

8

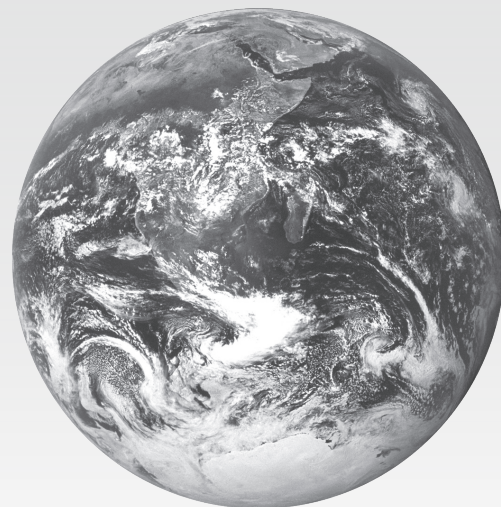
Matrices simétricas y formas cuadráticas

COMPETENCIAS POR DESARROLLAR

- 1 Modelar y resolver problemas de optimización restringida de la forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sujeta a $\mathbf{x}^T \mathbf{x} = 1$ y del tipo $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sujeta a las restricciones $\mathbf{x}^T \mathbf{x} = 1$ y $\mathbf{x}^T \mathbf{u}_1 = 0$, utilizando la diagonalización ortogonal de la matriz simétrica $\mathbf{A}$ de la forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$. Estos problemas surgen en economía, física, estadística, etcétera.

ACTIVIDADES DE APRENDIZAJE

- ◆ Identificar si una matriz es simétrica.
- ◆ Determinar si una matriz es ortogonal.
- ◆ Encontrar los valores propios de una matriz simétrica y una base ortogonal para el espacio propio que corresponde a cada uno de ellos.
- ◆ Diagonalizar una matriz simétrica.
- ◆ Encontrar una diagonalización ortogonal para una matriz simétrica.
- ◆ Construir una descomposición espectral de una matriz simétrica.
- ◆ Utilizar algún programa de cómputo para encontrar los valores propios de una matriz simétrica $\mathbf{A}$ y una base ortonormal para el núcleo de la matriz $\mathbf{A} - \lambda \mathbf{I}$, con λ valor propio de $\mathbf{A}$.
- ◆ Analizar las propiedades enunciadas en el teorema espectral para matrices simétricas.
- ◆ Encontrar la forma cuadrática correspondiente a una matriz simétrica $\mathbf{A}$.
- ◆ Expresar una forma cuadrática como $\mathbf{x}^T \mathbf{A} \mathbf{x}$ donde $\mathbf{A}$ es una matriz simétrica.
- ◆ Observar que la forma cuadrática asociada a una matriz diagonal no tiene términos de producto cruzado.
- ◆ Efectuar un cambio de variable que transforme una forma cuadrática en otra sin términos de producto cruzado y proporcionar la nueva forma cuadrática.
- ◆ Determinar si una forma cuadrática dada es positiva/negativa definida, indefinida, positiva/negativa semidefinida.
- ◆ Graficar algunas formas cuadráticas en $\mathbb{R}^2$ que correspondan a elipses o hipérbolas en posición estándar.
- ◆ Graficar algunas formas cuadráticas que correspondan a elipses o hipérbolas rotadas, haciendo un cambio de variable para llevarlas a la posición estándar, indicando cuáles son los ejes principales.
- ◆ Determinar el valor máximo de una forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sujeta a la restricción $\mathbf{x}^T \mathbf{x} = 1$, y un vector unitario en el cual se alcance este máximo, utilizando valores y vectores propios de la matriz simétrica $\mathbf{A}$.
- ◆ Determinar el valor mínimo de una forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sujeta a la restricción $\mathbf{x}^T \mathbf{x} = 1$, y un vector unitario en el cual se alcance este mínimo, utilizando los valores propios y los vectores propios de la matriz simétrica $\mathbf{A}$.
- ◆ Determinar el valor máximo de una forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$ sujeta a las restricciones $\mathbf{x}^T \mathbf{x} = 1$ y $\mathbf{x}^T \mathbf{u}_1 = 0$ donde $\mathbf{u}_1$ es un vector propio unitario correspondiente al mayor de los valores propios de la matriz simétrica $\mathbf{A}$.





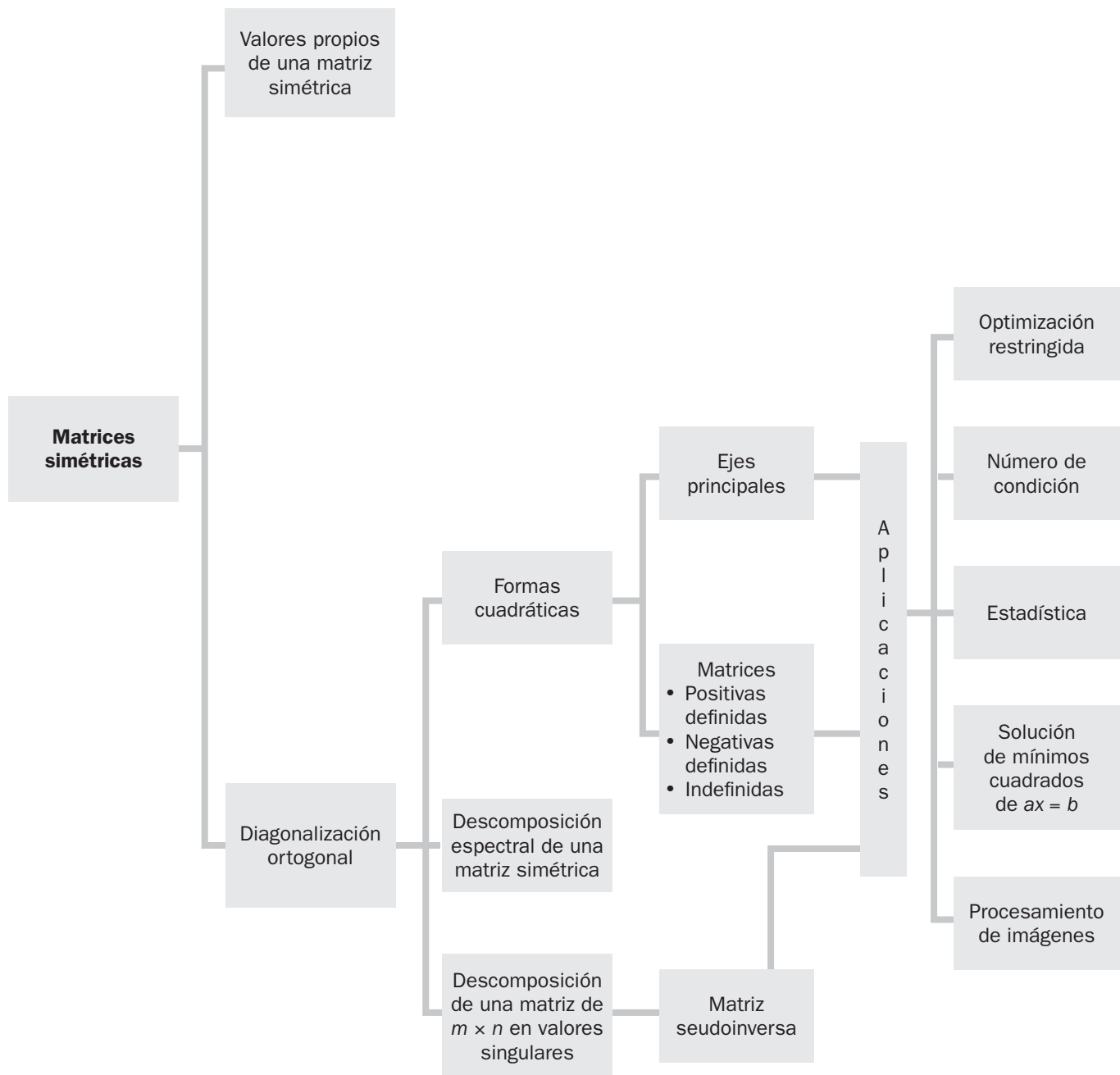
COMPETENCIAS POR DESARROLLAR

- 2 Modelar y resolver problemas de aplicación en estadística, física, etc., en los que se requiere la diagonalización ortogonal de una matriz o la descomposición en valores singulares.
- 3 Modelar y resolver problemas de aplicación en estadística, física, etc., en los que se requiere llevar a cabo el análisis de componentes principales.

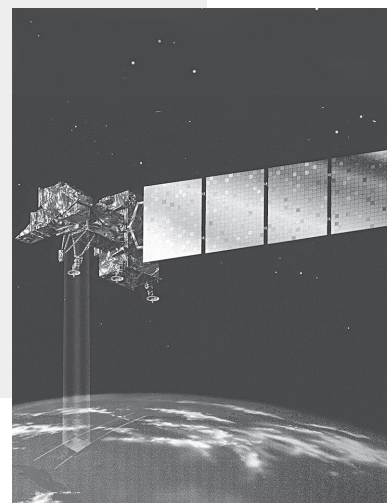
ACTIVIDADES DE APRENDIZAJE

- ◆ Encontrar un vector unitario $\mathbf{x}$ en donde se maximice $\|A\mathbf{x}\|$ con A una matriz de $m \times n$, y calcular esta longitud máxima.
- ◆ Para una matriz A de $m \times n$, encontrar los valores propios de $A^T A$ y los valores singulares de A .
- ◆ Maximizar la forma cuadrática $\mathbf{x}^T(A^T A)\mathbf{x}$ sujeta a la restricción $\|\mathbf{x}\| = 1$.
- ◆ Encontrar una descomposición en valores singulares de una matriz A de $m \times n$.
- ◆ Determinar el rango de una matriz si se conoce su descomposición en valores singulares.
- ◆ Utilizar algún paquete de cómputo para encontrar los valores singulares de una matriz, una descomposición en valores singulares y determinar su número de condición.
- ◆ Encontrar la descomposición en valores singulares reducida de una matriz A de $m \times n$ y la pseudoinversa de A . Resolver la ecuación $A\mathbf{x} = \mathbf{b}$ utilizando la pseudoinversa de A .
- ◆ Convertir una matriz de observaciones a la forma de desviación media y construir la matriz de covarianza muestral.
- ◆ Encontrar las componentes principales de un conjunto de datos.
- ◆ Dada una matriz de covarianza de datos encontrar la primera componente principal de los datos y calcular el porcentaje de la varianza total contenida en este componente.
- ◆ Si x_1, x_2 son las variables para un conjunto de datos bidimensional, encontrar una nueva variable y_1 de la forma $y_1 = c_1x_1 + c_2x_2$, con $c_1 + c_2 = 1$, tal que y_1 tenga la máxima varianza posible sobre los datos dados y determinar qué porcentaje de los datos se explica mediante y_1 .

ORGANIZADOR GRÁFICO



Procesamiento de imágenes multicanal



Alrededor del mundo, en poco más de 80 *minutos*, los dos satélites Landsat cruzan silenciosamente el firmamento en órbitas casi polares, grabando imágenes del terreno y las líneas costeras, en franjas de 185 kilómetros de ancho. Luego de 16 días, cada satélite habrá transitado por encima de casi todos los kilómetros cuadrados de la superficie de la Tierra, de manera que cualquier zona se puede monitorizar cada ocho días.

Las imágenes Landsat son útiles para muchos objetivos. Los diseñadores y planificadores urbanos las utilizan para estudiar la rapidez y dirección del crecimiento urbano, el desarrollo industrial y otros cambios en el uso del suelo. Esas imágenes permiten analizar la humedad del suelo en las zonas rurales, clasificar la vegetación de regiones remotas, y localizar arroyos y lagos tierra adentro. Los gobiernos pueden detectar y evaluar los daños de desastres naturales, como incendios forestales, flujos de lava, inundaciones y huracanes. Las agencias de protección ambiental pueden identificar contaminación por emisiones de humo y medir la temperatura del agua en lagos y ríos cerca de plantas de energía.

Los sensores satelitales toman siete imágenes simultáneas de cada región del suelo que se va a estudiar. Los sensores registran energía de bandas de longitud de onda separadas: tres en el espectro de luz visible y cuatro en las bandas infrarroja y térmica. Cada imagen se digitaliza y se almacena como un arreglo rectangular de números, donde

cada número indica la intensidad de la señal en un punto pequeño (o *pixel*) correspondiente en la imagen. Cada una de las siete imágenes es un canal de una *imagen multicanal* o *multiespectral*.

Las siete imágenes Landsat de una región fija, por lo general, contienen mucha información redundante, ya que algunos aspectos aparecen en varias imágenes. Sin embargo, otros aspectos, por su color o temperatura, reflejan luz que solo se registra en uno o dos de los sensores. Un objetivo del procesamiento de imágenes multiespectrales es el de visualizar los datos en una forma que permita extraer información de mejor calidad que cuando se estudia cada imagen por separado.

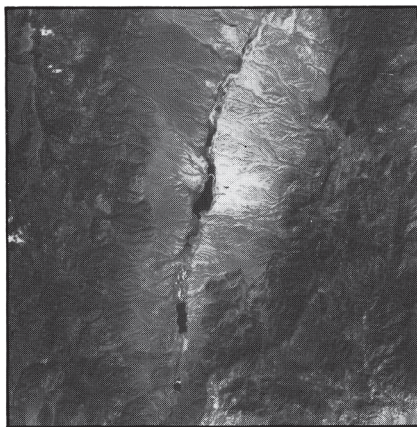
El *análisis de componentes principales* es una manera eficaz de suprimir información redundante y concentrar, en tan solo una o dos imágenes compuestas, la mayor parte de la información de los datos iniciales. A grandes rasgos, el objetivo es encontrar una combinación lineal especial de las imágenes, es decir, una lista de pesos que en cada pixel combinen los siete valores correspondientes de las imágenes en un nuevo valor. Los pesos se eligen de manera que permitan ampliar, en comparación con cualquiera de las imágenes originales, el rango de las intensidades de luz (la *varianza de la escena*) en la imagen compuesta (llamada la *primera componente principal*). También es posible obtener imágenes adicionales de *componentes*, mediante un criterio que se explicará en la sección 8.5.

En las fotografías tomadas sobre el Valle Railroad, Nevada, que se muestran a continuación se ilustra el análisis de componentes principales. En las fotografías *a)*, *b)* y *c)* se observan imágenes de tres bandas espectrales Landsat. La información total en las tres bandas se reacomoda en las tres imágenes de componentes principales en las fotografías *d)*, *e)* y *f)*. La primera componente *d)* presenta (o “explica”) el 93.5% de la varianza de la escena presente en los datos iniciales. De esta forma, los datos iniciales de tres canales

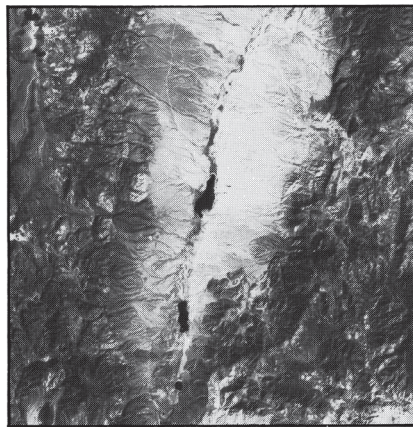
se reducen a datos de un canal, con un pérdida (en algún sentido) de solo el 6.5% de la varianza de la escena.

La empresa Earth Satellite Corporation, en Rockville, Maryland, que amablemente proporcionó las fotos que aquí se incluyen, está experimentando con imágenes de 224 bandas espectrales separadas. El análisis de componentes principales, esencial para esos conjuntos enormes de datos, por lo general reduce los datos a unas 15 componentes principales utilizables.

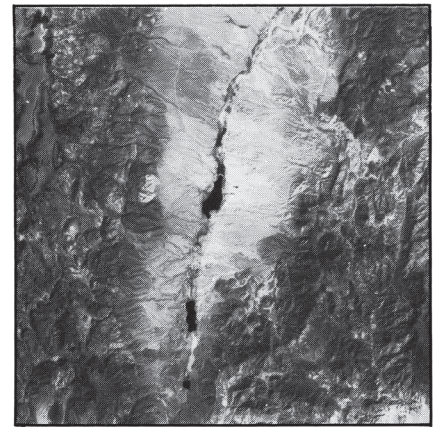
WEB



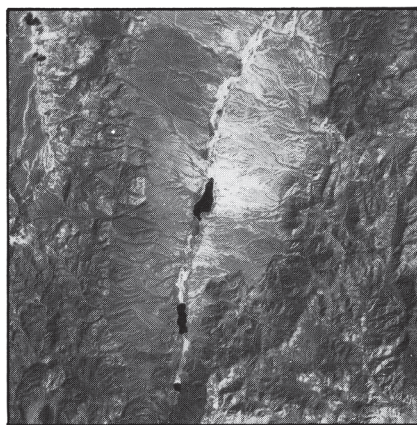
a) Banda espectral 1: Azul visible.



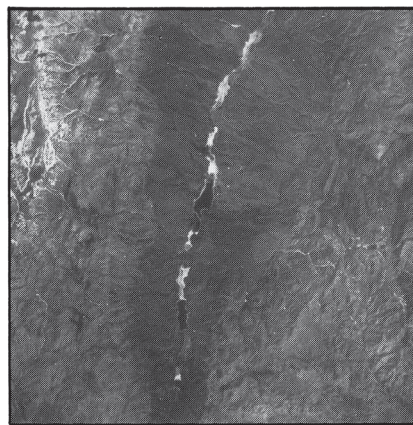
b) Banda espectral 4: Casi infrarrojo.



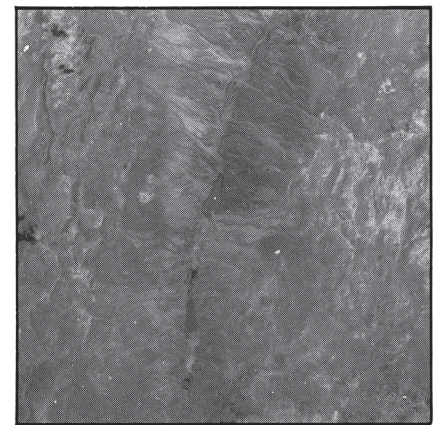
c) Banda espectral 7: Infrarrojo medio.



d) Componente principal 1: 93.5%.



e) Componente principal 2: 5.3%.



f) Componente principal 3: 1.2%.

En las aplicaciones, las matrices simétricas surgen, de una u otra manera, con mayor frecuencia que cualquier otra clase de matrices. La teoría es bella y rica, dependiendo en esencia de la diagonalización y la ortogonalidad, temas que se estudiaron en los capítulos 5 y 7, respectivamente. La diagonalización de una matriz simétrica, descrita en la sección 8.1, es el fundamento del análisis en las secciones 8.2 y 8.3 concernientes a formas cuadráticas. A la vez, la sección 8.3 es necesaria para las dos últimas secciones referentes a la descomposición en valores singulares y al procesamiento de imágenes descrito en el ejemplo introductorio. En este capítulo todos los vectores y las matrices tienen entradas reales.

8.1 DIAGONALIZACIÓN DE MATRICES SIMÉTRICAS

Una matriz **simétrica** es una matriz A tal que $A^T = A$. Esta matriz es necesariamente cuadrada. Sus entradas en la diagonal principal son arbitrarias, pero sus otras entradas se presentan por pares, en lados opuestos de la diagonal principal.

EJEMPLO 1 De las siguientes matrices, solo las primeras tres son simétricas:

$$\begin{aligned} \text{Simétricas: } & \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \\ \text{No simétricas: } & \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \quad \blacksquare \end{aligned}$$

Para comenzar el estudio de matrices simétricas, es útil revisar el proceso de diagonalización de la sección 7.3.

EJEMPLO 2 Si es posible, diagonalice la matriz $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

SOLUCIÓN La ecuación característica de A es

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

Los cálculos estándar producen una base para cada espacio propio:

$$\lambda = 8: \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = 6: \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}; \quad \lambda = 3: \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Estos tres vectores conforman una base para $\mathbb{R}^3$. De hecho, es fácil comprobar que $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ es una base *ortogonal* de $\mathbb{R}^3$. La experiencia del capítulo 5 sugiere que una base *ortonormal* sería útil en los cálculos, así que a continuación se presentan los vectores propios normalizados a uno:

$$\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Sean

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Entonces $A = PDP^{-1}$, como es usual. Pero esta vez, como P es cuadrada y tiene columnas ortonormales, resulta que P es una matriz *ortogonal*, y P^{-1} es simplemente P^T . (Véase la sección 5.2). $\blacksquare$

El teorema 1 explica por qué los vectores propios del ejemplo 2 son ortogonales, ya que corresponden a distintos valores propios.

TEOREMA 1

Si A es simétrica, entonces dos vectores propios cualesquiera de diferentes espacios propios son ortogonales.

