MSEL
Modelling in Science Education and Learning

Modelling in Science Education and Learning
Volume 8(2), 2015 doi: 10.4995/msel.2015.4010.
Instituto Universitario de Matemática Pura y Aplicada
Universitat Politècnica de València

Álgebra lineal y descomposición en valores singulares

Linear Algebra and Singular Valued Decomposition

Jose M. Calabuig Rodríguez, Lluís M. Garcia Raffi, Enrique A. Sánchez Pérez
Universitat de Politècnica de València
jmcalabu@mat.upv.es, lmgarcia@mat.upv.es, easancpe@mat.upv.es

Abstract

En este trabajo presentamos una propuesta entorno a cómo se puede utilizar la Descomposición en valores singulares de una matriz para desarrollar un temario de una asignatura de Álgebra lineal en un grado de ingeniería. Para ello introducimos algunas técnicas especiales, resultados y ejemplos.

In this work we present how we can use the Singular Valued Decomposition of a matrix as a main tool to develop a Linear algebra course in engineering. Some special techniques, results and examples are introduced in this setting.

Palabras clave: modelización, álgebra lineal, descomposición en valores singulares, DVS.
Keywords: modelling, linear algebra, singular valued decomposition, SVD.

1 Introducción

Bajo diferentes nombres, el *Álgebra Lineal* es una de las asignaturas que aparece en los programas de estudios de todos los grados de estudios técnicos. Con la implantación del *Espacio Europeo de Educación Superior* las asignaturas de matemáticas, en general, y el *Álgebra lineal* en particular, han sufrido un importante recorte en el número de créditos. Asimismo las asignaturas de *Modelización Matemática* que habitualmente se impartían en las antiguas ingenierías han desaparecido de los nuevos planes de estudios.

Una posibilidad para suplir la carencia en el número de horas que se imparten en matemáticas en los nuevos grados de ingeniería sería el cambio en la metodología de enseñanza-aprendizaje. En este trabajo presentamos algunas ideas para cambiar esta metodología en las asignaturas de *Álgebra lineal*. La idea se basa en dos aspectos centrales: por una parte, y desde el punto de vista del alumnado, se trataría de utilizar los conocimientos que tienen los alumnos que acceden a los grados universitarios de su etapa estudios de *Enseñanza Secundaria Obligatoria* (ESO) y de bachillerato; por otra parte, desde el punto de vista del profesorado, el cambio metodológico se centraría en la introducción de las definiciones, resultados, ejemplos,... entorno a un proyecto.

Nuestra propuesta concreta se basa en el uso de la llamada *Descomposición en valores singulares* (DVS) de una matriz. Esta descomposición tiene tres características que, bajo nuestro punto de vista, la hacen especialmente interesante:

- Desde un punto de vista metodológico tiene muchas aplicaciones sencillas a nivel de ingeniería que permitirán ver a los alumnos la importancia de las matemáticas en su formación como ingenieros. Sin duda la más conocida, y utilizada ya en algunos grados —véase por ejemplo Domínguez (2011)— es la compresión de imágenes digitales. Sin embargo existen otras aplicaciones como son el *Análisis Semántico Latente* (Latent Semantic Index, en inglés) utilizada en la conocida plataforma de entretenimiento online Netflix© (https://www.netflix.com/es/) y de la que hablaremos más adelante.

- Por otra parte, desde un punto de vista matemático, la *Descomposición en valores singulares* hace uso de gran parte de las técnicas de *Álgebra lineal* que a nivel básico puede necesitar un ingeniero como, por ejemplo, matrices, aplicaciones lineales, valores y vectores propios, diagonalización, espacio euclídeo,...

- Finalmente la implementación de dicha técnica a la hora de resolver algunos problemas de modelización concretos, como los comentados anteriormente, pueden servir como argmento para introducir al alumno en las asignaturas de *Métodos numéricos* y de *Programación*.

Además de esta sección de carácter introductorio, el presente trabajo consta de tres secciones más. En la primera se presenta la construcción geométrica de la *Descomposición en valores singulares*. Se introducen aquí los primeros temas de *Álgebra lineal* que se pueden trabajar en la asignatura; más concretamente se proponen algunas ideas entorno a: (1) los temas de matrices y sistemas de ecuaciones lineales (y el espacio $\mathbb{R}^n$); (2) las aplicaciones lineales entre espacios de tipo $\mathbb{R}^n$ prestando especial atención a la visión geométrica así como los conceptos de subespacio vectorial y (3) la construcción geométrica (en dos dimensiones) de la descomposición. En la sección siguiente nos centramos ya en el cálculo concreto de la *Descomposición en valores singulares*. Tras una primera parte preliminar en la sección nos centramos en: (4) la diagonalización y diagonalización ortogonal de matrices (pasando por el *Método de ortogonalización de Gram-Schmidt*); para terminar la sección con: (5) el cálculo explícito de la descomposición. En la última sección se presentan un par de propuestas para completar lo que sería una asignatura de *Álgebra lineal* con problemas de carácter numérico que se podrían implementar en sesiones de laboratorio. Así comenzamos con: (6) el *Teorema del rango aproximado* que nos conducirá a las nociones de normas matriciales; terminando el trabajo con la aplicación de la descomposición a la resolución del (7) *Problema lineal de los mínimos cuadrados*.

Si bien la metodología propuesta no se ha llevado a cabo íntegramente en una asignatura, parte de ella se ha implementado ya con éxito en las asignaturas de *Matemáticas II* (durante los cursos 2012/2013 y 2013/2014) y de *Complementos de Métodos Matemáticos para la nivelación* (durante el curso 2014/2015) en el Grado de Tecnologías Industriales de la Universitat Politécnica de València.

