

Valores singulares. ¿Qué son?.¿Para qué sirven?

por

Ion Zaballa, Universidad del País Vasco-Euskal Herriko Unibertsitatea

Los valores singulares son objetos muy útiles en Álgebra Lineal Numérica y han recibido mucha atención durante los últimos años. El objetivo es introducirlos a una audiencia general de manera elemental aprovechando su interpretación geométrica. Han recibido diferentes nombres a lo largo de su historia, que recordaremos brevemente. Se presentarán también algunas aplicaciones mostrando su interés en dos campos: en el análisis numérico y en el tratamiento de imágenes. Debido a su fuerte naturaleza geométrica, en la presentación oral se utilizó un *GUI* (Graphical User Interface) de MATLAB¹ del que se expondrán aquí algunas imágenes que, esperamos, sirvan para facilitar la comprensión de los conceptos.

1. Una breve reseña histórica

Los valores singulares aparecen en varias partes de la matemática. El nombre con el que los presentamos aquí (valores singulares) es el que se ha consolidado en la comunidad de analistas numéricos, pero han sido denominados de diferentes maneras a lo largo de su no muy larga historia. El resumen que se presenta aquí está basado en dos textos: un excelente “report” realizado por uno de los pioneros y más afamados algebristas numéricos de la actualidad, G. W. Stewart [12], y en la introducción del capítulo 3 del segundo de los dos enciclopédicos tomos sobre Análisis Matricial escritos por R. Horn and Ch. Johnson [5].

El origen de los valores singulares se encuentra en el intento de los geóme-

¹MATLAB es una marca registrada de The Math Works, Inc.

tras del siglo XIX por conseguir, en lenguaje actual, la reducción de una forma cuadrática a forma diagonal mediante cambios de base ortogonales. La primera contribución en este sentido parece ser de Eugene Beltrami, un geómetra diferencial italiano, que intentando promover entre sus estudiantes el gusto por el estudio de las formas bilineales escribió un artículo en la revista *Giornale di Matematiche ad Uso degli Studenti Delle Universita*, [2].

Beltrami comienza con una forma bilineal $f(x, y) = x^T A y$, donde A es una matriz real de orden n , y viene a demostrar que existen matrices ortogonales Q_1 y Q_2 , $n \times n$, tales que

$$Q_1^T A Q_2 = \Sigma = \text{Diag}(\sigma_1, \dots, \sigma_n). \quad (10.1)$$



Eugene Beltrami (1835-1899)

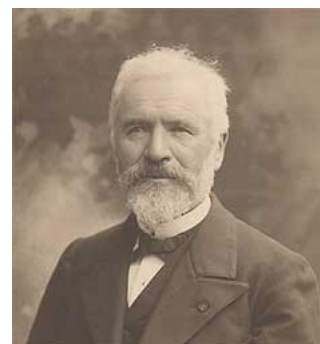
es una matriz diagonal de números reales positivos, los cuales son las raíces cuadradas de los valores propios de $A^T A$ (o de $A A^T$, dado que estas dos matrices simétricas son semejantes). En realidad, Beltrami propone, más que demuestra, la descomposición anterior cuando los elementos diagonales son todos distintos entre sí. En otras palabras, se puede considerar que Beltrami es el descubridor de lo que hoy en día llamaríamos la descomposición en valores singulares de matrices reales cuadradas no singulares con valores singulares distintos. Omite las posibles situaciones de degeneración que, como veremos, tienen importantes consecuencias.

Esta omisión no es debida a un intento de simplificar las cosas para sus estudiantes, sino que, tal y como se comprueba por las ideas que aporta, posiblemente sea consecuencia de no plantearse el problema en su completa generalidad.

Camille Jordan publicó un par de trabajos sobre formas bilineales, [7,8], un año más tarde que Beltrami, por lo que se le puede considerar codescubridor de los valores singulares.

Su aproximación es completamente diferente. De hecho, en el primero de los artículos citados, su preocupación es la de buscar el máximo y mínimo de la forma bilineal $P = x^T A y$ condicionados a que $\|x\|_2 = \|y\|_2 = 1$. Usando la condición de la anulación de la diferencial en los máximos y mínimos relativos, Jordan prueba que dicho máximo es la raíz cuadrada del mayor valor propio de la matriz

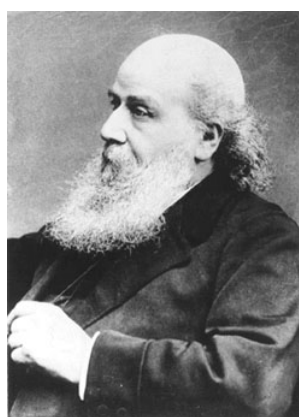
$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \quad (10.2)$$



Camille Jordan (1831-1921)

Es posiblemente la primera vez que se usa el hoy ampliamente utilizado proceso de *deflacción*, que consiste en reducir un problema sobre matrices al mismo problema con una matriz de orden una unidad inferior. La caracterización de los valores singulares de A como las raíces cuadradas positivas de los valores propios de la matriz en (10.2) es un hecho muy conocido y popular, aunque habitualmente atribuido a Wielandt (ver [3]) y Lanczos [9].

Desconocedor, aparentemente, de los trabajos de Beltrami y Jordan, J.J. Sylvester escribió una nota a pie de página y dos artículos sobre la diagonalización de formas bilineales mediante sustituciones ortogonales.



John Joseph Sylvester
(1814-1897)

La nota a pie de página apareció en un artículo de la revista *The Messenger of Mathematics* [13]. El artículo lleva por título “A new proof that a general quadric may be reduced to its canonical form (that is, a linear function of squares) by means of a real orthogonal substitution”, y en él Sylvester propone un algoritmo para reducir una forma cuadrática a forma diagonal. En la nota a pie de página señala que una iteración similar es posible para diagonalizar una forma bilineal y dice que lo ha enviado para su publicación en las *Comptes Rendues* [14]. Esto nos da una idea de que no conocía el artículo original de Jordan publicado más de una década antes.

En palabras de Stewart estos artículos de Sylvester “no se leen fácilmente. El estilo es opaco y Sylvester pontifica sin probar, dejando demasiados detalles al lector”. Aparentemente el autor estima sus resultados como importantes, pero, sin poner en duda que lo son, hoy en día se ven más como un redescubrimiento de resultados ya obtenidos por otros. En cualquier caso, Sylvester es el primero en dar nombre a los números reales positivos que aparecen en la forma canónica diagonal: *multiplicadores canónicos*.

Aunque hay nuevos resultados en el terreno del álgebra lineal ligados a los valores singulares, damos un salto en el tiempo para redescubrirlos en un área de la matemática bien distinta: las ecuaciones integrales, uno de los temas que recibió más atención durante las primeras décadas del siglo XX.