DEMOSTRACIÓN Sean $\mathbf{v}_1$ y $\mathbf{v}_2$ los vectores propios asociados a los valores propios diferentes λ_1 y λ_2 . Para demostrar que $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, calcule

$$\begin{aligned}\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 && \text{Puesto que } \mathbf{v}_1 \text{ es un vector propio} \\ &= (\mathbf{v}_1^T A^T) \mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) && \text{Debido a que } A^T = A \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) && \text{Porque } \mathbf{v}_2 \text{ es un vector propio} \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2\end{aligned}$$

Por lo tanto, $(\lambda_1 - \lambda_2) \mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Pero $(\lambda_1 - \lambda_2) \neq 0$, así que $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. ■

El tipo especial de diagonalización del ejemplo 2 es crucial para la teoría de matrices simétricas. Se dice que una matriz A de $n \times n$ es **diagonalizable ortogonalmente** si existen una matriz ortogonal P (con $P^{-1} = P^T$) y una matriz diagonal D tales que

$$A = PDP^T = PDP^{-1} \quad (1)$$

Esta diagonalización requiere n vectores propios ortonormales linealmente independientes. ¿Cuándo es posible esto? Si A es diagonalizable ortogonalmente como en la ecuación (1), entonces

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Por lo tanto, ¡ A es simétrica! El teorema 2 que se presenta a continuación, muestra que, a la inversa, cada matriz simétrica es diagonalizable ortogonalmente. La demostración es mucho más difícil y se omite; la idea principal para la demostración se presentará después del teorema 3.

TEOREMA 2

Una matriz A de $n \times n$ es diagonalizable ortogonalmente si y solo si A es una matriz simétrica.

Este teorema es sorprendente, ya que el trabajo realizado en el capítulo 7 sugeriría que es prácticamente imposible decir cuándo una matriz es diagonalizable. Pero este no es el caso para las matrices simétricas.

El siguiente ejemplo se refiere a una matriz cuyos valores propios no son todos distintos.

EJEMPLO 3 Diagonalice ortogonalmente la matriz $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, cuya ecuación característica es

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

SOLUCIÓN Los cálculos usuales producen bases para los espacios propios:

$$\lambda = 7: \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = -2: \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

Aunque $\mathbf{v}_1$ y $\mathbf{v}_2$ son linealmente independientes, no son ortogonales. Recuerde de la sección 5.2 que la proyección de $\mathbf{v}_2$ sobre $\mathbf{v}_1$ es $\frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1$, y que la componente de $\mathbf{v}_2$ ortogonal a $\mathbf{v}_1$ es

$$\mathbf{z}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} - \frac{-1/2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$$

Entonces $\{\mathbf{v}_1, \mathbf{z}_2\}$ es un conjunto ortogonal en el espacio propio para $\lambda = 7$. (Observe que $\mathbf{z}_2$ es una combinación lineal de los vectores propios $\mathbf{v}_1$ y $\mathbf{v}_2$, así que $\mathbf{z}_2$ está en el espacio propio. Esta construcción de $\mathbf{z}_2$ es justamente el proceso de Gram-Schmidt de la sección 5.4). Como el espacio propio tiene dos dimensiones (con la base $\mathbf{v}_1, \mathbf{v}_2$), el conjunto ortogonal $\{\mathbf{v}_1, \mathbf{z}_2\}$ es una *base ortogonal* para el espacio propio, de acuerdo con el teorema de la base. (Véase la sección 2.9 o la 4.5).

Al normalizar $\mathbf{v}_1$ y $\mathbf{z}_2$ se obtiene la siguiente base ortonormal del espacio propio para $\lambda = 7$:

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

Una base ortonormal del espacio propio para $\lambda = -2$ es

$$\mathbf{u}_3 = \frac{1}{\|2\mathbf{v}_3\|} 2\mathbf{v}_3 = \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

De acuerdo con el teorema 1, $\mathbf{u}_3$ es ortogonal a los vectores propios $\mathbf{u}_1$ y $\mathbf{u}_2$. Por lo tanto, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ es un conjunto ortonormal. Sean

$$P = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{bmatrix}, \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Entonces P diagonaliza ortogonalmente a A , y a $A = PDP^{-1}$. ■

En el ejemplo 3, el valor propio 7 tiene multiplicidad dos y el espacio propio es bidimensional. Este hecho no es fortuito, como lo muestra el siguiente teorema.

Teorema espectral

Algunas veces el conjunto de valores propios de una matriz A se denomina el *espectro* de A , y la siguiente descripción de los valores propios se conoce como *teorema espectral*.

TEOREMA 3

Teorema espectral para matrices simétricas

Una matriz simétrica A de $n \times n$ tiene las siguientes propiedades:

- a) A tiene n valores propios reales, contando las multiplicidades.
- b) La dimensión del espacio propio para cada valor propio λ es igual a la multiplicidad de λ como una raíz de la ecuación característica.
- c) Los espacios propios son mutuamente ortogonales, en el sentido de que los vectores propios correspondientes a diferentes valores propios son ortogonales.
- d) A es diagonalizable ortogonalmente.

El inciso a) se deduce del ejercicio 24 de la sección 7.5. El inciso b) se deduce fácilmente del inciso d). (Véase el ejercicio 31). El inciso c) es el teorema 1. A partir del inciso a), puede hacerse una demostración para d) utilizando el ejercicio 32 y la factorización de Schur analizada en el ejercicio complementario 16 del capítulo 5. Se omiten los detalles.

Descomposición espectral

Suponga que $A = PDP^{-1}$, donde las columnas de P son los vectores propios ortonormales $\mathbf{u}_1, \dots, \mathbf{u}_n$ de A , y los valores propios correspondientes $\lambda_1, \dots, \lambda_n$ están en la matriz diagonal D . Entonces, como $P^{-1} = P^T$,

$$\begin{aligned} A = PDP^T &= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= [\lambda_1 \mathbf{u}_1 \quad \cdots \quad \lambda_n \mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \end{aligned}$$

Utilizando la expansión columna-fila de un producto (teorema 10 de la sección 2.4), se puede escribir

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T \quad (2)$$

A esta representación de A se le llama **descomposición espectral** de A porque la divide en partes determinadas por su espectro (valores propios). Cada término en (2) es una matriz de $n \times n$ de rango 1. Por ejemplo, cada columna de $\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$ es un múltiplo de $\mathbf{u}_1$. Además, cada matriz $\mathbf{u}_j \mathbf{u}_j^T$ es una **matriz de proyección** en el sentido de que para cada $\mathbf{x}$ en $\mathbb{R}^n$, el vector $(\mathbf{u}_j \mathbf{u}_j^T) \mathbf{x}$ es la proyección ortogonal de $\mathbf{x}$ sobre el subespacio generado por $\mathbf{u}_j$. (Véase el ejercicio 35).

EJEMPLO 4 Construya una descomposición espectral de la matriz A que tiene la diagonalización ortogonal

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

SOLUCIÓN Denote las columnas de P como $\mathbf{u}_1$ y $\mathbf{u}_2$. Entonces,

$$A = 8\mathbf{u}_1 \mathbf{u}_1^T + 3\mathbf{u}_2 \mathbf{u}_2^T$$

Para comprobar esta descomposición de A , calcule

$$\begin{aligned} \mathbf{u}_1 \mathbf{u}_1^T &= \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix} \\ \mathbf{u}_2 \mathbf{u}_2^T &= \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix} \end{aligned}$$

y

$$8\mathbf{u}_1 \mathbf{u}_1^T + 3\mathbf{u}_2 \mathbf{u}_2^T = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = A \quad \blacksquare$$

NOTA NUMÉRICA

Cuando A es simétrica y no demasiado grande, los modernos algoritmos computacionales de gran desempeño calculan valores propios y vectores propios con gran precisión. Aplican una secuencia de transformaciones de semejanza sobre A que implican matrices ortogonales. Las entradas diagonales de las matrices transformadas convergen rápidamente a los valores propios de A . (Véase las notas numéricas de la sección 7.2). Por lo general, el empleo de matrices ortogonales evita el acumulamiento de errores numéricos durante el proceso. Cuando A es simétrica, la secuencia de matrices ortogonales se combina para formar una matriz ortogonal cuyas columnas son vectores propios de A .

Una matriz no simétrica no puede tener un conjunto completo de vectores propios ortogonales, pero el algoritmo continuará produciendo valores propios más o menos exactos. Después de eso, se requieren técnicas no ortogonales para calcular los vectores propios.

PROBLEMAS DE PRÁCTICA

1. Demuestre que si A es una matriz simétrica, entonces A^2 también es simétrica.
2. Demuestre que si A es diagonalizable ortogonalmente, entonces A^2 también lo es.

8.1 EJERCICIOS[†]

En los ejercicios 1 a 6, determine cuáles matrices son simétricas.

1. $\begin{bmatrix} 3 & 5 \\ 5 & -7 \end{bmatrix}$

2. $\begin{bmatrix} -3 & 5 \\ -5 & 3 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}$

4. $\begin{bmatrix} 0 & 8 & 3 \\ 8 & 0 & -2 \\ 3 & -2 & 0 \end{bmatrix}$

5. $\begin{bmatrix} -6 & 2 & 0 \\ 0 & -6 & 2 \\ 0 & 0 & -6 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}$

En los ejercicios 7 a 12, determine qué matrices son ortogonales. Si alguna es ortogonal, encuentre su inversa.

7. $\begin{bmatrix} .6 & .8 \\ .8 & -.6 \end{bmatrix}$

8. $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

9. $\begin{bmatrix} -5 & 2 \\ 2 & 5 \end{bmatrix}$

10. $\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$

11. $\begin{bmatrix} 2/3 & 2/3 & 1/3 \\ 0 & 1/\sqrt{5} & -2/\sqrt{5} \\ \sqrt{5}/3 & -4/\sqrt{45} & -2/\sqrt{45} \end{bmatrix}$

12. $\begin{bmatrix} .5 & .5 & -.5 & -.5 \\ -.5 & .5 & -.5 & .5 \\ .5 & .5 & .5 & .5 \\ -.5 & .5 & .5 & -.5 \end{bmatrix}$

En los ejercicios 13 a 22, diagonalice ortogonalmente las matrices, dando una matriz ortogonal P y una matriz diagonal D . Para ahorrarle tiempo, los valores propios en los ejercicios 17 a 22 son:

(17) 5, 2, -2; (18) 25, 3, -50; (19) 7, -2; (20) 13, 7, 1; (21) 9, 5, 1; (22) 2, 0.

13. $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

14. $\begin{bmatrix} 1 & 5 \\ 5 & 1 \end{bmatrix}$

15. $\begin{bmatrix} 16 & -4 \\ -4 & 1 \end{bmatrix}$

16. $\begin{bmatrix} -7 & 24 \\ 24 & 7 \end{bmatrix}$

17. $\begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 1 \end{bmatrix}$

18. $\begin{bmatrix} -2 & -36 & 0 \\ -36 & -23 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

19. $\begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$

20. $\begin{bmatrix} 7 & -4 & 4 \\ -4 & 5 & 0 \\ 4 & 0 & 9 \end{bmatrix}$

21. $\begin{bmatrix} 4 & 1 & 3 & 1 \\ 1 & 4 & 1 & 3 \\ 3 & 1 & 4 & 1 \\ 1 & 3 & 1 & 4 \end{bmatrix}$

22. $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$

23. Sean $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ y $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Compruebe que 2 es un valor propio de A y $\mathbf{v}$ un vector propio. Después, diagonalice A ortogonalmente.

24. Sean $A = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & 2 \\ -2 & 2 & 2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ y $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

Compruebe que $\mathbf{v}_1$ y $\mathbf{v}_2$ son vectores propios de A . Después diagonalice A ortogonalmente.

[†] Los ejercicios marcados con un asterisco deben trabajarse en colaboración con los compañeros de clase. Se sugiere formar equipos de dos o tres estudiantes.

En los ejercicios 25 y 26, marque cada enunciado como verdadero o falso. Justifique sus respuestas.

- *25. a) Una matriz de $n \times n$ que es diagonalizable ortogonalmente debe ser simétrica.
 b) Si $A^T = A$ y si los vectores $\mathbf{u}$ y $\mathbf{v}$ satisfacen $A\mathbf{u} = 3\mathbf{u}$ y $A\mathbf{v} = 4\mathbf{v}$, entonces $\mathbf{u} \cdot \mathbf{v} = 0$.
 c) Una matriz simétrica de $n \times n$ tiene n valores propios reales distintos.
 d) Para $\mathbf{v}$ diferente de cero en $\mathbb{R}^n$, la matriz $\mathbf{v}\mathbf{v}^T$ es una matriz de proyección.
- *26. a) Cada matriz simétrica es diagonalizable ortogonalmente.
 b) Si $B = PDP^T$, donde $P^T = P^{-1}$ y D es una matriz diagonal, entonces B es una matriz simétrica.
 c) Una matriz ortogonal es diagonalizable ortogonalmente.
 d) La dimensión de un espacio propio de una matriz simétrica es igual a la multiplicidad del valor propio correspondiente.
27. Suponga que A es una matriz simétrica de $n \times n$, y B es cualquier matriz de $n \times m$. Demuestre que B^TAB , B^TB y BB^T son matrices simétricas.
28. Demuestre que si A es una matriz simétrica de $n \times n$, entonces $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$ para toda $\mathbf{x}$, $\mathbf{y}$ en $\mathbb{R}^n$.
- *29. Suponga que A es invertible y diagonalizable ortogonalmente. Explique por qué A^{-1} también es diagonalizable ortogonalmente.
- *30. Suponga que tanto A como B son diagonalizables ortogonalmente y que $AB = BA$. Explique por qué AB también es diagonalizable ortogonalmente.
- *31. Sean $A = PDP^{-1}$, donde P es ortogonal y D es diagonal, y λ es un valor propio de A de multiplicidad k . Entonces λ se presenta k veces sobre la diagonal de D . Explique por qué la dimensión del espacio propio para λ es k .
- *32. Suponga que $A = PRP^{-1}$, donde P es ortogonal y R es triangular superior. Demuestre que si A es simétrica, entonces R es simétrica y, por lo tanto, realmente es una matriz diagonal.
33. Construya una descomposición espectral de A del ejemplo 2.
34. Construya una descomposición espectral de A del ejemplo 3.
- *35. Sea $\mathbf{u}$ el vector unitario en $\mathbb{R}^n$, y sea $B = \mathbf{u}\mathbf{u}^T$.
- a) Dada cualquier $\mathbf{x}$ en $\mathbb{R}^n$, calcule $B\mathbf{x}$ y demuestre que $B\mathbf{x}$ es la proyección ortogonal de $\mathbf{x}$ sobre $\mathbf{u}$, como se describió en la sección 5.2.
- b) Demuestre que B es una matriz simétrica y que $B^2 = B$.
- c) Demuestre que $\mathbf{u}$ es un vector propio de B . ¿Cuál es el valor propio correspondiente?
- *36. Sea B una matriz simétrica de $n \times n$ tal que $B^2 = B$. Cualquier matriz de este tipo se conoce como **matriz de proyección** (o **matriz de proyección ortogonal**). Para cualquier $\mathbf{y}$ en $\mathbb{R}^n$, sean $\hat{\mathbf{y}} = B\mathbf{y}$ y $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.
- a) Demuestre que $\mathbf{z}$ es ortogonal a $\hat{\mathbf{y}}$.
- b) Sea W el espacio columna de B . Demuestre que $\mathbf{y}$ es la suma de un vector en W y de un vector en $W^\perp$. ¿Por qué esto demuestra que $B\mathbf{y}$ es la proyección ortogonal de $\mathbf{y}$ sobre el espacio columna de B ?

[M] En los ejercicios 37 a 40, diagonalice ortogonalmente las matrices indicadas. Para practicar los métodos de esta sección, no utilice la rutina de vectores propios de su programa de matrices. En vez de ello, aplique el programa para encontrar los valores propios, y para cada valor propio λ , obtenga una base ortonormal para $\text{Nul}(A - \lambda I)$, como en los ejemplos 2 y 3.

$$37. \begin{bmatrix} 5 & 2 & 9 & -6 \\ 2 & 5 & -6 & 9 \\ 9 & -6 & 5 & 2 \\ -6 & 9 & 2 & 5 \end{bmatrix}$$

$$38. \begin{bmatrix} .38 & -.18 & -.06 & -.04 \\ -.18 & .59 & -.04 & .12 \\ -.06 & -.04 & .47 & -.12 \\ -.04 & .12 & -.12 & .41 \end{bmatrix}$$

$$39. \begin{bmatrix} .31 & .58 & .08 & .44 \\ .58 & -.56 & .44 & -.58 \\ .08 & .44 & .19 & -.08 \\ .44 & -.58 & -.08 & .31 \end{bmatrix}$$

$$40. \begin{bmatrix} 10 & 2 & 2 & -6 & 9 \\ 2 & 10 & 2 & -6 & 9 \\ 2 & 2 & 10 & -6 & 9 \\ -6 & -6 & -6 & 26 & 9 \\ 9 & 9 & 9 & 9 & -19 \end{bmatrix}$$

SOLUCIONES A LOS PROBLEMAS DE PRÁCTICA

- $(A^2)^T = (AA)^T = A^T A^T$, por una propiedad de la transpuesta. Por hipótesis, $A^T = A$. Así que $(A^2)^T = AA = A^2$, lo que demuestra que A^2 es simétrica.
- Si A es diagonalizable ortogonalmente, entonces A es simétrica de acuerdo con el teorema 2. Según el problema de práctica 1, A^2 es simétrica y, por lo tanto, es diagonalizable ortogonalmente (teorema 2).

8.2 FORMAS CUADRÁTICAS

Hasta ahora en este libro nos hemos concentrado en ecuaciones lineales, excepto por la sumas de cuadrados que se encontraron en el capítulo 5 al calcular $\mathbf{x}^T\mathbf{x}$. Estas sumas y expresiones más generales, llamadas *formas cuadráticas*, se presentan con frecuencia en aplicaciones del álgebra lineal a la ingeniería (en criterios de diseño y optimización) y el procesamiento de señales (como potencia de ruido de salida). Las aplicaciones también surgen, por ejemplo, en física (como energías potencial y cinética), geometría diferencial (como la curvatura normal de superficies), economía (como funciones de utilidad) y estadística (en elipsoides de confianza). Parte de las bases matemáticas para dichas aplicaciones fluye fácilmente de nuestro trabajo de matrices simétricas.

Una **forma cuadrática** en $\mathbb{R}^n$ es una función Q definida sobre $\mathbb{R}^n$ cuyo valor en un vector $\mathbf{x}$ de $\mathbb{R}^n$ se puede calcular mediante una expresión de la forma $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, donde A es una matriz simétrica de $n \times n$. La matriz A se denomina **matriz de la forma cuadrática**.

El ejemplo más sencillo de una forma cuadrática diferente de cero es $Q(\mathbf{x}) = \mathbf{x}^T I \mathbf{x} = \|\mathbf{x}\|^2$. Los ejemplos 1 y 2 revelan la conexión entre cualquier matriz simétrica A y la forma cuadrática $\mathbf{x}^T A \mathbf{x}$.

EJEMPLO 1 Sea $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Calcule $\mathbf{x}^T A \mathbf{x}$ para las siguientes matrices:

$$a) \quad A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \qquad b) \quad A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

SOLUCIÓN

$$a) \quad \mathbf{x}^T A \mathbf{x} = [x_1 \quad x_2] \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1 \quad x_2] \begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2$$

b) En A existen dos entradas -2 . Observe cómo participan en los cálculos. La entrada $(1, 2)$ en A está en negritas.

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= [x_1 \quad x_2] \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1 \quad x_2] \begin{bmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{bmatrix} \\ &= x_1(3x_1 - 2x_2) + x_2(-2x_1 + 7x_2) \\ &= 3x_1^2 - 2x_1x_2 - 2x_2x_1 + 7x_2^2 \\ &= 3x_1^2 - 4x_1x_2 + 7x_2^2 \end{aligned}$$

La presencia de $-4x_1x_2$ en la forma cuadrática del ejemplo 1b) se debe a las entradas -2 fuera de la diagonal en la matriz A . En contraste, la forma cuadrática asociada con la matriz diagonal A del ejemplo 1a) no tiene el *producto cruzado* x_1x_2 .

EJEMPLO 2 Para $\mathbf{x}$ en $\mathbb{R}^3$, sea $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Escriba esta forma cuadrática como $\mathbf{x}^T A \mathbf{x}$.

SOLUCIÓN Los coeficientes de x_1^2 , x_2^2 , x_3^2 están en la diagonal de A . Para que A sea simétrica, el coeficiente de $x_i x_j$ con $i \neq j$ debe dividirse en partes iguales entre las entradas (i, j) y (j, i) en A . El coeficiente de $x_1 x_3$ es 0. Es sencillo comprobar que:

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

EJEMPLO 3 Sea $Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2$. Calcule el valor de $Q(\mathbf{x})$ para $\mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -2 \end{bmatrix}$ y $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

SOLUCIÓN

$$Q(-3, 1) = (-3)^2 - 8(-3)(1) - 5(1)^2 = 28$$

$$Q(2, -2) = (2)^2 - 8(2)(-2) - 5(-2)^2 = 16$$

$$Q(1, -3) = (1)^2 - 8(1)(-3) - 5(-3)^2 = -20 \quad \blacksquare$$

En algunos casos, las formas cuadráticas son más fáciles de usar cuando no tienen productos cruzados, es decir, cuando la matriz de la forma cuadrática es una matriz diagonal. Por fortuna, el producto cruzado se puede eliminar realizando un adecuado cambio de variable.

Cambio de variable en una forma cuadrática

Si $\mathbf{x}$ representa un vector variable en $\mathbb{R}^n$, entonces un **cambio de variable** es una ecuación de la forma

$$\mathbf{x} = P\mathbf{y}, \quad \text{o, de manera equivalente,} \quad \mathbf{y} = P^{-1}\mathbf{x} \quad (1)$$

donde P es una matriz invertible y $\mathbf{y}$ es un nuevo vector variable en $\mathbb{R}^n$. Aquí $\mathbf{y}$ es el vector de coordenadas de $\mathbf{x}$ con respecto a la base de $\mathbb{R}^n$ determinada por las columnas de P . (Véase la sección 4.4).

