



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

COMPONENTES PRINCIPALES  
PARA DATOS DIRECCIONALES

T E S I S

QUE PARA OBTENER EL TÍTULO DE :

A C T U A R I A

P R E S E N T A :

CLAUDIA IRENE LARRACILLA CAMACHO



FACULTAD DE CIENCIAS  
UNAM

DIRECTORA DE TESIS:  
MAT. MARGARITA ELVIRA CHÁVEZ CANO

2006



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Esta tesis va dedicada especialmente a mi papá, Ricardo Larracilla†, y a mi papá Alfonso†: donde quiera que estén siempre los llevo en mis pensamientos con cariño y admiración.*

# AGRADECIMIENTOS

A la UNAM, de la que me siento muy orgullosa, por permitirme pertenecer a una gran comunidad y por brindarme una excelente formación tanto en el aspecto académico como en el personal.

A la vida por haberme dado una familia que me quiere y alienta a ser mejor cada día. Por haberme dado unos amigos que me apoyan cada día y por haber puesto en mi camino las experiencias que he vivido, ya sean buenas o malas que han hecho de mí lo que hoy soy.

A mi mamá, Irene Teresa, porque además de darme la vida y su cariño, ha tenido la fuerza, inteligencia y dedicación para hacerme una persona de bien cada día.

A mi hermana Jacqueline, por su apoyo y confianza en mí.

A mis tíos Jesús, Paz, Concepción, Manuel y Martín, así como a mis abuelitos Carmen y Alfonso<sup>†</sup> por todo su tiempo, cariño y ejemplo.

A mis primos Lupita, Luis, Charo y Daniel por todas esas tareas en las que me ayudaron y por su ejemplo; pero sobre todo, por ser como mis hermanos.

A mis tíos Agustín y Alfonso por su apoyo y ayuda en los momentos difíciles.

A mi amiga Mayra, pues a pesar de la distancia, siempre ha estado conmigo. Mayra: tú sabes que eres más que mi mejor amiga, ¡eres mi hermana!.

A mis tíos Beto y Lupe por todo su apoyo.

A mis amigos Paola, Ivonne, Rafael, Mireya, Alma, José Luis, Adriana, Ángeles, Yéssica, Felipe, Martha, Marcos, Daniel, David, Karina, Itzia y Cynthia por compartir conmigo los buenos y los malos momentos.

A mi directora de tesis Mat. Margarita Elvira Chávez Cano por su tiempo, paciencia, ayuda y apoyo así como por sus consejos para la realización de este trabajo, pero sobre todo, por su cariño y confianza. ¡Mil gracias!

A Olga por ser mi amiga y darme su apoyo y confianza.

A Gerardo y Carlos por su amistad.

A mis sinodales Dr. Carlos Díaz Ávalos, Act. Francisco Sánchez Villarreal, M. en C. Leticia Eugenia Gracia Medrano Valdelamar y Act. Jaime Vázquez Alamilla por su tiempo y cuidado para revisar mi trabajo, así como por sus consejos para mejorarlo.

Finalmente, quiero darle gracias a Manuel por su ayuda bibliográfica en temas de física.

# Contenido

<b>INTRODUCCIÓN</b>	<b>1</b>
<b>1 ANÁLISIS DE COMPONENTES PRINCIPALES</b>	<b>3</b>
1.1 Introducción . . . . .	3
1.2 Definición . . . . .	4
1.3 Obtención de las componentes principales a partir de los parámetros poblacionales . . . . .	5
1.3.1 Componentes principales a partir de la matriz de varianzas y covarianzas $\Sigma$ . . . . .	5
1.3.2 Componentes principales a partir de la matriz de correlaciones $\mathbf{P}$ . . . . .	10
1.4 Obtención de las componentes principales a partir de la matriz de datos . . . . .	12
1.4.1 Componentes principales a partir de la matriz de varianzas y covarianzas muestrales $\mathbf{S}$ . . . . .	12
1.4.2 Componentes principales a partir de la matriz de correlaciones muestrales $\mathbf{R}$ . . . . .	14
1.5 Cuántas componentes principales retener . . . . .	15
1.6 Información en las últimas componentes principales . . . . .	17
1.6.1 Valores propios iguales a cero . . . . .	17
1.6.2 Valores propios cercanos a cero . . . . .	17
1.6.3 Valores propios repetidos . . . . .	18
1.7 Algunos aspectos importantes de las componentes principales . . . . .	19
1.7.1 Scores o puntajes de las componentes principales . . . . .	19
1.7.2 Ortogonalidad . . . . .	20
1.7.3 Vectores de carga de componentes . . . . .	20
1.7.4 Correlación entre las variables originales y las componentes principales . . . . .	22
1.8 Interpretación de las componentes principales . . . . .	24
1.8.1 Patrones especiales en la matriz de varianzas y covarianzas muestrales ( $\mathbf{S}$ ) o en la matriz de correlaciones muestrales ( $\mathbf{R}$ ) . . . . .	24
1.8.2 Interpretación geométrica de las componentes principales . . . . .	25
1.9 Ejemplo . . . . .	31

<b>2</b>	<b>CONCEPTOS EN OTRAS ÁREAS RELACIONADOS CON ESTADÍSTICA ESFÉRICA</b>	<b>44</b>
2.1	Introducción . . . . .	44
2.2	Nociones básicas de Geometría . . . . .	45
2.2.1	Ángulos y algunos conceptos básicos de trigonometría . . . . .	45
2.2.2	Vectores . . . . .	50
2.2.3	Coordenadas rectangulares . . . . .	61
2.3	Conceptos importantes de física . . . . .	66
2.3.1	Centro de masa . . . . .	66
2.3.2	Momento de inercia . . . . .	67
<b>3</b>	<b>CONCEPTOS BÁSICOS DE ESTADÍSTICA ESFÉRICA</b>	<b>68</b>
3.1	Introducción . . . . .	68
3.2	Datos direccionales . . . . .	69
3.3	Sistemas coordenados en estadística circular . . . . .	70
3.3.1	Coordenadas polares . . . . .	70
3.4	Sistemas coordenados en estadística esférica . . . . .	71
3.4.1	Coordenadas esféricas . . . . .	71
3.4.2	Coordenadas geográficas . . . . .	72
3.5	Proyecciones más utilizadas en estadística esférica . . . . .	74
3.6	Rotación de vectores unitarios y de ejes . . . . .	78
3.7	Medidas descriptivas . . . . .	79
3.7.1	Dirección media . . . . .	79
3.7.2	Varianza . . . . .	85
<b>4</b>	<b>COMPONENTES PRINCIPALES PARA DATOS DIRECCIONALES</b>	<b>88</b>
4.1	Introducción . . . . .	88
4.2	Formas de distribuciones esféricas . . . . .	89
4.3	Momento de inercia y ejes principales . . . . .	91
4.4	Obtención de las direcciones principales a partir de la matriz <b>B</b> . . . . .	95
4.5	Obtención de las direcciones principales a partir de la matriz <b>T</b> . . . . .	101
4.6	Interpretación de las direcciones principales obtenidas a partir de la matriz <b>B</b> . . . . .	103
4.7	Interpretación de las direcciones principales obtenidas a partir de la matriz <b>T</b> . . . . .	104
4.8	Ejemplo . . . . .	105
	<b>CONCLUSIONES</b>	<b>112</b>
	<b>Bibliografía</b>	<b>114</b>

# INTRODUCCIÓN

Los datos direccionales en tres dimensiones surgen en muchas áreas de la experimentación científica y observacional. Como ejemplos de datos direccionales de varios campos, podemos enumerar muchos, entre ellos tenemos: en Astrofísica, las direcciones de llegada de los baños de rayos cósmicos; en Cristalografía, las direcciones de los ejes ópticos en cristales de cuarzo; en Meteorología podemos hablar de las direcciones del viento observadas en un lugar dado; en Oceanografía podemos mencionar las mediciones de las direcciones de las corrientes oceánicas; etc. Además, observaciones que no son en algún sentido orientaciones, algunas veces pueden ser reexpresadas en la forma de orientaciones y analizadas como datos direccionales.

Así, el presente trabajo tiene como objetivo mostrar cómo se emplea, en este tipo de observaciones, una de las técnicas de análisis multivariado más utilizadas: el análisis de componentes principales. Para ello consta de cuatro capítulos, los cuales resumiremos a continuación.

El capítulo 1 tiene como propósito presentar, a grandes rasgos, el análisis de componentes principales cuando se trabaja con datos que no tengan una dirección. Esta técnica de análisis multivariado tiene como propósitos principales, dado un conjunto de datos medidos en  $p$  variables, obtener una reducción en la dimensión de los mismos y su interpretación.

La estadística al igual que todas las ciencias, necesita apoyarse en conceptos de otras áreas, por ello, el capítulo 2 tiene como finalidad dar un breve resumen de conceptos muy básicos de geometría analítica y física, los cuales resultan de interés para poder comprender los conceptos más elementales de estadística esférica y para introducir el momento de inercia, concepto en el que se basa el desarrollo de los ejes principales.

El capítulo 3 se centra, principalmente en presentar conceptos de estadística esférica necesarios para poder trabajar con datos direccionales y para poder interpretar los resultados que arrojan las direcciones principales.

Finalmente, el capítulo 4 se centra en el desarrollo matemático de las direcciones principales, que, como veremos más adelante, a diferencia del análisis de componentes principales, el encontrar el eje principal de un conjunto de observaciones direccionales nos da información sobre la dispersión de las observaciones y de la forma de la distribución que sigue la población de la que fueron obtenidas.



# Capítulo 1

## ANÁLISIS DE COMPONENTES PRINCIPALES

### 1.1 Introducción

Cuando se toman medidas de personas o de objetos, el investigador frecuentemente incluye tantas variables como sea posible para evitar pasar por alto alguna que pudiera tener futura relevancia. Desafortunadamente, cuando la dimensión de  $p$  (número de variables) es grande, un conjunto de datos no sólo podría ser muy costoso de obtener, sino también podría ser difícil de manejar y de estudiar si no se comprime de alguna manera. En este caso existen muchas opciones para el investigador, una de ellas es el *análisis de componentes principales lineales*.

El *análisis de componentes principales* es una técnica multivariada muy utilizada en la práctica que data de principios del siglo XX. Originalmente fue introducida por Karl Pearson en 1901 y posteriormente en 1933 fue ampliamente desarrollada matemáticamente por Harold Hotelling.

Pearson se enfocó en el aspecto de una aproximación: dado un vector aleatorio  $p$ -variado (o un "sistema de puntos en el espacio", en la terminología de Pearson), encontrar una aproximación óptima en un subespacio lineal de menor dimensión.

Por otra parte, la reinención de esta técnica por Hotelling se enfocó en un aspecto diferente: suponiendo que  $\mathbf{X}$  es un vector aleatorio, entonces cómo podemos transformarlo en un nuevo vector aleatorio  $\mathbf{Y}$  cuyas componentes sean no correlacionadas e incluso independientes.

La idea básica de este método es describir la variación de un conjunto de variables

correlacionadas en términos de un conjunto de variables no correlacionadas, cada una de las cuales sea una combinación lineal particular de las variables originales. Las nuevas variables se obtienen en orden decreciente de acuerdo a la proporción de varianza que las variables aportan. El objetivo de este análisis es ver si las primeras componentes aportan la mayor parte de la variación en los datos originales. Si lo hacen, entonces proveen una reducción de la dimensión de las variables, que puede ser útil en la simplificación de análisis posteriores.

Una característica importante del análisis de componentes principales es que esta técnica está enfocada en la estructura central de una muestra de observaciones de  $p$  variables, esto es, ninguna de estas variables es indicada como dependiente, y no se supone ninguna agrupación de observaciones. Así, los propósitos más importantes de este análisis son: *reducción en la dimensión de los datos e interpretación de los mismos*, esto último se hará graficando los scores o puntajes de las primeras dos componentes para cada individuo, en los datos donde se reduce la dimensión a dos.

## 1.2 Definición

Como se indicó anteriormente, el análisis de componentes principales transforma el conjunto original de variables correlacionadas en otro conjunto de variables no correlacionadas que son combinaciones lineales de las variables originales que explican la mayor parte de la varianza del conjunto original.

Las componentes principales son obtenidas de tal manera que la primera componente principal, denotada por  $Y_1$ , explique la cantidad más grande de la variación total. Esto es,  $Y_1$  es la combinación lineal de las variables  $X_j$ ,  $j = 1, 2, \dots, p$

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$$

donde los valores  $a_{11}, a_{21}, \dots, a_{p1}$  han sido escogidos para maximizar la varianza de  $Y_1$ .

La segunda componente principal,  $Y_2$ , es aquella combinación lineal de las variables que no está correlacionada con  $Y_1$  y la cual acumula la máxima cantidad de la variación total restante que no ha sido considerada por  $Y_1$ . Por lo que, en general, la  $m$ -ésima componente principal es aquella combinación lineal de  $X$ 's,

$$Y_m = a_{1m}X_1 + a_{2m}X_2 + \dots + a_{pm}X_p$$

la cual tiene la varianza más grande de todas las combinaciones lineales restantes

que están no correlacionadas con todas las componentes principales obtenidas previamente.

## 1.3 Obtención de las componentes principales a partir de los parámetros poblacionales

### 1.3.1 Componentes principales a partir de la matriz de varianzas y covarianzas $\Sigma$

Para obtener las componentes principales primero supondremos que se tiene un vector de  $p$  variables correlacionadas  $\mathbf{X}$  con características distribucionales desconocidas, pero con

$$E[\mathbf{X}] = \boldsymbol{\mu} = (\mu_i)_{p \times 1} \quad y \quad Var[\mathbf{X}] = \boldsymbol{\Sigma} = (\sigma_{ij})_{p \times p},$$

donde  $\boldsymbol{\Sigma}$  es la matriz de varianzas y covarianzas. Haremos el desarrollo bajo estas hipótesis para después proceder a realizar el análisis con la matriz de datos.