En 1907, Erhard Schmidt (el del famoso procedimiento de Gram-Schmidt) publicó, [11], una teoría general de ecuaciones integrales reales con núcleos simétricos y no simétricos. Introduce los conceptos de *valor propio* y de *función propia* de un núcleo $K(x, y)$ continuo en $[a, b] \times [a, b]$ mediante las ecuaciones integrales

$$u(x) = \lambda \int_a^b K(x, y)v(y)dy \text{ y } v(y) = \lambda \int_a^b K(x, y)u(x)dx.$$

Y muestra, a continuación, que λ debe ser real porque λ^2 es un valor propio del núcleo simétrico definido positivo

$$H(x, y) = \int_a^b K(x, t)K(t, y)dt.$$



Erhard Schmidt
(1876-1959)

Si pensamos en $K(x, y)$ como lo análogo a la matriz A , entonces $H(x, y)$ es análogo a $A^T A$. De hecho, Schmidt transita, en dimensión infinita, por senderos similares a los de Beltrami.

Un poco después de la publicación del artículo de Schmidt, Bateman, [1], introduce la notación de *valores singulares* para referirse a unos números que son esencialmente los recíprocos de los valores propios definidos por Schmidt. Y Picard, [10], hace notar que para los núcleos simétricos los valores propios de Schmidt son reales y en este caso (pero no en general) los llama valores singulares.



Hermann Weyl
(1885-1955)

A pesar de todo, la denominación no se estabiliza, y entre los muchos matemáticos notables que han trabajado en las propiedades de los valores singulares, debemos nombrar a Hermann Weyl con importantes contribuciones a la teoría de perturbación de los mismos. Pero en cuanto a nomenclatura se refiere, todavía en su artículo [15] habla de “dos clases de valores propios” y en una traducción al inglés de un tratado ruso sobre operadores no autoadjuntos, Gohberg y Krein, [4], se refieren a los valores singulares como “ s -números” de un operador.

La fuerte implantación de los ordenadores y el consecuente apogeo del análisis numérico ha contribuido a que finalmente la denominación de valores singulares haya arraigado en la comunidad matemática.

2. ¿Qué son los valores singulares?

Podríamos responder a esta pregunta con una demostración: la del teorema de descomposición de una matriz u operador lineal en valores singulares y definirlos a partir del mismo tras probar su unicidad y unas cuantas propiedades más. Pero teniendo como objetivo la divulgación y como motivación la geometría podemos hacernos una perfecta idea de lo que son los valores singulares paseando por las

avenidas más amplias y mejor conocidas de la geometría.

Para responder a la pregunta ¿qué son los valores singulares? podemos formular otra pregunta: ¿cuál es la imagen de la circunferencia unidad (la de centro el origen de coordenadas y radio 1) por una aplicación lineal?. Posiblemente muchos de los potenciales lectores de estas notas sabrán responder a esta pregunta rápidamente. Para aquellos que no lo saben o albergan alguna duda, podemos diseñar, con ayuda de MATLAB, un pequeño experimento que nos aproximará a la respuesta con facilidad. De hecho, toda esta presentación irá acompañada de gráficas que explican, pero no demuestran, los resultados que iremos exponiendo y que aparecerán como resultado de una interacción con un GUI (Graphical User Interface) diseñado con MATLAB. La pantalla principal de este GUI se muestra en la Figura 1.

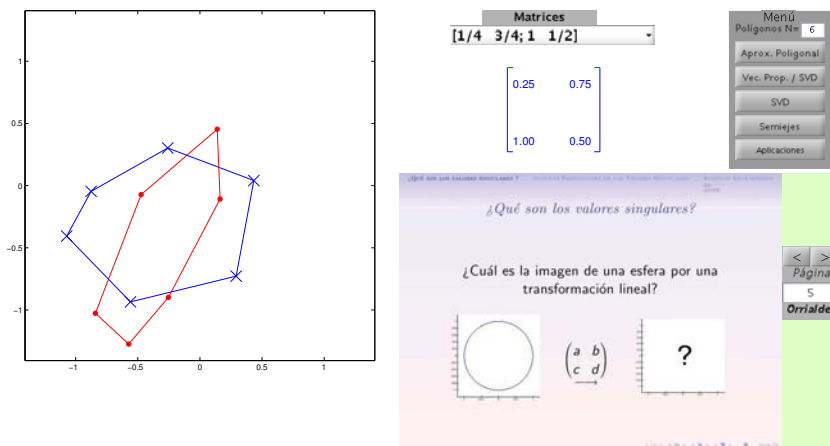


Figura 1: Pantalla principal del GUI con sus cuatro componentes. Arriba a la derecha el menú de opciones para presentar las propiedades principales de los valores singulares. Arriba en el centro la matriz con la que se hacen las operaciones y sobre la que se calculan los valores singulares; justo encima de la matriz dibujada hay un menú para elegir nuevas matrices. A la izquierda los ejes en los que se dibujarán las figuras geométricas que ponen de manifiesto las propiedades de los valores singulares. Y abajo a la derecha las páginas con las transparencias que se utilizaron para apoyar la presentación oral.

El experimento consiste en lo siguiente: lo que sí sabemos todos es que la imagen de un segmento recto por una aplicación lineal es otro segmento recto (quizá degenerado a un punto); y en consecuencia la imagen de un polígono es otro polígono. Esto es lo que se presenta en la ventana de gráficos de la Figura 1. Hemos elegido representar un polígono de $N = 6$ lados; hemos señalado los puntos de intersección de los lados con el símbolo \times y MATLAB ha hecho dos cosas: ha dibujado el hexágono correspondiente en azul (para quienes puedan verlo en color), y ha calculado las imágenes de los lados del hexágono produciendo

un hexágono en rojo (de nuevo para quien pueda verlo en color) y las intersecciones de los lados las ha marcado con el símbolo \bullet . Pero inmediatamente surge una pregunta ¿es esto siempre así?. Es decir, ¿es la imagen de un hexágono por una aplicación lineal siempre otro hexágono? ¿No puede ser un pentágono o un heptágono, por ejemplo?. Desde luego con el polígono de la Figura 2 uno podría tener sus dudas; y con los de la Figura 3 no cabe ninguna: la imagen del hexágono ha colapsado a un segmento. Claro que enseguida se aprecia que, a diferencia de los dos casos anteriores, la matriz del último ejemplo es singular: tiene sus dos columnas linealmente dependientes.

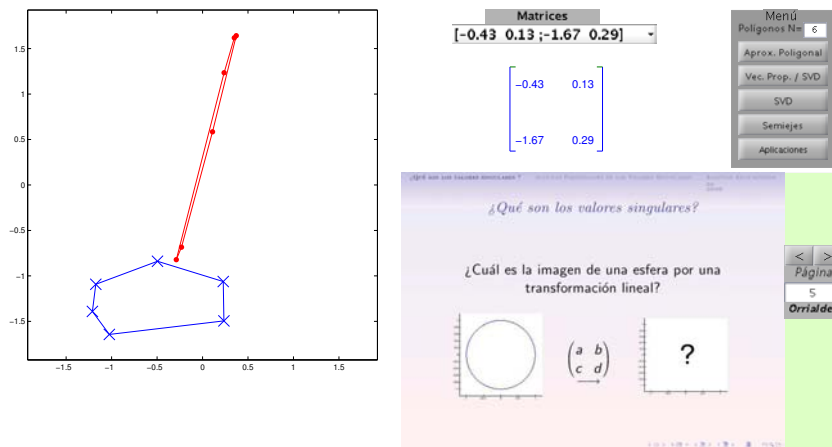


Figura 2: La imagen del hexágono por la matriz $\begin{bmatrix} -0,43 & 0,13 \\ -1,67 & 0,29 \end{bmatrix}$ es casi un segmento.