Si el cambio de variable (1) se efectúa en una forma cuadrática $\mathbf{x}^T A \mathbf{x}$, entonces

$$\mathbf{x}^T A \mathbf{x} = (P\mathbf{y})^T A (P\mathbf{y}) = \mathbf{y}^T P^T A P \mathbf{y} = \mathbf{y}^T (P^T A P) \mathbf{y} \quad (2)$$

y la nueva matriz de la forma cuadrática es $P^T A P$. Como A es simétrica, entonces el teorema 2 garantiza que existe una matriz *ortogonal* P tal que $P^T A P$ es una matriz diagonal D , y así la forma cuadrática en (2) se convierte en $\mathbf{y}^T D \mathbf{y}$. Esta es la estrategia del siguiente ejemplo.

EJEMPLO 4 Realice un cambio de variable que transforme la forma cuadrática del ejemplo 3 en una forma cuadrática sin productos cruzados.

SOLUCIÓN La matriz de la forma cuadrática del ejemplo 3 es

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

El primer paso es que A se diagonalice ortogonalmente. Sus valores propios resultan ser $\lambda = 3$ y $\lambda = -7$. Los vectores propios unitarios asociados son:

$$\lambda = 3: \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}; \quad \lambda = -7: \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Esos vectores son automáticamente ortogonales (porque corresponden a diferentes valores propios), de manera que brindan una base ortonormal para $\mathbb{R}^2$. Sean

$$P = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}$$

Entonces, $A = P D P^{-1}$ y $D = P^{-1} A P = P^T A P$, como se indicó antes. Un cambio de variable conveniente es

$$\mathbf{x} = P\mathbf{y}, \quad \text{donde } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{y} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

De esta forma,

$$\begin{aligned} x_1^2 - 8x_1x_2 - 5x_2^2 &= \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{P} \mathbf{y})^T \mathbf{A} (\mathbf{P} \mathbf{y}) \\ &= \mathbf{y}^T \mathbf{P}^T \mathbf{A} \mathbf{P} \mathbf{y} = \mathbf{y}^T \mathbf{D} \mathbf{y} \\ &= 3y_1^2 - 7y_2^2 \end{aligned}$$

Para ilustrar el significado de la igualdad de las formas cuadráticas del ejemplo 4, se puede calcular $Q(\mathbf{x})$ para $\mathbf{x} = (2, -2)$ empleando la nueva forma cuadrática. Primero, como $\mathbf{x} = \mathbf{P} \mathbf{y}$,

$$\mathbf{y} = \mathbf{P}^{-1} \mathbf{x} = \mathbf{P}^T \mathbf{x}$$

de manera que

$$\mathbf{y} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 6/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

Por lo tanto,

$$\begin{aligned} 3y_1^2 - 7y_2^2 &= 3(6/\sqrt{5})^2 - 7(-2/\sqrt{5})^2 = 3(36/5) - 7(4/5) \\ &= 80/5 = 16 \end{aligned}$$

Este es el valor de $Q(\mathbf{x})$ en el ejemplo 3 cuando $\mathbf{x} = (2, -2)$. Véase la figura 1.

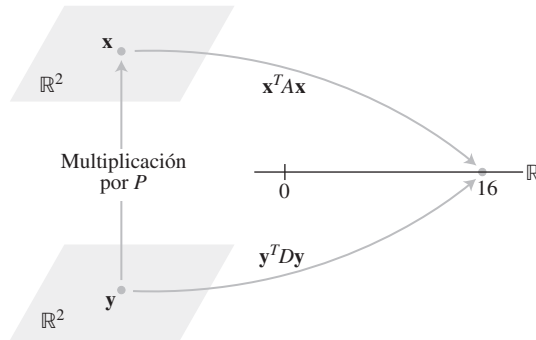


FIGURA 1 Cambio de variable en $\mathbf{x}^T \mathbf{A} \mathbf{x}$.

El ejemplo 4 ilustra el siguiente teorema. En esencia, la demostración del teorema se presentó antes del ejemplo 4.

TEOREMA 4

Teorema de los ejes principales

Sea A una matriz simétrica de $n \times n$. Luego, existe un cambio de variable ortogonal, $\mathbf{x} = \mathbf{P} \mathbf{y}$, que convierte la forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$ en una forma cuadrática $\mathbf{y}^T \mathbf{D} \mathbf{y}$ sin productos cruzados.

En el teorema, las columnas de $\mathbf{P}$ se llaman **ejes principales** de la forma cuadrática $\mathbf{x}^T \mathbf{A} \mathbf{x}$. El vector $\mathbf{y}$ es el vector de coordenadas de $\mathbf{x}$ respecto de la base ortonormal de $\mathbb{R}^n$ que dan esos ejes principales.

Ejes principales desde un punto de vista geométrico

Suponga que $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, donde A es una matriz simétrica invertible de 2×2 , y sea c una constante. Es posible demostrar que el conjunto de todas las $\mathbf{x}$ en $\mathbb{R}^2$ que satisfacen

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = c \quad (3)$$

corresponde a una elipse (o circunferencia), a una hipérbola, a dos rectas que se intersecan, a un solo punto, o quizá no contenga puntos. Si A es una matriz diagonal, la gráfica está en *posición estándar*, como se muestra en la figura 2. Si A no es una matriz diagonal, la gráfica de la ecuación (3) ha dado un giro respecto de la posición estándar, como se muestra en la

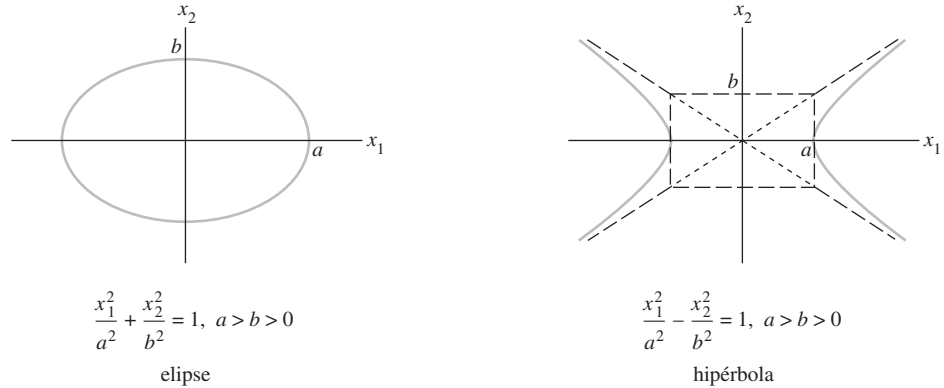


FIGURA 2 Una elipse y una hipérbola en posición estándar.

figura 3. Encontrar los *ejes principales* (determinados por los vectores propios de A) equivale a encontrar un nuevo sistema de coordenadas con respecto al cual la gráfica está en posición estándar.

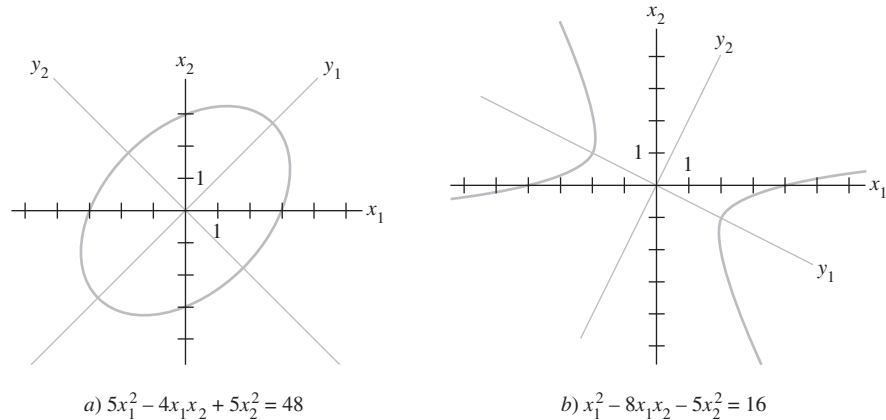


FIGURA 3 Una elipse y una hipérbola que *no* están en posición estándar.

La hipérbola en la figura 3b) es la gráfica de la ecuación $\mathbf{x}^T A \mathbf{x} = 16$, donde A es la matriz del ejemplo 4. El eje y_1 positivo en la figura 3b) está en la dirección de la primera columna de la matriz P del ejemplo 4, y el eje y_2 positivo se encuentra en la dirección de la segunda columna de P .

EJEMPLO 5 La elipse de la figura 3a) es la gráfica de la ecuación $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$. Encuentre un cambio de variable que elimine el producto cruzado de la ecuación.

SOLUCIÓN La matriz de la forma cuadrática es $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$. Los valores propios de A resultan ser 3 y 7, con los vectores propios unitarios correspondientes

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Sea $P = [\mathbf{u}_1 \ \mathbf{u}_2] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. Entonces, P diagonaliza ortogonalmente a A ,

de manera que el cambio de variable $\mathbf{x} = P\mathbf{y}$ produce la forma cuadrática $\mathbf{y}^T D\mathbf{y} = 3y_1^2 + 7y_2^2$. La figura 3a) muestra los nuevos ejes para este cambio de variable. ■

Clasificación de formas cuadráticas

Cuando A es una matriz de $n \times n$, la forma cuadrática $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ es una función de valores reales con dominio $\mathbb{R}^n$. La figura 4 presenta las gráficas de cuatro formas cuadráticas con dominio $\mathbb{R}^2$. Para cada punto $\mathbf{x} = (x_1, x_2)$ en el dominio de una forma cuadrática Q , la gráfica muestra el punto (x_1, x_2, z) , donde $z = Q(\mathbf{x})$. Observe que, excepto en $\mathbf{x} = \mathbf{0}$, todos los valores de $Q(\mathbf{x})$ son positivos en la figura 4a) y todos son negativos en la figura 4d). Las secciones transversales horizontales de las gráficas son elipses en la figura 4a) y 4d), y son hipérbolas en la figura 4c).

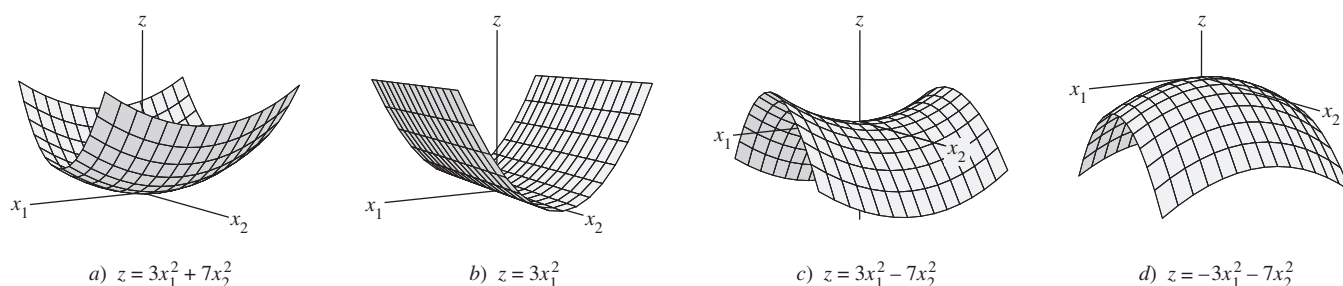


FIGURA 4 Gráficas de formas cuadráticas.

Los ejemplos sencillos de 2×2 de la figura 4 ponen de manifiesto las siguientes definiciones.

DEFINICIÓN

Una forma cuadrática Q es:

- a) **positiva definida** si $Q(\mathbf{x}) > 0$ para toda $\mathbf{x} \neq \mathbf{0}$,
- b) **negativa definida** si $Q(\mathbf{x}) < 0$ para toda $\mathbf{x} \neq \mathbf{0}$,
- c) **indefinida** si $Q(\mathbf{x})$ toma valores positivos y negativos.

También, se dice que Q es **positiva semidefinida** si $Q(\mathbf{x}) \geq 0$ para toda $\mathbf{x}$, y **negativa semidefinida** si $Q(\mathbf{x}) \leq 0$ para toda $\mathbf{x}$. Las formas cuadráticas en los incisos a) y b) de la figura 4 son semidefinidas positivas, pero la forma en a) se describe mejor como positiva definida.

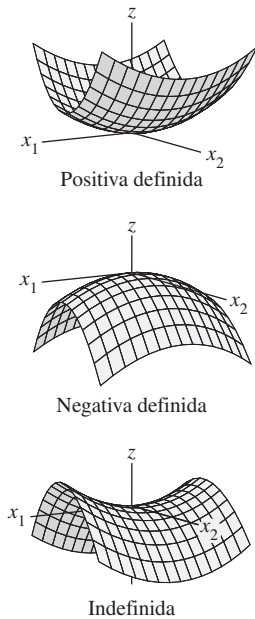
El teorema 5 caracteriza algunas formas cuadráticas en términos de valores propios.

TEOREMA 5

Formas cuadráticas y valores propios

Sea A una matriz simétrica de $n \times n$. Así, una forma cuadrática $\mathbf{x}^T A \mathbf{x}$ es:

- a) positiva definida si y solo si todos los valores propios de A son positivos,
- b) negativa definida si y solo si todos los valores propios de A son negativos, o
- c) indefinida si y solo si A tiene valores propios positivos y negativos.



DEMOSTRACIÓN De acuerdo con el teorema de los ejes principales, existe un cambio ortogonal de variable $\mathbf{x} = P\mathbf{y}$ tal que

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \quad (4)$$

donde $\lambda_1, \dots, \lambda_n$ son los valores propios de A . Como P es invertible, existe una correspondencia uno a uno entre todas las $\mathbf{x}$ y todas las $\mathbf{y}$ diferentes de cero. Por lo tanto, los valores de $Q(\mathbf{x})$ para $\mathbf{x} \neq \mathbf{0}$ coinciden con los valores de la expresión en el miembro derecho de la ecuación (4), la cual, como es evidente, está controlada por los signos de los valores propios $\lambda_1, \dots, \lambda_n$, en las tres maneras descritas en el teorema. ■

EJEMPLO 6 ¿ $Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3$ es positiva definida?

SOLUCIÓN A causa de todos los signos positivos, esta forma “parece” positiva definida. Pero la matriz de la forma es

$$A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

y los valores propios de A resultan ser 5, 2 y -1 . Así, Q es una forma cuadrática indefinida, y no positiva definida. ■

Con frecuencia, la clasificación de una forma cuadrática se realiza sobre la matriz de la forma. Por consiguiente, una **matriz A positiva definida** es una matriz *simétrica* para la cual la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ es positiva definida. Otros términos, tales como **matriz positiva semidefinida**, se definen de manera análoga.

WEB

NOTA NUMÉRICA

Una manera rápida de determinar si una matriz simétrica es positiva definida consiste en intentar factorizar A en la forma $A = \mathbb{R}^T \mathbb{R}$, donde $\mathbb{R}$ es triangular superior con entradas diagonales positivas. (Un enfoque consiste en utilizar un algoritmo ligeramente modificado para la factorización LU). Tal *factorización de Cholesky* es posible si y solo si A es positiva definida. Véase el ejercicio complementario 7 al final del capítulo 8.

PROBLEMA DE PRÁCTICA†

- Describa una matriz A positiva semidefinida en términos de sus valores propios.

WEB

8.2 EJERCICIOS†

1. Calcule la forma cuadrática $\mathbf{x}^T A \mathbf{x}$, cuando $A = \begin{bmatrix} 5 & 1/3 \\ 1/3 & 1 \end{bmatrix}$

y

a) $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ b) $\mathbf{x} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$ c) $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

2. Determine la forma cuadrática $\mathbf{x}^T A \mathbf{x}$, para $A = \begin{bmatrix} 4 & 3 & 0 \\ 3 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

y

a) $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ b) $\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$ c) $\mathbf{x} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$

3. Encuentre la matriz de la forma cuadrática. Suponga que $\mathbf{x}$ está en $\mathbb{R}^2$.

a) $10x_1^2 - 6x_1x_2 - 3x_2^2$ b) $5x_1^2 + 3x_1x_2$

4. Obtenga la matriz de la forma cuadrática. Suponga que $\mathbf{x}$ está en $\mathbb{R}^2$.

a) $20x_1^2 + 15x_1x_2 - 10x_2^2$ b) x_1x_2

† Se sugiere trabajar los problemas marcados con un punto primero en forma individual y, luego, discutirlos con todo el grupo y el profesor.

5. Determine la matriz de la forma cuadrática. Suponga que $\mathbf{x}$ está en $\mathbb{R}^3$.

a) $8x_1^2 + 7x_2^2 - 3x_3^2 - 6x_1x_2 + 4x_1x_3 - 2x_2x_3$
 b) $4x_1x_2 + 6x_1x_3 - 8x_2x_3$

6. Encuentre la matriz de la forma cuadrática. Suponga que $\mathbf{x}$ está en $\mathbb{R}^3$.

a) $5x_1^2 - x_2^2 + 7x_3^2 + 5x_1x_2 - 3x_1x_3$
 b) $x_3^2 - 4x_1x_2 + 4x_2x_3$

7. Realice un cambio de variable, $\mathbf{x} = P\mathbf{y}$, que transforme la forma cuadrática $x_1^2 + 10x_1x_2 + x_2^2$ en una forma cuadrática sin producto cruzado. Determine P y la nueva forma cuadrática.

8. Sea A la matriz de la forma cuadrática

$$9x_1^2 + 7x_2^2 + 11x_3^2 - 8x_1x_2 + 8x_1x_3$$

Es posible demostrar que los valores propios de A son 3, 9 y 15. Encuentre una matriz ortogonal P tal que el cambio de variable $\mathbf{x} = P\mathbf{y}$ transforme $\mathbf{x}^T A \mathbf{x}$ en una forma cuadrática sin productos cruzados. Determine P y la nueva forma cuadrática.

En los ejercicios 9 a 18, clasifique las formas cuadráticas. Después realice un cambio de variable, $\mathbf{x} = P\mathbf{y}$, que convierta la forma cuadrática en una que no incluya productos cruzados. Escriba la nueva forma cuadrática. Construya P utilizando los métodos de la sección 8.1.

9. $3x_1^2 - 4x_1x_2 + 6x_2^2$ 10. $9x_1^2 - 8x_1x_2 + 3x_2^2$
 11. $2x_1^2 + 10x_1x_2 + 2x_2^2$ 12. $-5x_1^2 + 4x_1x_2 - 2x_2^2$
 13. $x_1^2 - 6x_1x_2 + 9x_2^2$ 14. $8x_1^2 + 6x_1x_2$
 15. [M] $-2x_1^2 - 6x_2^2 - 9x_3^2 - 9x_4^2 + 4x_1x_2 + 4x_1x_3 + 4x_1x_4 + 6x_3x_4$
 16. [M] $4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2 + 3x_1x_2 + 3x_3x_4 - 4x_1x_4 + 4x_2x_3$
 17. [M] $x_1^2 + x_2^2 + x_3^2 + x_4^2 + 9x_1x_2 - 12x_1x_4 + 12x_2x_3 + 9x_3x_4$
 18. [M] $11x_1^2 - x_2^2 - 12x_1x_2 - 12x_1x_3 - 12x_1x_4 - 2x_3x_4$
 19. ¿Cuál es el valor más grande posible de la forma cuadrática $5x_1^2 + 8x_2^2$ si $\mathbf{x} = (x_1, x_2)$ y $\mathbf{x}^T \mathbf{x} = 1$, es decir, si $x_1^2 + x_2^2 = 1$? (Intente algunos ejemplos de $\mathbf{x}$).
 20. ¿Cuál es el valor más grande de la forma cuadrática $5x_1^2 - 3x_2^2$ si $\mathbf{x}^T \mathbf{x} = 1$?

En los ejercicios 21 y 22, las matrices son de $n \times n$ y los vectores están en $\mathbb{R}^n$. Marque cada enunciado como verdadero o falso. Justifique sus respuestas.

- *21. a) La matriz de una forma cuadrática es una matriz simétrica.
 b) Una forma cuadrática no tiene productos cruzados si y solo si la matriz de la forma cuadrática es una matriz diagonal.
 c) Los ejes principales de una forma cuadrática $\mathbf{x}^T A \mathbf{x}$ son vectores propios de A .
 d) Una forma cuadrática positiva definida Q satisface $Q(\mathbf{x}) > 0$ para toda $\mathbf{x}$ en $\mathbb{R}^n$.

- e) Si todos los valores propios de una matriz simétrica A son positivos, entonces la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ es positiva definida.

- f) Una factorización de Cholesky de una matriz simétrica A tiene la forma $A = R^T R$, para una matriz triangular superior R con entradas diagonales positivas.

- *22. a) La expresión $\|\mathbf{x}\|^2$ es una forma cuadrática.
 b) Si A es simétrica y P es una matriz ortogonal, entonces el cambio de variable $\mathbf{x} = P\mathbf{y}$ convierte $\mathbf{x}^T A \mathbf{x}$ en una forma cuadrática sin productos cruzados.
 c) Si A es una matriz simétrica de 2×2 , entonces el conjunto de $\mathbf{x}$ tal que $\mathbf{x}^T A \mathbf{x} = c$ (para una constante c) corresponde a un círculo, una elipse o una hipérbola.
 d) Una forma cuadrática indefinida es positiva semidefinida o negativa semidefinida.
 e) Si A es simétrica y la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ solo tiene valores negativos para $\mathbf{x} \neq \mathbf{0}$, entonces todos los valores propios de A son negativos.