2 Construcción geométrica de la DVS

La *Descomposición en valores singulares* se puede introducir de una manera sencilla e intuitiva utilizando geometría. Para ello nos planteamos el siguiente problema:

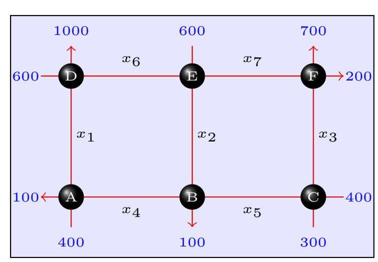
> *Fijada una matriz A de tamaño 2×2 queremos encontrar dos mallas perpendiculares (ortogonales) de forma que la matriz A transforme la primera malla en la segunda.*

Bajo la apariencia aparentemente sencilla del enunciado aparecen ya muchas matemáticas (en nuestro caso particular *Álgebra lineal*) escondidas que podemos comenzar a introducir a los alumnos. Citamos algunas de ellas:

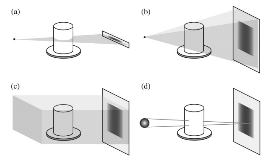
(1) Como paso preliminar empezaríamos recordando las nociones relacionadas con espacio $\mathbb{R}^2$ (o $\mathbb{R}^n$ con más generalidad) incluyendo los conocidos por los alumnos producto escalar y norma de un vector (con la correspondiente noción de ortogonalidad y ortonormalidad). En este repaso se pueden ya introducir los conceptos de sistema generador, dependencia e independencia lineal y base. Continuando con este *repaso de resultados conocidos* podríamos continuar con la teoría matricial. En este sentido podríamos comenzar con un repaso de la resolución de sistemas de ecuaciones lineales introduciendo algunos problemas de modelización. Citamos a continuación algunos de ellos. Ejemplos similares y otros distintos pueden encontrarse en los libros de Lay (2012) y Grossman (2008).

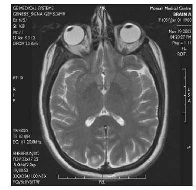
Interpolación Calcular, si existe, la ecuación de la parábola de $\mathbb{R}^2$ que pasa por $(1, 1)$, $(0, 2)$ y $(2, 0)$.

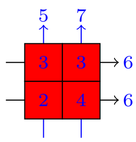
Flujo de carreteras
Consideremos el siguiente mapa de calles de una sola dirección en una determinada ciudad. Los números indican el flujo de entrada y de salida de vehículos en cada calle a una hora determinada. El ayuntamiento de la ciudad quiere realizar obras de alcantarillado entre A y B. La central de tráfico puede controlar el flujo de vehículos mediante la regulación de los semáforos, colocando agentes en los cruces o cerrando calles. Se trata de minimizar el tráfico entre A y B sin ocasionar congestiones en el resto de calles.

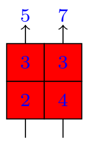


Tomografía (Procesamiento de imágenes por secciones)

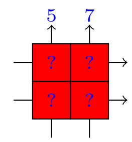






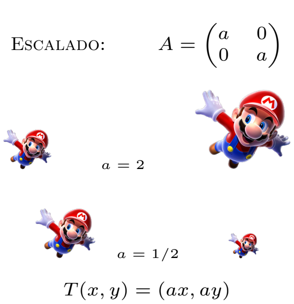


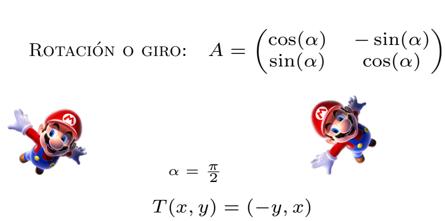
Imaginemos que tenemos una imagen formada tan sólo por 4 pixeles y que realizamos 2 proyecciones: una horizontal (es decir con un ángulo de 0º grados) y otra vertical (con un ángulo de 90º grados). ¿Es posible saber el valor de cada pixel a partir de los valores de las proyecciones?

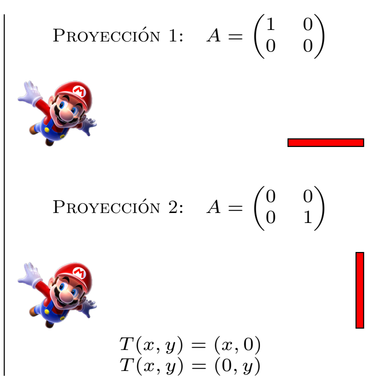


Una vez llegados a las matrices podríamos aquí introducir ya algunas técnicas como por ejemplo el *Método de Gauss-Jordan* para la obtención de la *Forma escalonada* a partir de operaciones elementales así como el rango de una matriz. Asimismo se podrían revisar las operaciones de matrices prestando especial atención a la conexión del producto de matrices con el producto escalar vía la traspuesta de una matriz y la conocida fórmula $(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{x}^T \cdot \boldsymbol{y}$ —donde escribimos los vectores $\boldsymbol{x}, \boldsymbol{y}$ como matrices columna—. Si bien el tema de determinantes es un tema clásico en este tipo de asignaturas éste es uno de los temas que, al menos a nivel práctico, suelen conocer de su etapa de estudios preuniversitarios. Es por esto que optamos por dedicar alguna sesión de problemas a hacer un breve repaso del cálculo de los determinantes y, sobretodo, a repasar las propiedades importantes de éstos.