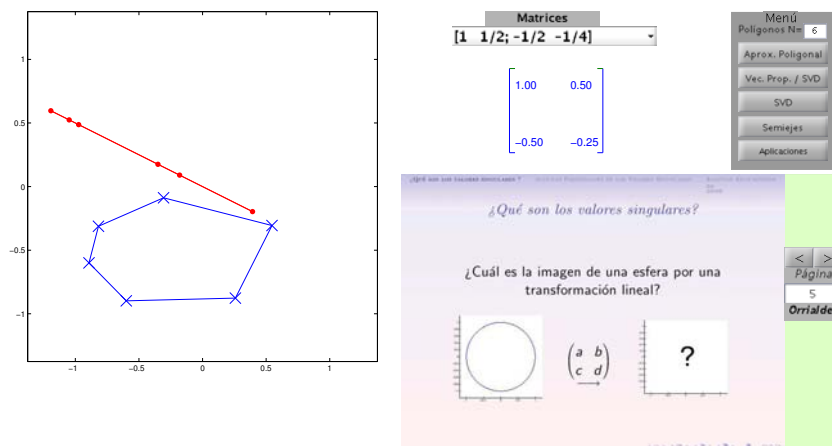


Figura 3: La imagen del hexágono por la matriz $\begin{bmatrix} 1 & 0,5 \\ -0,5 & -0,25 \end{bmatrix}$ es exactamente un segmento.

Bien, puede ser que un hexágono se convierta, mediante una transformación lineal, en un segmento, pero todavía queda por responder si se puede convertir en pentágono o cuadrilátero, o heptágono, etc.. Se deja a cargo del lector comprobar que las únicas posibilidades para la imagen de un hexágono son: un hexágono, un segmento o un punto. Claro que debemos advertir que para demostrarlo puede necesitar lápiz y papel y conocimientos básicos de Álgebra Lineal.

Sigamos con nuestro experimento. Admitamos que los polígonos por aplicaciones lineales se transforman en polígonos con el mismo número de lados, en segmentos o en puntos. ¿En qué nos ayuda esto para saber cuál es la imagen de la circunferencia unidad? La idea es inscribir un cuadrado en la circunferencia unidad y pedirle a MATLAB que dibuje su imagen por la matriz que elijamos; que luego duplique el número de lados del polígono inscrito en la circunferencia unidad (un octógono) y dibuje, de nuevo, su imagen. La impresión que obtendremos es que se va obteniendo la circunferencia unidad y podremos comprobar la figura que resulta en las imágenes. Lo hacemos con dos matrices:

$$y \begin{bmatrix} 1 & 0,5 \\ -0,5 & -0,25 \end{bmatrix}.$$

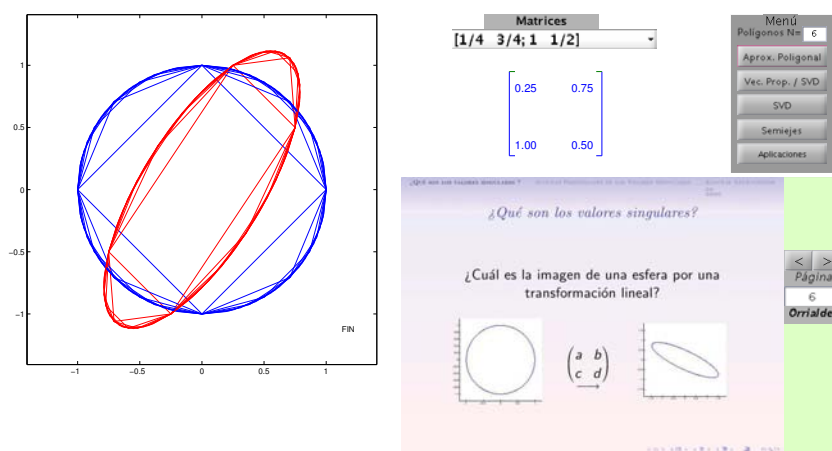


Figura 4: La imagen de la circunferencia unidad por la matriz $\begin{bmatrix} 0,25 & 0,75 \\ 1 & 0,5 \end{bmatrix}$.

Los resultados los tenemos en las Figuras 4 y 5. En la primera vemos que, como la matriz es no singular (sus dos columnas son linealmente independientes), la imagen de la circunferencia es una elipse centrada en el origen de coordenadas, mientras que en la segunda, como podíamos sospechar ya por los ejemplos anteriores, la imagen es un segmento. Podríamos añadir que si la matriz utilizada fuera la matriz cuyos elementos son todos cero, la imagen sería un punto: el $(0, 0)$.

Así pues podemos concluir, como se muestra en la transparencia de las Figuras 4 y 5, que la imagen de la circunferencia unidad por una transformación lineal es una elipse centrada en el origen de coordenadas; claro que ésta puede ser degenerada: un segmento o un punto. Aunque todos nuestros experimentos los haremos en dimensión 2, conviene decir aquí que para dimensiones superiores tenemos un resultado similar: la imagen de la esfera unidad en un espacio de dimensión n es una hiperelipse que puede ser degenerada; es decir, el número de semiejes puede ser inferior a n , y habitar, en consecuencia, en un espacio de dimensión menor que n .

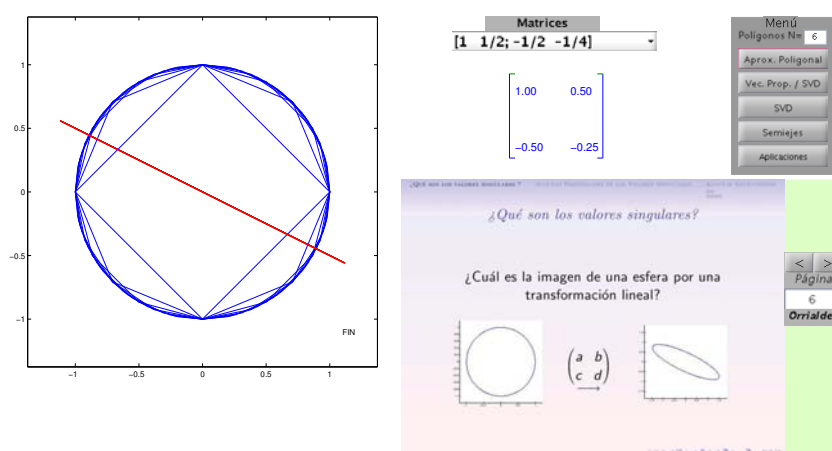


Figura 5: La imagen de la circunferencia unidad por la matriz $\begin{bmatrix} 1 & 0,5 \\ -0,5 & -0,25 \end{bmatrix}$.