Los ejercicios 23 y 24 muestran cómo clasificar una forma cuadrática $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, cuando $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ y $\det A \neq 0$, sin obtener los valores propios de A .

- *23. Si λ_1 y λ_2 son los valores propios de A , entonces el polinomio característico de A se puede escribir de dos maneras: $\det(A - \lambda I)$ y $(\lambda - \lambda_1)(\lambda - \lambda_2)$. Con base en este hecho, demuestre que $\lambda_1 + \lambda_2 = a + d$ (las entradas diagonales de A) y $\lambda_1 \lambda_2 = \det A$.
 *24. Compruebe los siguientes enunciados.
 a) Q es positiva definida si $\det A > 0$ y $a > 0$.
 b) Q es negativa definida si $\det A > 0$ y $a < 0$.
 c) Q es indefinida si $\det A < 0$.
 *25. Demuestre que si B es de $m \times n$, entonces $B^T B$ es positiva semidefinida; y si B es de $n \times n$ e invertible, entonces $B^T B$ es positiva definida.
 26. Demuestre que si una matriz A de $n \times n$ es positiva definida, entonces existe una matriz B positiva definida tal que $A = B^T B$. [Sugerencia: Escriba $A = P D P^T$, con $P^T = P^{-1}$. Construya una matriz C diagonal tal que $D = C^T C$, y sea $B = P C P^T$. Demuestre que B funciona].

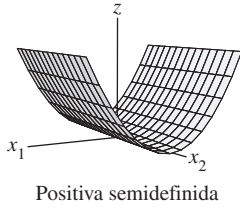
27. Sean A y B matrices simétricas de $n \times n$ cuyos valores propios son todos positivos. Demuestre que todos los valores propios de $A + B$ son positivos. [Sugerencia: Considere formas cuadráticas].

28. Sea A una matriz simétrica e invertible de $n \times n$. Demuestre que si la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ es positiva definida, entonces también lo es la forma cuadrática $\mathbf{x}^T A^{-1} \mathbf{x}$. [Sugerencia: Considere valores propios].

SOLUCIÓN AL PROBLEMA DE PRÁCTICA

Efectúe un cambio de variable ortogonal $\mathbf{x} = P\mathbf{y}$, y escriba

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$



como en la ecuación (4). Si un valor propio —por ejemplo, λ_i — fuera negativo, entonces $\mathbf{x}^T A \mathbf{x}$ sería negativo para la $\mathbf{x}$ correspondiente a $\mathbf{y} = \mathbf{e}_i$ (la i -ésima columna de I_n). Así, todos los valores propios de una forma cuadrática positiva semidefinida deben ser no negativos. A la inversa, si los valores propios son positivos, la expansión anterior indica que $\mathbf{x}^T A \mathbf{x}$ debe ser positiva semidefinida.

8.3 OPTIMIZACIÓN RESTRINGIDA

Es frecuente que ingenieros, economistas, científicos y matemáticos necesiten encontrar el valor máximo o mínimo de una forma cuadrática $Q(\mathbf{x})$ para $\mathbf{x}$ en algún conjunto específico. Por lo común, el problema se puede adaptar de tal manera que $\mathbf{x}$ varíe sobre un conjunto de vectores unitarios. Este *problema de optimización restringida* tiene una solución interesante y elegante. El ejemplo 6 y el análisis de la sección 8.5 ilustrarán cómo se presentan en la práctica este tipo de problemas.

El requisito de que un vector $\mathbf{x}$ en $\mathbb{R}^n$ sea unitario se puede establecer en varias formas equivalentes:

$$\|\mathbf{x}\| = 1, \quad \|\mathbf{x}\|^2 = 1, \quad \mathbf{x}^T \mathbf{x} = 1$$

y

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \quad (1)$$

La versión ampliada de la ecuación (1) de $\mathbf{x}^T \mathbf{x} = 1$ se utiliza comúnmente en las aplicaciones.

Cuando una forma cuadrática Q no tiene productos cruzados, es fácil encontrar el máximo y el mínimo de $Q(\mathbf{x})$ para $\mathbf{x}^T \mathbf{x} = 1$.

EJEMPLO 1 Obtenga los valores máximo y mínimo de $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ con la restricción $\mathbf{x}^T \mathbf{x} = 1$.

SOLUCIÓN Como x_2^2 y x_3^2 son no negativos, observe que

$$4x_2^2 \leq 9x_2^2 \quad \text{y} \quad 3x_3^2 \leq 9x_3^2$$

y por lo tanto,

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\leq 9x_1^2 + 9x_2^2 + 9x_3^2 \\ &= 9(x_1^2 + x_2^2 + x_3^2) \\ &= 9 \end{aligned}$$

siempre que $x_1^2 + x_2^2 + x_3^2 = 1$. Así, el valor máximo de $Q(\mathbf{x})$ no puede exceder 9 cuando $\mathbf{x}$ es un vector unitario. Además, $Q(\mathbf{x}) = 9$ cuando $\mathbf{x} = (1, 0, 0)$. Por consiguiente, 9 es el valor máximo de $Q(\mathbf{x})$ para $\mathbf{x}^T \mathbf{x} = 1$.

Para encontrar el valor mínimo de $Q(\mathbf{x})$, observe que

$$9x_1^2 \geq 3x_1^2, \quad 4x_2^2 \geq 3x_2^2$$

y en consecuencia,

$$Q(\mathbf{x}) \geq 3x_1^2 + 3x_2^2 + 3x_3^2 = 3(x_1^2 + x_2^2 + x_3^2) = 3$$

siempre que $x_1^2 + x_2^2 + x_3^2 = 1$. Además, $Q(\mathbf{x}) = 3$ cuando $x_1 = 0$, $x_2 = 0$, y $x_3 = 1$. Así, 3 es el valor mínimo de $Q(\mathbf{x})$ cuando $\mathbf{x}^T \mathbf{x} = 1$. ■

En el ejemplo 1 es fácil ver que la matriz de la forma cuadrática Q tiene valores propios 9, 4 y 3, y que los valores propios mayor y menor son iguales al máximo y mínimo (restringidos) de $Q(\mathbf{x})$, respectivamente. Se verá que lo mismo es válido para cualquier forma cuadrática.

EJEMPLO 2 Sean $A = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$, y $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ para $\mathbf{x}$ en $\mathbb{R}^2$. La figura 1 muestra la gráfica de Q . La figura 2 solo muestra la parte de la gráfica dentro de un cilindro; la intersección del cilindro con la superficie es el conjunto de puntos (x_1, x_2, z) tales que $z = Q(x_1, x_2)$ y $x_1^2 + x_2^2 = 1$. Las “alturas” de esos puntos son los valores restringidos de $Q(\mathbf{x})$. Geométricamente, el problema de optimización restringido es localizar los puntos más alto y más bajo en la curva de intersección.

Los dos puntos más altos sobre la curva están 7 unidades arriba del plano x_1x_2 , en $x_1 = 0$ y $x_2 = \pm 1$. Esos puntos corresponden al valor propio 7 de A y a los vectores propios $\mathbf{x} = (0, 1)$ y $-\mathbf{x} = (0, -1)$. De manera similar, los dos puntos más bajos sobre la curva están 3 unidades arriba del plano x_1x_2 , y corresponden al valor propio 3 y a los vectores propios $(1, 0)$ y $(-1, 0)$. ■

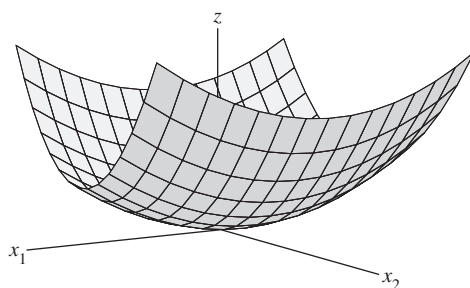


FIGURA 1 $z = 3x_1^2 + 7x_2^2$.

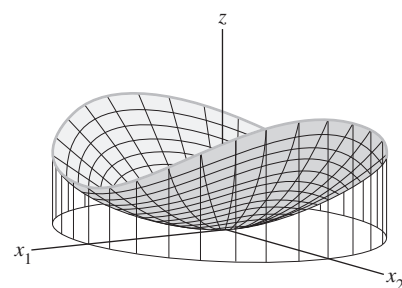


FIGURA 2 La intersección de $z = 3x_1^2 + 7x_2^2$ y el cilindro $x_1^2 + x_2^2 = 1$.

En la figura 2 cada punto de la curva de intersección tiene una coordenada z entre 3 y 7, y para cualquier número t entre 3 y 7, existe un vector unitario $\mathbf{x}$ tal que $Q(\mathbf{x}) = t$. En otras palabras, el conjunto de todos los posibles valores de $\mathbf{x}^T A \mathbf{x}$, para $\|\mathbf{x}\| = 1$, es el intervalo cerrado $3 \leq t \leq 7$.

Es posible demostrar que para cualquier matriz simétrica A , el conjunto de todos los posibles valores de $\mathbf{x}^T A \mathbf{x}$, para $\|\mathbf{x}\| = 1$, es un intervalo cerrado sobre el eje real. (Véase el ejercicio 13). Denote los puntos extremos izquierdo y derecho de este intervalo con m y M , respectivamente. Es decir, sean

$$m = \min \{ \mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1 \}, \quad M = \max \{ \mathbf{x}^T A \mathbf{x} : \|\mathbf{x}\| = 1 \} \quad (2)$$

En el ejercicio 12 se le pide demostrar que si λ es un valor propio de A , entonces $m \leq \lambda \leq M$. El siguiente teorema dice que m y M son en sí mismos valores propios de A , justo como en el ejemplo 2.¹

TEOREMA 6

Sea A una matriz simétrica, y m y M se definen como en la ecuación (2). Entonces M es el valor propio más grande λ_1 de A y m es el valor propio más pequeño de A . El valor de $\mathbf{x}^T A \mathbf{x}$ es M cuando $\mathbf{x}$ es un vector propio unitario $\mathbf{u}_1$ correspondiente a M . El valor de $\mathbf{x}^T A \mathbf{x}$ es m cuando $\mathbf{x}$ es un vector propio unitario correspondiente a m .

¹ Los términos *mínimo* y *máximo* en la ecuación (2), y *menor* y *mayor* en el teorema, se refieren al ordenamiento natural de los números reales, no a magnitudes.

DEMOSTRACIÓN A se diagonaliza ortogonalmente como PDP^{-1} . Se sabe que

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} \quad \text{cuando } \mathbf{x} = P\mathbf{y} \quad (3)$$

También,

$$\|\mathbf{x}\| = \|P\mathbf{y}\| = \|\mathbf{y}\| \quad \text{para toda } \mathbf{y}$$

porque $P^T P = I$ y $\|P\mathbf{y}\|^2 = (P\mathbf{y})^T (P\mathbf{y}) = \mathbf{y}^T P^T P \mathbf{y} = \mathbf{y}^T \mathbf{y} = \|\mathbf{y}\|^2$. En particular, $\|\mathbf{y}\| = 1$ si y solo si $\|\mathbf{x}\| = 1$. Así, $\mathbf{x}^T A \mathbf{x}$ y $\mathbf{y}^T D \mathbf{y}$ asumen el mismo conjunto de valores conforme $\mathbf{x}$ y $\mathbf{y}$ varían sobre el conjunto de todos los vectores unitarios.

Para simplificar la notación, suponga que A es una matriz de 3×3 con valores propios $a \geq b \geq c$. Arregle las columnas (los vectores propios) de P de manera que $P = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3]$ y

$$D = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

Dado cualquier vector unitario $\mathbf{y}$ en $\mathbb{R}^3$ con coordenadas y_1, y_2, y_3 , observe que

$$ay_1^2 = ay_1^2$$

$$by_2^2 \leq ay_2^2$$

$$cy_3^2 \leq ay_3^2$$

y se obtienen estas desigualdades:

$$\begin{aligned} \mathbf{y}^T D \mathbf{y} &= ay_1^2 + by_2^2 + cy_3^2 \\ &\leq ay_1^2 + ay_2^2 + ay_3^2 \\ &= a(y_1^2 + y_2^2 + y_3^2) \\ &= a\|\mathbf{y}\|^2 = a \end{aligned}$$

De esta manera, $M \leq a$, por definición de M . Sin embargo, $\mathbf{y}^T D \mathbf{y} = a$ cuando $\mathbf{y} = \mathbf{e}_1 = (1, 0, 0)$, así que en efecto $M = a$. De acuerdo con la ecuación (3), la $\mathbf{x}$ que corresponde a $\mathbf{y} = \mathbf{e}_1$ es el vector propio $\mathbf{u}_1$ de A , porque

$$\mathbf{x} = P\mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{u}_1$$

Por lo tanto, $M = a = \mathbf{e}_1^T D \mathbf{e}_1 = \mathbf{u}_1^T A \mathbf{u}_1$, lo que demuestra el enunciado sobre M . Un argumento similar demuestra que m es el valor propio más pequeño, c , y ese valor de $\mathbf{x}^T A \mathbf{x}$ se obtiene cuando $\mathbf{x} = P\mathbf{e}_3 = \mathbf{u}_3$. ■

EJEMPLO 3 Sea $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$. Encuentre el valor máximo de la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ sujeta a la restricción $\mathbf{x}^T \mathbf{x} = 1$, y obtenga un vector unitario en el que se alcance dicho valor máximo.

SOLUCIÓN Según el teorema 6, el valor máximo deseado es el valor propio más grande de A . La ecuación característica resulta ser

$$0 = -\lambda^3 + 10\lambda^2 - 27\lambda + 18 = -(\lambda - 6)(\lambda - 3)(\lambda - 1)$$

El valor propio más grande es 6.

El máximo restringido de $\mathbf{x}^T A \mathbf{x}$ se obtiene cuando $\mathbf{x}$ es un vector propio unitario

para $\lambda = 6$. Resuelva $(A - 6I)\mathbf{x} = 0$ y encuentre un vector propio $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Establezca $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$. ■

En el teorema 7 y en aplicaciones posteriores, los valores de $\mathbf{x}^T A \mathbf{x}$ se calculan con restricciones adicionales sobre el vector unitario $\mathbf{x}$.

TEOREMA 7

Sean A , λ_1 y $\mathbf{u}_1$ como en el teorema 6. Así, el valor máximo de $\mathbf{x}^T A \mathbf{x}$ sujeto a las restricciones

$$\mathbf{x}^T \mathbf{x} = 1, \quad \mathbf{x}^T \mathbf{u}_1 = 0$$

es el segundo valor propio más grande, λ_2 , y este máximo se logra cuando $\mathbf{x}$ es un vector propio $\mathbf{u}_2$ correspondiente a λ_2 .

El teorema 7 se demuestra mediante un argumento similar al anterior donde el teorema se reduce al caso en que la matriz de la forma cuadrática es diagonal. El siguiente ejemplo da una idea de la demostración para el caso de una matriz diagonal.

EJEMPLO 4 Encuentre el valor máximo de $9x_1^2 + 4x_2^2 + 3x_3^2$ con las restricciones $\mathbf{x}^T \mathbf{x} = 1$ y $\mathbf{x}^T \mathbf{u}_1 = 0$, donde $\mathbf{u}_1 = (1, 0, 0)$. Observe que $\mathbf{u}_1$ es un vector propio unitario correspondiente al valor propio más grande $\lambda = 9$ de la matriz de la forma cuadrática.

SOLUCIÓN Si las coordenadas de $\mathbf{x}$ son x_1, x_2, x_3 , entonces la restricción $\mathbf{x}^T \mathbf{u}_1 = 0$ significa simplemente que $x_1 = 0$. Para tal vector unitario, $x_2^2 + x_3^2 = 1$, y

$$\begin{aligned} 9x_1^2 + 4x_2^2 + 3x_3^2 &= 4x_2^2 + 3x_3^2 \\ &\leq 4x_2^2 + 4x_3^2 \\ &= 4(x_2^2 + x_3^2) \\ &= 4 \end{aligned}$$

Así, el máximo restringido de la forma cuadrática no excede 4. Y este valor se logra para $\mathbf{x} = (0, 1, 0)$, que es un vector propio para el segundo valor propio más grande de la matriz de la forma cuadrática. ■

EJEMPLO 5 Sean A la matriz del ejemplo 3 y $\mathbf{u}_1$ un vector propio unitario correspondiente al valor propio más grande de A . Encuentre el valor máximo de $\mathbf{x}^T A \mathbf{x}$ sujeto a las condiciones

$$\mathbf{x}^T \mathbf{x} = 1, \quad \mathbf{x}^T \mathbf{u}_1 = 0 \tag{4}$$

SOLUCIÓN A partir del ejemplo 3, el segundo valor propio más grande de A es $\lambda = 3$. Resuelva $(A - 3I)\mathbf{x} = \mathbf{0}$ para encontrar un vector propio, y normalícelo para obtener

$$\mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

El vector $\mathbf{u}_2$ es automáticamente ortogonal a $\mathbf{u}_1$ porque los vectores corresponden a diferentes valores propios. Así, el máximo de $\mathbf{x}^T A \mathbf{x}$ sujeto a las restricciones en la ecuación (4) es 3, que se obtiene cuando $\mathbf{x} = \mathbf{u}_2$. ■

El siguiente teorema generaliza el teorema 7 y, junto con el teorema 6, ofrece una útil caracterización de *todos* los valores propios de A . Se omite la demostración.

TEOREMA 8

Sea A una matriz simétrica de $n \times n$ con una diagonalización ortogonal $A = PDP^{-1}$, donde las entradas sobre la diagonal en D están ordenadas de tal manera que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y las columnas de P son los vectores propios unitarios correspondientes $\mathbf{u}_1, \dots, \mathbf{u}_n$. Entonces, para $k = 2, \dots, n$, el valor máximo de $\mathbf{x}^T A \mathbf{x}$ sujeto a las restricciones

$$\mathbf{x}^T \mathbf{x} = 1, \quad \mathbf{x}^T \mathbf{u}_1 = 0, \quad \dots, \quad \mathbf{x}^T \mathbf{u}_{k-1} = 0$$

es el valor propio λ_k , y este máximo se logra en $\mathbf{x} = \mathbf{u}_k$.

El teorema 8 será útil en las secciones 8.4 y 8.5. La siguiente aplicación solo requiere del teorema 6.

EJEMPLO 6 El gobierno de un condado planea reparar x cientos de millas de caminos públicos y puentes, y mejorar y cientos de acres de parques y áreas recreativas durante el próximo año. El condado debe decidir cómo distribuir sus recursos (fondos, equipo, mano de obra, etc.) entre esos dos proyectos. Si es más eficaz en términos de costos trabajar simultáneamente en ambos proyectos en lugar de trabajar en uno solo, entonces x y y podrían satisfacer una *restricción* del tipo

$$4x^2 + 9y^2 \leq 36$$

Véase la figura 3. Cada punto (x, y) en el *conjunto factible* sombreado representa una posible obra pública programada para el año. Los puntos sobre la curva de restricción, $4x^2 + 9y^2 = 36$, utilizan las cantidades máximas de recursos disponibles.

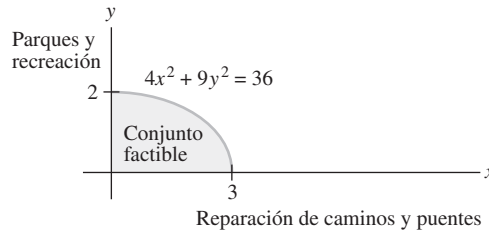


FIGURA 3 Programas de trabajos públicos.

Al elegir su programa de obras públicas, el condado quiere considerar las opiniones de sus residentes. Para medir el valor, o *utilidad*, que los residentes de un lugar asignarían a los diversos programas de trabajo (x, y) , algunas veces los economistas emplean una función de la forma

$$q(x, y) = xy$$

El conjunto de puntos (x, y) en el cual $q(x, y)$ es una constante se denomina *curva de indiferencia*. En la figura 4 se muestran tres de estas curvas. Los puntos sobre una curva de indiferencia corresponden a las alternativas que los residentes del condado, como grupo, encontrarían igualmente valiosas.² Encuentre el programa de obras públicas que maximice la función utilidad q .

SOLUCIÓN La ecuación de restricción $4x^2 + 9y^2 = 36$ no describe un conjunto de vectores unitarios, pero un cambio de variable puede resolver el problema. Rescriba la restricción en la forma

$$\left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1$$

² Las curvas de indiferencia se analizan en Michael D. Intriligator, Ronald G. Bodkin y Cheng Hsiao, *Econometric Models, Techniques, and Applications* (Upper Saddle River, NJ: Prentice-Hall, 1996).