(2) Tras este repaso (y quizás breve ampliación) de la teoría matricial y de sistema de ecuaciones lineales llegamos a una de las primeras propuestas de cambio metodológico de la asignatura. Habitualmente en los cursos de *Álgebra lineal* después del tema de matrices (y en algunos casos determinantes) se introduce la noción de espacio vectorial y a continuación el de aplicación lineal. Si bien es cierto que ésto proporciona a la asignatura un alto grado de generalidad pensamos que el paso de abstracción del espacio $\mathbb{R}^n$ a la noción de espacio vectorial presupone un grado de conocimiento y dominio del primero que no sabemos si tienen. En este sentido, y con el fin de afianzar el conocimiento del espacio $\mathbb{R}^n$, nuestra propuesta es introducir la noción de aplicación lineal entre dos espacios tipo $\mathbb{R}^n$ a partir de las matrices (este enfoque se puede encontrar, por ejemplo, en el libro de Otto Bretscher, 2014). Más concretamente una aplicación $T : \mathbb{R}^n \to \mathbb{R}^m$ decimos que es *lineal* si existe una matriz de tamaño $m \times n$ de forma que $T(\boldsymbol{x}) = A \cdot \boldsymbol{x}$ para todo $\boldsymbol{x} \in \mathbb{R}^n$. Además de la conexión directa de la noción con las matrices, esta definición proporciona una conexión geométrica que nos puede permitir afianzar la noción de aplicación lineal. Para ello podríamos introducir algunos ejemplos sencillos:



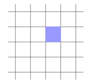




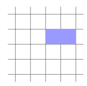
O centrándonos ya en la resolución de la cuestión con la que hemos iniciado esta sección:

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ y \end{pmatrix}$$

el cuadrado

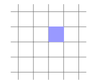


se transforma en

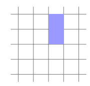


$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 2y \end{pmatrix}$$

el cuadrado

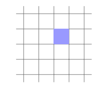


se transforma en

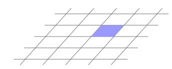


$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ y \end{pmatrix}$$

el cuadrado

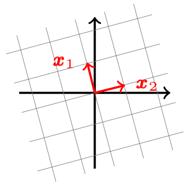


se transforma en

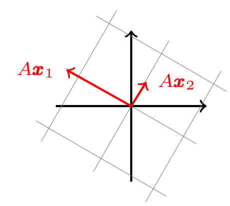


En este punto es donde podríamos ahora trabajar los conceptos de subespacio vectorial de $\mathbb{R}^n$, subespacio generado (span) e incluso el concepto de subespacio ortogonal que necesitaremos en la siguiente sección para la construcción de la *Descomposición en valores singulares*.

(3) Pasamos ya a la resolución de nuestra primera cuestión relativa a la *Descomposición en valores singulares*.



Elegimos dos vectores $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$ ortogonales y unitarios y buscamos una matriz A de forma que los vectores $A\boldsymbol{x}_1$ e $A\boldsymbol{x}_2$ también sean ortogonales (no necesariamente unitarios).



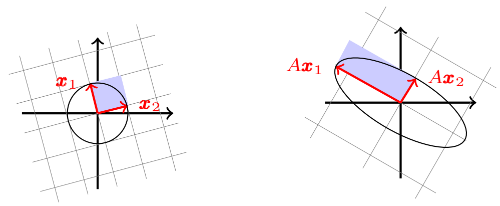
Denotamos por $\boldsymbol{y}_1$ y $\boldsymbol{y}_2$ a dos vectores unitarios en las direcciones de $A\boldsymbol{x}_1$ y $A\boldsymbol{x}_2$. Esto significa que $A\boldsymbol{x}_1 = \sigma_1\boldsymbol{y}_1$ y $A\boldsymbol{x}_2 = \sigma_2\boldsymbol{y}_2$. Los escalares σ_1 y σ_2 son los llamados *valores singulares de la matriz A*. Veamos entonces la descomposición:

- Si elegimos un vector $\boldsymbol{x}$, por el hecho de ser $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ una base, $\boldsymbol{x} = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2$. Pero por ser los vectores ortogonales los escalares α_1 y α_2 se determinan a partir del producto escalar mediante los llamados *Coeficientes de Fourier*:
$$\boldsymbol{x} = \frac{(\boldsymbol{x}|\boldsymbol{x}_1)}{(\boldsymbol{x}_1|\boldsymbol{x}_1)}\boldsymbol{x}_1 + \frac{(\boldsymbol{x}|\boldsymbol{x}_2)}{(\boldsymbol{x}_2|\boldsymbol{x}_2)}\boldsymbol{x}_2,$$
que al ser los vectores unitarios se reduce a $\boldsymbol{x} = (\boldsymbol{x}|\boldsymbol{x}_1)\boldsymbol{x}_1 + (\boldsymbol{x}|\boldsymbol{x}_2)\boldsymbol{x}_2$.
- Multiplicando por A: $A\boldsymbol{x} = (\boldsymbol{x}|\boldsymbol{x}_1)A\boldsymbol{x}_1 + (\boldsymbol{x}|\boldsymbol{x}_2)A\boldsymbol{x}_2 = (\boldsymbol{x}|\boldsymbol{x}_1)\sigma_1\boldsymbol{y}_1 + (\boldsymbol{x}|\boldsymbol{x}_2)\sigma_2\boldsymbol{y}_2$.
Y puesto que los vectores los escribimos en columna el producto escalar se puede escribir usando el producto de matrices (de momento son todos números reales) mediante la fórmula $(\boldsymbol{x}|\boldsymbol{x}_i) = (\boldsymbol{x}_i|\boldsymbol{x}) = \boldsymbol{x}_i^T \cdot \boldsymbol{x}$. Luego
$$A\boldsymbol{x} = (\boldsymbol{x}_1^T \cdot \boldsymbol{x})\sigma_1\boldsymbol{y}_1 + (\boldsymbol{x}_2^T \cdot \boldsymbol{x})\sigma_2\boldsymbol{y}_2 = \sigma_1\boldsymbol{y}_1(\boldsymbol{x}_1^T \cdot \boldsymbol{x}) + \sigma_2\boldsymbol{y}_2(\boldsymbol{x}_2^T \cdot \boldsymbol{x}),$$
$$A\boldsymbol{x} = \big(\sigma_1(\boldsymbol{y}_1 \cdot \boldsymbol{x}_1^T) + \sigma_2(\boldsymbol{y}_2 \cdot \boldsymbol{x}_2^T)\big) \cdot \boldsymbol{x}.$$