Ya estamos en condiciones de presentar nuestros objetos de estudio: los valores singulares.

Definición 2.1 Los *valores singulares* de una matriz A son las longitudes de los semiejes de la hiperelipse en que se convierte la esfera unidad por A .

Son, por lo tanto, números reales no negativos (porque se admiten semiejes de longitud 0): $s_1 \geq s_2 \geq \dots \geq 0$.

A cada matriz le podemos asociar sus valores singulares, pero puede haber, y en general hay, muchas matrices que produzcan la misma elipse como imagen de la circunferencia unidad. Todas ellas tendrán los mismo valores singulares. Se dice que son *ortogonalmente equivalentes*. (En el caso de matrices con valores complejos se habla de matrices unitariamente equivalentes, pero nosotros sólo trataremos el caso real). En la siguiente sección veremos a qué se debe la denominación “ortogonalmente”.

3. Algunas propiedades de los valores singulares

Podemos usar MATLAB para realizar otro experimento que nos muestra, una vez más, que la imagen de la circunferencia unidad por una matriz es una elipse. Pero el objetivo ahora es otro: poner de manifiesto esa propiedad que con tanto trabajo e ingenio demostraron Beltrami, Jordan y Sylvester: las formas bilineales pueden reducirse mediante sustituciones ortogonales a forma diagonal.

El experimento es el siguiente: fijada una matriz, dibujamos un par de vectores unitarios (de longitud 1) x , y que sean perpendiculares (ortogonales). Éstos aparecen en verde (para quien pueda verlo en color) en la Figura 6. MATLAB ha calculado sus imágenes por la matriz expuesta y ha dibujado los vectores Ax y Ay en azul.

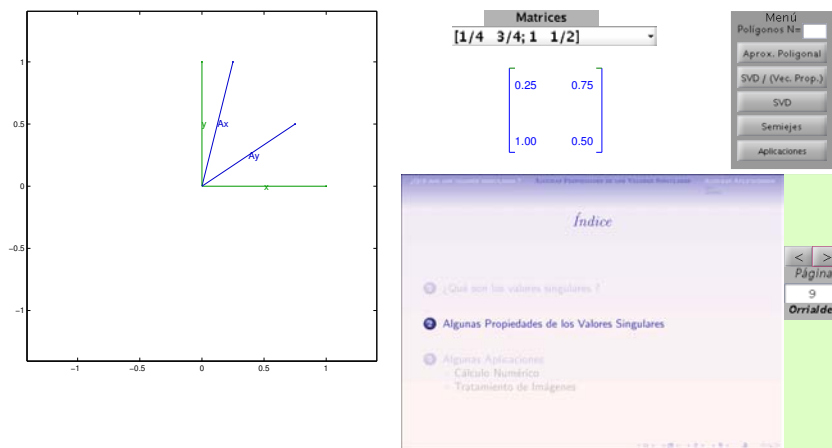


Figura 6: Vectores ortogonales y sus imágenes.

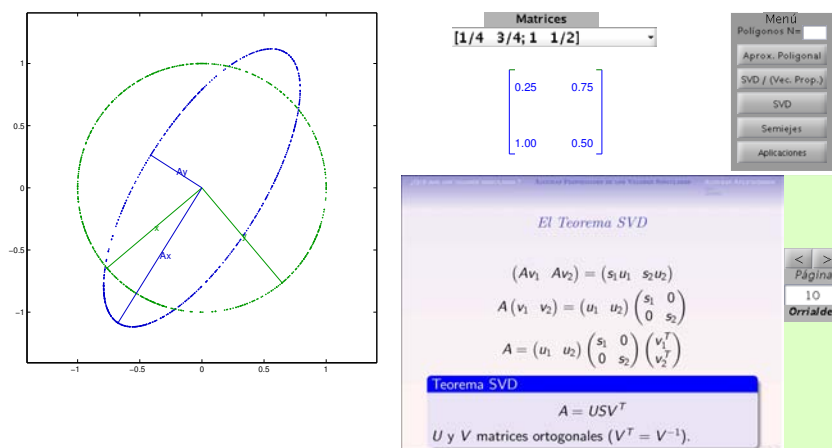


Figura 7: Una posición de x , y para la que Ax y Ay son los semiejes de la elipse.

Ahora, utilizando el ratón del ordenador podemos hacer girar los vectores x , y manteniendo siempre la ortogonalidad y su longitud. Como éstos están pegados al

origen de coordenadas ambos describirán la circunferencia unidad. Sus imágenes Ax , Ay , ya sabemos, deben describir una elipse con centro en el origen de coordenadas. Lo interesante es la siguiente observación que se pone de manifiesto en la Figura 7: hay una posición de los vectores x , y tal que sus imágenes son los semiejes de la elipse; es decir, tal que también Ax y Ay son perpendiculares. De hecho no hay una sino dos posiciones de los vectores x , y para los que Ax , Ay son los semiejes, pero en la Figura 7 sólo presentamos uno de los dos casos posibles.

Esta propiedad tiene una interpretación algebraica muy interesante. Por una parte, si s_1 y s_2 son los valores singulares de A , como son las longitudes de los semiejes, existen vectores u_1 y u_2 de longitud 1 tales que s_1u_1 y s_2u_2 son los vectores que forman los semiejes de la elipse; y por lo tanto u_1 y u_2 son ortogonales. Por otra parte, el experimento recién realizado nos dice que para ciertos vectores v_1 y v_2 ortogonales de longitud 1 (los vectores x , y de la Figura 7, Av_1 y Av_2 son los semiejes de la elipse. Entonces

$$Av_1 = s_1u_1 \quad \text{y} \quad Av_2 = s_2u_2.$$

Podemos poner esto en forma matricial:

$$A \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}.$$

Y como los vectores u_1 y u_2 son ortogonales tenemos que poniendo

$$Q_1^T = \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} \quad \text{y} \quad Q_2 = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

entonces $Q_1^T Q_1 = I_2$ y

$$Q_1^T A Q_2 = \text{Diag}(s_1, s_2),$$

que es el resultado que demostró Beltrami (10.1), y que hoy en día se conoce con el nombre de Teorema de Descomposición en Valores Singulares o, abreviadamente, Teorema SVD por sus iniciales en inglés.

Por otra parte, los valores singulares pueden caracterizarse también algebraicamente. En efecto, s_1 es el mayor valor singular y su significado geométrico es la longitud del mayor de los semiejes de la elipse en que se convierte la circunferencia unidad por una transformación lineal. Ahora bien, en la circunferencia unidad están representados todos los vectores de longitud 1 del plano. Es decir, la circunferencia unidad es el lugar geométrico de los vectores de longitud, o norma euclídea, igual a 1. Simbólicamente

$$\text{Circunferencia unidad} = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}.$$

Entre todos estos vectores hay uno, digamos v , tal que la longitud de Av es el semieje mayor de la elipse. Y precisamente el hecho de que sea la longitud del semieje mayor significa, además, que los transformados de todos los demás vectores de la circunferencia unidad por A nunca tienen longitud mayor que s_1 . Esto lo podemos escribir de la siguiente forma:

$$s_1 = \max_{\|x\|_2=1} \|Ax\|_2;$$

s_1 nos da la mayor longitud alcanzable por los vectores Ax para x en la circunferencia unidad.