Como hemos visto anteriormente, la primera componente  $Y_1$  debe ser una combinación lineal de las variables originales de tal manera que  $Y_1$  tenga la máxima varianza posible, entonces:

$$Y_1 = \mathbf{a}'_1 \mathbf{X}$$

donde:  $\mathbf{a}'_1 = (a_{11}, a_{21}, \dots, a_{p1})$ , y  $\mathbf{X}' = (X_1, X_2, \dots, X_p)$ , de esta forma:

$$Var[Y_1] = Var[\mathbf{a}'_1 \mathbf{X}] = \mathbf{a}'_1 \boldsymbol{\Sigma} \mathbf{a}_1,$$

por lo que podemos plantear el problema:

$$Max \quad \mathbf{a}'_1 \boldsymbol{\Sigma} \mathbf{a}_1 \quad \text{sujeto a} \quad \mathbf{a}'_1 \mathbf{a}_1 = 1$$

La restricción se plantea para tener unicidad en la solución además de limitar la varianza de  $Y_1$ , pues entre más grandes sean los elementos de  $\mathbf{a}_1$ , la varianza de  $Y_1$  crecerá arbitrariamente, además de que esta normalización preserva las distancias en el espacio de dimensión  $p$ .

Para resolver este problema, el método apropiado es el de multiplicadores de Lagrange. Este método consiste en plantear la siguiente función:

$$\varphi = \mathbf{a}'_1 \boldsymbol{\Sigma} \mathbf{a}_1 - \lambda (\mathbf{a}'_1 \mathbf{a}_1 - 1)$$

que lleva al problema de encontrar el valor máximo de la función anterior, por lo que derivando  $\varphi$  con respecto a  $\mathbf{a}_1$ , tenemos:

$$\frac{d\varphi}{d\mathbf{a}_1} = 2\boldsymbol{\Sigma}\mathbf{a}_1 - 2\lambda\mathbf{a}_1$$

e igualando al vector cero esta derivada simbólica tenemos:

$$\begin{aligned} 2\boldsymbol{\Sigma}\mathbf{a}_1 - 2\lambda\mathbf{a}_1 &= \mathbf{0} \Leftrightarrow \\ 2(\boldsymbol{\Sigma}\mathbf{a}_1 - \lambda\mathbf{a}_1) &= \mathbf{0} \Leftrightarrow \\ \boldsymbol{\Sigma}\mathbf{a}_1 - \lambda\mathbf{a}_1 &= \mathbf{0} \Leftrightarrow \\ (\boldsymbol{\Sigma} - \lambda\mathbf{I})\mathbf{a}_1 &= \mathbf{0} \end{aligned}$$

la última ecuación tiene solución no trivial si  $\boldsymbol{\Sigma} - \lambda\mathbf{I}$  es una matriz singular, por lo que su determinante debe ser igual a cero:

$$|\boldsymbol{\Sigma} - \lambda\mathbf{I}| = 0 \tag{1.1}$$

que es conocido como el polinomio característico de  $\boldsymbol{\Sigma}$  y su grado es  $p$ , pues éste es el orden de  $\boldsymbol{\Sigma}$ .

Como vimos

$$\begin{aligned} \boldsymbol{\Sigma}\mathbf{a}_1 - \lambda\mathbf{a}_1 &= \mathbf{0} \Leftrightarrow \\ \boldsymbol{\Sigma}\mathbf{a}_1 &= \lambda\mathbf{a}_1 \end{aligned}$$

por lo que para resolver el problema basta con encontrar los valores y vectores propios de  $\boldsymbol{\Sigma}$ .

Ahora, la varianza de  $Y_1$  va a estar dada por:

$$Var[Y_1] = Var[\mathbf{a}'_1\mathbf{X}] = \mathbf{a}'_1 Var[\mathbf{X}] \mathbf{a}_1 = \mathbf{a}'_1\boldsymbol{\Sigma}\mathbf{a}_1$$

por otro lado:

$$\begin{aligned} \boldsymbol{\Sigma}\mathbf{a}_1 - \lambda\mathbf{a}_1 &= \mathbf{0} \Leftrightarrow \\ \boldsymbol{\Sigma}\mathbf{a}_1 &= \lambda\mathbf{a}_1 \Leftrightarrow \\ \mathbf{a}'_1\boldsymbol{\Sigma}\mathbf{a}_1 &= \lambda\mathbf{a}'_1\mathbf{I}\mathbf{a}_1 \end{aligned}$$

y  $\lambda\mathbf{a}'_1\mathbf{I}\mathbf{a}_1 = \lambda$ , ya que  $\mathbf{a}'_1\mathbf{a}_1 = 1$ , lo que implica que:

$$\begin{aligned} Var[Y_1] &= \mathbf{a}'_1\boldsymbol{\Sigma}\mathbf{a}_1 = \lambda \\ \therefore Var[Y_1] &= \lambda \end{aligned}$$

entonces podemos observar que la varianza de  $Y_1$  es una raíz del polinomio 1.1, por lo que para formar a  $Y_1$  tomaremos el vector propio asociado al valor propio más grande.

Tenemos así la primera componente principal:

$$Y_1 = \mathbf{a}'_1 \mathbf{X}$$

Dado que las componentes principales son combinaciones lineales no correlacionadas, ordenadas de acuerdo a su varianza en orden decreciente, la segunda componente  $Y_2$  tendrá la segunda varianza más grande y no estará correlacionada con  $Y_1$ :

$$Y_2 = \mathbf{a}'_2 \mathbf{X}$$

donde:  $\mathbf{a}'_2 = (a_{12}, a_{22}, \dots, a_{p2})$ , y  $\mathbf{X}' = (X_1, X_2, \dots, X_p)$ , de esta forma, para poder encontrar  $\mathbf{a}_2$ , necesitamos obtener la varianza de  $Y_2$  y la covarianza de  $Y_1$  y  $Y_2$ :

$$Var [Y_2] = Var [\mathbf{a}'_2 \mathbf{X}] = \mathbf{a}'_2 \Sigma \mathbf{a}_2,$$

$$\begin{aligned} Cov [Y_1, Y_2] &= E [(Y_2 - E [Y_2]) (Y_1 - E [Y_1])'] \\ &= E [(\mathbf{a}'_2 \mathbf{X} - E [\mathbf{a}'_2 \mathbf{X}]) (\mathbf{a}'_1 \mathbf{X} - E [\mathbf{a}'_1 \mathbf{X}])'] \\ &= E [(\mathbf{a}'_2 \mathbf{X} - \mathbf{a}'_2 E [\mathbf{X}]) (\mathbf{a}'_1 \mathbf{X} - \mathbf{a}'_1 E [\mathbf{X}])'] \\ &= E [(\mathbf{a}'_2 \mathbf{X} - \mathbf{a}'_2 \boldsymbol{\mu}) (\mathbf{a}'_1 \mathbf{X} - \mathbf{a}'_1 \boldsymbol{\mu})'] \\ &= E [\mathbf{a}'_2 (\mathbf{X} - \boldsymbol{\mu}) (\mathbf{X} - \boldsymbol{\mu})' \mathbf{a}_1] \\ &= \mathbf{a}'_2 Var [\mathbf{X}] \mathbf{a}_1 \\ &= \mathbf{a}'_2 \Sigma \mathbf{a}_1 \end{aligned}$$

como queremos que  $Y_1$  y  $Y_2$  sean no correlacionadas:

$$Cov [Y_1, Y_2] = \mathbf{a}'_2 \Sigma \mathbf{a}_1 = 0$$

Por otra parte, del desarrollo para la primera componente principal teníamos que:

$$\Sigma \mathbf{a}_1 - \lambda \mathbf{a}_1 = \mathbf{0}$$

multiplicando por la izquierda por  $\mathbf{a}'_2$  tenemos que:

$$\begin{aligned} \mathbf{a}'_2 \Sigma \mathbf{a}_1 - \lambda \mathbf{a}'_2 \mathbf{a}_1 &= 0 && \Leftrightarrow \\ \mathbf{a}'_2 \Sigma \mathbf{a}_1 &= \lambda \mathbf{a}'_2 \mathbf{a}_1 && \text{pero } \mathbf{a}'_2 \Sigma \mathbf{a}_1 = 0 \\ \therefore \mathbf{a}'_2 \mathbf{a}_1 &= 0 \end{aligned}$$

podemos ver que una restricción que será planteada al problema, es que  $\mathbf{a}_2$  debe ser ortogonal a  $\mathbf{a}_1$  entonces el problema quedará planteado de la siguiente manera:

$$\text{Max } \mathbf{a}'_2 \Sigma \mathbf{a}_2 \quad \text{sujeto a } \mathbf{a}'_2 \mathbf{a}_2 = 1 \quad \text{y} \quad \mathbf{a}'_2 \mathbf{a}_1 = 0$$

nuevamente, utilizando multiplicadores de Lagrange definiremos la siguiente función:

$$\psi = \mathbf{a}'_2 \Sigma \mathbf{a}_2 - \lambda (\mathbf{a}'_2 \mathbf{a}_2 - 1) - \gamma \mathbf{a}'_2 \mathbf{a}_1$$

para resolver el problema basta con encontrar el máximo de  $\psi$ . Derivando  $\psi$  con respecto a  $\mathbf{a}_2$  e igualando al vector cero:

$$\begin{aligned} \frac{d\psi}{d\mathbf{a}_2} &= 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \gamma \mathbf{a}_1 & (1.2) \\ 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \gamma \mathbf{a}_1 &= \mathbf{0} \end{aligned}$$

multiplicando por  $\mathbf{a}'_1$  por la izquierda tenemos:

$$2\mathbf{a}'_1 \Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}'_1 \mathbf{a}_2 - \gamma \mathbf{a}'_1 \mathbf{a}_1 = 0$$

donde podemos observar que  $\gamma = 0$  porque  $\mathbf{a}'_1 \Sigma \mathbf{a}_2 = \mathbf{a}'_1 \mathbf{a}_2 = 0$  y  $\mathbf{a}'_1 \mathbf{a}_1 = 1$ , por lo que 1.2 puede ser reducida a:

$$\begin{aligned} \frac{d\psi}{d\mathbf{a}_2} &= 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 \\ 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 &= \mathbf{0} & \Leftrightarrow \\ 2(\Sigma \mathbf{a}_2 - \lambda \mathbf{a}_2) &= \mathbf{0} & \Leftrightarrow \\ (\Sigma - \lambda \mathbf{I}) \mathbf{a}_2 &= \mathbf{0} \end{aligned}$$

al igual que para obtener  $Y_1$ , solamente basta con tomar el segundo valor propio más grande  $\lambda_2$  y  $\mathbf{a}_2$  su vector propio correspondiente. Así, la segunda componente principal será:

$$Y_2 = \mathbf{a}'_2 \mathbf{X}$$

Si continuamos con este procedimiento la  $m$ -ésima componente principal estará dada por:

$$Y_m = \mathbf{a}'_m \mathbf{X} \quad (1.3)$$

donde  $\mathbf{a}'_m$  es el vector propio asociado al  $m$ -ésimo valor propio de  $\Sigma$  :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \dots \geq \lambda_p$$

y su varianza será:

$$Var [Y_m] = Var [\mathbf{a}'_m \mathbf{X}] = \mathbf{a}'_m Var [\mathbf{X}] \mathbf{a}_m = \mathbf{a}'_m \boldsymbol{\Sigma} \mathbf{a}_m$$

pero  $\mathbf{a}'_m \boldsymbol{\Sigma} \mathbf{a}_m = \lambda_m \mathbf{a}'_m \mathbf{I} \mathbf{a}_m = \lambda_m$ .

$$\therefore Var [Y_m] = \lambda_m$$

Cuando algunos de los valores propios de  $\boldsymbol{\Sigma}$  sean iguales, no existe problema, pues los vectores propios se escogen de tal manera que sean ortogonales.

Al tener todos los vectores propios podemos formar la matriz  $\mathbf{A}$  de  $p \times p$ , donde:  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p)$ , y al tomar los vectores  $\mathbf{a}_j$  de tal manera que  $\mathbf{a}'_j \mathbf{a}_j = 1$ , la matriz  $\mathbf{A}$  es ortogonal, es decir  $\mathbf{A} \mathbf{A}' = \mathbf{I}$ . También podemos formar el vector  $\mathbf{Y}$  de  $p \times 1$  de componentes principales. Entonces:

$$\mathbf{Y} = \mathbf{A}' \mathbf{X}$$

La matriz de varianzas y covarianzas de  $\mathbf{Y}$  de  $p \times p$  la denotaremos por  $\boldsymbol{\Lambda}$ , la cual es una matriz diagonal pues es la matriz de varianzas y covarianzas de las componentes principales que son variables no correlacionadas, por lo que su covarianza es igual a cero y los elementos sobre la diagonal son los valores propios que a su vez son la varianza de cada  $Y_j$ :

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}$$

Por otra parte:

$$\boldsymbol{\Lambda} = Var [\mathbf{Y}] = Var [\mathbf{A}' \mathbf{X}] = \mathbf{A}' Var [\mathbf{X}] \mathbf{A} = \mathbf{A}' \boldsymbol{\Sigma} \mathbf{A}$$

$$\therefore \boldsymbol{\Lambda} = \mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \tag{1.4}$$

Al despejar  $\boldsymbol{\Sigma}$  de 1.4 tendremos:

$$\begin{aligned} \boldsymbol{\Lambda} &= \mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \Leftrightarrow \\ \mathbf{A} \boldsymbol{\Lambda} &= \mathbf{A} \mathbf{A}' \boldsymbol{\Sigma} \mathbf{A} \end{aligned}$$

pero  $\mathbf{A}\mathbf{A}' = \mathbf{I}$ , por lo que  $\mathbf{A}\mathbf{\Lambda} = \mathbf{\Sigma}\mathbf{A}$ , y multiplicando por la derecha por  $\mathbf{A}'$ :

$$\begin{aligned}\mathbf{A}\mathbf{\Lambda}\mathbf{A}' &= \mathbf{\Sigma}\mathbf{A}\mathbf{A}' \Leftrightarrow \\ \mathbf{A}\mathbf{\Lambda}\mathbf{A}' &= \mathbf{\Sigma}\mathbf{I} \\ \therefore \mathbf{A}\mathbf{\Lambda}\mathbf{A}' &= \mathbf{\Sigma}\end{aligned}$$

Calculando la traza de la matriz  $\mathbf{\Sigma}$ , podemos ver que:

$$\begin{aligned}tr(\mathbf{\Sigma}) &= tr(\mathbf{A}\mathbf{\Lambda}\mathbf{A}') \\ &= tr(\mathbf{A}'\mathbf{A}\mathbf{\Lambda}) \\ &= tr(\mathbf{\Lambda}) \\ &= \sum_{i=1}^p \lambda_i = \sum_{i=1}^p Var[Y_i]\end{aligned}$$

lo que indica que la varianza total de las variables originales, expresada como la suma de varianzas, es igual a la suma de las varianzas de todas las  $p$  componentes principales. La igualdad anterior nos lleva a un resultado muy importante: *"la  $i$ -ésima componente principal aporta una proporción  $\lambda_i / \sum_{i=1}^p \lambda_i$  de la variación total de las variables originales"*, por lo que las primeras  $m$  componentes principales aportan una proporción de:

$$\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}$$

de la variación total.