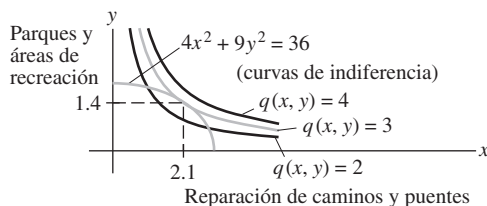


FIGURA 4 El programa de obras públicas óptimo es (2.1, 1.4).

y defina

$$x_1 = \frac{x}{3}, \quad x_2 = \frac{y}{2}, \quad \text{es decir,} \quad x = 3x_1 \quad y \quad y = 2x_2$$

Entonces, la ecuación de restricción se convierte en

$$x_1^2 + x_2^2 = 1$$

y la función utilidad se convierte en $q(3x_1, 2x_2) = (3x_1)(2x_2) = 6x_1x_2$. Sea $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Entonces, el problema es maximizar $Q(\mathbf{x}) = 6x_1x_2$ sujeto a $\mathbf{x}^T\mathbf{x} = 1$. Observe que $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$, donde

$$A = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$$

Los valores propios de A son ± 3 , con vectores propios $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ para $\lambda = 3$ y $\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ para

$\lambda = -3$. Así, el valor máximo de $Q(\mathbf{x}) = q(x_1, x_2)$ es 3, que se alcanza cuando $x_1 = 1/\sqrt{2}$ y $x_2 = 1/\sqrt{2}$.

En términos de las variables originales, el programa de trabajos públicos óptimo es $x = 3x_1 = 3/\sqrt{2} \approx 2.1$ cientos de millas de caminos y puentes, y $y = 2x_2 = \sqrt{2} \approx 1.4$ cientos de acres de parques y áreas de recreación. El programa de obras públicas óptimo es justo el punto donde se encuentran la curva de restricción y la curva de indiferencia $q(x, y) = 3$. Los puntos (x, y) con la utilidad más alta están sobre las curvas de indiferencia que no tocan la curva de restricción. Véase la figura 4. ■

PROBLEMAS DE PRÁCTICA

1. Sea $Q(\mathbf{x}) = 3x_1^2 + 3x_2^2 + 2x_1x_2$. Encuentre un cambio de variable que transforme Q en una forma cuadrática sin el producto cruzado, y determine la nueva forma cuadrática.
2. Con Q como en el problema 1, encuentre el valor máximo de $Q(\mathbf{x})$ con la restricción $\mathbf{x}^T\mathbf{x} = 1$, y determine un vector unitario en el que se logre el máximo.

8.3 EJERCICIOS†

En los ejercicios 1 y 2, encuentre el cambio de variable $\mathbf{x} = P\mathbf{y}$ que transforme la forma cuadrática $\mathbf{x}^T A \mathbf{x}$ en $\mathbf{y}^T D \mathbf{y}$ como se muestra.

1. $5x_1^2 + 6x_2^2 + 7x_3^2 + 4x_1x_2 - 4x_2x_3 = 9y_1^2 + 6y_2^2 + 3y_3^2$
2. $3x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3 = 5y_1^2 + 2y_2^2$

[Sugerencia: $\mathbf{x}$ y $\mathbf{y}$ deben tener el mismo número de coordenadas, de manera que la forma cuadrática que se muestra aquí debe tener un coeficiente cero para y_3^2].

En los ejercicios 3 a 6, encuentre: a) el valor máximo de $Q(\mathbf{x})$ con la restricción $\mathbf{x}^T\mathbf{x} = 1$, b) un vector unitario $\mathbf{u}$ donde se logre este máximo, y c) el máximo de $Q(\mathbf{x})$ con las condiciones $\mathbf{x}^T\mathbf{x} = 1$ y $\mathbf{x}^T\mathbf{u} = 0$.

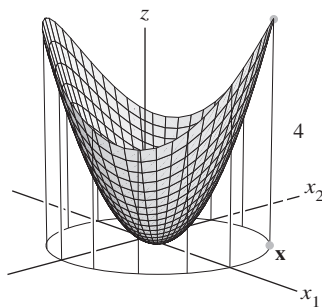
3. $Q(\mathbf{x}) = 5x_1^2 + 6x_2^2 + 7x_3^2 + 4x_1x_2 - 4x_2x_3$
(Véase el ejercicio 1).

4. $Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3$
(Véase el ejercicio 2).
5. $Q(\mathbf{x}) = 5x_1^2 + 5x_2^2 - 4x_1x_2$
6. $Q(\mathbf{x}) = 7x_1^2 + 3x_2^2 + 3x_1x_2$
7. Sea $Q(\mathbf{x}) = -2x_1^2 - x_2^2 + 4x_1x_2 + 4x_2x_3$. Determine un vector unitario $\mathbf{x}$ en $\mathbb{R}^3$ en el cual $Q(\mathbf{x})$ sea máximo, con la restricción $\mathbf{x}^T\mathbf{x} = 1$. [Sugerencia: Los valores propios de la matriz de la forma cuadrática Q son 2, -1 y -4].
8. Sea $Q(\mathbf{x}) = 7x_1^2 + x_2^2 + 7x_3^2 - 8x_1x_2 - 4x_1x_3 - 8x_2x_3$. Obtenga un vector unitario $\mathbf{x}$ en $\mathbb{R}^3$ en el que $Q(\mathbf{x})$ sea máximo, con la restricción $\mathbf{x}^T\mathbf{x} = 1$. [Sugerencia: Los valores propios de la matriz de la forma cuadrática Q son 9 y -3].
9. Encuentre el valor máximo de $Q(\mathbf{x}) = 7x_1^2 + 3x_2^2 - 2x_1x_2$, con la restricción $x_1^2 + x_2^2 = 1$. (No continúe para encontrar un vector donde se alcance el máximo).
10. Obtenga el valor máximo de $Q(\mathbf{x}) = -3x_1^2 + 5x_2^2 - 2x_1x_2$, con la restricción $x_1^2 + x_2^2 = 1$. (No continúe para encontrar un vector donde se alcance el máximo).
11. Suponga que $\mathbf{x}$ es un vector propio unitario de una matriz A correspondiente a un valor propio 3. ¿Cuál es el valor de $\mathbf{x}^T A \mathbf{x}$?

- *12. Sea λ cualquier valor propio de una matriz simétrica A . Justifique el enunciado hecho en esta sección de que $m \leq \lambda \leq M$, donde m y M se definen como en la ecuación (2). [Sugerencia: Encuentre un $\mathbf{x}$ tal que $\lambda = \mathbf{x}^T A \mathbf{x}$].
- *13. Considere que A es una matriz simétrica de $n \times n$, y que M y m representan los valores máximo y mínimo de la forma cuadrática $\mathbf{x}^T A \mathbf{x}$; denote los vectores propios unitarios correspondientes como $\mathbf{u}_1$ y $\mathbf{u}_n$. Los siguientes cálculos demuestran que dado cualquier número t entre M y m , existe un vector unitario $\mathbf{x}$ tal que $t = \mathbf{x}^T A \mathbf{x}$. Compruebe que $t = (1 - \alpha)m + \alpha M$ para algún número α entre 0 y 1. Después deje que $\mathbf{x} = \sqrt{1 - \alpha}\mathbf{u}_n + \sqrt{\alpha}\mathbf{u}_1$, y demuestre que $\mathbf{x}^T\mathbf{x} = 1$ y $\mathbf{x}^T A \mathbf{x} = t$.

[M] En los ejercicios 14 a 17, siga las instrucciones para los ejercicios 3 a 6.

14. $x_1x_2 + 3x_1x_3 + 30x_1x_4 + 30x_2x_3 + 3x_2x_4 + x_3x_4$
15. $3x_1x_2 + 5x_1x_3 + 7x_1x_4 + 7x_2x_3 + 5x_2x_4 + 3x_3x_4$
16. $4x_1^2 - 6x_1x_2 - 10x_1x_3 - 10x_1x_4 - 6x_2x_3 - 6x_2x_4 - 2x_3x_4$
17. $-6x_1^2 - 10x_2^2 - 13x_3^2 - 13x_4^2 - 4x_1x_2 - 4x_1x_3 - 4x_1x_4 + 6x_3x_4$



El valor máximo de $Q(\mathbf{x})$ sujeto a $\mathbf{x}^T\mathbf{x} = 1$ es 4.

SOLUCIONES A LOS PROBLEMAS DE PRÁCTICA

1. La matriz de la forma cuadrática es $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. Es fácil encontrar los valores propios, 4 y 2, y los vectores propios unitarios correspondientes, $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ y $\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. Así que el cambio deseado de variable es $\mathbf{x} = P\mathbf{y}$, donde $P = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. (Aquí un error común es olvidar normalizar los vectores propios). La nueva forma cuadrática es $\mathbf{y}^T D \mathbf{y} = 4y_1^2 + 2y_2^2$.
2. El máximo de $Q(\mathbf{x})$ para un vector unitario $\mathbf{x}$ es 4, y el máximo se logra en el vector propio unitario $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. [Una respuesta incorrecta común es $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Este vector maximiza la forma cuadrática $\mathbf{y}^T D \mathbf{y}$ en vez de $Q(\mathbf{x})$].

8.4 DESCOMPOSICIÓN EN VALORES SINGULARES

Los teoremas de diagonalización de las secciones 7.3 y 8.1 intervienen en muchas aplicaciones interesantes. Por desgracia, como se sabe, no todas las matrices se pueden factorizar como $A = PDP^{-1}$ con D diagonal. Sin embargo, ¡una factorización $A = QDP^{-1}$ es posible para cualquier matriz A de $m \times n$! Una factorización especial de este tipo, llamada *descomposición en valores singulares*, es una de las más útiles factorizaciones matriciales en álgebra lineal aplicada.

La descomposición en valores singulares se basa en la siguiente propiedad de la diagonalización ordinaria que se puede imitar para matrices rectangulares: los valores absolutos de los valores propios de una matriz simétrica A miden las cantidades que A estira o comprime

ciertos vectores (los vectores propios). Si $A\mathbf{x} = \lambda\mathbf{x}$ y $\|\mathbf{x}\| = 1$, entonces

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\| = |\lambda| \quad (1)$$

Si λ_1 es el valor propio con la mayor magnitud, entonces un vector propio unitario $\mathbf{v}_1$ correspondiente identifica una dirección en la cual el efecto de estiramiento de A es máximo. Es decir, la longitud de $A\mathbf{x}$ se maximiza cuando $\mathbf{x} = \mathbf{v}_1$, y $\|A\mathbf{v}_1\| = |\lambda_1|$, por la ecuación (1). Esta descripción de $\mathbf{v}_1$ y $|\lambda_1|$ tiene un análogo para matrices rectangulares que conducirán a la descomposición en valores singulares.

EJEMPLO 1 Si $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, entonces la transformación lineal $\mathbf{x} \mapsto A\mathbf{x}$ mapea la esfera unitaria $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ en $\mathbb{R}^3$ sobre una elipse en $\mathbb{R}^2$, como se muestra en la figura 1. Encuentre un vector unitario $\mathbf{x}$ en el que la longitud $\|A\mathbf{x}\|$ se maximiza, y calcule esa longitud máxima.

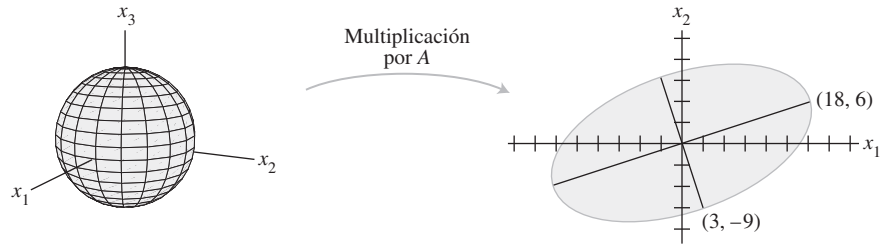


FIGURA 1 Una transformación de $\mathbb{R}^3$ a $\mathbb{R}^2$.

SOLUCIÓN La cantidad $\|A\mathbf{x}\|^2$ se maximiza con la misma $\mathbf{x}$ que maximiza $\|A\mathbf{x}\|$, y $\|A\mathbf{x}\|^2$ es más fácil de estudiar. Observe que

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T (A\mathbf{x}) = \mathbf{x}^T A^T A \mathbf{x} = \mathbf{x}^T (A^T A) \mathbf{x}$$

Además, $A^T A$ es una matriz simétrica, ya que $(A^T A)^T = A^T A^{TT} = A^T A$. Así que ahora el problema es maximizar la forma cuadrática $\mathbf{x}^T (A^T A) \mathbf{x}$ sujeta a la restricción $\|\mathbf{x}\| = 1$. Por el teorema 6 de la sección 8.3, el valor máximo es el valor propio más grande λ_1 de $A^T A$. Además, el valor máximo se alcanza en un vector propio unitario de $A^T A$ correspondiente a λ_1 .

Para la matriz A en este ejemplo,

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Los valores propios de $A^T A$ son $\lambda_1 = 360$, $\lambda_2 = 90$ y $\lambda_3 = 0$. Los vectores propios unitarios correspondientes son, respectivamente,

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

El valor máximo de $\|A\mathbf{x}\|^2$ es 360, que se alcanza cuando $\mathbf{x}$ es el vector unitario $\mathbf{v}_1$. El vector $A\mathbf{v}_1$ es el punto sobre la elipse de la figura 1 más alejado del origen, a saber,

$$A\mathbf{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}$$

Para $\|\mathbf{x}\| = 1$, el valor máximo de $\|A\mathbf{x}\|$ es $\|A\mathbf{v}_1\| = \sqrt{360} = 6\sqrt{10}$. ■

El ejemplo 1 sugiere que el efecto de A sobre la esfera unitaria en $\mathbb{R}^3$ está relacionado con la forma cuadrática $\mathbf{x}^T (A^T A) \mathbf{x}$. De hecho, el comportamiento geométrico general de la transformación $\mathbf{x} \mapsto A\mathbf{x}$ se refleja en esta forma cuadrática, como se verá más adelante.

Los valores singulares de una matriz de $m \times n$

Sea A una matriz de $m \times n$. $A^T A$ es simétrica y se puede diagonalizar ortogonalmente. Sea $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ una base ortonormal para $\mathbb{R}^n$ que consiste en vectores propios de $A^T A$, y sean $\lambda_1, \dots, \lambda_n$ los valores propios asociados de $A^T A$. Por lo tanto, para $1 \leq i \leq n$,

$$\begin{aligned}\|\mathbf{A}\mathbf{v}_i\|^2 &= (\mathbf{A}\mathbf{v}_i)^T \mathbf{A}\mathbf{v}_i = \mathbf{v}_i^T A^T \mathbf{A}\mathbf{v}_i \\ &= \mathbf{v}_i^T (\lambda_i \mathbf{v}_i) && \text{Puesto que } \mathbf{v}_i \text{ es un vector propio de } A^T A \\ &= \lambda_i && \text{Porque } \mathbf{v}_i \text{ es un vector unitario}\end{aligned}\quad (2)$$

De esta forma, todos los valores propios de $A^T A$ son no negativos. Al volver a numerar, si es necesario, se puede suponer que los valores propios están arreglados de manera que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

Los **valores singulares** de A son las raíces cuadradas de los valores propios de $A^T A$, que se denotan con $\sigma_1, \dots, \sigma_n$, y se arreglan en orden decreciente. Es decir, $\sigma_i = \sqrt{\lambda_i}$ para $1 \leq i \leq n$. De acuerdo con la ecuación (2), *los valores singulares de A son las longitudes de los vectores $\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_n$.*

EJEMPLO 2 Sea A la matriz del ejemplo 1. Como los valores propios de $A^T A$ son 360, 90 y 0, entonces los valores singulares de A son

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0$$

A partir del ejemplo 1, el primer valor singular de A es el máximo de $\|\mathbf{A}\mathbf{x}\|$ sobre todos los vectores unitarios, y se alcanza el máximo en el vector propio unitario $\mathbf{v}_1$. El teorema 7 de la sección 8.3 indica que el segundo valor singular de A es el máximo de $\|\mathbf{A}\mathbf{x}\|$ sobre todos los vectores unitarios que sean *ortogonales* a $\mathbf{v}_1$, y este máximo se logra en el segundo vector propio unitario, $\mathbf{v}_2$ (ejercicio 22). Para $\mathbf{v}_2$ del ejemplo 1,

$$\mathbf{A}\mathbf{v}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

Este punto está sobre el eje menor de la elipse de la figura 1, así como $\mathbf{A}\mathbf{v}_1$ está sobre el eje mayor. (Véase la figura 2). Los primeros dos valores singulares de A son las longitudes de los semiejes mayor y menor de la elipse. ■

En la figura 2, no es casual que $\mathbf{A}\mathbf{v}_1$ y $\mathbf{A}\mathbf{v}_2$ sean ortogonales, como muestra el siguiente teorema.

TEOREMA 9

Suponga que $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ es una base ortonormal de $\mathbb{R}^n$ que consiste en vectores propios de $A^T A$, arreglados de tal forma que los valores propios correspondientes de $A^T A$ satisfacen $\lambda_1 \geq \dots \geq \lambda_n$, y suponga que A tiene r valores singulares diferentes de cero. Entonces, $\{\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_r\}$ es una base ortogonal para $\text{Col } A$, y $\text{rango } A = r$.

DEMOSTRACIÓN Como $\mathbf{v}_i$ y $\lambda_j \mathbf{v}_j$ son ortogonales para $i \neq j$,

$$(\mathbf{A}\mathbf{v}_i)^T (\mathbf{A}\mathbf{v}_j) = \mathbf{v}_i^T A^T \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^T (\lambda_j \mathbf{v}_j) = 0$$

Por lo tanto, $\{\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_n\}$ es un conjunto ortogonal. Además, ya que las longitudes de los vectores $\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_n$ son los valores singulares de A , y puesto que hay r valores singulares diferentes de cero, $\mathbf{A}\mathbf{v}_i \neq \mathbf{0}$ si y solo si $1 \leq i \leq r$. Así que $\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_r$ son vectores linealmente

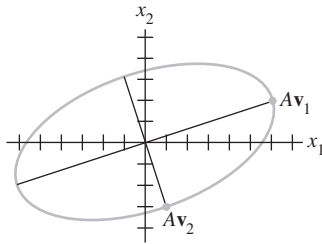


FIGURA 2

independientes, y están en Col A . Finalmente, para cualquier $\mathbf{y}$ en Col A , por ejemplo, $\mathbf{y} = A\mathbf{x}$, se puede escribir $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, y

$$\begin{aligned}\mathbf{y} &= A\mathbf{x} = c_1A\mathbf{v}_1 + \cdots + c_rA\mathbf{v}_r + c_{r+1}A\mathbf{v}_{r+1} + \cdots + c_nA\mathbf{v}_n \\ &= c_1A\mathbf{v}_1 + \cdots + c_rA\mathbf{v}_r + 0 + \cdots + 0\end{aligned}$$

Así, $\mathbf{y}$ está en Gen $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$, lo que demuestra que $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ es una base (ortogonal) para Col A . Por lo tanto, rango $A = \dim \text{Col } A = r$. ■

NOTA NUMÉRICA

En algunos casos, el rango de A puede ser muy sensible a pequeños cambios en las entradas de A . El método evidente de contar el número de columnas pivote en A no funciona muy bien si A se reduce por filas con una computadora. Con frecuencia, el error por redondeo genera una forma escalonada con rango completo.

En la práctica, la manera más confiable de estimar el rango de una matriz A grande consiste en contar el número de valores singulares diferentes de cero. En este caso, se supone que los valores singulares distintos de cero y extremadamente pequeños son iguales a cero para todos los fines prácticos, y el *rango efectivo* de la matriz es el número que se obtiene al contar los valores singulares restantes diferentes de cero.¹

Descomposición en valores singulares

La descomposición de A implica a una matriz Σ “diagonal” de $m \times n$ de la forma

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{array}{l} \leftarrow m - r \text{ filas} \\ \leftarrow n - r \text{ columnas} \end{array} \quad (3)$$

donde D es una matriz diagonal de $r \times r$ para alguna r que no exceda el mínimo de m y n . [Si r es igual a m o n o a ambos, entonces no aparecen algunas (o todas) las matrices cero].

TEOREMA 10

Descomposición en valores singulares

Sea A una matriz de $m \times n$ con rango r . Entonces existe una matriz Σ de $m \times n$ como la de la ecuación (3) para la cual las entradas diagonales en D son los primeros r valores singulares de A , $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, y existen una matriz ortogonal U de $m \times m$ y una matriz ortogonal V de $n \times n$ tales que

$$A = U\Sigma V^T$$

Cualquier factorización $A = U\Sigma V^T$, con U y V ortogonales, Σ como en la ecuación (3), y entradas diagonales positivas en D , se llama **descomposición en valores singulares** (o **DVS**) de A . Las matrices U y V no están determinadas de forma única por A , pero las entradas diagonales en D son necesariamente los valores singulares de A . Véase el ejercicio 19. Las columnas de U en esta descomposición se llaman **vectores singulares izquierdos** de A , y las columnas de V se llaman **vectores singulares derechos** de A .

¹ En general, no es sencillo estimar el rango. Para un análisis de los aspectos sutiles implicados, véase Philip E. Gill, Walter Murray y Margaret H. Wright, *Numerical Linear Algebra and Optimization*, vol.1 (Redwood City, CA: Addison-Wesley, 1991), sección 5.8.