Esta caracterización proporciona ciertas propiedades interesantes de s_1 que cualquier persona mínimamente relacionada con las propiedades del \max puede demostrar sin dificultad:

1. Para cualquier matriz A , $s_1(A) \geq 0$ y $s_1(A) = 0$ si y sólo si $A = 0$.
2. Para cualquier matriz A y cualquier número α $s_1(\alpha A) = |\alpha|s_1(A)$.
3. Para cualesquiera matrices A y B , $s_1(A + B) \leq s_1(A) + s_1(B)$

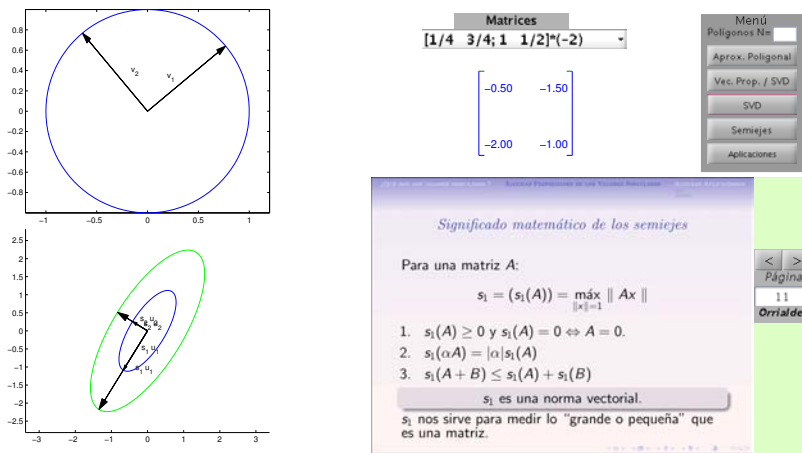


Figura 8: Las imágenes de la circunferencia unidad por dos matrices. La proporcionalidad entre éstas se transmite a las imágenes.

La Figura 8 nos muestra una imagen sugerente de la segunda propiedad. En ella se han dibujado las imágenes de la circunferencia unidad por las matrices

$$A = \begin{bmatrix} 0,25 & 0,75 \\ 1 & 0,5 \end{bmatrix} \quad \text{y} \quad B = -2A = \begin{bmatrix} -0,5 & -1,5 \\ -2 & -1 \end{bmatrix}$$

y se observa que una elipse es el “doble” que la otra.

Es bien conocido que las tres condiciones caracterizan las normas. En este caso definen **una norma** en el espacio de las matrices. Las normas sirven para medir lo grande o pequeña que son las matrices y la distancia entre ellas. En otras palabras, el mayor valor singular de una matriz es una norma (conocida con el nombre de norma de Hilbert, espectral o de operador) y nos sirve para tener una idea de lo grande o pequeña que es una matriz.

¿Hay alguna propiedad similar para el menor valor singular?. Desde luego, los argumentos empleados más arriba para mostrar que $s_1 = \max_{\|x\|_2=1} \|Ax\|_2$, sirven para ver que $s_n = \min_{\|x\|_2=1} \|Ax\|_2$. Pero s_n no es una norma de matriz. No obstante, hay una propiedad muy importante ligada al valor singular más pequeño que tiene que ver también con distancias. Para acercarnos a esta propiedad, y para dimensión 2, hemos diseñado un nuevo experimento. Tomemos como referencia la Figura 9. En ella ha desaparecido el menú de matrices y en su lugar aparece una sola matriz:

$$A = \begin{bmatrix} 0,8 & -0,6 \\ 0,6 & 0,8 \end{bmatrix}.$$

A su izquierda hay dos ejes. En el superior aparecen dos vectores: las columnas de la matriz A . Y en el inferior la elipse que es imagen por la matriz A de la circunferencia unidad con los semiejes s_1u_1 y s_2u_2 . En esta figura la imagen de la circunferencia unidad por la matriz A es una circunferencia. Las matrices que dejan invariante la circunferencia unidad son las matrices ortogonales. En efecto, las dos columnas de A son ortogonales y por eso los vectores dibujados en los ejes de arriba son perpendiculares y de longitud 1.

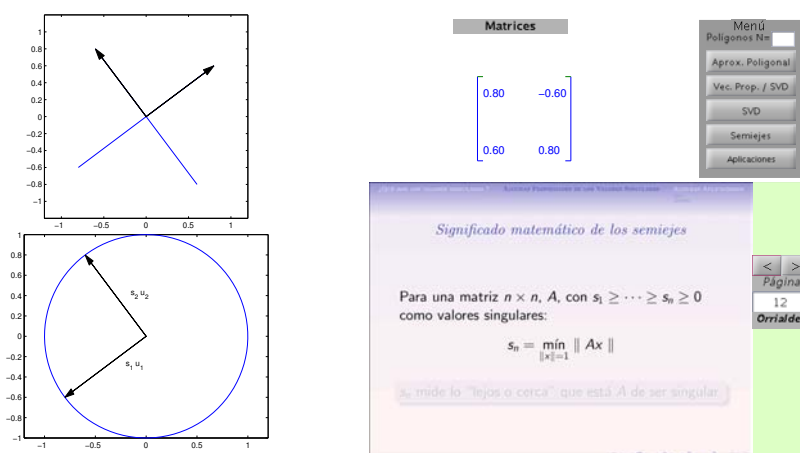


Figura 9: La circunferencia unidad es invariante por la matriz $A = \begin{bmatrix} 0,8 & -0,6 \\ 0,6 & 0,8 \end{bmatrix}$.

El experimento consiste en lo siguiente: a cada golpe de ratón uno de los vectores girará una cierta cantidad de grados hacia el otro, y éste permanecerá siem-

pre fijo. El vector que gira será siempre unitario. La matriz reflejará este hecho, la primera columna permanecerá fija y la segunda nos dará las coordenadas del nuevo vector que ha girado. En los ejes de abajo, se dibujará la elipse que es imagen de la circunferencia unidad por la nueva matriz con los correspondientes semiejes. En el experimento en directo se observa que a medida que el vector va girando al encuentro del otro, la elipse pasa de ser una circunferencia a una elipse cada vez más delgada (con el semieje menor cada vez más pequeño) hasta que en el límite la elipse se colapsa en un segmento. En las Figuras 10 y 11 presentamos dos momentos de este proceso: uno intermedio y el final.

