B \end{pmatrix} \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2 \geq \sigma_1^2 + w^*w = \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2^2 = (\sigma_1^2 + w^*w)^{1/2} \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2$$

luego $\|S\|_2 \geq (\sigma_1^2 + w^*w)^{1/2}$. Como U_1 y V_1 son unitarias, sabemos que $\|S\|_2 = \|A\|_2$ (si no resulta obvio, ver la Sec.4), de donde $w = 0$. Si $n = 1$ o $m = 1$ listo. Si no, B describe la acción de A en el subespacio ortogonal a v_1. Por hipótesis inductiva, B tiene descomposición SVD, $B = U_2\Sigma_2V_2^*$, y con ella se verifica que:

$$A = U_1 \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix}^* V_1^*$$

y haciendo los productos se tiene la SVD de A.

Para la unicidad, si A es cuadrada y todos los valores singulares son distintos, entonces A^*A y AA^* tienen todos sus autovalores distintos (si A no fuera cuadrada, alguno de los dos productos va a tener el 0 como autovalor repetido). De manera que los autoespacios correspondientes tienen dimensión 1. □

3. Ejemplo 1

- Ejemplo 1: Encontrar la SVD de

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

Formamos primero el producto:

$$AA^* = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

La ecuación característica queda $(17-\lambda)^2 - 8^2 = 0$, con soluciones $\lambda_1 = 25$ y $\lambda_2 = 9$, de manera que tenemos $\sigma_1 = 5$ y $\sigma_2 = 3$. Sabemos que Σ tiene la misma forma de A, y estos valores en la diagonal, así que:

$$\Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

Para obtener U encontramos los autovectores:

$$\begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

Para $\lambda = 25$ tenemos $17x + 8y = 25x$ de donde se tiene $x = y$ y un vector puede ser $(1,1)$. Para $\lambda = 9$ tenemos $17x + 8y = 9x$ de donde se tiene $x = -y$ y un vector puede ser $(1,-11)$. Para que la U nos quede unitaria, los dividimos por su respectiva norma y tenemos:

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Ahora para obtenere V tenemos que resolver el sistema $AV = U\Sigma$, que en este caso queda:

$$AV = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 5 & 3 & 0 \\ 5 & -3 & 0 \end{pmatrix}$$

Es decir, las columnas del lado derecho son los u_j multiplicados por el respectivo σ_j (o cero cuando se acabaron los u_j). Ahora vamos a resolver para obtener los v_j, observando antes dos cosas: dado que queremos una V unitaria, los v_j tienen que ser ortogonales y de norma 1.

Que son ortogonales sale *solo*; en este caso es claro porque son autovectores de A^*A (hermitiana) de autovalores distintos (25,9,0). Más abajo damos una justificación más general.

Que sean de norma unitaria lo vamos pidiendo cuando resolvemos cada sistema, que en principio es compatible indeterminado. En este caso, por ejemplo, tenemos que si llamamos v_1 a la primer columna de V, debe cumplirse:

$$Av_1 = \sigma_1 u_1$$

o sea:

$$\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} \end{pmatrix}$$

que como sistema es:

$$\begin{cases} 3x+2y+2z = \frac{5}{\sqrt{2}} \\ 2x+3y-2z = \frac{5}{\sqrt{2}} \end{cases}$$

Sumando las dos ecuaciones tenemos $x+y=\sqrt{2}$ y entonces es fácil encontrar una solución que además sea de norma unitaria, y es $(x,y,z) = (\sqrt{2}/2, \sqrt{2}/2, 0)$.

Para la segunda columna tenemos:

$$\begin{cases} 3x+2y+2z = \frac{3}{\sqrt{2}} \\ 2x+3y-2z = -\frac{3}{\sqrt{2}} \end{cases}$$

Sumando tenemos $x+y=0$, así que $x=-y$ y con esto $x+2z=3/\sqrt{2}$ es decir $z=\frac{3}{2\sqrt{2}}-\frac{x}{2}$. Además por normalización tenemos $x^2+y^2+z^2=1$ o sea $2x^2+z^2=1$. Con ello nos queda que:

$$2x^2+\left(\frac{3}{2\sqrt{2}}-\frac{x}{2}\right)^2 = \frac{9}{4}x^2-\frac{3\sqrt{2}}{4}x+\frac{9}{8} = 1$$

Esta última ecuación tiene una raíz doble $x=\sqrt{2}/6$ y con ello tenemos la segunda columna de V:

$$\begin{pmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{6} \\ \frac{2\sqrt{2}}{3} \end{pmatrix}$$

El sistema que corresponde a la última columna es homogéneo, así que ahí normalizar no es problema (encontramos cualquier solución, dividimos por la norma y sigue siendo solución porque el sistema es homogéneo). Por ejemplo $(1,-1,-1/2)$ que normalizado da $(2/3,-2/3,-1/3)$.