### 1.3.2 Componentes principales a partir de la matriz de correlaciones $\mathbf{P}$ .

Generalmente, extraer las componentes de la matriz de varianzas y covarianzas  $\mathbf{\Sigma}$ , preserva la intención del análisis de componentes principales, especialmente si las componentes serán usadas en futuros cálculos. Sin embargo, en algunos casos, las componentes principales tendrán una mejor interpretación si es utilizada la matriz de correlaciones  $\mathbf{P}$ . Además de lo anterior existe otro motivo: *la varianza de una variable no es independiente de la escala*.

La razón de lo anterior es simple: imaginemos un conjunto de variables  $X_1, X_2, \dots, X_p$ , las cuales son de diferentes tipos, es decir, están medidas en diferentes



unidades. En un caso así, la estructura de las componentes principales derivadas de  $\Sigma$  dependerá sobre todo de la selección arbitraria de las unidades de medición; adicionalmente, si hay grandes diferencias entre las varianzas de  $X_1, X_2, \dots, X_p$ , las variables que tengan las varianzas más grandes tenderán a dominar las primeras componentes principales. Esta dificultad se puede modificar de tal manera que las componentes obtenidas no se vean afectadas, esto se puede conseguir estandarizando las variables para que cada una tenga varianza igual a uno, esto significa que estamos utilizando la matriz de correlaciones  $\mathbf{P}$  en lugar de la matriz de varianzas y covarianzas  $\Sigma$ .

Para encontrar las componentes principales primero obtendremos la matriz  $\mathbf{P}$  a partir de la matriz  $\Sigma$ . Primero definiremos como  $\mathbf{D}$  a la matriz diagonal cuyos elementos sobre la diagonal son el recíproco de las desviaciones estándar de las variables  $X_i$ :

$$\mathbf{D} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

de esta forma, la matriz  $\mathbf{P}$  estará dada por:

$$\mathbf{P} = \mathbf{D}\Sigma\mathbf{D}$$

así cada término de esta matriz será de la forma:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$$

que son los coeficientes de correlación entre las variables  $X_i$  y  $X_j$ , por lo tanto la matriz  $\mathbf{P}$  será:

$$\mathbf{P} = \mathbf{D}\Sigma\mathbf{D} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix},$$

La derivación matemática es la misma, de tal forma que las componentes principales resultan ser combinaciones lineales de los vectores propios de  $\mathbf{P}$ . Sin embargo, no son las mismas que se obtendrían si utilizáramos  $\Sigma$ . Cuando se elige analizar  $\mathbf{P}$  en lugar de  $\Sigma$ , resulta ser una decisión arbitraria para considerar a las variables igualmente importantes.

En la matriz de correlación, los términos sobre la diagonal son todos iguales a uno. De esta forma:

$$\text{tr}(\mathbf{P}) = p = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{Var}[Y_i]$$

Entonces la proporción que aporta a la variación total la  $j$ -ésima componente es  $\lambda_j/p$ .

El decidir cuál de las matrices se debe utilizar para obtener las componentes principales dependerá en gran parte de la naturaleza del problema: si existen diferencias importantes entre las unidades de medición de cada variable o entre sus varianzas.

## 1.4 Obtención de las componentes a partir de la matriz de datos

En la práctica resulta muy difícil contar con la matriz de varianzas y covarianzas  $\Sigma$ . Por consiguiente, a partir de los datos conocidos será necesario estimar dos matrices que nos servirán para obtener las componentes principales: la matriz de varianzas y covarianzas muestrales  $\mathbf{S}$  y la matriz de correlaciones muestrales  $\mathbf{R}$ . Para ambos casos, supondremos que los valores de las  $p$  variables  $X_1, X_2, \dots, X_p$  son obtenidas a partir de observaciones de  $n$  individuos, numerados desde 1 hasta  $n$ . Sea  $x_{ij}$  el valor de la variable  $j$  obtenida del individuo  $i$ . La matriz de datos de  $n \times p$  puede ser escrita como:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

### 1.4.1 Componentes principales a partir de la matriz de varianzas y covarianzas muestrales $\mathbf{S}$

Sean:

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad \text{la media muestral de la variable } x_j$$

y

$$s_{ij} = \frac{\sum_{r=1}^n (x_{ri} - \bar{x}_i)(x_{rj} - \bar{x}_j)}{n - 1} \quad \text{la covarianza muestral de las variables } x_i \text{ y } x_j$$

El denominador  $n - 1$  es elegido de tal forma que  $s_{ij}$  sea un estimador insesgado de la correspondiente covarianza poblacional, denotada por  $\sigma_{ij}$ . Entonces la matriz  $\mathbf{S}$  tiene la expresión siguiente:

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}$$

Como  $s_{ij}$  es un estimador insesgado para  $\sigma_{ij}$ ,  $\forall i$  y  $j$ , decimos que  $\mathbf{S}$  es un estimador insesgado para la matriz de varianzas y covarianzas  $\Sigma$ .  $\mathbf{S}$  es simétrica y sus elementos sobre la diagonal son las varianzas muestrales de las variables. Las desviaciones estándar están dadas por  $\sqrt{s_{ii}}$ . Las cantidades dependen de las escalas utilizadas para medir las variables y generalmente son difíciles de interpretar como una medida de asociación entre una pareja dada de variables.

La matriz de varianzas y covarianzas muestral puede ser expresada en términos de los vectores de observación:

$$\begin{aligned} \mathbf{S} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})' \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' - n \bar{\mathbf{x}} \bar{\mathbf{x}}' \right) \end{aligned}$$

Lo anterior lo podemos expresar en términos de la matriz de datos como:

$$\mathbf{S} = \frac{1}{n-1} (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}})' (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'),$$

donde

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}, \quad \bar{\mathbf{x}}' = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p).$$

El problema de encontrar las componentes principales es determinar la combinación lineal  $\sum_{j=1}^p a_{jk} x_j$  de las variables  $X_1, X_2, \dots, X_p$  que tenga la varianza muestral máxima. La varianza muestral de  $\sum_{j=1}^p a_j x_j$  es, en términos vectoriales,  $\mathbf{a}'\mathbf{S}\mathbf{a}$ . Siguiendo el procedimiento mencionado en la sección anterior, los elementos de  $\mathbf{a}$  deben satisfacer las  $p$  ecuaciones lineales simultáneas:

$$(\mathbf{S} - \lambda \mathbf{I}) \mathbf{a} = \mathbf{0}$$

donde  $\lambda$  es un multiplicador de Lagrange. El valor de  $\lambda$  es seleccionado de tal forma que

$$|\mathbf{S} - \lambda\mathbf{I}| = 0$$

Aquí  $\lambda$  es el valor propio más grande de  $\mathbf{S}$ , la matriz de varianzas y covarianzas muestrales, siendo  $\mathbf{a}$  el correspondiente vector propio. El proceso de obtener las siguientes componentes se sigue de la misma manera como se obtuvieron a partir de la matriz de varianzas y covarianzas poblacionales  $\Sigma$ . Dado que  $\mathbf{S}$  es una matriz semidefinida positiva, los valores propios son todos no negativos y representan las varianzas estimadas de las diferentes componentes. Es importante tener en cuenta que las componentes principales obtenidas de esta matriz se ven como las componentes principales y no como estimadores de las cantidades correspondientes obtenidas de  $\Sigma$ .

### 1.4.2 Componentes principales a partir de la matriz de correlaciones muestrales $\mathbf{R}$

Como ya hemos explicado anteriormente, la varianza de una variable no es independiente de la escala. Cuando las variables son medidas en unidades muy diferentes, los efectos de la escala pueden influir en la composición de las componentes principales obtenidas, es decir, las componentes de  $\mathbf{S}$  serán dominadas por las variables con mayor varianza y las variables con menor varianza contribuirán muy poco. Para tales casos, como ya mencionamos, obtendremos la matriz de correlaciones muestrales  $\mathbf{R}$ . La matriz de correlaciones muestrales de  $p \times p$ , se puede obtener a partir de la matriz de varianzas y covarianzas muestrales,  $\mathbf{S}$ :

$$\mathbf{R} = \mathbf{D}^*\mathbf{S}\mathbf{D}^*$$

donde  $\mathbf{D}^*$  es la matriz diagonal:

$$\mathbf{D}^* = \begin{pmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{pmatrix}$$

$$\therefore \mathbf{R} = \mathbf{D}^* \mathbf{S} \mathbf{D}^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

donde:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}} \sqrt{s_{jj}}}$$

Para extraer las componentes principales de la matriz de correlaciones muestrales se sigue el mismo procedimiento que para la matriz  $\mathbf{S}$ .

Aunque el procedimiento para obtener las componentes principales de  $\mathbf{R}$  o de  $\mathbf{S}$  es el mismo, hay algunas diferencias en la interpretación. Como ya vimos, en la matriz de correlación, los términos sobre la diagonal son todos iguales a uno. De esta manera, la suma de estos términos, es decir, la suma de las varianzas de las observaciones estandarizadas será igual a  $p$ . Entonces la suma de los valores propios de  $\mathbf{R}$  será también igual a  $p$ , y la proporción que aporta a la variación total la  $j$ -ésima componente será  $\lambda_j/p$ . Además, los coeficientes  $a_{ij}$  son valores normalizados, los cuales deben ser aplicados a los scores de las variables estandarizadas cuando se calculan los scores de las componentes.

## 1.5 Cuántas componentes principales retener

Como ya hemos mencionado anteriormente, el análisis de componentes principales es frecuentemente utilizado con el propósito de generar un conjunto reducido de variables que acumulen la mayor parte de la varianza de los datos originales, y que pueda ser utilizado en análisis subsecuentes. Lo anterior nos lleva a decidir cuántas componentes hay que retener con el objeto de resumir adecuadamente la dimensión de los datos. Desgraciadamente, no existe un método aceptado universalmente para hacerlo. La decisión depende del juicio del investigador. Algunos procedimientos para determinar cuántas componentes hay que retener han sido recomendados. Las reglas más comunes que han sido propuestas son:

1. Incluya sólo las suficientes componentes para explicar el porcentaje significativo de la variación total. Entre el 70% y el 90% son sugeridas por algunos autores, aunque éstas generalmente se volverán más pequeñas mientras  $p$  o  $n$  aumenten.<sup>1</sup>

---

<sup>1</sup>Everitt (1991)

En este método el reto consiste en seleccionar un porcentaje adecuado. Si se pretende un porcentaje muy alto, se corre el riesgo de incluir componentes que son específicas de la muestra o específicas de una variable. Por específicas de una muestra significa que una componente no se pueda generalizar a la población o a otras muestras. Una componente de una variable específica es dominada por una sola variable y no representa un resumen de muchas variables.<sup>2</sup>

2. Excluya aquellas componentes que procedan de valores propios que sean menores que el promedio, es decir, menores que 1 si las componentes fueron obtenidas de la matriz de correlaciones  $\mathbf{R}$ . Esta regla fue anteriormente propuesta por Kaiser (1958), pero Jolliffe (1972) ha sugerido, a base de simulaciones, que una regla más apropiada sería excluir aquéllas cuyos valores propios asociados sean menores que 0.7.<sup>3</sup>
3. Un tercer método consiste en realizar una gráfica *scree*, la cual se construye graficando  $i$  versus  $\lambda_i$  (sugerido por Cattell en 1965) y seleccionar como el número de componentes el valor de  $i$  correspondiente a un "codo" en la curva, esto es, el punto que se localiza donde los valores propios "grandes" terminan y comienzan los valores propios "pequeños".<sup>4</sup> Así, como se muestra en la gráfica, lo más apropiado es retener dos componentes, pues a partir de la tercera comienza una disminución notable en los valores propios.

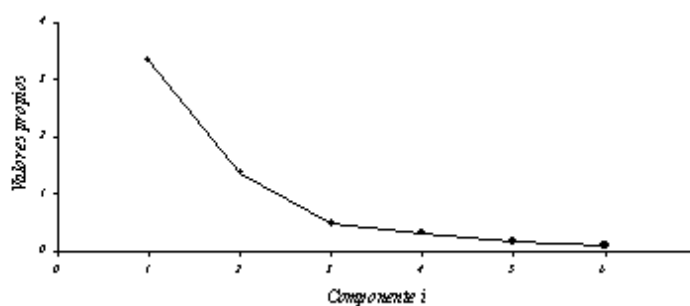


Fig. 1.5.1 Gráfica scree.

---

<sup>2</sup>Rencher (1995)

<sup>3</sup>Otra vez se supone que se ha utilizado la matriz de correlaciones  $\mathbf{R}$

<sup>4</sup>Jolliffe (1986), señala una alternativa a este modo de graficar los valores propios: graficar  $\log(\lambda_i)$  versus  $i$ .

## 1.6 Información en las últimas componentes principales

Como ya hemos visto, las componentes principales son de gran utilidad para reducir la dimensión de los datos y así facilitar el análisis de la información que puedan proveer; asimismo hemos enumerado técnicas para seleccionar un número adecuado de componentes que conserven la mayor parte de información posible sin necesidad de ocupar un número de componentes que resulte difícil de manejar. Sin embargo, las últimas componentes pueden contener información que pudiera resultar útil en algunas aplicaciones.