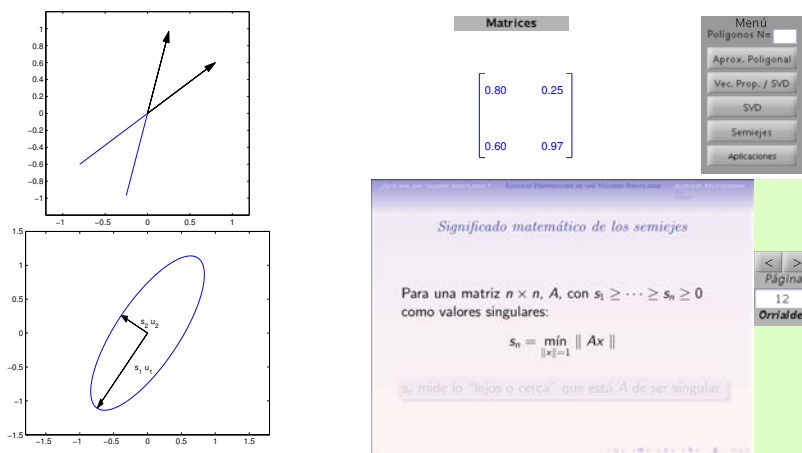


Figura 10: Los vectores columna de la matriz $\begin{bmatrix} 0,8 & 0,25 \\ 0,6 & 0,97 \end{bmatrix}$ y la elipse correspondiente.

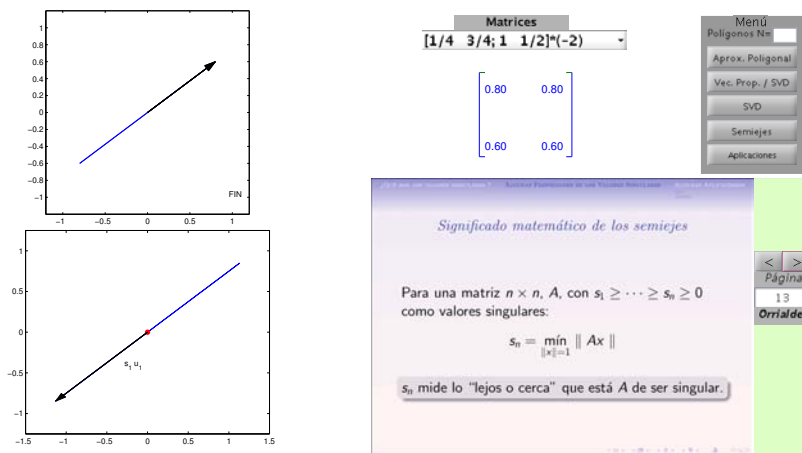


Figura 11: Cuando los dos vectores de la matriz son linealmente dependientes la elipse es un segmento.

Cuanto más se parecen las dos columnas de A más pequeño es s_2 . Lo grande o pequeño que sea este número nos da una idea de si las columnas de A son “muy linealmente independientes” o “muy poco linealmente independientes”. En otras palabras, s_2 nos da una idea de lo cerca o lejos que está A de ser singular. De hecho este es un resultado muy importante que podemos escribir, para dimensión n , de la siguiente manera:

$$s_n = \min_{\text{rang} B \leq n-1} \|A - B\|_2 \quad (10.3)$$

Las implicaciones de este resultado en análisis numérico son muy importantes. Apenas veremos una en la próxima sección.

4. Algunas aplicaciones de los valores singulares

Presentamos un par de campos donde los valores singulares tienen aplicaciones. Dentro de las matemáticas son fundamentales en cálculo numérico, y en especial, en álgebra lineal numéricas. Fuera de las matemáticas, mostramos su utilidad en el tratamiento y compresión de imágenes. Hay algoritmos muy eficientes para la compresión de imágenes que no usan los valores singulares, pero mostraremos aquí esta utilidad más que nada como una curiosidad.

Hay muchos otros ámbitos en los que los valores singulares aportan conocimiento. Basta, por ejemplo, escribir *singular values and google* en este buscador de internet para comprobar la multitud de citas disponibles.

4.1 Cálculo numérico

Los valores singulares son los objetos fundamentales para calcular el rango de una matriz. Todo paquete de cálculo numérico que se precie (MATLAB, MATHEMATICA, MAPLE, SCILAB, ...) usa los valores singulares para determinar el rango de las matrices y si son fiables los resultados que dichos paquetes devuelven al resolver problemas tales como la resolución de sistemas lineales, los ajustes por mínimos cuadrados o el cálculo de valores propios.

Nos centraremos en el rango. Por el Teorema SVD sabemos que existen matrices ortogonales Q_1 y Q_2 tales que

$$Q_1^T A Q_2 = \Sigma = \text{Diag}(s_1, \dots, s_n).$$

Esto implica que $\text{rang} A = \text{rang} \Sigma$ y éste es el número de valores singulares distintos de cero.

Este es un resultado teórico muy interesante pero cualquier persona con ciertos conocimientos de cálculo numérico podría argumentar que hay otros métodos más económicos de calcular el rango de una matriz. Y es cierto, pero no más fiables.

En primer lugar, una de las consecuencias de la caracterización de s_n en (10.3) es que el conjunto de las matrices de rango completo es abierto porque todas las matrices que estén a una distancia de A menor que s_n tienen rango completo. Pero es que además este conjunto también es denso. Consecuencia: como cualquier cálculo sobre una matriz sufre, cuando menos, de los errores de redondeo, rara vez se calcula con la matriz dada exactamente sino con alguna aproximada. En el supuesto de que la matriz dada sea singular, podría perfectamente darse el caso de que la aproximada fuera no singular porque la densidad indica que tan cerca como queramos de la matriz dada tenemos matrices no singulares. El cálculo del rango se vuelve, por este tipo de consideraciones, una verdadera pesadilla.

¿En qué ayudan los valores singulares?. Uno de los resultados aportados por Hermann Weyl es el siguiente resultado de perturbación: si $\text{rang}A = k$ y $\tilde{A} = A + E$ entonces $\|E\|_2^2 \geq s_{k+1}^2 + \dots + s_n^2$. Es decir, puede que $\text{rang}\tilde{A} = n$ pero si los últimos valores $n - k$ valores singulares son muy pequeños entonces ¡ojo! puede que la matriz real tenga rango k . Un buen paquete de cálculo numérico debería poner en guardia al usuario sobre una situación de este tipo para que éste pueda tomar las decisiones pertinentes. No es posible diseñar un algoritmo para calcular el rango de una matriz que pueda dar el resultado correcto con absoluta fiabilidad en todas las situaciones posibles. Por ejemplo, el siguiente código de MATLAB calcula los valores singulares de una matriz 20×20 de números escogidos aleatoriamente en el intervalo $(0, 1)$ con distribución normal. No se muestra la matriz producida, y se muestra sólo el valor singular s_{20} , para ahorrar espacio:

```
Tomamos una matriz aleatoriamente
A=randn(20);
echo off
Calculemos sus valores singulares
s=svd(A);
s(20)
ans =
    0.0255
echo off
Según este resultado el rango debe ser 20

rank(A)
ans =
    20
```

El último valor singular es $s(20) = 0,0255$. Una de dos, la matriz tiene rango 20 o es producto de una perturbación $A + E$ con $\|E\|_2 \geq 0,0255$, con lo que no se

parecería nada a la matriz original. Lo más probable es que la matriz tenga rango 20, y esa es la respuesta que produce MATLAB.