Así tenemos $A = \sigma_1(\boldsymbol{y}_1 \cdot \boldsymbol{x}_1^T) + \sigma_2(\boldsymbol{y}_2 \cdot \boldsymbol{x}_2^T)$, que se puede escribir en forma matricial

$$A = \underbrace{(\boldsymbol{y}_1, \boldsymbol{y}_2)}_{U}\underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}}_{\Sigma}\underbrace{\begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \end{pmatrix}}_{V^T}.$$

Con esta construcción queda clara la interpretación gráfica de los vectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ e $\{\boldsymbol{y}_1, \boldsymbol{y}_2\}$ pero, ¿cuál es la interpretación geométrica de los valores singulares σ_1 y σ_2? Una posible respuesta vendría dada por la fórmula $A\boldsymbol{x}_i = \sigma_i\boldsymbol{y}_i, i = 1, 2$. Sin embargo podemos completar nuestro dibujo inicial



Observamos que en el círculo la función $\|A\boldsymbol{x}\|$ tiene su máximo en $\boldsymbol{x}_1$ y su mínimo en $\boldsymbol{x}_2$. Y como para $i = 1, 2$

$$\|A\boldsymbol{x}_i\|^2 = \|\sigma_i \boldsymbol{y}_i\|^2 = \sigma_i^2,$$

entonces los valores singulares son las medidas de los semiejes de la elipse.

3 Cálculo de la DVS

Como ya se ha comentado en la introducción de este trabajo la elección de la *Descomposición en valores singulares* como proyecto de trabajo para desarrollar una asignatura de *Álgebra lineal* no se debe únicamente a su visión geométrica y de conexión con las matrices. Desde un punto de vista práctico esta descomposición tiene diferentes aplicaciones. Por ejemplo, recientemente (2008), la plataforma Netflix que es una "plataforma de entretenimiento mediante tarifa plana mensual streaming (flujo) multimedia (principalmente, películas y series de televisión) bajo demanda por Internet y de DVD-por-correo" (véase `https://es.wikipedia.org/wiki/Netflix`) convocó y resolvió en El premio del millón de dólares de Netflix.

NETFLIX

http://www.netflixprize.com/index

La citada empresa ofreció un premio de 1 millón de dólares para cualquier persona que pudiera mejorar la precisión de su sistema de recomendación de películas en un 10%. Se utilizaron técnicas muy sofisticadas (lejos del alcance de este trabajo) pero en el corazón de todas ellas estuvo la descomposición de DVS[*].

Otra aplicación íntimamente ligada a la descomposición es el llamado *Análisis Semántico Latente*. Esta teoría intenta resolver los problemas de sinonimia (y homonimia) y polisemia en los buscadores como por ejemplo Google. Más concretamente imaginemos que estamos interesados en buscar una determinada palabra en nuestro buscador, por ejemplo la palabra `banco`. Si nuestra palabra tiene múltiples significados (polisemia) una búsqueda literal (letra a letra) de nuestra palabra produciría muchos resultados posibles (`Banco de Santander`, `banco de peces`, `banco de niebla`, `banco de jardín`, `banco de sangre`, `banco de trabajo`). Por otra parte si estamos buscando un banco para un jardín quizás podamos estar interesados en que el buscador nos dé los resultados de otras palabras diferentes pero con un mismo (o similar) significado (sinonimia). En nuestro caso por ejemplo `asiento`, `silla`, `taburete`, `sillón`, `butaca`, `hamaca`,... El *Análisis Semántico Latente* proporciona una herramienta de búsqueda por conceptos o definiciones (en contraposición con la búsqueda literal). La base matemática para esta teoría es, de nuevo, la *Descomposición en valores singulares*. El lector interesado en esta aplicación puede consultar el *Example 5.12.4* del libro de Meyer (Meyer, 2000).

[*] El premio se lo llevó el grupo **BellKor's Pragmatic Chaos** de *AT&T (American Telephone and Telegraph)* el 1 de septiembre de 2009 (`https://en.wikipedia.org/wiki/Netflix_Prize`).

Modelling in Science Education and Learning
http://polipapers.upv.es/index.php/MSEL

Tras esta pequeña incursión en las aplicaciones de la *Descomposición en valores singulares* (existen otras muchas como *La reducción de ruido en el Análisis de datos*, *La compresión de imágenes digitales*,...) en esta sección nos centramos en el segundo problema:

Fijada una matriz A, ¿cómo se calcula de Descomposición en valores singulares de A?