Como se señaló anteriormente, los valores propios sirven como varianzas de las componentes principales, lo que implica que las últimas componentes tienen las varianzas más pequeñas. Así podemos considerar varios casos cuando las varianzas de las componentes son pequeñas.

### 1.6.1 Valores propios iguales a cero

Si algunas de las variables originales son linealmente dependientes, entonces, algunos de los valores propios de  $\Sigma$  serán cero. La dimensión del espacio en donde están las observaciones es igual al rango de  $\Sigma$ , y está dado por  $(p - k)$  donde  $k$  es igual al número de valores propios iguales a cero. Si hay  $k$  valores propios iguales a cero, entonces podemos encontrar  $k$  restricciones linealmente independientes sobre las variables. Estas restricciones son llamadas *relaciones estructurales*.

Si tenemos una muestra de una población más grande, entonces también existe una dependencia lineal exacta en la muestra. Si observamos a los valores propios de  $\mathbf{S}$ , entonces los estimadores muestrales de los valores propios iguales a cero de  $\Sigma$  deberán ser cero también.

### 1.6.2 Valores propios cercanos a cero

La ocurrencia de dependencia lineal exacta es rara, excepto cuando uno introduce de manera deliberada algunas variables redundantes. Un problema práctico más importante es detectar la dependencia lineal aproximada. Si el valor propio más pequeño es muy cercano a cero, entonces la  $p$ -ésima componente principal,  $\mathbf{a}'_p \mathbf{X}$ , es casi constante, y la dimensión de  $\mathbf{X}$  es "casi" menor que  $p$ . Si los últimos  $(p - r)$

valores propios se considera que son pequeños, entonces la "eficiencia" de restringir la dimensión a  $r$  está dada por  $\sum_{i=1}^r \lambda_i / \sum_{i=1}^p \lambda_i$ . Las componentes principales que corresponden a los valores propios pequeños son variables para las cuales los elementos de la población tienen valores casi iguales.

Si  $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_p$  son "pequeños", se pierde poca información reemplazando los valores de las correspondientes componentes principales con sus medias, las cuales se han elegido que sean cero usando la ecuación  $\mathbf{Y} = \mathbf{A}'(\mathbf{X} - \bar{\mathbf{x}})$ . De esta forma, se pueden aproximar los scores o puntajes para el  $m$ -ésimo individuo por:

$\mathbf{y}'_m = (y_{m1}, y_{m2}, \dots, y_{mr}, 0, \dots, 0)$  y aproximar las correspondientes observaciones originales por:  $\mathbf{x}_m = \mathbf{A}\mathbf{y}_m + \bar{\mathbf{x}}$ . Podemos aproximar también  $\mathbf{S}$  por  $\sum_{i=1}^r \lambda_i \mathbf{a}_i \mathbf{a}'_i$ . En ambas fórmulas usamos únicamente las primeras  $r$  componentes y es práctica usual tomar las primeras pocas columnas de  $\mathbf{A}$  que corresponden a los valores propios que se consideran como "grandes".

### 1.6.3 Valores propios repetidos

Algunas veces algunos de los valores propios de  $\Sigma$  serán iguales. Si tenemos que  $\lambda_{q+1} = \dots = \lambda_{q+k}$ , entonces  $\lambda = \lambda_{q+1}$  por lo que se dice que es una raíz de multiplicidad  $k$ . Los vectores propios correspondientes a las raíces múltiples no son únicos ya que uno puede elegir cualquier conjunto ortonormal en el subespacio apropiado de  $k$  dimensiones. Las componentes principales correspondientes tendrán la misma varianza. El problema que existe en la práctica es que las raíces correspondientes a la matriz de varianzas y covarianzas muestral, generalmente no serán iguales, así que la multiplicidad, por lo general, no se observa en el caso muestral. En lugar de esto, encontramos que muestras diferentes darán estimadores de los valores propios completamente diferentes y por lo tanto éstos no deben ser considerados como "variables características". Esto es algo que debemos tener en cuenta cuando se encuentra que los valores propios son aproximadamente iguales. Se han desarrollado pruebas para la igualdad de valores propios pero éstas requieren la hipótesis de normalidad y son válidas solamente para muestras grandes. Un caso importante ocurre cuando los últimos  $k$  valores propios son iguales. En este caso, la variación en las últimas  $k$  dimensiones se dice que es *esférica*. Entonces las últimas  $k$  componentes principales pueden ser consideradas como medida de variabilidad no específica y las características esenciales de  $\mathbf{X}$  están representadas por las primeras  $(p - k)$  componentes principales. Por supuesto que, si los últimos  $k$  valores característicos son iguales y son cero, entonces



los comentarios anteriores correspondientes se aplican a este caso.

## 1.7 Algunos aspectos importantes de las componentes principales

A continuación presentaremos algunas consideraciones que debemos tener en cuenta cuando se realiza un análisis de componentes principales.

### 1.7.1 Scores o puntajes de las componentes principales

Como ya hemos señalado anteriormente, un análisis de componentes principales es el primer paso para realizar un estudio detallado de un conjunto de datos. Para conseguir este objetivo, el investigador deseará obtener los scores de las componentes.

Los *scores o puntajes de las componentes principales* son las ubicaciones de las observaciones con respecto a sus ejes componentes principales. Para obtener los scores individuales de las componentes principales obtenidas simplemente se puede aplicar la fórmula 1.3 a los valores de las variables individuales. Sin embargo, es más conveniente que las componentes principales tengan media cero.

#### Scores o puntajes de las componentes corregidas por la media

Para obtener los scores de las componentes principales cuya media sea cero, es necesario sumar un vector apropiado de constantes de tal forma que las componentes principales tengan todas media cero. Si la media muestral es  $\bar{\mathbf{x}}$ , y  $\mathbf{A}$  es la matriz de vectores propios para la matriz de varianzas y covarianzas muestral  $\mathbf{S}$ , entonces la transformación usual es:

$$\mathbf{Y} = \mathbf{A}'(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1}') \quad (1.5)$$

Ésta consiste de una traslación seguida por una rotación ortogonal. Usando esta ecuación para la observación  $\mathbf{x}_m$ , para el individuo  $m$ -ésimo, se encuentra que:

$$\mathbf{y}_m = \mathbf{A}'(\mathbf{x}_m - \bar{\mathbf{x}})$$

En la cual, los elementos de  $\mathbf{y}_m$  son llamados los scores o puntajes para el individuo  $m$ -ésimo. Es esencial notar que si  $\mathbf{A}$  denota la matriz de vectores propios para la matriz de correlación muestral, entonces no será necesario realizar la transformación

anterior puesto que ya la hemos aplicado al realizar la estandarización de los datos, por lo que la ecuación anterior debería ser utilizada únicamente después de estandarizar las observaciones  $(\mathbf{x}_m - \bar{\mathbf{x}})$ , de tal forma que cada variable tenga varianza unitaria.

De la ecuación 1.5 vemos que multiplicando por la matriz  $\mathbf{A}$  por la izquierda:

$$\begin{aligned}\mathbf{Y} &= \mathbf{A}'(\mathbf{X} - \bar{\mathbf{x}}) \\ \mathbf{AY} &= \mathbf{AA}'(\mathbf{X} - \bar{\mathbf{x}}) \quad \text{y} \quad \mathbf{AA}' = \mathbf{I} \\ \therefore \quad \mathbf{AY} &= \mathbf{X} - \bar{\mathbf{x}}\end{aligned}$$

### 1.7.2 Ortogonalidad

Es fácil notar que la restricción para los vectores propios de ser ortogonales da lugar a una matriz  $\mathbf{A}$  ortogonal. El usar una rotación ortogonal es que la suma de los cuadrados de las desviaciones para cada individuo alrededor del vector de medias permanecerá sin cambio por la rotación. Para demostrarlo denotaremos por  $\tilde{\mathbf{X}}$  a la matriz de  $(n \times p)$  de datos corregidos por la media y la matriz de los scores o puntajes de  $(n \times p)$  por  $\mathbf{Y}$ , tenemos que:

$$\begin{aligned}\mathbf{Y} &= \tilde{\mathbf{X}}\mathbf{A} \\ \mathbf{YY}' &= \tilde{\mathbf{X}}\mathbf{A}(\tilde{\mathbf{X}}\mathbf{A})' = \tilde{\mathbf{X}}\mathbf{AA}'\tilde{\mathbf{X}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}'\end{aligned}$$

Los términos de la diagonal de las matrices de  $(n \times n)$   $\mathbf{YY}'$  y  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}'$  dan las sumas de cuadrados para cada individuo.

### 1.7.3 Vectores de carga de componentes

Como vimos anteriormente, los vectores  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$ , los cuales definen las componentes principales son usados para formar la matriz  $\mathbf{A}$ , y los valores propios,  $\lambda_1, \lambda_2, \dots, \lambda_p$ , arreglados en una matriz diagonal,  $\mathbf{\Lambda}$ ; asimismo, vimos que la matriz de varianzas y covarianzas  $\mathbf{S}$  de las variables originales está dada por:

$$\mathbf{S} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}' \tag{1.6}$$

cuando las componentes son obtenidas de la matriz  $\mathbf{S}$ .

Si cambiamos la escala de los vectores  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$  de tal forma que la suma de los cuadrados de sus elementos es igual al correspondiente valor propio,  $\lambda_i$ , en lugar de

la unidad, es decir, calculando  $\mathbf{c}_i = \sqrt{\lambda_i} \mathbf{a}_i$ , así  $\mathbf{c}'_i \mathbf{c}_i = (\sqrt{\lambda_i} \mathbf{a}_i)' \sqrt{\lambda_i} \mathbf{a}_i = \lambda_i \mathbf{a}'_i \mathbf{a}_i = \lambda_i$ , además podemos reescribir a la ecuación 1.6 de la siguiente manera:

$$\mathbf{S} = \mathbf{C}\mathbf{C}'$$

donde  $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p]$ . Los elementos de  $\mathbf{C}$  son tales que los coeficientes de las componentes más importantes están con la escala cambiada (escalados) para ser generalmente más grandes que aquéllos de las componentes menos importantes, lo que parece intuitivamente razonable.

Los nuevos vectores ( $\mathbf{c}_i$ ) o vectores escalados tienen dos diferentes interpretaciones directas. Primero, supongamos que cambiamos la escala de las componentes de tal forma que todas tengan varianzas unitaria. Esto puede ser hecho por:  $\mathbf{Y}^* = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y}$ . Entonces, la transformación inversa  $\mathbf{X} = \mathbf{A}\mathbf{Y}$  (suponiendo que  $\mathbf{X}$  tiene vector de medias  $\mathbf{0}$ ) se transforma en:  $\mathbf{X} = \mathbf{A}\mathbf{\Lambda}^{\frac{1}{2}} \mathbf{Y}^* = \mathbf{C}\mathbf{Y}^*$ . Comparando esta ecuación con el modelo de análisis de factores (otra técnica de análisis multivariado), se dice que los elementos de  $\mathbf{C}$  son análogos a los coeficientes llamados cargas factoriales y por lo tanto son llamados *cargas de componentes*.

### El problema de cambiar la escala de las componentes principales

Es importante darse cuenta que las componentes principales de un conjunto de variables dependen de las escalas utilizadas para medir las variables, por ejemplo suponga que se tienen medidas de  $n$  tortugas: su altura en pulgadas, su edad en años y su peso en onzas para tener un vector  $\mathbf{X}$ . Denotaremos la matriz de varianzas y covarianzas muestral por  $\mathbf{S}_x$ , sus valores propios por  $\{\lambda_i\}$  y sus vectores propios por  $\{\mathbf{a}_i\}$ . Si transformamos a nuevas coordenadas de tal forma que:

$$\mathbf{Z}' = (\text{altura en centímetros, edad en meses, peso en gramos})$$

entonces  $\mathbf{z} = \mathbf{K}\mathbf{x}$ . Donde  $\mathbf{K}$  es la matriz diagonal:

$$\begin{pmatrix} \frac{1}{2.54} & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & \frac{1}{28.35} \end{pmatrix}$$

pues hay 2.54 centímetros en una pulgada, 12 meses en un año y 28.35 gramos en una onza.

La matriz de varianzas y covarianzas de las nuevas variables estará dada por:

$$\mathbf{S}_z = \mathbf{K}\mathbf{S}_x\mathbf{K}$$

ya que  $\mathbf{K}' = \mathbf{K}$ . Los valores y vectores propios de  $\mathbf{S}_z$  generalmente serán diferentes de los de  $\mathbf{S}_x$  y serán denotados por  $\{\lambda_i^*\}$  y  $\{\mathbf{a}_i^*\}$ . Sin embargo, estos valores y vectores propios no dan las mismas componentes principales cuando se transforman de regreso a las variables originales. Las componentes principales cambian a menos que:

1. Todos los elementos diagonales de  $\mathbf{K}$  sean los mismos, de tal forma que  $\mathbf{K} = c\mathbf{I}$ , donde  $c$  es un escalar constante. Esto significa que todas las variables son escaladas de la misma forma.
2. Las variables correspondientes a elementos diagonales desiguales de  $\mathbf{K}$  son no correlacionados. En particular, si todos los elementos de  $\mathbf{K}$  son desiguales, entonces  $\mathbf{S}_x$  debe ser una matriz diagonal, en cuyo caso no tiene sentido llevar a cabo un análisis de componentes principales ya que las variables son todas no correlacionadas.

#### 1.7.4 Correlación entre las variables originales y las componentes principales

El utilizar la correlación entre las variables originales y las componentes principales resulta muy útil cuando se realiza la interpretación de las componentes.