Ahora bien, consideremos este otro ejemplo con matrices de Hilbert. Las matrices de Hilbert son de la siguiente forma:

$$H = \text{hilb}(n), \quad \text{con} \quad h_{ij} = \frac{1}{i+j-1}.$$

Por ejemplo

$$\text{hilb}(3) = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

Son matrices no singulares y se conoce una fórmula para el elemento que ocupa la posición (i, j) de la inversa de $\text{hilb}(n)$:

$$(H^{-1})_{ij} = (-1)^{i+j}(i+j-1) \binom{n+i-1}{n-j} \binom{n+j-1}{n-i} \binom{i+j-2}{i-1}^2.$$

El siguiente código de MATLAB tiene como objetivo calcular los valores singulares y el rango de una matriz de Hilbert de dimensión 12:

Formamos la matriz de Hilbert 12x12

```
H=hilb(12);
```

```
echo off
```

Calculamos sus valores singulares

```
s=svd(H);
```

```
ans =
```

```
Columns 1 through 4
```

```
1.795372059561997e+00    3.802752459550370e-01
```

```
4.473854875218111e-02    3.722312237891171e-03
```

```
Columns 5 through 8
```

```
2.330890890217861e-04    1.116335748323230e-05
```

```
4.082376110364057e-07    1.122861066357240e-08
```

```
Columns 9 through 12
```

```
2.251964521323359e-10    3.111347477116982e-12
```

```
2.650150683111359e-14    1.055555522518574e-16
```

El rango es 12, pero

```
rank(H)
```

```
ans =
```

```
11
```


Vemos que la respuesta de MATLAB es incorrecta. Claro $s_{12} = 1,055 \cdot 10^{-16}$, es prácticamente cero. De hecho MATLAB no puede distinguir este número de cero porque hace redondeo en la 16ª cifra decimal devolviendo como respuesta que el rango es 11. Es como si, teniendo en cuenta el resultado de Weyl, tomara la decisión (en realidad, la decisión es del programador de MATLAB) de que la matriz calculada es una perturbación de una matriz de rango 11. En cualquier caso, es el usuario el que debe tomar la decisión oportuna. Los valores singulares pueden ayudar a tomar esa decisión.

4.2 Tratamiento de imágenes



Figura 12: Reproducción del grabado *Melencolia I* de Albrecht Dürer

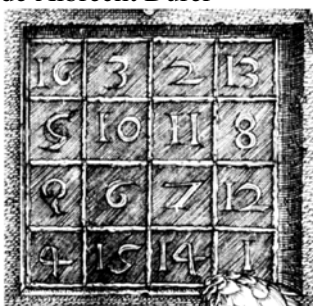


Figura 13: Detalle del cuadrado mágico del grabado *Melencolia*.

En esta última parte, vamos a tratar muy superficialmente el tratamiento de imágenes digitales. Tomaremos como ejemplo el sugerente grabado *Melencolia* de Albrecht Dürer (Figura 12). Hay cantidad de información sobre esta obra maestra de Dürer en la web. La elección se debe a la manifiesta afición del artista a las matemáticas y en el hecho de que este grabado reúne un gran número de símbolos matemáticos. Así, se puede distinguir una esfera, una semicircunferencia o un poliedro. Y sobre todo, en la parte superior derecha, un cuadrado mágico cuyo detalle se presenta en la Figura 13.

Hay muchas leyendas en torno a los cuadrados mágicos y toda una tradición de esoterismo. Hay cuadrados mágicos de todas las dimensiones a partir de $n = 3$ y, aparte de las propiedades mágicas que se les atribuyen, tienen propiedades matemáticas interesantes. La propia definición ya es una: suman lo mismo todas sus filas, todas las columnas, la diagonal y la contradiagonal.

Pero lo que nos interesa a nosotros ahora es que cada una de estas imágenes digitales no es más que una matriz de “pixels”. En concreto, el cuadrado mágico está almacenado en mi ordenador como una matriz de tamaño 359×371 pixels, y el grabado, en el que centraremos nuestra atención, es una matriz de tamaño 736×566 pixels. El objetivo es comprimir esta imagen sin perder mucha resolución. Es decir, tratar de conseguir la misma imagen, quizá con unos tonos de grises

menores, con muchos menos datos.

Llamemos A a la matriz de tamaño 736×566 que contiene los datos necesarios para obtener la imagen de la Figura 12. Se necesitan, por tanto 416576 números en punto flotante (digamos, datos) para conseguir la imagen. ¿Cómo se pueden usar los valores singulares para obtener esta misma imagen con menos datos?. La idea es la siguiente: De acuerdo con el Teorema SVD, existen matrices ortogonales U y V tales que $A = U\Sigma V^T$, siendo U una matriz de tamaño 736×736 , V de tamaño 566×566 , ambas ortogonales, y

$$\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}, \quad S = \text{Diag}(s_1, s_2, \dots, s_{566}),$$

con $s_1 \geq s_2 \geq \dots \geq s_{566}$ los valores singulares de A .

El número de filas especificadas como cero en Σ es, cuando menos, $170 = 736 - 566$, así que las últimas 170 últimas columnas de U no juegan ningún papel. Vamos a denotar como U_1 la submatriz de U formada por sus primeras 566 columnas. Podemos escribir

$$A = U_1 S V^T. \quad (10.4)$$

Si u_i representa la i -ésima columna de U_1 y v_i^T la i -ésima fila de V^T entonces

$$A = \sum_{i=1}^{566} s_i u_i v_i^T. \quad (10.5)$$

Esto es fácil de ver. Lo mejor es un ejemplo con números más pequeños. Supongamos que A es 4×3 entonces

$$\begin{aligned} A &= \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \end{bmatrix} \begin{bmatrix} s_1 & & \\ & s_2 & \\ 0 & 0 & s_3 \\ & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix} = \\ & \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} s_1 & & \\ & s_2 & \\ & & s_3 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix} = \\ & \begin{bmatrix} s_1 u_1 & s_2 u_2 & s_3 u_3 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix} = s_1 u_1 v_1^T + s_2 u_2 v_2^T + s_3 u_3 v_3^T. \end{aligned}$$

Volviendo a nuestra matriz original, cada una de las matrices $s_i u_i v_i^T$ es de tamaño 736×566 , pero sólo se necesitan $736 + 566 + 1 = 1303$ datos para formarlas: los 736 elementos de u_i , los 566 de v_i^T y 1 de s_i . A la matriz $s_i u_i v_i^T$ se le

llama el i -ésimo **modo** de A . Así $s_1 u_1 v_1^T$ es el primer modo de A , $s_2 u_2 v_2^T$ es el segundo modo, etc. Estos modos no están determinados de forma única por A , pero eso ahora no tiene importancia. Lo que sí es importante es que algunos de estos modos son más dominantes que otros para formar la matriz A . Esto no es difícil de comprender. Empecemos con un ejemplo sencillo. En la Figura 14 el vector dibujado tiene componentes $(0,1,1)$. La segunda componente es dominante respecto de la primera de modo que la dirección dominante de este vector es la del eje de ordenadas. Si la primera componente fuera del orden, digamos, de 10^{-3} no podríamos distinguir gráficamente dicho vector del vector $(0,1)$.