Así tenemos la V completa:

$$V = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{6} & -\frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & -\frac{1}{3} \end{pmatrix}$$

Y el ejemplo se completa comprobando que efectivamente hemos encontrado la descomposición es decir que $A = U\Sigma V^*$:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{6} & -\frac{\sqrt{2}}{6} & \frac{2\sqrt{2}}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

- **Observación sobre la obtención de U y V:** Retomemos el problema, por ejemplo, de hallar V una vez que tenemos U.

  A partir de la factorización $A = U\Sigma V^*$ que implica $A^* = V\Sigma^* U^*$, podemos multiplicar a derecha por U obteniendo $A^*U = V\Sigma^*$ (y de donde se puede ver que para $1 \leq j \leq \min\{m,n\}$ se tiene la igualdad (recordar que Σ es real)

  $$A^* u_j = \sigma_j v_j \tag{3}$$

  donde u_j y v_j son los vectores columnas de U y de V. De aquí se ve que, para los $\sigma_j \neq 0$ se tiene $v_j = \frac{1}{\sigma_j} A^* u_j$ (y es fácil ver que son ortonormales). Con ello tenemos una forma de rápida de hallar dichas columnas de V dada U. Si la matriz A es cuadrada ($m = n$) y de rango completo (todos los $\sigma_j \neq 0$), listo.

  Pero queda la pregunta de qué hacer con los de $\sigma_j = 0$ o con el caso en que haya *menos* u_j que v_j (o sea $m < n$, como en nuestro ejemplo) o más ($m > n$).

  Para ver que pasa volvamos al sistema $AV = U\Sigma$, pensando el caso $m < n$ como en nuestro ejemplo. Mirando como es Σ:

  $$\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \sigma_m & \cdots & 0 & 0 \end{pmatrix}$$

  queda claro que, tanto si $\sigma_j = 0$ o si $j > m$ (o sea, *"se acabaron los u"*) se tiene que $Av_j = 0$ (el vector columna de ceros). O sea que los v_j correspondientes están en el *núcleo* de A ($v_j \in Ker(A)$).

  Por otro lado, la igualdad 3 nos dice que los v_j con $\sigma_j \neq 0$ son una base de la imagen de A^* (los u_j son una base de $\mathbb{C}^m$). Por último, es fácil ver que la imágen de A^* y $Ker(A)$ son complementos ortogonales en C^n.

  Si hubiera habido *más* u_j de los necesarios (o sea $m > n$) está claro que no puede haber más de n que correspondan a $\sigma_j \neq 0$ (por una cuestión de rango). Así que se hace igual, se obtienen los v_j para los $\sigma_j \neq 0$ y si faltan se buscan en el núcleo de A.

  Con todo ello tenemos una nueva versión del método para hallar la SVD:

1. Encontrar Σ y U como antes.
2. Para los $\sigma_j \neq 0$ encontrar los v_j según:

$$v_j = \frac{1}{\sigma_j} A^* u_j$$

3. Si ya tenemos todos los v_j, listo. Si no, buscar una base ortonormal de $Ker(A)$ y unirla a los ya encontrados.

Naturalmente, si huberamos empezado por hallar la V, los papeles de A y A^* se invierten: en el paso (2) tenemos $u_j = \frac{1}{\sigma_j} A v_j$ y en el último llegado el caso hay que buscar una base de $Ker(A^*)$.

Nótese que de esta forma prácticamente hemos vuelto a probar la existencia de las SVD, y de hecho esta es la forma en que la prueba el muy buen libro de Rainer Kress [1]; la demostración que dimos más arriba es la de Trefethen y Bau [2].