La correlación entre la variable  $X_i$  y la componente principal  $Y_j$  estará dada por:

$$\rho_{X_i, Y_j} = \frac{Cov[X_i, Y_j]}{\sqrt{Var[X_i]} \sqrt{Var[Y_j]}}$$

Por un lado, sabemos que:

$$\mathbf{Y} = \mathbf{A}'\mathbf{X}$$

como  $\mathbf{A}$  es una matriz ortogonal, multiplicando ambos lados de la igualdad por  $\mathbf{A}$ :

$$\begin{aligned} \mathbf{AY} &= \mathbf{AA}'\mathbf{X} \\ \therefore \mathbf{X} &= \mathbf{AY} \end{aligned}$$

entonces:

$$X_i = a_{i1}Y_1 + a_{i2}Y_2 + \dots + a_{ip}Y_p = \sum_{k=1}^p a_{ik}Y_k$$

así la covarianza entre  $X_i$  y  $Y_j$  se podrá expresar de la siguiente manera:

$$\begin{aligned}
Cov [X_i, Y_j] &= Cov \left[ \sum_{k=1}^p a_{ik} Y_k, Y_j \right] \\
&= E \left[ \left( \sum_{k=1}^p a_{ik} Y_k - E \left[ \sum_{k=1}^p a_{ik} Y_k \right] \right) (Y_j - E [Y_j]) \right] \\
&= E \left[ \left( \sum_{k=1}^p a_{ik} Y_k - \sum_{k=1}^p E [a_{ik} Y_k] \right) (Y_j - E [Y_j]) \right] \\
&= E \left[ \left( \sum_{k=1}^p a_{ik} Y_k - \sum_{k=1}^p a_{ik} E [Y_k] \right) (Y_j - E [Y_j]) \right] \\
&= E \left[ \left( \sum_{k=1}^p a_{ik} (Y_k - E [Y_k]) \right) (Y_j - E [Y_j]) \right] \\
&= E \left[ \sum_{k=1}^p a_{ik} (Y_k - E [Y_k]) (Y_j - E [Y_j]) \right] \\
&= E \left[ \sum_{k \neq j}^p a_{ik} Cov [Y_k, Y_j] + a_{ij} Var [Y_j] \right]
\end{aligned}$$

pero como  $Cov [Y_k, Y_j] = 0$  y  $Var [Y_j] = \lambda_j$ , tenemos que:

$$Cov [X_i, Y_j] = a_{ij} Var [Y_j] = a_{ij} \lambda_j$$

por otro lado,  $Var [X_j] = \sigma_{jj}$ , entonces la correlación va estar dada por:

$$\rho_{X_i, Y_j} = \frac{Cov [X_i, Y_j]}{\sqrt{Var [X_i]} \sqrt{Var [Y_j]}} = \frac{a_{ij} \lambda_j}{\sqrt{\sigma_{ii}} \sqrt{\lambda_j}} = \frac{a_{ij} \sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}$$

Como señalamos en el apartado anterior, la matriz  $\mathbf{C}$  tiene dos diferentes interpretaciones directas, una de ellas ya fue mencionada anteriormente, y la segunda interpretación directa para la matriz  $\mathbf{C}$  surge cuando la matriz de correlación  $\mathbf{P}$  de  $\mathbf{X}$  ha sido analizada de tal forma que  $\mathbf{P} = \mathbf{C}\mathbf{C}'$ , esto es, en el caso en que las componentes hayan sido obtenidas a través de la matriz de correlaciones, las variables  $X_i$  fueron estandarizadas, por lo que  $\sqrt{Var [X_i]} = 1$  y la correlación estará dada por:

$$\rho_{X_i, Y_j} = \frac{Cov [X_i, Y_j]}{\sqrt{Var [X_i]} \sqrt{Var [Y_j]}} = \frac{a_{ij} \lambda_j}{(1) (\sqrt{\lambda_j})} = a_{ij} \sqrt{\lambda_j}$$

Así la matriz de correlación está dada por:

$$Corr [\mathbf{Y}, \mathbf{X}] = \mathbf{A}\mathbf{\Lambda}^{\frac{1}{2}} = \mathbf{C}$$

Así cuando  $\mathbf{C}$  es calculada a partir de una matriz de correlaciones  $\mathbf{P}$ , sus elementos miden las correlaciones entre las componentes y las variables estandarizadas, y estos elementos son entonces llamados de la misma forma: *correlaciones entre las componentes y las variables originales*.

## 1.8 Interpretación de las componentes principales

Hemos señalado que las componentes principales obtenidas a partir de  $\mathbf{R}$  serán distintas a aquellas obtenidas a partir de la matriz  $\mathbf{S}$ . En vista de lo anterior, al momento de escoger entre las componentes principales obtenidas a partir de la matriz de varianzas y covarianzas muestrales  $\mathbf{S}$  y la matriz de correlaciones muestrales  $\mathbf{R}$ , es necesario tener presente que éstas tendrán diferentes interpretaciones. Si las variables tienen varianzas muy diferentes, es mejor, para fines de interpretación, utilizar  $\mathbf{R}$  en vez de  $\mathbf{S}$ .

### 1.8.1 Patrones especiales en la matriz de varianzas y covarianzas muestrales ( $\mathbf{S}$ ) o en la matriz de correlaciones muestrales ( $\mathbf{R}$ )

Si existen ciertos patrones en los elementos de las matrices  $\mathbf{S}$  o  $\mathbf{R}$ , será posible deducir la estructura de las componentes principales. Así, como hemos mencionado en varias ocasiones, que si hay variables con varianzas mucho más grandes que el resto, éstas aportarán la mayor parte de la variabilidad en la primera componente principal.

También es muy importante tener en cuenta, al momento de llevar a cabo un análisis de componentes principales, que las variables estén correlacionadas, pues puede suceder que una o más componentes principales dupliquen una o más variables, es decir, la o las variables serán componentes principales por sí mismas. A continuación demostraremos este hecho.

Suponga que tenemos  $p$  variables no correlacionadas, entonces la matriz de va-

rianzas y covarianzas muestrales estará dada de la siguiente manera:

$$\mathbf{S} = \begin{pmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{pp} \end{pmatrix}$$

y el polinomio característico será:

$$0 = |\mathbf{S} - \lambda \mathbf{I}| = \prod_{i=1}^p (s_{ii} - \lambda)$$

cuyas soluciones serán:

$$\lambda_i = s_{ii}, \quad i = 1, \dots, p$$

Los correspondientes vectores propios normalizados tendrán un 1 en la  $i$ -ésima posición y 0's en las demás posiciones:

$$\mathbf{a}'_i = (0, \dots, 0, 1, 0, \dots, 0)$$

Entonces la  $i$ -ésima componente será:

$$y_i = \mathbf{a}'_i \mathbf{x} = x_i$$

En la práctica, difícilmente existirán correlaciones muestrales iguales a cero, pero si las correlaciones son todas pequeñas, las variables serán duplicadas en gran parte por las componentes principales.

Cuando todas las covarianzas o correlaciones son positivas, todos los elementos del primer vector propio  $\mathbf{a}_1$  serán positivos. Dado que los vectores propios  $\mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_p$  son ortogonales a  $\mathbf{a}_1$ , estos vectores tendrán elementos positivos y negativos. Cuando todos los elementos de  $\mathbf{a}_1$  son positivos, la primera componente principal es un promedio ponderado de las variables y algunas veces se toma como una medida de *tamaño*. Asimismo, los coeficientes positivos y negativos de las siguientes componentes principales pueden ser considerados como medida de *forma*.

## 1.8.2 Interpretación geométrica de las componentes principales

Es fácil mostrar, en términos geométricos, que la primera componente principal define la línea de mejor ajuste, en términos de mínimos cuadrados de las observaciones

$p$ -dimensionales de la muestra. Estas observaciones pueden ser representadas en una dimensión tomando su proyección sobre esta línea, es decir, encontrando el score de su primera componente principal. Si los puntos son colineales, esta representación podría contener la mayor variación en los datos, y la matriz de varianzas y covarianzas muestrales podría tener un solo valor propio distinto de cero, pues como ya vimos anteriormente valores propios iguales a cero indican dependencia lineal.

Por otra parte, al graficar los scores individuales correspondientes a las dos primeras componentes individuales (cuando la dimensión de los datos haya sido posible reducirla a dos), podremos ser capaces de identificar datos atípicos o grupos de individuos.

### Gráficas biplot y $h$ -plot

La gráfica biplot es una técnica introducida por Gabriel (1971) que describe tanto las relaciones que existen entre las observaciones  $p$ -dimensionales  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  y las relaciones entre las variables. Este tipo de gráfica está basada en el hecho de que cualquier matriz  $\mathbf{B}_{n \times m}$  de rango  $r$  puede ser factorizada de manera no única como:

$$\mathbf{B} = \mathbf{G}\mathbf{H}' \quad (1.7)$$

donde  $\mathbf{G}$  y  $\mathbf{H}$  son matrices de  $n \times r$  y  $m \times r$ , respectivamente, ambas de rango  $r$ . Entonces

$$b_{ij} = \mathbf{g}'_i \mathbf{h}_j,$$

donde  $\mathbf{g}'_i$  y  $\mathbf{h}'_j$  son los renglones de  $\mathbf{G}$  y  $\mathbf{H}$ , respectivamente por lo que tenemos cada elemento de la matriz  $\mathbf{B}$  representado por vectores de dimensión  $r$ . Esta factorización asigna cada  $\mathbf{g}_i$  a un reglón de  $\mathbf{B}$  y cada  $\mathbf{h}_j$  es asignado a una columna de  $\mathbf{B}$ . Si  $r = 2$ , podemos graficar los  $n + m$  vectores en dos dimensiones y obtener una gráfica biplot.<sup>5</sup> Para  $r > 2$  puede ser posible aproximar  $\mathbf{B}$  por una matriz de rango 2 y la correspondiente biplot explicará a  $\mathbf{B}$  por sí misma.

De 1.7 podemos escribir  $\mathbf{B} = (\mathbf{G}\mathbf{R}') (\mathbf{H}\mathbf{R}^{-1})'$  para alguna matriz no singular  $\mathbf{R}$ , sin embargo, el hecho de que  $\mathbf{B}$  no pueda ser factorizada de manera única, nos puede llevar a encontrar distintas biplots, por tal motivo, una factorización muy útil es la descomposición de la matriz  $\mathbf{B}$  en valores singulares :

$$\mathbf{B} = \mathbf{L}_r \mathbf{\Delta}_r \mathbf{M}'_r = \sum_{i=1}^r \delta_i \mathbf{l}_i \mathbf{m}'_i, \quad (1.8)$$

---

<sup>5</sup>Nombrado de esta manera por Gabriel (1971)



donde  $\mathbf{L}_r$  es una matriz de  $n \times r$  de rango  $r$  cuyas columnas ortogonales ( $\mathbf{L}'_r \mathbf{L}_r = \mathbf{I}_r$ ) son los vectores propios asociados a los valores propios de la matriz  $\mathbf{B}\mathbf{B}'$ ,  $\mathbf{M}_r$  es una matriz de  $m \times r$  de rango  $r$  cuyas columnas  $\mathbf{m}_i$  son los vectores propios asociados a los valores propios de la matriz  $\mathbf{B}'\mathbf{B}$ . La matriz diagonal,  $\mathbf{\Delta}_r = \text{diag}(\delta_1, \delta_2, \dots, \delta_r)$ , con  $\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$  son los valores singulares positivos de  $\mathbf{B}$ , esto es, son las raíces cuadradas positivas de los valores propios distintos de cero de la matriz  $\mathbf{B}\mathbf{B}'$  o de la matriz  $\mathbf{B}'\mathbf{B}$ . Se puede ver de 1.8 que:

$$\mathbf{B}'\mathbf{B}\mathbf{M} = \mathbf{M}\mathbf{\Delta}\mathbf{L}'\mathbf{L}\mathbf{\Delta}\mathbf{M}'\mathbf{M} = \mathbf{M}\mathbf{\Delta}^2$$

o

$$\mathbf{B}'\mathbf{B}\mathbf{m}_i = \delta_i^2 \mathbf{m}_i,$$

por lo tanto, las columnas de  $\mathbf{M}$  son los vectores propios de  $\mathbf{B}'\mathbf{B}$ . Así,

$$\mathbf{B} = \mathbf{L}(\mathbf{\Delta}\mathbf{M}') = \mathbf{G}\mathbf{H}',$$

ahora verificaremos la existencia de la factorización sugerida en 1.7 para esta elección particular de  $\mathbf{G}$  y  $\mathbf{H}$ ,

$$\mathbf{H}\mathbf{H}' = (\mathbf{\Delta}\mathbf{M}')'(\mathbf{\Delta}\mathbf{M}') = \mathbf{M}\mathbf{\Delta}'\mathbf{\Delta}\mathbf{M}' = \mathbf{M}\mathbf{\Delta}^2\mathbf{M}' = \sum_{i=1}^r \delta_i^2 \mathbf{m}_i \mathbf{m}'_i = \mathbf{B}'\mathbf{B} \quad (1.9)$$

$$\mathbf{G}\mathbf{G}' = \mathbf{L}\mathbf{L}' \quad (1.10)$$

$$\begin{aligned} &= \mathbf{B}(\mathbf{\Delta}\mathbf{M}')^{-1} \left[ \mathbf{B}(\mathbf{\Delta}\mathbf{M}')^{-1} \right]' \\ &= (\mathbf{B}\mathbf{M}\mathbf{\Delta}^{-1}) (\mathbf{B}\mathbf{M}\mathbf{\Delta}^{-1})' \\ &= \mathbf{B}\mathbf{M}\mathbf{\Delta}^{-1} (\mathbf{\Delta}^{-1})' \mathbf{M}'\mathbf{B}' \\ &= \mathbf{B}\mathbf{M}(\mathbf{\Delta}^{-1})^2 \mathbf{M}'\mathbf{B}' \\ &= \mathbf{B}(\mathbf{M}\mathbf{\Delta}^2\mathbf{M}')^{-1} \mathbf{B}' \\ &= \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}' \end{aligned} \quad (1.11)$$

Ahora la factorización 1.7 es única y nos lleva a una transformación ortogonal ya que se impuso la condición de que  $\mathbf{G}'\mathbf{G} = \mathbf{L}'\mathbf{L} = \mathbf{I}_r$ .