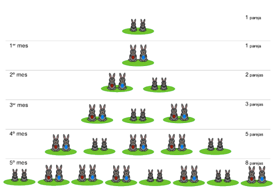
Enumeramos algunos pasos que podemos seguir que nos permitirán ir avanzando (y completando) en lo que sería un temario estándard de una asignatura de *Álgebra lineal* de un grado de ingeniería.

(4) Uno de los temas clásicos en los temarios de ingeniería es la útil *Diagonalización* de matrices cuadradas (así como su correspondiente versión utilizando la estructura euclídea, la llamada *Diagonalización ortogonal*). A partir de la fórmula obtenida anteriormente

$$A = \underbrace{(\boldsymbol{y}_1, \boldsymbol{y}_2)}_{U} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \end{pmatrix}}_{V^T},$$

podemos dar la definición de *Diagonalización de una matriz A cuadrada* como $A = PDP^{-1}$ siendo D una matriz diagonal y P una matriz invertible (siempre que existan). En este sentido cabe destacar la existencia de muchos ejemplos que podemos utilizar en los que aparece no sólo este tipo de factorizaciones (además de la *Clasificación de cónicas y cuádricas*) sino la propia noción de valor y vector propio.

Un primer ejemplo sería el *Algoritmo PageRank*. Gran parte del éxito de GOOGLE se debe al citado algoritmo que proporciona un orden de importancia a las diferentes páginas web a partir de un cierto tipo de vector asociado a una matriz. El análisis de dicho algoritmo proporciona un buen tema práctico para un curso de *Álgebra lineal*. Debido a su complejidad (y a la falta de tiempo) pondremos aquí otro sencillo (pero muy famoso) ejemplo que podríamos usar: la conocida *sucesión de Fibonacci* (véase por ejemplo http://www.youtube.com/watch?v=DKGsBUxRcV0). Imaginemos que una pareja de conejos tarda un mes en alcanzar la edad fértil, y a partir de ese momento cada vez engendra otra pareja de conejos, que a su vez (tras llegar a la edad de la fertilidad) engendrarán cada mes una pareja de conejos. ¿Cuántos conejos habrá al cabo de un determinado número de meses?



En forma matricial

$$\begin{pmatrix} u_{n+1} \\ u_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_n \\ u_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2 \begin{pmatrix} u_{n-1} \\ u_{n-2} \end{pmatrix} = \ldots = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} u_1 \\ u_0 \end{pmatrix}.$$

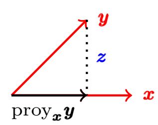
¿Cuánto vale A^n?¿Qué pasaría si A fuese diagonal?
A no es diagonal, pero... ¿podemos relacionarla con alguna matriz diagonal?

El lector interesado puede encontrar otras aplicaciones de esta teoría por ejemplo, en el estudio de los juegos de mesa (Calabuig et al., 2013).

Tras el estudio de la diagonalización de matrices podemos centrarnos en el caso ortogonal. Para ello podríamos "revisitar" nuestro concepto de aplicación lineal (recordemos que en nuestro caso se limita a una transformación determinada por una matriz) para llegar al concepto de matriz/transformación ortogonal (matriz cuadrada P, que cumple que $P^T P = I$, donde I es la matriz identidad) como aquella que mantiene las distancias vía la fórmula

$$(P\boldsymbol{x}|P\boldsymbol{y}) = (P\boldsymbol{x})^T \cdot (P\boldsymbol{y}) = \boldsymbol{x}^T \cdot (P^T \cdot P) \cdot \boldsymbol{y} = \boldsymbol{x}^T \cdot \underbrace{(P^T \cdot P)}_{I} \cdot \ \boldsymbol{y} = (\boldsymbol{x}|\boldsymbol{y}).$$

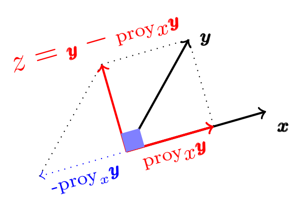
En este punto aparece de manera natural la conexión entre matrices simétricas (aquellas matrices A que cumplen que $A^T = A$) con las ortogonales a través de la caracterización: *una matriz A es simétrica si y, sólo si, es diagonalizable ortogonalmente, es decir, existen P ortogonal y D diagonal tal que $A = PDP^T$.* Aquí se pueden introducir algunas de las propiedades importantes de las matrices simétricas (por ejemplo que sus valores propios son reales y que vectores propios correspondientes a valores propios distintos son vectores ortogonales) así como las de las matrices ortogonales —a saber en este caso que, la inversa de una matriz ortogonal coincide con su traspuesta, las columnas (y filas) de una matriz ortogonal forman un conjunto *ortonormal* (y recíprocamente) y, como hemos comentado las matrices ortogonales conservan el producto escalar y, por lo tanto la norma, la distancia y el ángulo—. En la parte final de este apartado llegamos a otro de los temas clásicos en las asignaturas de *Álgebra lineal* y que de nuevo aparece aquí de una manera natural. Como acabamos de decir todo matriz simétrica (recordemos que de momento todas nuestras matrices son reales) se puede diagonalizar ortogonalemente, pero...¿cómo se obtiene la matriz ortogonal P? Como ha de ser ortogonal sus columnas han de ser vectores ortonormales. La matriz P sabemos que está formada por los vectores propios de A y según acabamos de comentar los vectores propios de valores propios distintos siempre son ortogonales pero...¿son ortogonales dos vectores propios de un mismo valor propio? La respuesta sabemos que es negativa lo que nos lleva al problema de construir dos vectores ortogonales a partir de dos que no los son y, en general, al bien conocido *Método de ortogonalización de Gram-Schmidt*. Para ello introducimos la noción de proyección ortogonal.