4. Aplicaciones y Otras Observaciones

1. **Resolución de Sistemas Lineales:** Dado el sistema $Ax = b$ tenemos que $U\Sigma V^* x = b$, y multiplicando cada miembro por U^* resulta $\Sigma y = \tilde{b}$ donde hemos llamado $y = V^* x$ y $\tilde{b} = U^* b$.

   Pero el sistema $\Sigma y = \tilde{b}$ es de resolución trivial porque Σ es diagonal. Y para obtener x basta observar que al ser V unitaria resulta $Vy = x$, así que no hay más que multiplicar.

2. **Compresión de Información:** De manera muy informal (que más abajo haremos más precisa en ciertos casos) las descomposicón SVD nos sirve para *comprimir información*.

   Esto se deriva de la intuición geométrica que le dimos: la matriz A transforma una circunferencia en una elipse, lo cual pensado a partir de $A = U\Sigma V^*$ lo interpretamos como: (i) primero una rotación rígida o reflexión, dada por V^*; (ii) *estiramientos* dados por los coeficientes de Σ; (iii) otra transformación rígida dada por U.

   Ahora, dado que los σ_j tienen distinta magnitud, el *efecto* de A será mayor en las direcciones dadas por los de mayor valor, es decir los primeros $(\sigma_1, \sigma_2, ...)$. Si nos quedamos sólo con las direcciones donde A tiene mayor efecto y *tiramos* las otras, entonces no se perderá demasiada información, pero habremos almacenado muchos menos datos.

   En este sentido es posible usar la SVD para *comprimir* la información implícita en una matriz y esto tiene aplicaciones en inteligencia artificial (aprendizaje de diccionarios), compresión de imágenes (ej. 22 de la guía) y mínimos cuadrados (ver más abajo).

3. **Relación entre la SVD y $\|A\|_2$:** Como U y V son unitarias, conservan la norma 2, es decir, $\|Ux\|_2 = \|x\|_2$ para cualquier $x \in \mathbb{C}^m$ y $\|Vy\|_2 = \|y\|_2$ $\forall y \in \mathbb{C}^n$. Lo propio sucede naturalmente con las conjugadas hermitianas U^* y V^*.

   Esto implica entonces que, dada la descomposición SVD, $A = U\Sigma V^*$, resulte $\|A\|_2 = \|\Sigma\|_2$. Para verlo, sea x con $\|x\|_2 = 1$; tenemos $\|Ax\|_2 = \|U\Sigma V^* x\|_2 = \|\Sigma V^* x\|_2 \leq \|\Sigma\|_2 \|V^* x\|_2 = \|\Sigma\|_2 \|x\|_2 = \|\Sigma\|_2$. De donde se sigue que $\|A\|_2 \leq \|\Sigma\|_2$. Como a su vez $U^*AV = \Sigma$ se puede encontrar la otra desigualdad y resulta $\|A\|_2 = \|\Sigma\|_2$.

   Es fácil ver por definición que $\|\Sigma\|_2 = \sigma_1$. Con ello se tiene probada la fórmula $\|A\|_2 = \sqrt{\rho(A^*A)}$ a partir de la descomposición SVD y la igualdad práctica $\|A\|_2 = \sigma_1$

4. *$Cond_2(A)$* **y SVD:** sea $A \in \mathbb{C}^{n\times n}$ inversible. A partir de $A = U\Sigma V^*$ deducimos $A^{-1} = V\Sigma^{-1}U^*$ y como en el apartado anterior, se ve que $\|A^{-1}\|_2 = \|\Sigma^{-1}\|_2 = 1/\sigma_n$.

   Con ello, resulta $Cond_2(A) = \sigma_1/\sigma_n$ que también habíamos visto de otras formas.

5. **Forma Reducida:** la igualdad $AV = U\Sigma$ se puede interpretar como que A toma los elementos de una base ortonormal (las columnas de V) y los manda a los generadores de la imagen de A que son los vectores columnas de U *estirados o contraídos* por los σ_j correspondientes.

   Si $m > n$ (más filas que columnas), al hacer el producto $U\Sigma$ se *pierden* las últimas columnas de U. Entonces, hubiera sido lo mismo quedarse con una $\hat{\Sigma}$ cuadrada (tirando las filas nulas) y una $\hat{U} \in \mathbb{C}^{m\times n}$ (tirando las columnas correspondientes). Esta es la *forma reducida* de la SVD y también podemos pensar a la SVD a partir de *completar* aquella.