Por otro lado, para encontrar una aproximación de rango  $s$  ( $s < r$ ) para  $\mathbf{B}$ , utilizaremos el criterio de la norma de Frobenius que nos lleva a la aproximación:

$$\mathbf{B}_{(s)} = \sum_{i=1}^s \delta_i \mathbf{l}_i \mathbf{m}'_i,$$

donde  $\mathbf{B}_{(s)}$  es el valor de la matriz  $\mathbf{C}_{n \times m}$  de rango  $s$  que minimiza a:

$$\|\mathbf{B} - \mathbf{C}\| = \sum_{i=1}^n \sum_{j=1}^m (b_{ij} - c_{ij})^2$$

La bondad de ajuste está dada por:

$$\rho_s^{(2)} = 1 - \frac{\|\mathbf{B} - \mathbf{B}_{(s)}\|^2}{\|\mathbf{B}\|^2} = \frac{\sum_{i=1}^s \delta_i^2}{\sum_{i=1}^r \delta_i^2} \quad (1.12)$$

Otra propiedad muy útil de  $\mathbf{B}_{(s)}$  es la siguiente:

**Lema 1.1** *Si la suma de los renglones de  $\mathbf{B}$  es cero, es decir,  $\mathbf{B}'\mathbf{1}_n = \mathbf{0}$ , entonces, en términos de la norma de Frobenius,  $\mathbf{B}_{(s)}$  es la matriz de rango  $s$  cuyas diferencias entre columnas son las que mejor aproximan a las diferencias entre columnas de  $\mathbf{B}$ .*

**Demostración.** *Sea  $\mathbf{C}$  una matriz de  $n \times m$  tal que  $\mathbf{C}'\mathbf{1}_n = \mathbf{0}$ , y sean  $\mathbf{b}_j$  y  $\mathbf{c}_j$  las columnas de  $\mathbf{B}$  y  $\mathbf{C}$ . Si  $\mathbf{u}_j = \mathbf{b}_j - \mathbf{c}_j$ , entonces*

$$\begin{aligned} \sum_{j=1}^m \sum_{k=1}^m \|\mathbf{b}_j - \mathbf{b}_k - (\mathbf{c}_j - \mathbf{c}_k)\|^2 &= \sum_{j=1}^m \sum_{k=1}^m \|\mathbf{u}_j - \mathbf{u}_k\|^2 = \sum_{j=1}^m \sum_{k=1}^m (\mathbf{u}_j^2 - 2\mathbf{u}_j' \mathbf{u}_k + \mathbf{u}_k^2) \\ &= m \sum_{j=1}^m \mathbf{u}_j^2 - 2 \sum_{j=1}^m \sum_{k=1}^m \mathbf{u}_j' \mathbf{u}_k + m \sum_{k=1}^m \mathbf{u}_k^2 \\ &= 2m \sum_{j=1}^m \mathbf{u}_j^2 - 2 \sum_{j=1}^m \sum_{k=1}^m \mathbf{u}_j' \mathbf{u}_k \\ &= 2m \sum_{j=1}^m \mathbf{u}_j^2 - 2m \sum_{j=1}^m \mathbf{u}_j' \left( \frac{1}{m} \sum_{k=1}^m \mathbf{u}_k \right) \\ &= 2m \left[ \sum_{j=1}^m \mathbf{u}_j^2 - \bar{\mathbf{u}}' \sum_{j=1}^m \mathbf{u}_j \right] \\ &= 2m \left[ \sum_{j=1}^m \mathbf{u}_j^2 - \bar{\mathbf{u}}' \sum_{j=1}^m \mathbf{u}_j - m\bar{\mathbf{u}}^2 + m\bar{\mathbf{u}}^2 \right] \\ &= 2m \left[ \sum_{j=1}^m \mathbf{u}_j^2 - \bar{\mathbf{u}}' \sum_{j=1}^m \mathbf{u}_j - \bar{\mathbf{u}}' \sum_{j=1}^m \mathbf{u}_j + m\bar{\mathbf{u}}^2 \right] \\ &= 2m \left[ \sum_{j=1}^m \mathbf{u}_j^2 - 2\bar{\mathbf{u}}' \sum_{j=1}^m \mathbf{u}_j + m\bar{\mathbf{u}}^2 \right] \\ &= 2m \sum_{j=1}^m (\mathbf{u}_j^2 - 2\bar{\mathbf{u}}' \mathbf{u}_j + \bar{\mathbf{u}}^2) \\ &= 2m \sum_{j=1}^m \|\mathbf{u}_j - \bar{\mathbf{u}}\|^2 \end{aligned}$$

$$\therefore \sum_{j=1}^m \sum_{k=1}^m \|\mathbf{b}_j - \mathbf{b}_k - (\mathbf{c}_j - \mathbf{c}_k)\|^2 = 2m \sum_{j=1}^m \|\mathbf{u}_j\|^2$$

ya que  $\mathbf{B}'\mathbf{1}_n = \mathbf{C}'\mathbf{1}_n = \mathbf{0}$  implica que  $\bar{\mathbf{u}} = \mathbf{0}$ . Ahora

$$\sum_{j=1}^m \|\mathbf{u}_j\|^2 = \sum_{i=1}^n \sum_{j=1}^m (b_{ij} - c_{ij})^2 = \|\mathbf{B} - \mathbf{C}\|^2$$

el cual es minimizado cuando  $\mathbf{C} = \mathbf{B}_{(s)}$ . ■

Gabriel aplicó toda la teoría anterior para la matriz de datos  $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_n)'$ , donde  $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$  y  $m = r = p$ . Entonces  $\tilde{\mathbf{X}}'\mathbf{1}_n = \mathbf{0}$  y

$$\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \sum_{i=1}^n \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' = \mathbf{Q} = (n-1)\mathbf{S},$$

Si  $\delta_1^2 \geq \delta_2^2 \geq \dots \geq \delta_p^2$  son los valores propios de  $\mathbf{Q}$ , entonces de 1.10 y 1.9,  $\mathbf{G} = (\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_p)$  y  $\mathbf{H} = (\delta_1 \mathbf{m}_1, \delta_2 \mathbf{m}_2, \dots, \delta_p \mathbf{m}_p)$ , donde  $\mathbf{G}$  es una matriz de  $n \times p$  y  $\mathbf{H}$  es una matriz de  $p \times p$ . Asimismo, de 1.9 y 1.11,  $\mathbf{H}\mathbf{H}' = \mathbf{Q}$  y  $\mathbf{G}\mathbf{G}' = \frac{\tilde{\mathbf{X}}\mathbf{S}^{-1}\tilde{\mathbf{X}}}{n-1}$ . Para quitar  $(n-1)$  de  $\mathbf{G}\mathbf{G}'$ , sean

$$\mathbf{G} = (n-1)^{\frac{1}{2}} (\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_p) \quad \text{y} \quad \mathbf{H} = (n-1)^{-\frac{1}{2}} (\delta_1 \mathbf{m}_1, \delta_2 \mathbf{m}_2, \dots, \delta_p \mathbf{m}_p)$$

Entonces tenemos que:

$$\mathbf{H}\mathbf{H}' = \mathbf{S} \quad \text{y} \quad \mathbf{G}\mathbf{G}' = \tilde{\mathbf{X}}\mathbf{S}^{-1}\tilde{\mathbf{X}}' \quad (1.13)$$

como y los renglones  $\mathbf{g}'_i$  y  $\mathbf{h}'_j$  de  $\mathbf{G}$  y  $\mathbf{H}$ , respectivamente, tienen las siguientes propiedades:

1. De 1.13,  $\mathbf{g}'_\alpha \mathbf{g}_\beta = \tilde{\mathbf{x}}'_\alpha \mathbf{S}^{-1} \tilde{\mathbf{x}}_\beta$ , y la distancia entre los puntos representados por  $\mathbf{g}_\alpha$  y  $\mathbf{g}_\beta$ , está dada por:

$$\begin{aligned} \|\mathbf{g}_\alpha - \mathbf{g}_\beta\| &= \sqrt{\mathbf{g}'_\alpha \mathbf{g}_\alpha + \mathbf{g}'_\beta \mathbf{g}_\beta - 2\mathbf{g}'_\alpha \mathbf{g}_\beta} \\ &= \sqrt{(\tilde{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\beta)' \mathbf{S}^{-1} (\tilde{\mathbf{x}}_\alpha - \tilde{\mathbf{x}}_\beta)} \\ &= \sqrt{[\mathbf{x}_\alpha - \bar{\mathbf{x}} - (\mathbf{x}_\beta - \bar{\mathbf{x}})]' \mathbf{S}^{-1} [\mathbf{x}_\alpha - \bar{\mathbf{x}} - (\mathbf{x}_\beta - \bar{\mathbf{x}})]} \\ &= \sqrt{(\mathbf{x}_\alpha - \mathbf{x}_\beta)' \mathbf{S}^{-1} (\mathbf{x}_\alpha - \mathbf{x}_\beta)} \end{aligned}$$

que es la distancia de Mahalanobis entre las observaciones  $\mathbf{x}_\alpha$  y  $\mathbf{x}_\beta$ .

2. La covarianza muestral entre las variables  $j$  y  $k$  es

$$s_{jk} = \mathbf{h}'_j \mathbf{h}_k,$$

y la varianza muestral es  $s_{jj} = \|\mathbf{h}_j\|^2$ .

3. La correlación muestral entre las variables  $j$  y  $k$  es el coseno del ángulo entre  $\mathbf{h}_j$  y  $\mathbf{h}_k$ :

$$r_{jk} = \frac{\mathbf{h}'_j \mathbf{h}_k}{\|\mathbf{h}_j\| \|\mathbf{h}_k\|}$$

4. Note que:

$$\begin{aligned} \|\mathbf{h}_j - \mathbf{h}_k\|^2 &= \mathbf{h}'_j \mathbf{h}_j - 2\mathbf{h}'_j \mathbf{h}_k + \mathbf{h}'_k \mathbf{h}_k \\ &= s_{jj} - 2s_{jk} + s_{kk} \end{aligned}$$

es la varianza muestral de la diferencia entre las variables  $j$  y  $k$ .

Si aproximamos  $\mathbf{B}$  por  $\mathbf{B}_{(2)}$ , entonces podemos usar la correspondiente gráfica biplot de  $n+p$  vectores  $\mathbf{g}_i$  y  $\mathbf{h}_j$  para visualizar información aproximada de las distancias entre observaciones y sobre varianzas, covarianzas y correlaciones muestrales entre las variables. La aproximación  $\mathbf{B}_{(2)}$  es satisfactoria si 1.12 es cercana a uno.

En conclusión, al graficar las filas de la matriz  $\mathbf{G}_{(2)}$  equivale a proyectar las observaciones sobre el plano de las dos componentes principales, es decir, graficar los puntajes corregidos por la media. Lo anterior es posible demostrarlo utilizando la relación entre las componentes principales y los vectores propios de la matriz  $\mathbf{Q} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}'$ . Si  $\mathbf{h}_i$  es un vector propio de  $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$  con valor propio  $\delta_i^2$ ,

$$\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\mathbf{h}_i = \delta_i^2 \mathbf{h}_i \quad (1.14)$$

y multiplicando por  $\tilde{\mathbf{X}}$  ambos miembros de 1.14, tenemos

$$\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\mathbf{h}_i = \delta_i^2 \tilde{\mathbf{X}}\mathbf{h}_i$$

es decir,  $\mathbf{y}_i = \tilde{\mathbf{X}}\mathbf{h}_i$  es un vector propio de la matriz  $\tilde{\mathbf{X}}\tilde{\mathbf{X}}'$  con el mismo valor propio  $\delta_i^2$ . Así, tenemos que las columnas de la matriz  $\mathbf{G}_{(2)}$  son los puntajes de las 2 primeras componentes principales corregidas por la media.

Por otro lado, se puede ver que dada  $\mathbf{S}$ , es posible encontrar los dos valores propios más grandes  $\lambda_1$  y  $\lambda_2$ , donde  $\lambda_i = \frac{\delta_i^2}{n-1}$ , con sus correspondientes vectores propios  $\mathbf{a}_1 = (n-1)^{-\frac{1}{2}} \delta_1 \mathbf{m}_1$  y  $\mathbf{a}_2 = (n-1)^{-\frac{1}{2}} \delta_2 \mathbf{m}_2$  se puede formar la matriz

$$\mathbf{H}_{(2)} = \left( \sqrt{\lambda_1} \mathbf{a}_1, \sqrt{\lambda_2} \mathbf{a}_2 \right) = (\mathbf{c}_1, \mathbf{c}_2)$$

donde los  $p$  renglones  $\mathbf{h}_j$  de  $\mathbf{H}_{(2)}$ , tienen aproximadamente las propiedades 2 - 4 anteriores. Las 2 columnas de la matriz son graficadas como vectores en el biplot. Corsten y Gabriel llamaron a la gráfica de los vectores  $\mathbf{h}_j$  la gráfica *h-plot* de  $\mathbf{S}$  y demostraron cómo las matrices de varianzas y covarianzas muestrales pueden ser comparadas a través de su gráficas *h-plot* y propusieron como medida de bondad de ajuste a

$$\rho_2 = \frac{\lambda_1^2 + \lambda_2^2}{\sum_{j=1}^p \lambda_j^2}$$

Cabe destacar que el análisis de correspondencias se basa en la teoría que da origen a la gráfica biplot.

Por todo lo anterior, resulta de gran utilidad graficar los vectores cuyas entradas son  $(c_{i1}, c_{i2})$ ,  $i = 1, \dots, p$ , esto es, graficar los vectores  $[\rho_{X_i, Y_1}, \rho_{X_i, Y_2}]$ , para encontrar grupos de variables, si es que éstos existen.

## 1.9 Ejemplo

Con la finalidad de ilustrar el análisis de componentes principales utilizaremos las dimensiones de esqueletos de zancudos de agua. Para ello, se hará uso del paquete estadístico S-PLUS.

Los zancudos de agua son insectos pequeños que viven sobre la superficie de lagos y estanques, como se puede observar en la figura 1.9.1.



Fig. 1.9.1 Zancudo de agua.

Estos insectos crecen a lo largo de seis etapas. Durante la transición de una etapa a otra, ellos mudan su piel, la cual es al mismo tiempo su esqueleto. La piel desprendida puede ser medida con mucha precisión bajo un microscopio; y el mismo animal puede ser seguido a lo largo de las seis etapas de crecimiento hasta que alcanza la edad adulta.