$$\text{proy}_{\boldsymbol{x}}\boldsymbol{y} = \frac{(\boldsymbol{y}|\boldsymbol{x})}{(\boldsymbol{x}|\boldsymbol{x})}\boldsymbol{x} = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|^2}\boldsymbol{x}.$$

$$\boldsymbol{z} = \boldsymbol{y} - \text{proy}_{\boldsymbol{x}}\boldsymbol{y} = \boldsymbol{y} - \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|^2}\boldsymbol{x}.$$

Esto nos conduce al método (a la izquierda vemos la representación gráfica para el caso de dos vectores):



Si $\boldsymbol{x} = \boldsymbol{0}$ definimos $\text{proy}_x\boldsymbol{y} = \boldsymbol{0}$.

Fijados $\boldsymbol{x}$ e $\boldsymbol{y}$, el vector $\boldsymbol{x}$ es ortogonal a $\boldsymbol{z} = \boldsymbol{y} - \text{proy}_x\boldsymbol{y}$.

Método de ortogonalización de Gram-Schmidt

Fijada una familia de vectores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$ construimos la nueva familia

$$\boldsymbol{y}_1 = \boldsymbol{x}_1$$

$$\boldsymbol{y}_2 = \boldsymbol{x}_2 - \text{proy}_{\boldsymbol{y}_1}\boldsymbol{x}_2$$

$$\boldsymbol{y}_3 = \boldsymbol{x}_3 - \text{proy}_{\boldsymbol{y}_1}\boldsymbol{x}_3 - \text{proy}_{\boldsymbol{y}_2}\boldsymbol{x}_3$$

$$\vdots$$

$$\boldsymbol{y}_p = \boldsymbol{x}_p - \text{proy}_{\boldsymbol{y}_1}\boldsymbol{x}_p - \text{proy}_{\boldsymbol{y}_2}\boldsymbol{x}_p - \ldots - \text{proy}_{\boldsymbol{y}_{p-1}}\boldsymbol{x}_p.$$

Entonces $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p$ es una familia ortogonal.

(5) En este punto ya podemos hacer la construcción de la *Descomposición en valores singulares* de una matriz A (cualquiera, no necesariamente cuadrada). Siguiendo de nuevo nuestra fórmula

$$A = \underbrace{(\boldsymbol{y}_1, \boldsymbol{y}_2)}_{U}\underbrace{\begin{pmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{pmatrix}}_{\Sigma}\underbrace{\begin{pmatrix}\boldsymbol{x}_1^T\\ \boldsymbol{x}_2^T\end{pmatrix}}_{V^T},$$

y teniendo en cuenta que U y V son matrices ortogonales (pues están formadas por columnas cuyos vectores son ortonormales) podemos ya definir la *Descomposición en valores singulares de una matriz A* como $A = U\Sigma V^T$ siendo U y V matrices ortogonales y Σ una matriz del mismo tamaño que A cuya diagonal está formada por los valores singulares de

A (que habitualmente se ordenan de mayor a menor). La primera gran diferencia a observar es que, a diferencia de la *Diagonalización* esta descomposición *siempre existe*. Para el cálculo podríamos seguir el siguiente método constructivo a través de unas propiedades sencillas de demostrar que permiten ir afianzando los resultados estudiados hasta el momento:

- En primer lugar dada la matriz A de tamaño $m \times n$ entonces la nueva matriz $B = A^T A$ cuadrada de tamaño n es una matriz simétrica luego puede diagonalizarse ortogonalmente. Existen entonces $P = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = V$ ortogonal y $D = (\lambda_1, \ldots, \lambda_n)$ diagonal tal que $A^T A = PDP^T$. Más aún, es fácil probar que, sus valores propios son números reales no negativos. Suponemos los valores propios de $A^T A$ ordenados en forma decreciente, considerando la posibilidad de tener algunos nulos
$$\lambda_1 \geq \ldots \geq \lambda_r > 0 = \lambda_{r+1} = \ldots = \lambda_n.$$
Así los valores singulares de A son las raíces cuadradas (positivas) de los valores propios de $A^T A$. Los denotamos por $\sigma_i = \sqrt{\lambda_i}$ para $1 \leq i \leq n$. Como antes los ordenamos
$$\sigma_1 \geq \ldots \geq \sigma_r > 0 = \sigma_{r+1} = \ldots = \sigma_n.$$
- Ahora podemos probar que el conjunto $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r\}$ definido por $\boldsymbol{y}_i = \sigma_i^{-1} A \boldsymbol{x}_i$, es un conjunto ortonormal de $\mathbb{R}^m$. Más aún si $\widetilde{U} = \big(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r\big)$, las ecuaciones
$$\widetilde{U}^T \boldsymbol{z} = \mathbf{0},$$
definen el subespacio ortogonal de $\mathsf{span}\big(\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r\}\big)$, $\mathsf{span}\big(\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r\}\big)^{\perp}$.
Sea $\{\boldsymbol{y}_{r+1}, \ldots, \boldsymbol{y}_m\}$ una base ortonormal de $\mathsf{span}\big(\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r\}\big)^{\perp}$. Entonces la matriz cuadrada de tamaño m
$$U = \big(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r, \boldsymbol{y}_{r+1}, \ldots, \boldsymbol{y}_m\big),$$
es ortogonal.
- Sea Σ la matriz diagonal $m \times n$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$. Con esta construcción toda matriz A de tamaño $m \times n$ se puede factorizar en la forma, $A = U \Sigma V^T$, siendo U cuadrada de tamaño m y V cuadrada de tamaño n ambas matrices ortogonales y Σ una matriz diagonal formada por los valores singulares de A. Además $rang(A) = r$ (que es el número de valores singulares no nulos, donde $rang(A)$ representa el rango de A). Nótese que las matrices U y V de la DVS no son únicas.