DEMOSTRACIÓN Sean λ_i y $\mathbf{v}_i$ como en el teorema 9, así $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ es una base ortogonal para Col A . Normalice cada $A\mathbf{v}_i$ para obtener una base ortonormal $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, donde

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

y

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad (1 \leq i \leq r) \quad (4)$$

Ahora se extiende $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ a una base ortonormal $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ de $\mathbb{R}^m$, y sean

$$U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m] \quad \text{y} \quad V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$$

Por construcción, U y V son matrices ortogonales. También, a partir de la ecuación (4),

$$AV = [A\mathbf{v}_1 \ \cdots \ A\mathbf{v}_r \ \mathbf{0} \ \cdots \ \mathbf{0}] = [\sigma_1 \mathbf{u}_1 \ \cdots \ \sigma_r \mathbf{u}_r \ \mathbf{0} \ \cdots \ \mathbf{0}]$$

Sea D la matriz diagonal con entradas diagonales $\sigma_1, \dots, \sigma_r$, con Σ como en la ecuación (3) anterior. Entonces,

$$\begin{aligned} U\Sigma &= [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & & & & 0 \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_r & 0 \\ \hline & & & & 0 \end{array} \right] \\ &= [\sigma_1 \mathbf{u}_1 \ \cdots \ \sigma_r \mathbf{u}_r \ \mathbf{0} \ \cdots \ \mathbf{0}] \\ &= AV \end{aligned}$$

Puesto que V es una matriz ortogonal, entonces $U\Sigma V^T = AVV^T = A$. ■

Los siguientes dos ejemplos se concentran en la estructura interna de una descomposición en valores singulares. Un algoritmo eficiente y numéricamente estable para esta descomposición utilizaría un enfoque diferente. Véase la nota numérica al final de la sección.

EJEMPLO 3 Utilice los resultados de los ejemplos 1 y 2 para construir una descomposición en valores singulares de $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

SOLUCIÓN Una construcción se puede dividir en tres pasos.

Paso 1. Encuentre una diagonalización ortogonal de $A^T A$. Es decir, determine los valores propios de $A^T A$ y un conjunto ortonormal correspondiente de vectores propios. Si A tuviera solo dos columnas, los cálculos podrían realizarse a mano. Matrices mucho más grandes, por lo general, requieren de un programa de matrices.² Sin embargo, para la matriz A en cuestión, los datos propios para $A^T A$ se dan en el ejemplo 1.

Paso 2. Obtenga V y Σ . Acomode los valores propios de $A^T A$ en orden decreciente. En el ejemplo 1, los valores propios ya están listados en orden descendente: 360, 90 y 0. Los vectores propios unitarios correspondientes, $\mathbf{v}_1$, $\mathbf{v}_2$ y $\mathbf{v}_3$, son los vectores singulares derechos de A . Utilizando el ejemplo 1, se construye

$$V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

² Véase la *Guía de estudio* si se desea consultar las instrucciones adecuadas para software y la calculadora gráfica. MATLAB, por ejemplo, puede dar los valores propios y los vectores propios mediante una sola instrucción, e.g.

Las raíces cuadradas de los valores propios son los valores singulares:

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

Los valores singulares diferentes de cero son las entradas diagonales de D . La matriz Σ tiene el mismo tamaño que A , con D en su esquina superior izquierda y ceros en las entradas restantes.

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \quad \Sigma = [D \quad 0] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

Paso 3. Construya U . Cuando A tiene rango r , las primeras r columnas de U son los vectores normalizados obtenidos de $A\mathbf{v}_1, \dots, A\mathbf{v}_r$. En este ejemplo, A tiene dos valores singulares diferentes de cero, de manera que $\text{rango } A = 2$. Recuerde de la ecuación (2) y del párrafo anterior al ejemplo 2 que $\|A\mathbf{v}_1\| = \sigma_1$ y $\|A\mathbf{v}_2\| = \sigma_2$. Por lo tanto,

$$\begin{aligned} \mathbf{u}_1 &= \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix} \\ \mathbf{u}_2 &= \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix} \end{aligned}$$

Observe que $\{\mathbf{u}_1, \mathbf{u}_2\}$ ya es una base para $\mathbb{R}^2$, por lo que no se necesitan vectores adicionales para U , y $U = [\mathbf{u}_1 \quad \mathbf{u}_2]$. La descomposición en valores singulares de A es

$$A = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$$

$\uparrow$ $\uparrow$ $\uparrow$
 U Σ V^T

EJEMPLO 4 Encuentre una descomposición en valores singulares de $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$.

SOLUCIÓN Primero, calcule $A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. Los valores propios de $A^T A$ son 18 y 0, con los vectores propios unitarios correspondientes

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Esos vectores unitarios forman las columnas de V :

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Los valores singulares son $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ y $\sigma_2 = 0$. Como solo existe un valor singular distinto de cero, la “matriz” D se puede escribir como un número individual. Es decir, $D = 3\sqrt{2}$. La matriz Σ tiene el mismo tamaño que A , con D en su esquina superior izquierda:

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Para construir U , primero se obtienen $A\mathbf{v}_1$ y $A\mathbf{v}_2$:

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

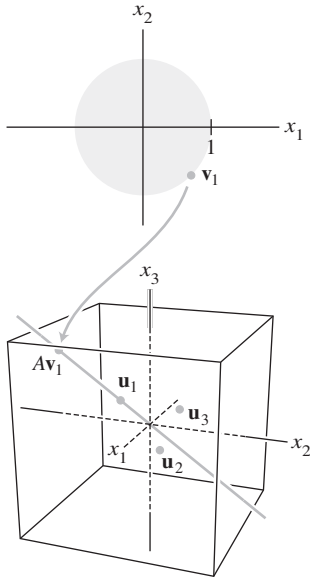


FIGURA 3

Para validar los cálculos, compruebe que $\|Av_1\| = \sigma_1 = 3\sqrt{2}$. Desde luego, $Av_2 = \mathbf{0}$ porque $\|Av_2\| = \sigma_2 = 0$. Hasta ahora, la única columna encontrada para U es

$$\mathbf{u}_1 = \frac{1}{3\sqrt{2}} Av_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

Las otras columnas de U se obtienen ampliando el conjunto $\{\mathbf{u}_1\}$ a una base ortonormal para $\mathbb{R}^3$. En este caso, se requieren dos vectores unitarios ortogonales $\mathbf{u}_2$ y $\mathbf{u}_3$ que sean ortogonales a $\mathbf{u}_1$. (Véase la figura 3). Cada vector debe satisfacer $\mathbf{u}_1^T \mathbf{x} = 0$, lo que es equivalente a la ecuación $x_1 - 2x_2 + 2x_3 = 0$. Una base para el conjunto solución de esta ecuación es

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

(Compruebe que $\mathbf{w}_1$ y $\mathbf{w}_2$ son ortogonales a $\mathbf{u}_1$). Aplique el proceso de Gram-Schmidt (con normalizaciones) a $\{\mathbf{w}_1, \mathbf{w}_2\}$, para obtener

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

Finalmente, establezca $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3]$, tome Σ y V^T de las ecuaciones anteriores, y escriba

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

■

Aplicaciones de la descomposición en valores singulares

Como se mencionó antes, la DVS se utiliza con frecuencia para estimar el rango de una matriz. Ahora se describirán brevemente otras aplicaciones numéricas, y en la sección 8.5 se expone una aplicación al procesamiento de imágenes.

EJEMPLO 5 (El número de condición) La mayoría de los cálculos numéricos que implican una ecuación $A\mathbf{x} = \mathbf{b}$ son tan confiables como es posible cuando se utiliza la DVS de A . Las dos matrices ortogonales U y V no afectan las longitudes de vectores o los ángulos entre vectores (teorema 7 de la sección 5.2). Cualquier posible inestabilidad en los cálculos numéricos se identifica en Σ . Si los valores singulares de A son extremadamente grandes o pequeños, los errores por redondeo son casi inevitables, pero un análisis de errores se facilita si se conocen las entradas en Σ y V .

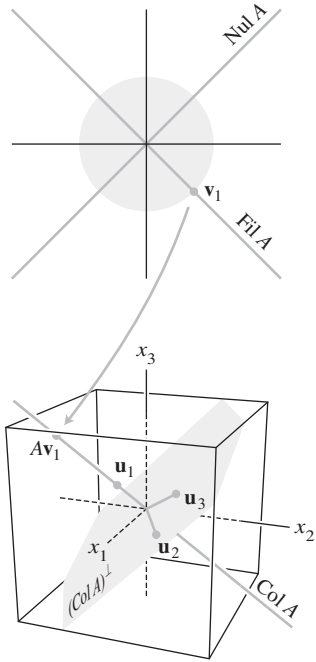
Si A es una matriz invertible de $n \times n$, entonces la razón σ_1/σ_n de los valores singulares más grande y más pequeño da el **número de condición** de A . Los ejercicios 41 a 43 de la sección 2.3 mostraron cómo el número de condición afecta la sensibilidad de una solución de $A\mathbf{x} = \mathbf{b}$ ante cambios (o errores) en las entradas de A . (En realidad, un “número de condición” de A se puede calcular en diversas formas, pero la definición que se presenta aquí se utiliza ampliamente para estudiar $A\mathbf{x} = \mathbf{b}$).

■

EJEMPLO 6 (Bases para subespacios fundamentales) Dada una DVS para una matriz A de $m \times n$, sean $\mathbf{u}_1, \dots, \mathbf{u}_m$ los vectores singulares izquierdos, $\mathbf{v}_1, \dots, \mathbf{v}_n$ los vectores singulares derechos, $\sigma_1, \dots, \sigma_n$ los valores singulares, y r el rango de A . De acuerdo con el teorema 9,

$$\{\mathbf{u}_1, \dots, \mathbf{u}_r\} \quad (5)$$

es una base ortonormal para $\text{Col } A$.



Los subespacios fundamentales en el ejemplo 4.

Recuerde del teorema 3 de la sección 5.1 que $(\text{Col } A)^\perp = \text{Nul } A^T$. Por lo tanto,

$$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\} \quad (6)$$

es una base ortonormal para $\text{Nul } A^T$.

Puesto que $\|\mathbf{A}\mathbf{v}_i\| = \sigma_i$ para $1 \leq i \leq n$, y σ_i es 0 si y solo si $i > r$, por lo tanto los vectores $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ generan un subespacio de $\text{Nul } A$ de dimensión $n - r$. De acuerdo con el teorema del rango, $\dim \text{Nul } A = n - \text{rango } A$. De ello se deduce que

$$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\} \quad (7)$$

es una base ortonormal para $\text{Nul } A$, de acuerdo con el teorema de la base (en la sección 4.5).

A partir de (5) y (6), el complemento ortogonal de $\text{Nul } A^T$ es $\text{Col } A$. Intercambiando A y A^T , observe que $(\text{Nul } A)^\perp = \text{Col } A^T = \text{Fil } A$. En consecuencia, a partir de (7),

$$\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \quad (8)$$

es una base ortonormal para $\text{Fil } A$.

La figura 4 resume las ecuaciones (5) a (8), pero muestra la base ortogonal $\{\sigma_1 \mathbf{u}_1, \dots, \sigma_r \mathbf{u}_r\}$ para $\text{Col } A$ en vez de la base normalizada, con la finalidad de recordarnos que $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ para $1 \leq i \leq r$. Las bases ortonormales explícitas para los cuatro subespacios fundamentales determinados por A son útiles en algunos cálculos, particularmente en problemas de optimización restringidos. ■

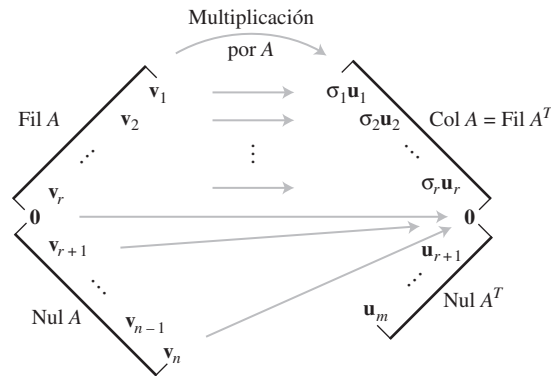


FIGURA 4 Los cuatro subespacios fundamentales y la acción de A .

Los cuatro subespacios fundamentales y el concepto de valores singulares proporcionan los enunciados finales del teorema de la matriz invertible. (Recuerde que se omitieron los enunciados del teorema acerca de A^T , ya que, de otra forma, el número de enunciados se duplicaría). Los otros enunciados se presentaron en las secciones 2.3, 2.9, 3.2, 4.6 y 7.2.

TEOREMA

Teorema de la matriz invertible (concluido)

Sea A una matriz de $n \times n$. Entonces, cada uno de los siguientes enunciados es equivalente al enunciado de que A es una matriz invertible.

- $u)$ $(\text{Col } A)^\perp = \{\mathbf{0}\}$.
- $v)$ $(\text{Nul } A)^\perp = \mathbb{R}^n$.
- $w)$ $\text{Fil } A = \mathbb{R}^n$.
- $x)$ A tiene n valores singulares diferentes de cero.

EJEMPLO 7 (DVS reducida y la pseudoinversa de A) Cuando Σ contiene filas o columnas de ceros, es posible una descomposición más compacta de A. Utilizando la notación ya establecida, sea $r = \text{rango } A$, y U y V se particionan en submatrices cuyos primeros bloques contienen r columnas:

$$U = [U_r \quad U_{m-r}], \quad \text{donde } U_r = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r]$$

$$V = [V_r \quad V_{n-r}], \quad \text{donde } V_r = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r]$$

Entonces U_r es de $m \times r$, y V_r es de $n \times r$. (Para simplificar la notación, considere U_{m-r} o V_{n-r} aun cuando tal vez alguna de ellas no tenga columnas). Por lo tanto, la multiplicación de matrices particionadas indica que

$$A = [U_r \quad U_{m-r}] \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T \quad (9)$$

Esta factorización de A se llama **descomposición en valores singulares reducida** de A. Como las entradas diagonales en D son diferentes de cero, D es invertible. La siguiente matriz se llama la **pseudoinversa** (o **inversa de Moore-Penrose**) de A:

$$A^+ = V_r D^{-1} U_r^T \quad (10)$$

Al final del capítulo, los ejercicios complementarios 12 a 14 exploran algunas de las propiedades de la descomposición en valores singulares reducida y de la pseudoinversa. ■

EJEMPLO 8 (Solución por mínimos cuadrados) Dada la ecuación $A\mathbf{x} = \mathbf{b}$, utilice la pseudoinversa de A en la ecuación (10) para definir

$$\hat{\mathbf{x}} = A^+ \mathbf{b} = V_r D^{-1} U_r^T \mathbf{b}$$

Entonces, a partir de la DVS en la ecuación (9),

$$\begin{aligned} A\hat{\mathbf{x}} &= (U_r D V_r^T)(V_r D^{-1} U_r^T \mathbf{b}) \\ &= U_r D D^{-1} U_r^T \mathbf{b} \quad \text{Porque } V_r^T V_r = I_r \\ &= U_r U_r^T \mathbf{b} \end{aligned}$$

De (5) se deduce que $U_r U_r^T \mathbf{b}$ es la proyección ortogonal $\hat{\mathbf{b}}$ de $\mathbf{b}$ sobre Col A. (Véase el teorema 10 de la sección 7.3). Por lo tanto, $\hat{\mathbf{x}}$ es una solución por mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$. En efecto, esta $\hat{\mathbf{x}}$ tiene la longitud más pequeña en comparación con todas las soluciones por mínimos cuadrados de $A\mathbf{x} = \mathbf{b}$. Véase el ejercicio complementario 14. ■

NOTA NUMÉRICA

Los ejemplos 1 a 4 y los ejercicios ilustran el concepto de valores singulares y sugieren cómo efectuar cálculos a mano. En la práctica, se debería evitar el cálculo de $A^T A$, porque cualquier error en las entradas de A se eleva al cuadrado en las entradas de $A^T A$. Existen métodos iterativos rápidos que producen los valores singulares y los vectores singulares de A con muchos decimales.

Lecturas adicionales

Horn, Roger A. y Charles R. Johnson, *Matrix Analysis* (Cambridge: Cambridge University Press, 1990).

Long, Cliff, "Visualization of Matrix Singular Value Decomposition". *Mathematics Magazine* **56** (1983), pp. 161-167.

Moler, C. B. y D. Morrison, "Singular Value Analysis of Cryptograms". *Amer. Math. Monthly* **90** (1983), pp. 78-87.

Strang, Gilbert, *Linear Algebra and Its Applications*, 4a. ed. (Belmont, CA: Brooks/Cole, 2005).

Watkins, David S., *Fundamentals of Matrix Computations* (Nueva York: Wiley, 1991), pp. 390-398, 409-421.

PROBLEMA DE PRÁCTICA

Dada una descomposición en valores singulares, $A = U\Sigma V^T$, encuentre una DVS de A^T . ¿Cómo están relacionados los valores singulares de A y A^T ?

8.4 EJERCICIOS[†]

En los ejercicios 1 a 4, obtenga los valores singulares de las matrices.

1. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$ 2. $\begin{bmatrix} -5 & 0 \\ 0 & 0 \end{bmatrix}$

3. $\begin{bmatrix} \sqrt{6} & 1 \\ 0 & \sqrt{6} \end{bmatrix}$ 4. $\begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}$

En los ejercicios 5 a 12, encuentre una DVS para cada matriz. [Sugerencia: En el ejercicio 11, una elección para U es

$$\begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{bmatrix}. \text{ En el ejercicio 12, una columna de } U$$

puede ser $\begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.

5. $\begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix}$ 6. $\begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$

7. $\begin{bmatrix} 2 & -1 \\ 2 & 2 \end{bmatrix}$ 8. $\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$

9. $\begin{bmatrix} 7 & 1 \\ 0 & 0 \\ 5 & 5 \end{bmatrix}$ 10. $\begin{bmatrix} 4 & -2 \\ 2 & -1 \\ 0 & 0 \end{bmatrix}$

11. $\begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}$ 12. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$

13. Encuentre la DVS de $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$. [Sugerencia: Trabaje con A^T].

14. En el ejercicio 7, obtenga un vector unitario $\mathbf{x}$ en el cual $A\mathbf{x}$ tiene longitud máxima.

*15. Suponga que la siguiente factorización es una DVS de una matriz A , con las entradas en U y V redondeadas a dos decimales.

$$A = \begin{bmatrix} .40 & -.78 & .47 \\ .37 & -.33 & -.87 \\ -.84 & -.52 & -.16 \end{bmatrix} \begin{bmatrix} 7.10 & 0 & 0 \\ 0 & 3.10 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ \times \begin{bmatrix} .30 & -.51 & -.81 \\ .76 & .64 & -.12 \\ .58 & -.58 & .58 \end{bmatrix}$$

a) ¿Cuál es el rango de A ?

b) Utilice esta descomposición de A , sin hacer cálculos, y escriba una base para $\text{Col } A$ y una base para $\text{Nul } A$. [Sugerencia: Primero escriba las columnas de V].

16. Repita el ejercicio 15 para la siguiente DVS de una matriz A de 3×4 :

$$A = \begin{bmatrix} -.86 & -.11 & -.50 \\ .31 & .68 & -.67 \\ .41 & -.73 & -.55 \end{bmatrix} \begin{bmatrix} 12.48 & 0 & 0 & 0 \\ 0 & 6.34 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \times \begin{bmatrix} .66 & -.03 & -.35 & .66 \\ -.13 & -.90 & -.39 & -.13 \\ .65 & .08 & -.16 & -.73 \\ -.34 & .42 & -.84 & -.08 \end{bmatrix}$$

En los ejercicios 17 a 24, A es una matriz de $m \times n$ con una descomposición en valores singulares $A = U\Sigma V^T$, donde U es una matriz ortogonal de $m \times m$, Σ es una matriz "diagonal" de $m \times n$ con r entradas positivas y sin entradas negativas, y V es una matriz ortogonal de $n \times n$. Justifique sus respuestas.

17. Suponga que A es cuadrada e invertible. Encuentre una descomposición en valores singulares de A^{-1} .

18. Demuestre que si A es cuadrada, entonces $|\det A|$ es el producto de los valores singulares de A .

19. Demuestre que las columnas de V son vectores propios de $A^T A$, que las columnas de U son vectores propios de $A A^T$, y que las entradas diagonales de Σ son los valores singulares de A . [Sugerencia: Utilice la DVS para calcular $A^T A$ y $A A^T$].

*20. Demuestre que si A es una matriz positiva definida de $n \times n$, entonces una diagonalización ortogonal $A = P D P^T$ es una descomposición en valores singulares de A .

- *21. Demuestre que si P es una matriz ortogonal de $m \times m$, entonces PA y A tienen los mismos valores singulares.
- *22. Justifique el enunciado del ejemplo 2 referente a que el segundo valor singular de una matriz A es el máximo de $\|A\mathbf{x}\|$ conforme $\mathbf{x}$ varía sobre todos los vectores unitarios ortogonales a $\mathbf{v}_1$, siendo $\mathbf{v}_1$ un vector singular derecho correspondiente al primer valor singular de A . [Sugerencia: Utilice el teorema 7 de la sección 8.3].
23. Sean $U = [\mathbf{u}_1 \cdots \mathbf{u}_m]$ y $V = [\mathbf{v}_1 \cdots \mathbf{v}_n]$, donde $\mathbf{u}_j$ y $\mathbf{v}_j$ son como en el teorema 10. Demuestre que
- $$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$
24. Utilizando la notación del ejercicio 23, demuestre que $A^T \mathbf{u}_j = \sigma_j \mathbf{v}_j$ para $1 \leq j \leq r = \text{rango } A$.
- *25. Sea $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ una transformación lineal. Describa cómo encontrar una base $\mathcal{B}$ para $\mathbb{R}^n$ y una base $\mathcal{C}$ para $\mathbb{R}^m$ tal que la matriz para T respecto de $\mathcal{B}$ y $\mathcal{C}$ sea una matriz “diagonal” de $m \times n$.