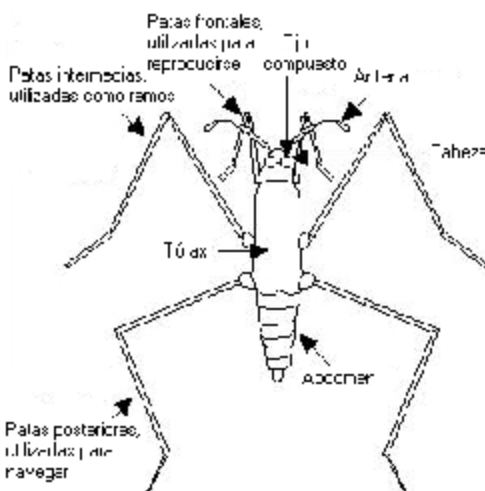


Fig. 1.9.2 Anatomía del zancudo de agua.

En la siguiente tabla se muestran datos para  $n = 88$  zancudos hembra de la especie *Limnopygia canaliculatus* que fueron obtenidos por el Dr. C.P. Klingenberg en la Universidad de Alberta, Edmonton, Canadá. Las seis variables medidas fueron el promedio de la longitud del fémur ( $F_1, F_2, F_3$ ) y el promedio la longitud de la tibia ( $T_1, T_2, T_3$ ) de las patas posteriores para las primeras tres etapas. Todas las medidas están en  $mm/1000$ .

	$F_1$	$T_1$	$F_2$	$T_2$	$F_3$	$T_3$		$F_1$	$T_1$	$F_2$	$T_2$	$F_3$	$T_3$
1	577	440	943	609	1488	919	6	568	438	929	616	1508	943
2	579	440	907	595	1547	949	7	584	430	930	622	1512	946
3	578	435	912	585	1499	883	8	582	440	917	600	1522	930
4	577	443	909	609	1505	917	9	579	448	958	637	1558	958
5	578	434	936	629	1517	943	10	565	427	902	597	1548	958

	$F_1$	$T_1$	$F_2$	$T_2$	$F_3$	$T_3$		$F_1$	$T_1$	$F_2$	$T_2$	$F_3$	$T_3$
11	570	425	894	572	1494	908	50	580	425	906	597	1467	900
12	595	431	927	613	1542	935	51	537	391	873	562	1404	838
13	501	391	806	542	1349	832	52	543	422	853	573	1408	891
14	588	449	965	630	1593	967	53	569	421	884	576	1434	865
15	572	423	915	580	1508	902	54	591	431	967	608	1567	933
16	580	424	922	590	1548	918	55	566	422	912	582	1491	903
17	561	414	950	613	1545	936	56	551	400	878	567	1476	882
18	570	416	923	602	1511	919	57	588	422	942	595	1526	907
19	565	436	906	612	1488	911	58	569	419	904	578	1491	893
20	576	426	910	582	1465	859	59	597	439	960	619	1571	958
21	558	420	920	593	1531	904	60	548	414	899	574	1501	882
22	549	427	883	600	1440	899	61	579	430	927	610	1520	958
23	562	423	921	615	1551	971	62	537	396	884	583	1468	878
24	549	410	882	577	1474	899	63	589	450	921	617	1517	933
25	569	411	917	573	1502	889	64	559	428	905	606	1503	958
26	567	413	903	572	1520	898	65	574	423	926	576	1523	894
27	594	462	996	666	1631	1014	66	576	422	910	601	1468	910
28	555	417	893	576	1478	888	67	566	426	931	611	1521	942
29	564	436	907	607	1492	929	68	582	438	949	618	1560	930
30	597	441	952	603	1490	914	69	566	420	900	592	1498	923
31	587	452	947	631	1512	917	70	552	412	894	575	1501	913
32	570	413	897	591	1485	887	71	567	424	907	593	1498	917
33	581	420	953	609	1552	951	72	585	435	958	612	1539	935
34	574	419	920	585	1493	897	73	568	423	902	579	1462	874
35	597	444	949	626	1554	936	74	592	430	931	605	1507	901
36	593	447	950	619	1519	937	75	582	427	930	617	1528	948
37	546	406	902	586	1470	867	76	569	421	902	585	1467	892
38	565	422	929	598	1484	897	77	572	419	917	594	1475	896
39	571	428	940	618	1505	923	78	551	418	896	591	1451	908
40	584	443	923	611	1516	936	79	586	443	931	613	1487	898
41	571	423	922	613	1485	928	80	571	410	907	591	1464	902
42	573	424	913	602	1506	944	81	550	402	894	563	1461	867
43	527	401	880	574	1424	886	82	567	429	923	619	1493	963
44	571	431	933	612	1577	976	83	570	414	905	583	1503	902
45	522	396	894	570	1468	867	84	566	425	927	613	1499	935
46	576	436	916	602	1497	916	85	583	437	939	607	1547	948
47	564	402	902	583	1451	852	86	575	426	925	602	1500	903
48	580	431	939	599	1551	925	87	577	413	901	583	1499	888
49	562	412	911	589	1484	883	88	585	448	949	616	1568	957

Como es usual en estudios de crecimiento, usamos la transformación logarítmica de variables, por lo que el análisis fue hecho con las variables  $x_1 = 100 \ln F_1$ ,  $x_2 = 100 \ln T_1$ ,  $x_3 = 100 \ln F_2$ ,  $x_4 = 100 \ln T_2$ ,  $x_5 = 100 \ln F_3$  y  $x_6 = 100 \ln T_3$ . A continuación la tabla con los valores de las variables transformadas.

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
1	635.784	608.677	684.907	641.182	730.519	682.329
2	636.130	608.677	681.014	638.856	734.407	685.541
3	635.957	607.535	681.564	637.161	731.255	678.333
4	635.784	609.357	681.235	641.182	731.655	682.111
5	635.957	607.304	684.162	644.413	732.449	684.907
6	634.212	608.222	683.411	642.325	731.854	684.907
7	636.990	606.379	683.518	643.294	732.119	685.224
8	636.647	608.677	682.111	639.693	732.778	683.518
9	636.130	610.479	686.485	645.677	735.116	686.485
10	633.683	605.678	680.461	639.192	734.472	686.485
11	634.564	605.209	679.571	634.914	730.921	681.124
12	638.856	606.611	683.195	641.836	734.084	684.055
13	621.661	596.871	669.208	629.527	720.712	672.383
14	637.673	610.702	687.213	644.572	737.337	687.420
15	634.914	604.737	681.892	636.303	731.854	680.461
16	636.303	604.973	682.655	638.012	734.472	682.220
17	632.972	602.587	685.646	641.836	734.278	684.162
18	634.564	603.069	682.763	640.026	732.053	682.329
19	633.683	607.764	680.904	641.673	730.519	681.454
20	635.611	605.444	681.344	636.647	728.961	675.577
21	632.436	604.025	682.437	638.519	733.368	680.683
22	630.810	605.678	678.333	639.693	727.240	680.128
23	633.150	604.737	682.546	642.162	734.666	687.833
24	630.810	601.616	678.219	635.784	729.574	680.128
25	634.388	601.859	682.111	635.089	731.455	679.010
26	634.036	602.345	680.572	634.914	732.647	680.017
27	638.688	613.556	690.375	650.129	739.695	692.166
28	631.897	603.309	679.459	635.611	729.845	678.897
29	633.505	607.764	681.014	640.853	730.787	683.411
30	639.192	608.904	685.857	640.192	730.653	681.783
31	637.502	611.368	685.330	644.731	732.119	682.111
32	634.564	602.345	679.906	638.182	730.317	678.784
33	636.475	604.025	685.961	641.182	734.730	685.751



	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
34	635.263	603.787	682.437	637.161	730.854	679.906
35	639.192	609.582	685.541	643.935	734.859	684.162
36	638.519	610.256	685.646	642.811	732.581	684.268
37	630.262	600.635	680.461	637.332	729.302	676.504
38	633.683	604.501	683.411	639.359	730.250	679.906
39	634.739	605.912	684.588	642.649	731.655	682.763
40	636.990	609.357	682.763	641.510	732.383	684.162
41	634.739	604.737	682.655	641.836	730.317	683.303
42	635.089	604.973	681.674	640.026	731.721	685.013
43	626.720	599.396	677.992	635.263	726.123	678.672
44	634.739	606.611	683.841	641.673	736.328	688.346
45	625.767	598.141	679.571	634.564	729.166	676.504
46	635.611	607.764	682.002	640.026	731.122	682.002
47	633.505	599.645	680.461	636.819	728.001	674.759
48	636.303	606.611	684.482	639.526	734.666	682.979
49	633.150	602.102	681.454	637.843	730.250	678.333
50	636.303	605.209	680.904	639.192	729.097	680.239
51	628.600	596.871	677.194	633.150	724.708	673.102
52	629.711	604.501	674.876	635.089	724.993	679.234
53	634.388	604.263	678.446	635.611	726.822	676.273
54	638.182	606.611	687.420	641.017	735.692	683.841
55	633.859	604.501	681.564	636.647	730.720	680.572
56	631.173	599.146	677.765	634.036	729.709	678.219
57	637.673	604.501	684.801	638.856	733.041	681.014
58	634.388	603.787	680.683	635.957	730.720	679.459
59	639.192	608.450	686.693	642.811	735.947	686.485
60	630.628	602.587	680.128	635.263	731.389	678.219
61	636.130	606.379	683.195	641.346	732.647	686.485
62	628.600	598.141	678.446	636.819	729.166	677.765
63	637.843	610.925	682.546	642.487	732.449	683.841
64	632.615	605.912	680.793	640.688	731.522	686.485
65	635.263	604.737	683.087	635.611	732.844	679.571
66	635.611	604.501	681.344	639.859	729.166	681.344
67	633.859	605.444	683.626	641.510	732.712	684.801
68	636.647	608.222	685.541	642.649	735.244	683.518
69	633.859	604.025	680.239	638.351	731.189	682.763
70	631.355	602.102	679.571	635.437	731.389	681.674
71	634.036	604.973	681.014	638.519	731.189	682.111
72	637.161	607.535	686.485	641.673	733.889	684.055

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
73	634.212	604.737	680.461	636.130	728.756	677.308
74	638.351	606.379	683.626	640.523	731.788	680.351
75	636.647	605.678	683.518	642.487	733.171	685.435
76	634.388	604.263	680.461	637.161	729.097	679.347
77	634.914	603.787	682.111	638.688	729.641	679.794
78	631.173	603.548	679.794	638.182	728.001	681.124
79	637.332	609.357	683.626	641.836	730.452	680.017
80	634.739	601.616	681.014	638.182	728.893	680.461
81	630.992	599.645	679.571	633.328	728.688	676.504
82	634.036	606.146	682.763	642.811	730.854	687.005
83	634.564	602.587	680.793	636.819	731.522	680.461
84	633.859	605.209	683.195	641.836	731.255	684.055
85	636.819	607.993	684.482	640.853	734.407	685.435
86	635.437	605.444	682.979	640.026	731.322	680.572
87	635.784	602.345	680.351	636.819	731.255	678.897
88	637.161	610.479	685.541	642.325	735.756	686.380

Como ya vimos anteriormente, antes de llevar a cabo el análisis de componentes principales, es necesario verificar que las variables de estudio estén correlacionadas, esto lo haremos a través de la matriz de correlaciones:

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.782 & 0.807 & 0.681 & 0.703 & 0.577 \\ 0.782 & 1.000 & 0.708 & 0.808 & 0.646 & 0.714 \\ 0.807 & 0.708 & 1.000 & 0.818 & 0.842 & 0.701 \\ 0.681 & 0.808 & 0.818 & 1.000 & 0.700 & 0.832 \\ 0.703 & 0.646 & 0.842 & 0.700 & 1.000 & 0.809 \\ 0.577 & 0.714 & 0.701 & 0.832 & 0.809 & 1.000 \end{pmatrix}$$

Así, todas las correlaciones son mayores que 0.570, por lo que es factible hacer el análisis de componentes principales a estos datos.

Haremos el análisis utilizando la matriz de varianzas y covarianzas ya que los datos están medidos en las mismas unidades y no existen muchas diferencias en sus valores por lo que no habrán diferencias importantes en las varianzas; así, tenemos el vector de medias:

$$\bar{\mathbf{x}}' = (634.515, 605.194, 682.102, 639.299, 731.473, 681.790)$$

y la matriz de varianzas y covarianzas muestrales:

$$\mathbf{S} = \begin{pmatrix} 9.247 & 7.972 & 7.274 & 6.894 & 6.110 & 6.252 \\ 7.972 & 11.248 & 7.032 & 9.016 & 6.190 & 8.521 \\ 7.274 & 7.032 & 8.782 & 8.071 & 7.128 & 7.400 \\ 6.894 & 9.016 & 8.071 & 11.082 & 6.663 & 9.862 \\ 6.110 & 6.190 & 7.128 & 6.663 & 8.171 & 8.232 \\ 6.252 & 8.521 & 7.400 & 9.862 & 8.232 & 12.680 \end{pmatrix}$$