Para una lista de propiedades importantes e interesantes de esta construcción remitimos de nuevo al lector al libro de Meyer (2000).

4 La DVS y los métodos numéricos

En esta última sección presentamos la conexión de la descomposición con los métodos numéricos. Esto se podría implementar en algunas sesiones de laboratorio de la asignatura. Presentamos para ello dos resultados a utilizar: el *Teorema del rango aproximado* y la conocida *Aproximación lineal mediante mínimos cuadrados*.

Uno de los resultados importantes de la *Descomposición en valores singulares* es el *Teorema del rango aproximado*. A nivel teórico este resultado proporciona la respuesta a la siguiente pregunta de este trabajo:

> *Fijada una matriz A con rango r. ¿Es posible encontrar una matriz A_k con rango $k < r$ que esté lo más cerca posible (en un sentido a determinar) de A?*

Como ya vimos en la segunda sección (para el caso de matrices cuadradas de tamaño 2) la *Descomposición en valores singulares* de una matriz A de tamaño $m \times n$ se puede escribir como la suma de matrices (de rango uno) vía la fórmula $A = \sigma_1(\boldsymbol{y}_1 \cdot \boldsymbol{x}_1^T) + \ldots + \sigma_r(\boldsymbol{y}_r \cdot \boldsymbol{x}_r^T)$, donde r es el rango de A, $\boldsymbol{y}_j$ denota la columna j-ésima de U y $\boldsymbol{x}_j$ denota la columna j-ésima de V. Desde un punto de vista computacional para almacenar los datos de A necesitaríamos mn variables mientras que, como para cada expresión de la forma $\sigma\big(y \cdot \boldsymbol{x}^T\big)$ serían necesarias $m+n+1$ variables, entonces a la vista de la fórmula anterior el número de entradas para almacenar los datos de A sería $min\big(mn, (m+n+1)r\big)$. Así la propia descomposición tendría ya su utilidad si el rango de A es muy pequeño.

Sin embargo esta teoría tiene un *plus adicional*. El *Teorema del rango aproximado* responde a la cuestión planteada arriba de una manera positiva y además garantiza que las matrices buscadas son exactamente las descritas mediante la expresión:

$$A_k = \sigma_1(\boldsymbol{y}_1 \cdot \boldsymbol{x}_1^T) + \ldots + \sigma_k(\boldsymbol{y}_k \cdot \boldsymbol{x}_k^T), \quad k < r.$$

El lector puede encontrar la aplicación de este resultado en el tratamiento de compresión de imágenes digitales descrito, por ejemplo, en Domínguez (2011). Centramos nuestra atención en el sentido de la expresión “estar lo más cerca posible” de nuestra pregunta a resolver.

(6) A partir de la norma de $\mathbb{R}^n$ podemos introducir la noción de *Norma de A inducida por la norma de $\mathbb{R}^n$* (trabajamos aquí con la norma euclídea) vía

$$\|A\| = \max\{\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} : \boldsymbol{x} \neq \boldsymbol{0}\} = \max\{\|A\boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\}.$$

Vamos a probar que para el cálculo de la $\|A\|$ se puede utilizar la *Descomposición en valores singulares* de A. En efecto:

- Sabemos que $A = U\Sigma V^T$ donde U y V son matrices ortogonales. Probamos en primer lugar que $\|A\| = \|\Sigma\|$. Para ello recordamos que una de las características de las matrices ortogonales es que mantienen la norma. Esto significa que $\|A\boldsymbol{x}\| = \|U\Sigma V^T\boldsymbol{x}\| = \|\Sigma V^T\boldsymbol{x}\|$ pues U es ortogonal. Ahora bien, como V es ortogonal también lo es V^T luego $\|V^T\boldsymbol{x}\| = \|\boldsymbol{x}\|$ y

$$\begin{aligned}\|A\| &= \max\{\|A\boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\} = \max\{\|\Sigma V^T\boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\} \\ &= \max\{\|\Sigma V^T\boldsymbol{x}\|_2 : \|V^T\boldsymbol{x}\| = 1\} = \|\Sigma\|.\end{aligned}$$

- Veamos ahora $\|\Sigma\| = \sigma_1$. Por una parte si tomamos $\|\boldsymbol{x}\| = 1$ entonces

$$\|\Sigma\boldsymbol{x}\| = \sqrt{\sum_{j=1}^{r} \big|\sigma_j x_j\big|^2} \leq \sqrt{\sum_{j=1}^{r} \big|\sigma_1 x_j\big|^2} = \sigma_1\sqrt{\sum_{j=1}^{r} \big|x_j\big|^2} \leq \sigma_1\|\boldsymbol{x}\| = \sigma_1.$$

De esta manera tenemos $\|\Sigma\| \leq \sigma_1$. Pero si tomamos $\boldsymbol{x}_0 = \mathbf{e}_1 = (1, 0, \ldots, 0)$ como $\|\boldsymbol{x}_0\| = 1$ entonces

$$\|\Sigma\| = \max\{\|\Sigma\boldsymbol{x}\| : \|\boldsymbol{x}\| = 1\} \geq \|\Sigma\boldsymbol{x}_0\| = \sigma_1.$$

En definitiva, $\|A\| = \sigma_1$, que es el mayor valor singular de A.