6. **Aproximación de rango bajo:** la matriz Σ (que tiene el mismo rango que A ya que U y V son unitarias) se puede pensar como una suma de matrices de rango 1:

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & \\ 0 & \sigma_2 & \cdots & \\ \vdots & \vdots & \ddots & \end{pmatrix} = \sum_{j=1}^{\text{mín}\{m,n\}} \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \cdots & & \\ 0 & \cdots & \sigma_j & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \end{pmatrix}$$

   Si nos quedamos con la suma de ν de estas matrices para $\nu \leq rg(A)$ (o sea tiramos las filas desde la $\nu + 1$ en adelante) tendremos una matriz de rango ν. Con ella tenemos una *aproximación* de A con rango ν, dada muliplicando por U a izquierda y por V^* a derecha:

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j u_j v_j^* \tag{4}$$

Existen otra muchas formas de aproximar A por matrices de menor rango, por ejemplo, simplemente reemplazando filas de A por ceros, o haciendo lo propio en la descomposición QR, etc. De todas las posibles, la aproximación dada por 4 es la *mejor*, al menos en $\|\|_2$.

Teorema 4.0.1 (Aproximación de rango Bajo)**.** Sea $r = rg(A)$ y sea $0 \leq \nu \leq r$ y sea A_ν dada por 4. Entonces:

$$\|A - A_\nu\|_2 = \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ rg(B) \leq \nu}} \|A - B\|_2$$

Nótese que $\|A - A_\nu\|_2 = \sigma_{\nu+1}$ (entendido como 0 si $\nu = \text{mín}\{m,n\}$).

Demostración. Supongamos que existe una B con $rg(B) \leq \nu$ y $\|A - B\|_2 < \sigma_{\nu+1}$. Entonces para cada $x \in Ker(B)$ se cumple:

$$\|Ax\|_2 = \|(A-B)x\|_2 \leq \|(A-B)\|_2 \|x\|_2 < \sigma_{\nu+1}\|x\|_2$$

Y sabemos que $dim(ker(B)) = n - \nu$. Por otro lado, como $AV = U\Sigma$, entonces $Av_j = \sigma_j u_j$ y entonces para $j = 1, ..., \nu + 1$ tenemos $\|Av_j\|_2 = \|\sigma_j u_j\|_2 = \sigma_j \geq \sigma_{\nu+1}$, o sea un subespacio de dimensión $\nu + 1$ donde se tiene dicha desigualdad. Por una cuestión de dimensión, este espacio y $ker(B)$ tendrían que tener intersección no nula, pero esto es absurdo. □

Con esta noción de mejor aproximación, podemos hacer un poco más precisa la idea de que la descomposición SVD permite *comprimir* información.

Queda claro de la fórmula dada en 4 que para $j > \nu$ las columnas u_j y v_j no juegan ningún papel. Por lo tanto, podemos tiraralas y quedarnos con matrices $\hat{U} \in \mathbb{C}^{m\times\nu}$ y $\hat{V} \in \mathbb{C}^{n\times\nu}$, y el producto seguirá igual, esto es $A_\nu = \hat{U}\Sigma_\nu\hat{V}^*$.

Almacenando estas matrices en lugar de las anteriores habremos *comprimido* la información. Por supuesto, si bien la aproximación es la mejor de rango ν en norma 2, cuán buena sea dependerá del rango elegido (ν y de la necesidad del usuario de la información). Los ejercicios 21 y 22 de la práctica tratan este tema.

7. **Distancia a las matrices singulares:** dada la forma de Σ es bastante obvio que si $p = \text{mín}\{m,n\}$, entonces σ_p (o sea el valor singular más chico) cumple $\sigma_p = \text{mín}_{\|x\|_2=1} \|\Sigma x\|_2$. Pero como $\|\Sigma x\|_2 = \|Ax\|_2$ se tiene $\sigma_p = \text{mín}_{\|x\|_2=1} \|Ax\|_2$.