[M] Calcule una DVS para cada matriz en los ejercicios 26 y 27. Informe las entradas matriciales finales con dos decimales. Utilice el método de los ejemplos 3 y 4.

$$26. A = \begin{bmatrix} -18 & 13 & -4 & 4 \\ 2 & 19 & -4 & 12 \\ -14 & 11 & -12 & 8 \\ -2 & 21 & 4 & 8 \end{bmatrix}$$

$$27. A = \begin{bmatrix} 6 & -8 & -4 & 5 & -4 \\ 2 & 7 & -5 & -6 & 4 \\ 0 & -1 & -8 & 2 & 2 \\ -1 & -2 & 4 & 4 & -8 \end{bmatrix}$$

28. [M] Determine los valores singulares de la matriz de 4×4 en el ejercicio 9 de la sección 2.3, y calcule el número de condición σ_1/σ_4 .
29. [M] Calcule los valores singulares de la matriz de 5×5 en el ejercicio 10 de la sección 2.3, y determine el número de condición σ_1/σ_5 .

SOLUCIÓN AL PROBLEMA DE PRÁCTICA

Si $A = U\Sigma V^T$, donde Σ es de $m \times n$, entonces $A^T = (V^T)^T \Sigma^T U^T = V \Sigma^T U^T$. Esto es una DVS de A^T , ya que V y U son matrices ortogonales y Σ^T es una matriz “diagonal” de $n \times m$. Como Σ y Σ^T tienen las mismas entradas diagonales diferentes de cero, A y A^T tienen los mismos valores singulares distintos de cero. [Nota: Si A es de $2 \times n$, entonces AA^T es solo de 2×2 y sus valores propios pueden ser más fáciles de calcular (a mano) que los valores propios de $A^T A$].

8.5 APLICACIONES AL PROCESAMIENTO DE IMÁGENES Y ESTADÍSTICA

Las fotografías satelitales en la introducción del capítulo dan un ejemplo de datos mutidimensionales, o *multivariados*; es decir, se trata de información organizada de manera que cada dato en el conjunto de datos esté identificado con un punto (vector) en $\mathbb{R}^n$. El principal objetivo de esta sección es explicar una técnica, llamada *análisis de componentes principales*, empleada para analizar datos multivariados. Los cálculos ilustrarán el uso de la diagonalización ortogonal y la descomposición en valores singulares.

El análisis de componentes principales es aplicable a cualquier conjunto de datos consistente en listas de mediciones realizadas a una colección de objetos o individuos. Por ejemplo, considere un proceso químico que produce un material plástico. Para monitorizar el proceso, se toman 300 muestras del material producido, y cada muestra se somete a una batería de ocho pruebas, tales como punto de fusión, densidad, ductilidad, resistencia a la tensión, etcétera. El informe del laboratorio para cada muestra es un vector en $\mathbb{R}^8$, y el conjunto de tales vectores forma una matriz de 8×300 , llamada **matriz de observaciones**.

En términos generales, se puede decir que los datos de control del proceso son octodimensionales. Los siguientes dos ejemplos describen datos que se pueden visualizar gráficamente.

EJEMPLO 1 Un ejemplo de datos bidimensionales está dado por un conjunto de pesos y alturas de N estudiantes de licenciatura. Sea $\mathbf{x}_j$ el **vector de observaciones** en $\mathbb{R}^2$ que lista el peso y la altura del j -ésimo estudiante. Si w denota el peso y h la altura, entonces la matriz

de observaciones tiene la forma

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \end{bmatrix}$$

$\uparrow \quad \uparrow \quad \quad \uparrow$
 $\mathbf{X}_1 \quad \mathbf{X}_2 \quad \quad \mathbf{X}_N$

El conjunto de vectores de observación se puede visualizar como una *gráfica de dispersión* bidimensional. Véase la figura 1. ■



FIGURA 1 Una gráfica de dispersión de vectores de observaciones $\mathbf{X}_1, \dots, \mathbf{X}_N$.

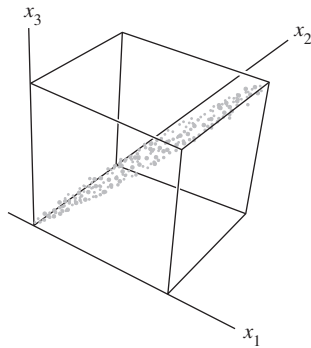


FIGURA 2 Gráfica de dispersión de datos espectrales para una imagen satelital.

EJEMPLO 2 Las primeras tres fotografías del Valle Railroad, Nevada, que se presentan en la introducción al capítulo, se pueden ver como *una* imagen de la región, con *tres componentes espectrales*, ya que se hicieron mediciones simultáneas de la región con tres distintas longitudes de onda. Cada fotografía da diferente información sobre la misma región física. Por ejemplo, el primer pixel en la esquina superior izquierda de cada fotografía corresponde al mismo lugar sobre el suelo (aproximadamente 30 metros por 30 metros). Cada pixel se asocia con un vector de observaciones en $\mathbb{R}^3$ que lista las intensidades de la señal para ese pixel en las tres bandas espectrales.

Por lo común, una imagen es de 2000×2000 píxeles, por lo que existen 4 millones de píxeles en la imagen. Los datos para la imagen forman una matriz con 3 filas y 4 millones de columnas (con las columnas arregladas en cualquier orden conveniente). En este caso, el carácter “multidimensional” de los datos se refiere a las tres dimensiones *espectrales* más que a las dos dimensiones *espaciales* que naturalmente pertenecen a cualquier fotografía. Los datos se pueden visualizar como un cúmulo de 4 millones de puntos en $\mathbb{R}^3$, quizá como en la figura 2. ■

Media y covarianza

Como preparación para el análisis de componentes principales, sea $[\mathbf{X}_1 \cdots \mathbf{X}_N]$ una matriz de observaciones de $p \times N$, tal como la que se describió anteriormente. La **media muestral**, $\mathbf{M}$, de los vectores de observaciones $\mathbf{X}_1, \dots, \mathbf{X}_N$ está dada por

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

Para los datos de la figura 1, la media muestral es el punto en el “centro” de la gráfica de dispersión. Para $k = 1, \dots, N$, sea

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$$

Las columnas de la matriz de $p \times N$

$$B = [\hat{\mathbf{X}}_1 \quad \hat{\mathbf{X}}_2 \quad \cdots \quad \hat{\mathbf{X}}_N]$$

tienen una media muestral cero, y se dice que B está en la **forma de desviación media**. Cuando la media muestral se resta de los datos de la figura 1, la gráfica de dispersión resultante tiene la forma que se observa en la figura 3.

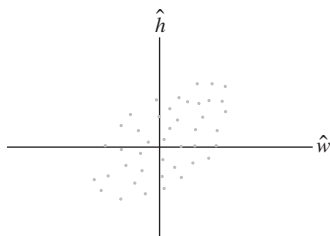


FIGURA 3 Datos de peso-altura en la forma de desviación media.

La **matriz de covarianza (muestral)** es la matriz S de $p \times p$ definida por

$$S = \frac{1}{N-1} BB^T$$

Como cualquier matriz de la forma BB^T es positiva semidefinida, entonces S lo es. (Véase el ejercicio 25 de la sección 8.2 con B y B^T intercambiadas).

EJEMPLO 3 Se realizan tres mediciones en cada uno de los cuatro individuos en una muestra aleatoria de una población. Los vectores de observaciones son

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix}, \quad \mathbf{X}_4 = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix}$$

Calcule la media muestral y la matriz de covarianza.

SOLUCIÓN La media muestral es

$$\mathbf{M} = \frac{1}{4} \left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 20 \\ 16 \\ 20 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix}$$

Restando la media muestral de $\mathbf{X}_1, \dots, \mathbf{X}_4$ se obtiene

$$\hat{\mathbf{X}}_1 = \begin{bmatrix} -4 \\ -2 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_2 = \begin{bmatrix} -1 \\ -2 \\ 8 \end{bmatrix}, \quad \hat{\mathbf{X}}_3 = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_4 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

y

$$B = \begin{bmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{bmatrix}$$

La matriz de covarianza muestral es

$$\begin{aligned} S &= \frac{1}{3} \begin{bmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{bmatrix} \begin{bmatrix} -4 & -2 & -4 \\ -1 & -2 & 8 \\ 2 & 4 & -4 \\ 3 & 0 & 0 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 30 & 18 & 0 \\ 18 & 24 & -24 \\ 0 & -24 & 96 \end{bmatrix} = \begin{bmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{bmatrix} \quad \blacksquare \end{aligned}$$

Para analizar las entradas en $S = [s_{ij}]$, considere que $\mathbf{X}$ representa un vector que varía sobre el conjunto de vectores de observaciones y denote las coordenadas de $\mathbf{X}$ como $x_1, \dots, x_p$. Entonces x_1 , por ejemplo, es un escalar que varía sobre el conjunto de las primeras coordenadas de $\mathbf{X}_1, \dots, \mathbf{X}_N$. Para $j = 1, \dots, p$, la entrada diagonal s_{jj} en S es la **varianza** de x_j .

La varianza de x_j mide la dispersión de los valores de x_j . (Véase el ejercicio 13). En el ejemplo 3, la varianza de x_1 es 10 y la varianza de x_3 es 32. El hecho de que 32 sea mayor que 10 indica que el conjunto de las terceras entradas en los vectores de respuesta tiene una dispersión más amplia que el conjunto de las primeras entradas.

La **varianza total** de los datos es la suma de las varianzas en la diagonal de S . En general, la suma de las entradas diagonales de una matriz cuadrada S se denomina **traza** de la matriz, y se representa como $\text{tr}(S)$. Así,

$$\{\text{varianza total}\} = \text{tr}(S)$$

La entrada s_{ij} de S para $i \neq j$ es la **covarianza** de x_i y x_j . Observe que en el ejemplo 3, la covarianza entre x_1 y x_3 es 0 porque la entrada (1, 3) de S es 0. Los especialistas en estadística dicen que x_1 y x_3 **no están correlacionadas**. El análisis de los datos multivariados en $\mathbf{X}_1, \dots, \mathbf{X}_N$ se simplifica notablemente cuando la mayoría de las variables (o todas) $x_1, \dots, x_p$ no están correlacionadas, es decir, cuando la matriz de covarianza de $\mathbf{X}_1, \dots, \mathbf{X}_N$ es diagonal o casi diagonal.

Análisis de componentes principales

Para simplificar, suponga que la matriz $[\mathbf{X}_1 \cdots \mathbf{X}_N]$ ya está en forma de desviación media. El objetivo del análisis de componentes principales es encontrar una matriz ortogonal de $p \times p$, $P = [\mathbf{u}_1 \cdots \mathbf{u}_p]$ que determine un cambio de variable, $\mathbf{X} = P\mathbf{Y}$, o

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

con la propiedad de que las nuevas variables $y_1, \dots, y_p$ no están correlacionadas y están en orden de varianza decreciente.

El cambio ortogonal de variable $\mathbf{X} = P\mathbf{Y}$ significa que cada vector de observación $\mathbf{X}_k$ recibe un “nuevo nombre”, $\mathbf{Y}_k$, tal que $\mathbf{X}_k = P\mathbf{Y}_k$. Observe que $\mathbf{Y}_k$ es el vector de coordenadas de $\mathbf{X}_k$ con respecto a las columnas de P , y $\mathbf{Y}_k = P^{-1}\mathbf{X}_k = P^T\mathbf{X}_k$ para $k = 1, \dots, N$.

No es difícil comprobar que para cualquier P ortogonal, la matriz de covarianza de $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ es P^TSP (ejercicio 11). Así, la matriz ortogonal P deseada es aquella que hace diagonal a P^TSP . Sea D una matriz diagonal con los valores propios $\lambda_1, \dots, \lambda_p$ de S sobre la diagonal, ordenados de tal manera que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, y sea P una matriz ortogonal cuyas columnas son los vectores propios unitarios correspondientes $\mathbf{u}_1, \dots, \mathbf{u}_p$. De esta forma, $S = PDP^T$ y $P^TSP = D$.

Los vectores propios unitarios $\mathbf{u}_1, \dots, \mathbf{u}_p$ de la matriz de covarianza S se llaman **componentes principales** de los datos (en la matriz de observaciones). La **primera componente principal** es el vector propio correspondiente al mayor valor propio de S , la **segunda componente principal** es el vector propio asociado al segundo mayor valor propio, y así sucesivamente.

La primera componente principal $\mathbf{u}_1$ determina la nueva variable y_1 de la siguiente manera. Sean $c_1, \dots, c_p$ las entradas en $\mathbf{u}_1$. Como $\mathbf{u}_1^T$ es la primera fila de P^T , entonces la ecuación $\mathbf{Y} = P^T\mathbf{X}$ indica que

$$y_1 = \mathbf{u}_1^T \mathbf{X} = c_1x_1 + c_2x_2 + \cdots + c_px_p$$

Así, y_1 es una combinación lineal de las variables originales $x_1, \dots, x_p$, utilizando como pesos las entradas en el vector propio $\mathbf{u}_1$. En forma similar, $\mathbf{u}_2$ determina la variable y_2 , y así sucesivamente.

EJEMPLO 4 Los datos iniciales para la imagen multispectral del Valle Railroad (ejemplo 2) consistían en 4 millones de vectores en $\mathbb{R}^3$. La matriz de covarianza asociada es¹

$$S = \begin{bmatrix} 2382.78 & 2611.84 & 2136.20 \\ 2611.84 & 3106.47 & 2553.90 \\ 2136.20 & 2553.90 & 2650.71 \end{bmatrix}$$

¹ Los datos del ejemplo 4 y de los ejercicios 5 y 6 fueron proporcionados por Earth Satellite Corporation de Rockville, Maryland.

Encuentre las componentes principales de los datos, y liste la nueva variable determinada por la primera componente principal.

SOLUCIÓN Los valores propios de S y las componentes principales asociadas (los vectores propios unitarios) son

$$\begin{aligned}\lambda_1 &= 7614.23 & \lambda_2 &= 427.63 & \lambda_3 &= 98.10 \\ \mathbf{u}_1 &= \begin{bmatrix} .5417 \\ .6295 \\ .5570 \end{bmatrix} & \mathbf{u}_2 &= \begin{bmatrix} -.4894 \\ -.3026 \\ .8179 \end{bmatrix} & \mathbf{u}_3 &= \begin{bmatrix} .6834 \\ -.7157 \\ .1441 \end{bmatrix}\end{aligned}$$

Utilizando dos decimales para simplificar, la variable para la primera componente principal es

$$y_1 = .54x_1 + .63x_2 + .56x_3$$

Esta ecuación se utilizó para crear la fotografía d) en la introducción del capítulo. Las variables x_1, x_2, x_3 son las intensidades de la señal en las tres bandas espectrales. Los valores de x_1 , convertidos a una escala de grises entre negro y blanco, produjeron la fotografía a). De manera similar, los valores de x_2 y x_3 generaron las fotografías b) y c), respectivamente. En cada pixel de la fotografía d), el valor en la escala de grises se calcula a partir de y_1 , una combinación lineal ponderada de x_1, x_2, x_3 . En este sentido, la fotografía d) “muestra” la primera componente principal de los datos. ■

En el ejemplo 4, la matriz de covarianza para los datos transformados, utilizando las variables y_1, y_2, y_3 , es

$$D = \begin{bmatrix} 7614.23 & 0 & 0 \\ 0 & 427.63 & 0 \\ 0 & 0 & 98.10 \end{bmatrix}$$

Aunque D es evidentemente más sencilla que la matriz de covarianza original S , aún no es notorio el mérito de construir las nuevas variables. Sin embargo, las varianzas de las variable y_1, y_2, y_3 aparecen sobre la diagonal de D , y sin duda la primer varianza en D es mucho mayor que las otras dos. Como se verá, este hecho permitirá verificar que, en esencia, los datos son unidimensionales más que tridimensionales.

Reducción de la dimensión de datos multivariados

El análisis de componentes principales es potencialmente valioso en aplicaciones donde la mayor parte de la variación, o el rango dinámico, en los datos se debe a variaciones en *tan solo unas cuantas* de las nuevas variables, $y_1, \dots, y_p$.

Es posible demostrar que un cambio ortogonal de variables, $\mathbf{X} = \mathbf{P}\mathbf{Y}$, no modifica la varianza total de los datos. (A grandes rasgos, esto es cierto porque la multiplicación por la izquierda por P no altera las longitudes de los vectores ni los ángulos entre ellos. Véase el ejercicio 12). Esto significa que si $S = \mathbf{P}\mathbf{D}\mathbf{P}^T$, entonces

$$\left\{ \begin{array}{l} \text{varianza total} \\ \text{de } x_1, \dots, x_p \end{array} \right\} = \left\{ \begin{array}{l} \text{varianza total} \\ \text{de } y_1, \dots, y_p \end{array} \right\} = \text{tr}(D) = \lambda_1 + \dots + \lambda_p$$

La varianza de y_j es λ_j , y el cociente $\lambda_j/\text{tr}(S)$ mide la fracción de la varianza total que se “explica” o se “capta” con y_j .

EJEMPLO 5 Calcule los diversos porcentajes de varianza de los datos multiespectrales de Valle Railroad que se muestran en las fotografías de las componentes principales, d), e), f), presentadas en la introducción del capítulo.

SOLUCIÓN La varianza total de los datos es

$$\text{tr}(D) = 7614.23 + 427.63 + 98.10 = 8139.96$$

[Verifique que este número también sea igual a $\text{tr}(S)$]. Los porcentajes de la varianza total explicados por las componentes principales son

Primera componente	Segunda componente	Tercera componente
$\frac{7614.23}{8139.96} = 93.5\%$	$\frac{427.63}{8139.96} = 5.3\%$	$\frac{98.10}{8139.96} = 1.2\%$

En un sentido, el 93.5% de la información recopilada por Landsat para la región de Valle Railroad se presenta en la fotografía *d*), el 5.3% en *e*) y solamente el 1.2% restante en *f*). ■

Los cálculos del ejemplo 5 revelan que los datos prácticamente no tienen varianza en la tercera (nueva) coordenada. Todos los valores de y_3 son cercanos a cero. Geométricamente, los puntos de datos están cercanos al plano $y_3 = 0$, y sus ubicaciones se pueden determinar más o menos exactamente conociendo solo los valores de y_1 y y_2 . De hecho, y_2 también tiene una varianza relativamente pequeña, lo que significa que los puntos están aproximadamente sobre una recta, y los datos, en esencia, son unidimensionales. Véase la figura 2, en la que los datos semejan un palo de paleta.

Caracterizaciones de las variables de componentes principales

Si $y_1, \dots, y_p$ se originan de un análisis de componentes principales de una matriz de observaciones de $p \times N$, entonces la varianza de y_1 es tan grande como sea posible en el siguiente sentido: si $\mathbf{u}$ es cualquier vector unitario y si $y = \mathbf{u}^T \mathbf{X}$, entonces la varianza de los valores de y cuando $\mathbf{X}$ varía sobre los datos originales $\mathbf{X}_1, \dots, \mathbf{X}_N$ es $\mathbf{u}^T \mathbf{S} \mathbf{u}$. De acuerdo con el teorema 8 de la sección 8.3, el valor máximo de $\mathbf{u}^T \mathbf{S} \mathbf{u}$, sobre todos los vectores unitarios $\mathbf{u}$, es el mayor valor propio λ_1 de S , y esta varianza se alcanza cuando $\mathbf{u}$ es el vector propio correspondiente $\mathbf{u}_1$. De manera similar, el teorema 8 indica que y_2 tiene la máxima varianza posible entre todas las variables $y = \mathbf{u}^T \mathbf{X}$ que *no están correlacionadas* con y_1 . De igual forma, y_3 tiene la máxima varianza posible entre todas las variables no correlacionadas con y_1 y y_2 , y así sucesivamente.

NOTA NUMÉRICA

La descomposición en valores singulares es la herramienta más importante para efectuar el análisis de componentes principales en aplicaciones prácticas. Si B es una matriz de observaciones de $p \times N$ en la forma de desviación media, y si $A = (1/\sqrt{N-1})B^T$, entonces $A^T A$ es la matriz de covarianza, S . Los cuadrados de los valores singulares de A son los p valores propios de S , y los vectores singulares derechos de A son las componentes principales de los datos.

Como se mencionó en la sección 8.4, el cálculo iterativo de la DVS de A es más rápido y preciso que una descomposición de valores propios de S . Esto es particularmente cierto, por ejemplo, en el procesamiento de imágenes hiperespectrales (con $p = 224$) que se mencionó en la introducción del capítulo. El análisis de componentes principales se completa en segundos en estaciones de trabajo especializadas.

Lectura adicional

Lillesand, Thomas M. y Ralph W. Kiefer, *Remote Sensing and Image Interpretation*, 4a. ed. (Nueva York: John Wiley, 2000).

PROBLEMAS DE PRÁCTICA[‡]

La siguiente tabla lista los pesos y las estaturas de cinco jóvenes:

Joven	#1	#2	#3	#4	#5
Peso (lb)	120	125	125	135	145
Altura (in.)	61	60	64	68	72

- 1. Encuentre la matriz de covarianza para los datos.
- 2. Realice un análisis de componentes principales de los datos para encontrar un solo *índice de tamaño* que explique la mayor parte de la variación en los datos.