Este estudio se puede completar con otra norma matricial: la llamada *Norma de Frobenius*, que está muy relacionada con la *Descomposición en valores singulares* y que se define mediante la fórmula:

$$\|A\|_F = \sqrt{\sum_{j,k} |a_{j,k}|^2}.$$

Es sencillo demostrar que: i) $\|A\|_F^2 = \text{traza}(A^T A)$, ii) $\|PA\|_F = \|A\|_F$ si P es ortogonal y iii) $\|A\|_F = \|\Sigma\|_F$. Todo esto proporciona la respuesta final a nuestra pregunta:

Teorema del rango aproximado
Sea A una matriz real de rango r. De entre todas las matrices de rango $k \leq r$, la matriz $A_k = \sigma_1(\boldsymbol{y}_1 \cdot \boldsymbol{x}_1^T) + \ldots + \sigma_k(\boldsymbol{y}_k \cdot \boldsymbol{x}_k^T)$, es la que minimiza el error que es, además,

$$\|A - A_k\| = \sigma_{k+1}, \qquad \|A - A_k\|_F = \sqrt{\textstyle\sum_{j \geq k+1} |\sigma_j|^2}.$$

Para finalizar este trabajo presentamos ahora otro de los métodos clásicos de estudio en los grados de ingeniería: el *Método de aproximación de los mínimos cuadrados*. Así, nuestro último problema a resolver en este trabajo es:

Fijada una matriz A de tamaño $m \times n$ y un sistema de ecuaciones lineales $A \cdot \boldsymbol{x} = \boldsymbol{b}$ incompatible (sin solución). ¿Cuál sería el vector $\widehat{\boldsymbol{x}}$ de forma que el nuevo vector $\widehat{\boldsymbol{b}} = A \cdot \widehat{\boldsymbol{x}}$ esté lo más cercano posible a $\boldsymbol{b}$?

(7) Habitualmente en los grados de ingeniería el *Problema de aproximación de los mínimos cuadrados* se suele resolver utilizando las ecuaciones normales o vía la factorización QR. El uso de la *Descomposición en valores singulares* proporciona una forma alternativa de resolver este problema. Para ello consideremos la factorización $A = U \cdot \Sigma \cdot V^T$ y supongamos que A tenga rango r. Entonces

$$A\boldsymbol{x} = \boldsymbol{b} \Longrightarrow U \cdot \Sigma \cdot V^T \cdot \boldsymbol{x} = \boldsymbol{b} \Longrightarrow \Sigma \cdot V^T \cdot \boldsymbol{x} = U^T \cdot \boldsymbol{b}.$$

Como A tiene rango r entonces $\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$ y la respuesta a nuestra pregunta viene determinada por la fórmula $\widehat{\boldsymbol{x}} = V \cdot \Delta \cdot U^T \cdot \boldsymbol{b}$, con $\Delta = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$. Un estudio de los diferentes métodos puede consultarse en el libro de Ascher y Greig (Ascher & Greif, 2011).

Conclusiones

El método de *Descomposición en valores singulares* de una matriz enlaza directamente con un gran número de contenidos básicos del *Álgebra Lineal* que se imparten en la mayoría de asignaturas de los primeros cursos de los grados de ingeniería. Este método abre la puerta a numerosas aplicaciones que permiten conectar las matemáticas con los intereses de los alumnos a la hora de cursar un grado de ingeniería. En este artículo, no sólo hemos pretendido presentar las aplicaciones del método sino también mostrar como los resultados de la teoría se pueden imbricar en el desarrollo de los mismos. En contraposición con los ejemplos de diagonalización ortogonal, mucho más estándar pero más limitados a aplicaciones puramente académicas en la mayoría de temarios y libros de texto, la *Descomposición en Valores singulares* pone el foco en éstos y otros muchos conceptos básicos del Espacio Euclídeo al servicio de problemas prácticos más complejos, más conectados con la realidad, que pueden motivar a nuestros alumnos a una visión de las matemáticas que va más allá de la resolución de problemas, dando no sólo sentido práctico sino también real o “metamatemático” a conceptos como transformación, norma, proximidad, error, etc.

Referencias

Ascher, U.M., Greif, C. (2011).
A First Course in Numerical Methods.
Siam. Computational Science & Engineering.

Bretscher, O. (2014).
Linear Algebra with Applications. Fifth Edition.
Pearson New International Edition.

Calabuig, J. M., García-Raffi, L. M., Sánchez-Pérez, E. A. (2013).
Álgebra lineal y juegos de mesa.
Modelling in Science Education and Learning 6(15), 185–196.

Domínguez Jiménez, M. E. (2011).
Matrices: un modelo para las fotografías digitales.
Modelling in Science Education and Learning 4(13), 169–180.

Grossman, S. I. (2008).
Álgebra lineal. Sexta Edición.
Mc Graw Hill.

Lay, D. C. (2012).
Álgebra lineal y sus aplicaciones. Cuarta Edición
Pearson. Always Learning.

Meyer, C. D. (2000).
Matrix Analysis and Applied Algebra.
Siam. Computational Science & Engineering.