Ahora consideremos el caso de $A \in \mathbb{C}^{n\times n}$ y sea B cualquier matriz singular del mismo tamaño. Dado x de norma 1 en el núcleo de B (existe porque es singular), tenemos $\sigma_n \leq \|Ax\|_2 = \|Ax - Bx\|_2 = \|(A-B)x\|_2$. Luego $\|(A-B)\|_2 \geq \sigma_n$. Si $\sigma_n = 0$ es obvio que el mínimo se realiza, tomando $B = A$. Si el valor singular más chico no es nulo, tomemos v un vector de norma unitaria tal que $\|Av\|_2 = \sigma_n$ y definamos $u = (1/\sigma_n)Av$. La matriz $A - \sigma_n uv^*$ sirve.

De manera que el valor singular más pequeño nos da la distancia de A al conjunto de matrices singulares. Tal vez de allí viene el nombre de valores singulares, habría que googlearlo.

5. Ejemplo 2

- Ejemplo 2:
  1. Encontrar para cada $n \in \mathbb{N}$ la SVD de
  $$A_n = \begin{pmatrix} -n & 1 & 1 & n \\ n & 1 & 1 & -n \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix}$$
  2. Calcular $\| A_n \|_2$ y $Cond_2(A_n)$
  3. Probar que $Cond_\infty(A_n) \to \infty$ con $n \to \infty$

Para hallar la SVD nos conviene en este caso usar $A_nA_n^*$, ya que toma una forma por bloques que la hace más fácil de diagonalizar. Tenemos:

$$A_nA_n^* = \begin{pmatrix} -n & 1 & 1 & n \\ n & 1 & 1 & -n \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} -n & n & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ n & -n & 1 & 1 \end{pmatrix} =$$
$$= \begin{pmatrix} 2n^2+2 & -2n^2+2 & 0 & 0 \\ -2n^2+2 & 2n^2+2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \tag{5}$$

El polinomio característico resulta $(4-\lambda)^2\left((2n^2+2-\lambda)^2 - (2n^2-2)^2\right)$ cuyas raíces son $\lambda = 4n^2$ y 4, esta última es raíz triple.

Si $n = 1$ son todos los autovalores iguales y AA^* es diagonal, así que U puede ser la identidad y despejar V es trivial. Tomemos el caso $n \geq 2$.

Ya tenemos Σ (tomando raíz a los autovalores y armando la diagonal) y es:

$$\Sigma = \begin{pmatrix} 2n & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Para hallar U empezamos con el autovalor $4n^2$ y nos queda el sistema:

$$\begin{cases} (2n^2+2)x + (-2n^2+2)y = 4n^2x \\ (-2n^2+2)x + (2n^2+2)y = 4n^2y \\ 4z = 4n^2z \\ 4w = 4n^2w \end{cases}$$

De las últimas 2, como $n \geq 2$ se tiene $z = w = 0$ y sumando las dos primeras concluimos que $x = -y$, con lo cual tenemos un primer vector singular:

$$u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix}$$

Para $\lambda = 4$ si hacemos un sistema analogo al anterior, es claro que z y w pueden ser cualquiera y que sumando de nuevo las dos primeras llegamos a la conclusión $x = y$ con lo cual podemos conseguir 3 vectores ortonormales:

$$u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix} \quad u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad u_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Ahora que ya tenemos U podemos hallar V por cualquiera de los métodos que mencionamos en el ejemplo obteniendo:

$$V = \begin{pmatrix} -\frac{\sqrt{2}}{2} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{2}}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{2}}{2} & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Para el punto (ii) tenemos, según hemos visto, $\|A_n\|_2 = \sigma_1 = 2n$ y $Cond_2(A_n) = \sigma_1/\sigma_n = 2n/2 = n$.

Para el punto (iii) se puede usar la desigualdad entre la $Cond_\infty$ y la $Cond_2$ que usamos en la práctica (ojo no confundir la n de A_n con la dimensión del espacio, que es lo que siempre me pasa a mí) o también usando que dada la descomposición SVD $A = U\Sigma V^*$ si A es inversible entonces Σ lo es y $A^{-1} = V\Sigma^{-1}U^*$, y escribiendo la matriz (o al menos una fila) es facil acotar la norma infinito de A^{-1} por abajo.

Referencias

[1] Kress, R., *Numerical Analysis*, Springer GTM 181, 1998.

[2] Trefethen, L. & Bau, D., *Numerical Linear Algebra*, SIAM, 1997.

[3] Stoer, J. & Bulirsch, R., *Introduction to Numerical Analysis*, Springer-Verlag, 1996.