8.5 EJERCICIOS[†]

En los ejercicios 1 y 2, convierta la matriz de observaciones a la forma de desviación media y construya la matriz de covarianza de la muestra.

1.
$$\begin{bmatrix} 19 & 22 & 6 & 3 & 2 & 20 \\ 12 & 6 & 9 & 15 & 13 & 5 \end{bmatrix}$$

2.
$$\begin{bmatrix} 1 & 5 & 2 & 6 & 7 & 3 \\ 3 & 11 & 6 & 8 & 15 & 11 \end{bmatrix}$$

3. Encuentre las componentes principales de los datos del ejercicio 1.
4. Encuentre las componentes principales de los datos del ejercicio 2.
5. [M] Se tomó una imagen Landsat con tres componentes espectrales de la base Homestead de la Fuerza Aérea, en Florida (después de que la base sufrió los estragos provocados por el huracán Andrew en 1992). La matriz de covarianza de los datos se muestra a continuación. Determine la primera componente principal de los datos, y calcule el porcentaje de la varianza total que está contenida en esa componente.

$$S = \begin{bmatrix} 164.12 & 32.73 & 81.04 \\ 32.73 & 539.44 & 249.13 \\ 81.04 & 249.13 & 189.11 \end{bmatrix}$$

6. [M] La siguiente matriz de covarianza se obtuvo de una imagen Landsat del río Columbia, Washington, utilizando datos de tres bandas espectrales. Sean x_1, x_2, x_3 las componentes espectrales de cada pixel en la imagen. Encuentre una nueva variable de la forma $y_1 = c_1x_1 + c_2x_2 + c_3x_3$ que tenga la máxima varianza posible, sujeta a la restricción $c_1^2 + c_2^2 + c_3^2 = 1$. ¿Qué porcentaje de la varianza total en los datos se explica mediante y_1 ?

$$S = \begin{bmatrix} 29.64 & 18.38 & 5.00 \\ 18.38 & 20.82 & 14.06 \\ 5.00 & 14.06 & 29.21 \end{bmatrix}$$

- * 7. Sean x_1, x_2 las variables para los datos bidimensionales del ejercicio 1. Determine una nueva variable y_1 de la forma $y_1 = c_1x_1 + c_2x_2$, con $c_1^2 + c_2^2 = 1$, tal que y_1 tenga la máxima varianza posible sobre los datos proporcionados. ¿Qué parte de la varianza en los datos se explica por y_1 ?
8. Repita el ejercicio 7 para los datos del ejercicio 2.

- * 9. Suponga que se aplican tres pruebas a una muestra aleatoria de estudiantes de licenciatura. Sean $\mathbf{X}_1, \dots, \mathbf{X}_N$ los vectores de observaciones en $\mathbb{R}^3$ que listan las tres calificaciones de cada estudiante, y para $j = 1, 2, 3$, denote con x_j la calificación de un estudiante en el j -ésimo examen. Suponga que la matriz de covarianza de los datos es

$$S = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{bmatrix}$$

Sea y un “índice” de desempeño de estudiantes, con $y = c_1x_1 + c_2x_2 + c_3x_3$ y $c_1^2 + c_2^2 + c_3^2 = 1$. Seleccione c_1, c_2, c_3 de manera que la varianza de y sobre el conjunto de datos sea lo más grande posible. [Sugerencia: Los valores propios de la matriz de covarianza de la muestra son $\lambda = 3, 6$ y 9].

10. [M] Repita el ejercicio 9 con $S = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 11 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

11. A partir de los datos multivariados $\mathbf{X}_1, \dots, \mathbf{X}_N$ (en $\mathbb{R}^p$) en la forma de desviación media, sea P una matriz de $p \times p$, y defina $\mathbf{Y}_k = P^T \mathbf{X}_k$ para $k = 1, \dots, N$.

- a) Demuestre que $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ están en la forma de desviación media. [Sugerencia: Deje que $\mathbf{w}$ sea el vector en $\mathbb{R}^N$ con un 1 en cada entrada. Entonces $[\mathbf{X}_1 \cdots \mathbf{X}_N]\mathbf{w} = \mathbf{0}$ (el vector cero en $\mathbb{R}^p$).
- b) Demuestre que si la matriz de covarianza de $\mathbf{X}_1, \dots, \mathbf{X}_N$ es S , entonces la matriz de covarianza de $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ es $P^T S P$.

12. Sea $\mathbf{X}$ un vector que varía sobre las columnas de una matriz de observaciones de $p \times N$, en tanto que P es una matriz ortogonal de $p \times p$. Demuestre que el cambio de variable $\mathbf{X} = P\mathbf{Y}$ no cambia la varianza total de los datos. [Sugerencia: De acuerdo con el ejercicio 11, es suficiente demostrar que $\text{tr}(P^T S P) = \text{tr}(S)$. Utilice una propiedad de la traza mencionada en el ejercicio 25 de la sección 7.4].

13. La matriz de covarianza de la muestra es una generalización de una fórmula para la varianza de una muestra de N mediciones escalares, por ejemplo, $t_1, \dots, t_N$. Si m es el promedio de $t_1, \dots, t_N$, entonces la *varianza de la muestra* está dada por

$$\frac{1}{N-1} \sum_{k=1}^n (t_k - m)^2 \quad (1)$$

Muestre cómo la matriz de covarianza de la muestra, S , definida antes del ejemplo 3, se puede escribir en una forma similar a (1). [Sugerencia: Utilice multiplicación de matrices particio-

nadas para escribir a S como $1/(N-1)$ por la suma de N matrices de tamaño $p \times p$. Para $1 \leq k \leq N$, escriba $\mathbf{X}_k - \mathbf{M}$ en vez de $\hat{\mathbf{X}}_k$.

SOLUCIONES A LOS PROBLEMAS DE PRÁCTICA

1. Primero arregle los datos en la forma de desviación media. Es fácil ver que el vector de la media muestral es $\mathbf{M} = \begin{bmatrix} 130 \\ 65 \end{bmatrix}$. Reste $\mathbf{M}$ de los vectores de observaciones (las columnas de la tabla) y obtenga

$$B = \begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}$$

Entonces la matriz de covarianza de la muestra es

$$\begin{aligned} S &= \frac{1}{5-1} \begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix} \begin{bmatrix} -10 & -4 \\ -5 & -5 \\ -5 & -1 \\ 5 & 3 \\ 15 & 7 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} 400 & 190 \\ 190 & 100 \end{bmatrix} = \begin{bmatrix} 100.0 & 47.5 \\ 47.5 & 25.0 \end{bmatrix} \end{aligned}$$

2. Los valores propios de S son (con dos decimales)

$$\lambda_1 = 123.02 \quad \text{y} \quad \lambda_2 = 1.98$$

El vector propio unitario a λ_1 es $\mathbf{u} = \begin{bmatrix} .900 \\ .436 \end{bmatrix}$. (Como S es de 2×2 , los cálculos se pueden hacer a mano si no se cuenta con un programa de matrices). Para el *índice de tamaño*, sea

$$y = .900\hat{w} + .436\hat{h}$$

donde $\hat{w}$ y $\hat{h}$ son el peso y la altura, respectivamente, en la forma de desviación media. La varianza de este índice sobre el conjunto de datos es 123.02. Puesto que la varianza total es $\text{tr}(S) = 100 + 25 = 125$, el índice de tamaño explica prácticamente toda la varianza de los datos (el 98.4% de esta).

La figura 4 muestra los datos originales para el problema de práctica 1 y la recta determinada por la primera componente principal $\mathbf{u}$. (En forma vectorial paramétrica, la recta es $\mathbf{x} = \mathbf{M} + t\mathbf{u}$). Es posible demostrar que la recta es la mejor aproximación a los

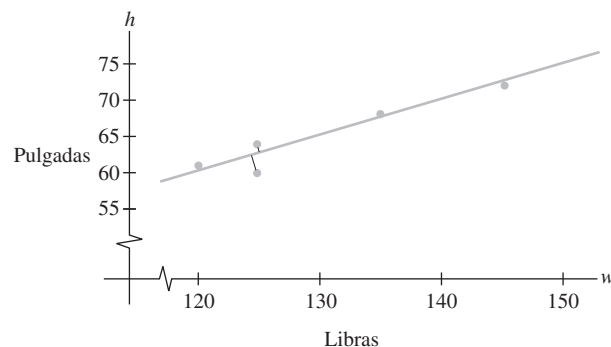


FIGURA 4 Una recta de regresión ortogonal determinada por la primera componente principal de los datos.

datos, en el sentido de que la suma de los cuadrados de las distancias *ortogonales* a la recta se minimiza. En efecto, el análisis de componentes principales es equivalente a la *regresión ortogonal*, pero eso es otra historia. Quizá se presente otra vez.

CAPÍTULO 8 EJERCICIOS COMPLEMENTARIOS

1. Marque cada enunciado como verdadero o falso. Justifique sus respuestas. En cada inciso, A representa una matriz de $n \times n$.
 - a) Si A es diagonalizable ortogonalmente, entonces A es simétrica.
 - b) Si A es una matriz ortogonal, entonces A es simétrica.
 - c) Si A es una matriz ortogonal, entonces $\|Ax\| = \|x\|$ para toda x en $\mathbb{R}^n$.
 - d) Los ejes principales de una forma cuadrática $x^T Ax$ pueden ser columnas de cualquier matriz P que diagonaliza a A .
 - e) Si P es una matriz de $n \times n$ con columnas ortogonales, entonces $P^T = P^{-1}$.
 - f) Si todo coeficiente en una forma cuadrática es positivo, entonces la forma cuadrática es definida positiva.
 - g) Si $x^T Ax > 0$ para alguna x , entonces la forma cuadrática $x^T Ax$ es positiva definida.
 - h) Con un cambio adecuado de variable, cualquier forma cuadrática se puede convertir en una forma cuadrática sin productos cruzados.
 - i) El valor más grande de una forma cuadrática $x^T Ax$, para $\|x\| = 1$, es la entrada más grande en la diagonal de A .
 - j) El valor máximo de una forma cuadrática positiva definida $x^T Ax$ es el mayor valor propio de A .
 - k) Una forma cuadrática positiva definida se puede convertir en una forma negativa definida mediante un adecuado cambio de variable $x = Pu$, para alguna matriz ortogonal P .
 - l) Una forma cuadrática indefinida es aquella cuyos valores propios no están definidos.
 - m) Si P es una matriz ortogonal de $n \times n$, entonces el cambio de variable $x = Pu$ convierte $x^T Ax$ en una forma cuadrática cuya matriz es $P^{-1}AP$.
 - n) Si U es de $m \times n$ con columnas ortogonales, entonces $UU^T x$ es la proyección ortogonal de x sobre $\text{Col } U$.
 - o) Si B es de $m \times n$ y x es un vector unitario en $\mathbb{R}^n$, entonces $\|Bx\| \leq \sigma_1$, donde σ_1 es el primer valor singular de B .
 - p) Una descomposición en valores singulares de una matriz B de $m \times n$ se puede escribir como $B = P\Sigma Q$, donde P es una matriz ortogonal de $m \times m$, Q es una matriz ortogonal de $n \times n$, y Σ es una matriz “diagonal” de $m \times n$.
 - q) Si A es de $n \times n$, entonces A y $A^T A$ tienen los mismos valores singulares.
 2. Sean $\{u_1, \dots, u_n\}$ una base ortonormal para $\mathbb{R}^n$, y $\lambda_1, \dots, \lambda_n$ cualesquiera escalares reales. Defina

$$A = \lambda_1 u_1 u_1^T + \dots + \lambda_n u_n u_n^T$$
 - a) Demuestre que A es simétrica.
 - b) Pruebe que $\lambda_1, \dots, \lambda_n$ son los valores propios de A .
 3. Sea A una matriz simétrica de $n \times n$ y rango r . Explique por qué la descomposición espectral de A representa a esta como la suma de r matrices de rango 1.
 4. Sea A una matriz simétrica de $n \times n$.
 - a) Demuestre que $(\text{Col } A)^\perp = \text{Nul } A$. [Sugerencia: Véase la sección 5.1].
 - b) Demuestre que cada y en $\mathbb{R}^n$ se puede escribir en la forma $y = \hat{y} + z$, con $\hat{y}$ en $\text{Col } A$ y z en $\text{Nul } A$.
 5. Demuestre que si v es un vector propio de una matriz A de $n \times n$, y v corresponde a un valor propio de A distinto de cero, entonces v está en $\text{Col } A$. [Sugerencia: Utilice la definición de un vector propio].
 6. Sea A una matriz simétrica de $n \times n$. Aplique el ejercicio 5 y una base de vectores propios en $\mathbb{R}^n$ con la finalidad de dar una segunda demostración de la descomposición en el ejercicio 4b).
 7. Demuestre que una matriz A de $n \times n$ es positiva definida si y solo si A admite una *factorización de Cholesky*, a saber, $A = R^T R$ para alguna matriz triangular superior invertible R cuyas entradas diagonales son todas positivas. [Sugerencia: Utilice una factorización QR y el ejercicio 26 de la sección 8.2].
 8. Con base en el ejercicio 7, demuestre que si A es positiva definida, entonces A tiene una factorización LU, $A = LU$, donde U tiene pivotes positivos en su diagonal. (Lo contrario también es verdadero).
- Si A es de $m \times n$, entonces la matriz $G = A^T A$ es la *matriz de Gram* de A . En este caso, las entradas de G son los productos interiores de las columnas de A . (Véase los ejercicios 9 y 10).
9. Demuestre que la matriz de Gram de cualquier matriz A es positiva semidefinida, con el mismo rango que A . (Véase los ejercicios de la sección 5.5).
 10. Demuestre que si una matriz G de $n \times n$ es positiva semidefinida y tiene rango r , entonces G es la matriz de Gram de alguna matriz A de $r \times n$. Esto se llama *factorización reveladora del rango* de G . [Sugerencia: Considere la descomposición espectral de G , y primero escriba G como BB^T para una matriz B de $n \times r$].
 11. Pruebe que cualquier matriz A de $n \times n$, admite una *descomposición polar* de la forma $A = PQ$, donde P es una matriz positiva semidefinida con el mismo rango que A , y donde Q es una matriz ortogonal de $n \times n$. [Sugerencia: Utilice la descomposición en valores singulares, $A = U\Sigma V^T$, y observe que $A = (U\Sigma U^T)(UV^T)$]. Esta descomposición se utiliza, por ejemplo, en ingeniería mecánica para modelar la deformación de un material. La matriz P describe la elongación o compresión del material (en las direcciones de los vectores propios de P), y Q describe la rotación del material en el espacio.

Los ejercicios 12 a 14 se refieren a una matriz A de $m \times n$ con una descomposición en valores singulares reducida, $A = U_r D V_r^T$, y la pseudoinversa $A^+ = V_r D^{-1} U_r^T$.

12. Compruebe las propiedades de A^+ :

- Para toda $\mathbf{y}$ en $\mathbb{R}^m$, $AA^+\mathbf{y}$ es la proyección ortogonal de $\mathbf{y}$ sobre $\text{Col } A$.
- Para toda $\mathbf{x}$ en $\mathbb{R}^n$, $A^+A\mathbf{x}$ es la proyección ortogonal de $\mathbf{x}$ sobre $\text{Fil } A$.
- $AA^+A = A$ y $A^+AA^+ = A^+$.

13. Suponga que la ecuación $A\mathbf{x} = \mathbf{b}$ es consistente, y sea $\mathbf{x}^+ = A^+\mathbf{b}$. De acuerdo con el ejercicio 23 de la sección 5.3, existe exactamente un vector $\mathbf{p}$ en $\text{Fil } A$ tal que $A\mathbf{p} = \mathbf{b}$. Los siguientes pasos prueban que $\mathbf{x}^+ = \mathbf{p}$ y $\mathbf{x}^+$ es la *solución de longitud mínima* de $A\mathbf{x} = \mathbf{b}$.

- Demuestre que $\mathbf{x}^+$ está en $\text{Fil } A$. [Sugerencia: Escriba $\mathbf{b}$ como $A\mathbf{x}$ para alguna $\mathbf{x}$, y utilice el ejercicio 12].
- Demuestre que $\mathbf{x}^+$ es una solución de $A\mathbf{x} = \mathbf{b}$.
- Demuestre que si $\mathbf{u}$ es cualquier solución de $A\mathbf{x} = \mathbf{b}$, entonces $\|\mathbf{x}^+\| \leq \|\mathbf{u}\|$, con la igualdad solo si $\mathbf{u} = \mathbf{x}^+$.

14. Dada cualquier $\mathbf{b}$ en $\mathbb{R}^m$, adapte el ejercicio 13 para demostrar que $A^+\mathbf{b}$ es la *solución por mínimos cuadrados de longitud mínima*. [Sugerencia: Considere la ecuación $A\mathbf{x} = \mathbf{b}$, donde $\hat{\mathbf{b}}$ es la proyección ortogonal de $\mathbf{b}$ sobre $\text{Col } A$].

[M] En los ejercicios 15 y 16, construya la pseudoinversa de A . Comience aplicando un programa de matrices para obtener la DVS de A , o, si no tiene un programa de ese tipo, comience con una diagonalización ortogonal de A^TA . Utilice la pseudoinversa para resolver $A\mathbf{x} = \mathbf{b}$, para $\mathbf{b} = (6, -1, -4, 6)$, y sea $\hat{\mathbf{x}}$ la solución. Haga un cálculo para comprobar que $\hat{\mathbf{x}}$ está en $\text{Fil } A$. Encuentre un vector $\mathbf{u}$ distinto de cero en $\text{Nul } A$, y compruebe que $\|\hat{\mathbf{x}}\| < \|\hat{\mathbf{x}} + \mathbf{u}\|$, lo que debe ser cierto de acuerdo con el ejercicio 13c).

$$\mathbf{15.} \quad A = \begin{bmatrix} -3 & -3 & -6 & 6 & 1 \\ -1 & -1 & -1 & 1 & -2 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & -1 & 1 & -1 \end{bmatrix}$$

$$\mathbf{16.} \quad A = \begin{bmatrix} 4 & 0 & -1 & -2 & 0 \\ -5 & 0 & 3 & 5 & 0 \\ 2 & 0 & -1 & -2 & 0 \\ 6 & 0 & -3 & -6 & 0 \end{bmatrix}$$

AUTOEVALUACIÓN

1. Sea $A = \begin{bmatrix} 13 & 10 & 4 \\ 10 & 12 & 6 \\ 4 & 6 & 5 \end{bmatrix}$. Diagonalice ortogonalmente la matriz A dando la matriz ortogonal P y la matriz diagonal D .

2. Determine la forma cuadrática $\mathbf{x}^\top A \mathbf{x}$ de la matriz $A = \begin{bmatrix} 9 & 3 & 0 \\ 3 & 1 & 5 \\ 0 & 5 & -2 \end{bmatrix}$ y el valor $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

3. Encuentre la matriz de la forma cuadrática $3x_1^2 + 5x_2^2 + 2x_3^2 - 12x_1x_2 + x_2x_3$.

4. Dada la forma cuadrática $Q(x) = 13x_1^2 + 12x_2^2 + 5x_3^2 + 20x_1x_2 + 8x_1x_3 + 12x_2x_3$, encuentre una matriz P de cambio de coordenadas, que al cambiar la variable $\mathbf{x} = P\mathbf{y}$ transforme a la forma cuadrática en otra que no tenga términos con productos cruzados.

5. Encuentre el valor máximo de la forma cuadrática $Q(x) = 14x_1^2 + 14x_2^2 + 18x_3^2 + 26x_1x_2 + 18x_1x_3 + 18x_2x_3$ sujeta a la restricción $\mathbf{x}^\top \mathbf{x} = 1$.

6. Encuentre un vector unitario para el cual la forma cuadrática $\mathbf{x}^\top A \mathbf{x}$ alcance su máximo, sujeta a la restricción $\mathbf{x}^\top \mathbf{x} = 1$.

7. Encuentre el valor máximo de la forma cuadrática $Q(x) = 14x_1^2 + 14x_2^2 + 18x_3^2 + 26x_1x_2 + 18x_1x_3 + 18x_2x_3$ sujeta a la restricción $\mathbf{x}^\top \mathbf{x} = 1$ y $\mathbf{x}^\top \mathbf{u} = 0$, donde $\mathbf{u}$ es un vector propio unitario correspondiente al valor propio más grande de la matriz de la forma cuadrática.

8. Encuentre los valores singulares de la matriz $A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & -3 \end{bmatrix}$.

9. Encuentre una descomposición en valores singulares de la matriz $A = \begin{bmatrix} 4 & 0 & -4 \\ 4 & 4 & 4 \end{bmatrix}$.

10. Convierta la matriz de observaciones $M = \begin{bmatrix} 9 & 12 & 0 & 3 \\ 7 & 4 & 10 & 7 \\ 13 & 1 & 1 & 1 \end{bmatrix}$ en su forma de desviación media y construya la matriz de covarianza de muestras.

11. Dada la matriz de covarianza $S = \begin{bmatrix} 30 & -12 & 12 \\ -12 & 6 & 0 \\ 12 & 0 & 36 \end{bmatrix}$, determine el porcentaje de la varianza total que está contenido en la primera componente principal.