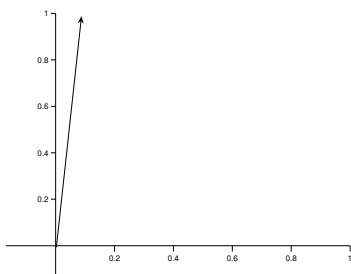


Figura 14: Vector con la segunda componente dominante.

Algo parecido ocurre con la matriz A : las matrices $u_i v_i^T$ son matrices de rango 1 de forma que determinan subespacios de dimensión 1, que podemos identificar con direcciones en \mathbb{R}^{566} . Además, debido a que los vectores u_i y v_i son ortogonales, estas direcciones son mutuamente ortogonales. El valor s_i nos da la coordenada de A en la dirección $u_i v_i^T$. Si hay valores singulares mucho mayores que otros, las direcciones de estos valores singulares son más determinantes en la formación de A que las de los valores singulares pequeños.

Veamos cómo se concreta todo esto con la matriz A que contiene los datos de la imagen de la Figura 12. El siguiente código de MATLAB calcula los valores singulares de A y matrices U y V que cumplen la condición (10.4), y los dibuja usando una escala logarítmica porque hay valores singulares muy grandes:

```
>> [U,S,V]=svd(X,0);
>> VT=V';
>> semilogy(diag(S),'b-o');
>> set(gca,'XGrid','on','YGrid','on');
```

El resultado es la gráfica de la Figura 15. En ella se aprecia que los 100 primeros valores singulares dominan a los 100 últimos, por ejemplo. Aquellos son más determinantes en la formación de la matriz A que éstos.

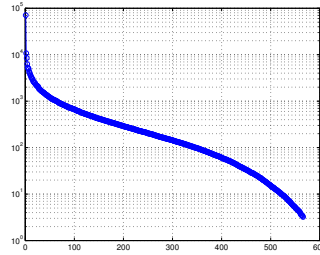


Figura 15: Vector con la segunda componente dominante.

Aproximando A por la suma de sus primeros modos conseguimos matrices que toman en consideración las direcciones fundamentales de A . Las imágenes de estas matrices deberán parecerse a las de A a medida que el número de modos que tomemos en consideración aumente. Sin embargo, con pocos modos conseguimos resultados bastante buenos. Observemos, antes de exponer los resultados, que, por ejemplo, 150 modos requieren (recordemos que cada modo precisa de 1303 datos) $150 \times 1303 = 195450$ datos; menos de la mitad de los datos necesarios para formar la matriz A completa.

En las siguientes figuras se presentan dos imágenes: la de la izquierda es la imagen original (Figura 12) formada con la matriz A y la de la derecha la de la matriz obtenida sumando los primeros m modos de A , con $m = 10, 50, 100$ y 150 .



Figura 16: Imagen de *Melencolia* usando 10 modos: 13030 datos.



Figura 17: Imagen de *Melencolia* usando 50 modos: 65150 datos. ($\approx 15\%$)



Figura 18: Imagen de *Melencolia* usando 100 modos: 130300 datos.($\approx 31\%$)



Figura 19: Imagen de *Melencolia* usando 150 modos: 195450 datos.($\approx 47\%$)

Con 100 modos la resolución es bastante buena, pero con 150 modos (menos de la mitad de los datos de la matriz original) no se distingue fácilmente el original de la copia.

Antes de acabar, es preciso insistir de nuevo que hay algoritmos más económicos para el tratamiento de imágenes y que no utilizan los valores singulares. Véase, por ejemplo, [6].

5. Nota final

El GUI de MATLAB que sirvió para hacer la presentación de la conferencia cuyo contenido se recoge en estas notas está disponible en la página web del autor:

www.ehu.es/izaballa/paseo_por_la_geometria05.

Se ha de tener en cuenta que, dada su complejidad, el GUI no está completamente depurado y una utilización incorrecta puede causar problemas. En la misma página web hay disponibles unas notas para su utilización. Por supuesto, la presencia de MATLAB corriendo por debajo del GUI es imprescindible.

Agradecimiento: Estas notas, la conferencia y el GUI de MATLAB no habrían sido realizadas si la invitación de los organizadores, profesora Marta Macho y

profesor Raúl Ibañez, de los magníficos ciclos que bajo la nombre de *Paseos por la geometría* animan a profesores y estudiantes a acercarse a la matemática en su vertiente divulgativa. Gracias Marta y Raúl por vuestra consideración y paciencia.

Bibliografía

- [1] H. Nateman, *A formula for the solving function of a certain integral equation of the second kind*, Transactions of the Cambridge Philosophical Society 20, 179-187, 1908
- [2] E. Beltrami, *Sulle funzioni bilineari*, Giornale di Matematiche ad Uso degli Studenti Delle Università 11, 98-106, 1873. Traducido al inglés por D. Boley: Technical Report, University of Minnesota, Department of Computer Science, 90-37, 1990.
- [3] K. Fan, A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proceedings of the American Mathematical Society 6, 111-116, 1955.
- [4] I. C. Gohberg, M. G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*, American Mathematical Society, 1969.
- [5] R. Horn, Ch. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.
- [6] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.
- [7] C. Jordan, *Mémoire sur les formes bilinéaires*, Journal de Mathématiques Pures et Appliquées 19, 35-54, 1874.
- [8] C. Jordan, *Sur la réduction des formes bilinéaires*, Comptes Rendues de l'Académie des Sciences 78, 614-617, 1874.
- [9] C. Lanczos, *Linear systems in self-adjoint form*, American Mathematical Monthly 65, 665-679, 1948.
- [10] É. Picard, *Sur un théorème général relatif aux intégrales de première espèce et sur quelques problèmes de physique mathématique*, Rendicongi del Circolo Matematico di Palermo 25, 79-97, 1910.
- [11] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener*, Mathematische Annalen 63, 433-476, 1907.
- [12] G.W. Stewart, *On the Early History of the Singular Value Decomposition*, Report disponible mediante ftp anónimo en la dirección tales.cs.umd.edu en el directorio pub/reports.
- [13] J.J. Sylvester, *A new proof that a general quadric may be reduced to its canonical form (that is, a linear function of squares) by means of a real orthogonal substitution*, The Messenger of Mathematics 19, 1-5, 1889.
- [14] J.J. Sylvester, *Sur la réduction biorthogonale d'une forme linéo-linéaire à sa forme canonique*, Comptes Rendues de l'Académie des Sciences 108, 651-653,

1889.

[14] H. Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proceedings of the National Academy of Sciences 35, 408-411, 1949.

Ion Zaballa

Universidad del País Vasco-Euskal Herri-
ko Unibertsitatea

Departamento de Matemática Aplicada,
Estadística e Investigación Operativa

Apdo. Correos 644. 48080 Bilbao

e-mail: ion.zaballa@ehu.es

www.ehu.es/jizaballa