Para encontrar las componentes principales debemos tomar los valores y vectores propios de  $\mathbf{S}$ , los valores propios son:

$$\lambda_1 = 48.1025, \quad \lambda_2 = 5.2481, \quad \lambda_3 = 3.9874, \quad \lambda_4 = 2.2613, \quad \lambda_5 = 1.1446 \quad \text{y} \quad \lambda_6 = 0.4654$$

y los correspondientes vectores propios son:

$$\begin{aligned} \mathbf{a}'_1 &= (0.3683, 0.4273, 0.3863, 0.4429, 0.3593, 0.4553) \\ \mathbf{a}'_2 &= (-0.6281, -0.3023, -0.1775, 0.1679, 0.1564, 0.6557) \\ \mathbf{a}'_3 &= (-0.1646, 0.6030, -0.4171, 0.3202, -0.5735, 0.0622) \\ \mathbf{a}'_4 &= (-0.1995, -0.3386, 0.4718, 0.6548, -0.3102, -0.3132) \\ \mathbf{a}'_5 &= (-0.6247, 0.4847, 0.2578, -0.1145, 0.3974, -0.3706) \\ \mathbf{a}'_6 &= (-0.1126, 0.1131, 0.5969, -0.4810, -0.5133, 0.3514) \end{aligned}$$

por lo que la matriz de vectores propios queda de la siguiente manera:

$$\mathbf{A} = \begin{pmatrix} 0.3683 & -0.6281 & -0.1646 & -0.1995 & -0.6247 & -0.1126 \\ 0.4273 & -0.3023 & 0.6030 & -0.3386 & 0.4847 & 0.1131 \\ 0.3863 & -0.1775 & -0.4171 & 0.4718 & 0.2578 & 0.5969 \\ 0.4429 & 0.1679 & 0.3202 & 0.6548 & -0.1145 & -0.4810 \\ 0.3593 & 0.1564 & -0.5735 & -0.3102 & 0.3974 & -0.5133 \\ 0.4553 & 0.6557 & 0.0622 & -0.3132 & -0.3706 & 0.3514 \end{pmatrix}$$

entonces las componentes principales serán:

$$\mathbf{y} = \mathbf{A}'\mathbf{x}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} 0.3683 & 0.4273 & 0.3863 & 0.4429 & 0.3593 & 0.4553 \\ -0.6281 & -0.3023 & -0.1775 & 0.1679 & 0.1564 & 0.6557 \\ -0.1646 & 0.603 & -0.4171 & 0.3202 & -0.5735 & 0.0622 \\ -0.1995 & -0.3386 & 0.4718 & 0.6548 & -0.3102 & -0.3132 \\ -0.6247 & 0.4847 & 0.2578 & -0.1145 & 0.3974 & -0.3706 \\ -0.1126 & 0.1131 & 0.5969 & -0.481 & -0.5133 & 0.3514 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

Se puede observar en los vectores propios que todos los elementos del vector  $\mathbf{a}_1$  son positivos por lo que podemos considerar a la primera componente principal como un promedio ponderado de las variables, así aquellos insectos que tengan medidas mucho mayores en las patas posteriores, tendrán puntajes más altos que el resto. Los elementos del vector propio  $\mathbf{a}_2$  nos indican que aquellos insectos con valores altos en las variables  $x_1$ ,  $x_2$  y  $x_3$  tendrán puntajes altos negativos en la segunda componente principal, se puede decir que aquellos insectos que eran más grandes en la primera etapa tendrán puntajes negativos altos en la segunda componente; mientras que aquellos que eran más pequeños en la primera etapa tendrán puntajes altos positivos en la segunda componente. Asimismo, aquellos insectos que tengan valores altos en las variables  $x_4$ ,  $x_5$  y  $x_6$  tendrán valores altos positivos en la segunda componente, esto es, se habrán desarrollado más que el resto en la tercera etapa.

Para escoger cuántas componentes retener, realizaremos la gráfica scree:

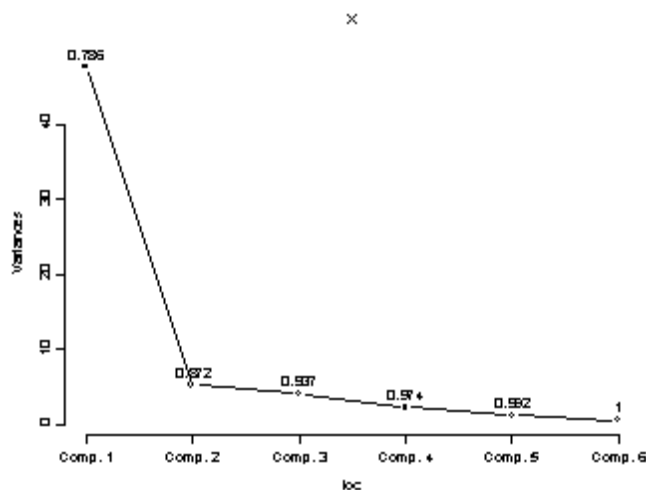


Fig. 1.9.3 Grafica scree.

y tomaremos la proporción de varianza que aportan:

$$\sum_{i=1}^6 \lambda_i = 48.1025 + 5.2481 + 3.9874 + 2.2613 + 1.1446 + 0.4654 = 61.2093$$

$\lambda_1$  aporta el:  $\frac{48.1025}{61.2093} \times 100\% = 78.5869\%$  de la variación total en los datos, mientras que  $\lambda_1$  y  $\lambda_2$  aportan:  $\frac{48.1025+5.2481}{61.2093} \times 100\% = 87.1609\%$  de la varianza total. Podemos observar a través de ambos métodos que la primera componente principal aporta cerca del 79% de proporción de la varianza total y que las dos primeras componentes aportan el 87% de la varianza total, que es bastante alta y suficiente para realizar el análisis con las dos primeras componentes.

Así tenemos que los puntajes correspondientes para cada uno de los 88 zancudos:

	<i>Comp. 1</i>	<i>Comp. 2</i>	<i>Comp. 3</i>	<i>Comp. 4</i>	<i>Comp. 5</i>	<i>Comp. 6</i>
1	3.776	-1.828	-1.906	-1.250	-0.824	1.698
2	4.230	0.969	-0.697	4.390	-0.226	-0.413
3	-1.275	-4.178	-0.624	1.584	-1.534	-0.294
4	2.957	-1.347	-3.182	0.996	-0.739	-1.077
5	6.264	1.145	-1.448	-2.039	0.700	-0.561
6	4.584	1.653	-2.274	-0.540	-0.644	0.602
7	5.530	0.859	-0.839	-1.113	2.080	-0.346
8	3.708	-0.990	-1.231	2.289	-0.192	-0.093
9	10.818	1.328	-1.339	-1.532	-1.661	-0.256
10	2.435	4.197	0.349	3.243	0.204	-0.669
11	-3.396	-0.845	0.072	3.701	0.146	0.643
12	5.720	-1.030	0.860	0.688	1.836	-1.441
13	-25.748	3.387	-4.932	0.815	-0.999	-0.272
14	12.497	0.938	0.654	0.213	-1.656	-0.250
15	-1.924	-1.390	1.514	1.688	-0.462	0.556
16	1.482	-0.620	2.763	1.929	-0.098	-0.667
17	2.898	3.548	3.445	-2.911	-0.559	0.167
18	0.141	1.060	1.631	-1.149	0.943	-0.310
19	0.885	-0.013	-3.473	-0.686	-0.930	-1.101
20	-4.688	-5.542	-0.491	-0.327	-0.849	-0.166
21	-1.304	1.039	1.907	-0.217	-2.071	-0.685
22	-4.716	1.164	-4.925	-0.888	-0.466	-0.379
23	4.641	5.858	0.774	0.372	0.553	-0.526
24	-7.389	2.121	0.068	1.073	0.158	-0.224
25	-4.605	-1.447	3.504	0.723	0.030	0.699
26	-4.312	-0.282	3.189	2.342	-0.149	-0.298

	<i>Comp. 1</i>	<i>Comp. 2</i>	<i>Comp. 3</i>	<i>Comp. 4</i>	<i>Comp. 5</i>	<i>Comp. 6</i>
27	20.781	3.290	-0.303	-1.530	-1.762	-0.370
28	-6.326	-0.087	0.030	1.090	-0.887	0.096
29	1.487	1.266	-3.162	0.460	-0.545	-0.071
30	4.857	-4.709	-0.657	-0.423	0.580	2.122
31	7.770	-3.093	-3.274	-2.092	-1.475	-0.543
32	-4.326	-1.119	0.692	-0.487	1.195	-1.565
33	5.522	1.860	3.655	-0.807	1.185	0.765
34	-2.223	-1.795	1.558	0.133	0.366	0.640
35	9.276	-2.012	-0.133	-0.446	-0.028	-1.113
36	8.089	-2.287	-1.558	-0.339	0.014	0.850
37	-8.206	0.204	1.078	-2.659	-1.345	-0.813
38	-1.368	-0.917	0.223	-2.027	-0.727	0.733
39	3.342	0.429	-0.389	-2.717	-0.177	0.177
40	5.332	-0.862	-2.161	1.169	0.128	-0.111
41	1.499	1.137	-1.027	-1.917	1.529	0.157
42	1.830	2.056	-0.242	0.852	1.754	0.309
43	-12.065	3.819	-1.084	-1.573	-0.491	1.360
44	7.141	4.579	1.524	1.709	-0.223	-0.158
45	-12.546	3.454	2.278	-2.209	-2.978	0.280
46	1.756	-1.242	-1.859	0.618	-0.234	0.012
47	-8.924	-2.967	1.736	-2.961	0.972	-0.990
48	3.973	-0.657	2.116	0.928	-0.985	0.049
49	-4.732	-0.796	1.349	-1.522	-0.149	-0.470
50	-1.404	-2.321	-1.446	-0.225	1.776	-0.189
51	-16.740	-0.685	0.627	-2.475	0.370	0.171
52	-10.213	1.113	-5.596	2.162	0.344	0.602
53	-7.673	-3.954	-2.128	0.629	0.696	-0.052
54	7.221	-1.383	3.709	-0.472	-0.487	0.649
55	-2.744	-0.645	0.579	1.010	-0.390	0.908
56	-10.080	1.196	2.183	1.113	0.737	-0.715
57	1.924	-2.591	3.152	-0.344	0.652	0.312
58	-4.007	-1.451	1.019	1.393	0.021	0.182
59	10.188	-0.369	1.871	0.428	0.523	0.246
60	-6.751	0.548	1.575	0.771	-2.407	-0.306
61	4.990	2.039	-0.267	0.702	1.661	0.667
62	-10.364	3.079	1.475	-2.195	-0.193	-1.352
63	6.543	-1.869	-3.312	1.252	-0.077	-0.776
64	1.873	4.528	-2.001	1.057	0.682	0.470
65	-1.690	-2.367	2.915	1.675	-1.355	0.742

	<i>Comp. 1</i>	<i>Comp. 2</i>	<i>Comp. 3</i>	<i>Comp. 4</i>	<i>Comp. 5</i>	<i>Comp. 6</i>
66	-0.968	-0.903	-1.193	-0.881	2.032	0.103
67	3.250	2.604	0.192	-0.885	-0.047	0.370
68	7.034	-0.580	0.942	-0.654	-1.497	-0.785
69	-1.539	1.530	-0.100	1.190	1.002	-0.227
70	-5.256	2.631	1.484	1.983	-0.275	0.354
71	-0.992	0.595	-0.333	0.866	0.231	0.013
72	6.619	-0.886	1.336	-0.843	-0.460	0.996
73	-5.360	-3.276	-0.724	0.388	-0.489	0.346
74	2.508	-3.727	0.431	-0.707	0.911	-0.645
75	5.222	1.454	0.376	-0.497	1.773	-0.465
76	-3.990	-1.680	-0.670	0.332	0.589	0.318
77	-2.288	-1.525	0.187	-1.194	0.847	0.333
78	-4.870	1.839	-2.113	-0.689	0.311	0.898
79	3.355	-4.195	-2.699	-1.281	-0.612	-0.256
80	-3.893	-0.328	0.701	-1.138	2.560	0.315
81	-10.698	-0.565	2.353	0.003	-0.394	0.702
82	4.193	3.808	-2.181	-0.943	1.650	1.017
83	-3.287	-0.290	1.939	0.968	0.8356	-0.381
84	2.264	2.090	-0.739	-1.661	0.518	0.415
85	6.367	0.394	0.642	1.319	-0.169	0.504
86	0.499	-1.511	0.123	-1.049	-0.079	-0.252
87	-3.920	-1.972	2.046	0.766	1.356	-1.223
88	9.531	0.317	-0.116	1.480	-1.450	0.312

Por otro lado, tenemos los vectores de carga de componentes:

	<i>Comp. 1</i>	<i>Comp. 2</i>	<i>Comp. 3</i>	<i>Comp. 4</i>	<i>Comp. 5</i>	<i>Comp. 6</i>
$x_1$	0.368	-0.628	0.165	0.200	0.625	-0.113
$x_2$	0.427	-0.302	-0.603	0.339	-0.485	0.113
$x_3$	0.386	-0.177	0.417	-0.472	-0.258	0.597
$x_4$	0.443	0.168	-0.320	-0.655	0.114	-0.481
$x_5$	0.359	0.156	0.573	0.310	-0.397	-0.513
$x_6$	0.455	0.656	0.000	0.313	0.371	0.351

y las correlaciones entre las componentes principales y las variables originales:

	<i>Comp. 1</i>	<i>Comp. 2</i>	<i>Comp. 3</i>	<i>Comp. 4</i>	<i>Comp. 5</i>	<i>Comp. 6</i>
$x_1$	0.840	-0.473	0.108	0.099	0.220	-0.025
$x_2$	0.883	-0.207	-0.359	0.152	-0.155	0.023
$x_3$	0.904	-0.137	0.281	-0.239	-0.093	0.137
$x_4$	0.923	0.116	-0.192	-0.296	0.037	-0.099
$x_5$	0.872	0.125	0.401	0.163	-0.149	-0.122
$x_6$	0.887	0.422	-0.035	0.132	0.111	0.067

Al observar las correlaciones entre las variables originales se puede ver que las variables  $x_2$  y  $x_3$  están muy correlacionadas, al igual que las variables  $x_4$  y  $x_5$ , esto nos indica que la información que está aportando cada pareja de variables es muy similar, por lo que podemos considerar a estas parejas como dos grupos de variables y por otra parte están las variables  $x_1$  y  $x_6$ ; lo anterior también es posible observarlo en la gráfica biplot que presenta la salida de S-PLUS, fig. 1.9.4, en donde se puede observar claramente cómo el ángulo que forman los vectores de las variables  $x_2$  y  $x_3$  y el ángulo formado por los vectores de las variables  $x_4$  y  $x_5$  son muy cercanos a  $0^\circ$ , por lo que el coseno de estos ángulos es muy cercano a 1 lo que nos indica una alta correlación entre cada pareja de variables, mientras que se puede notar que el ángulo que forman las variables  $x_1$  y  $x_6$  es muy cercano a  $90^\circ$ , así el coseno de este ángulo estará cercano a 0, lo que implica que no existe una asociación lineal entre estas dos variables. También, en la misma gráfica, es posible observar que la longitud de los vectores que representan a cada una de las seis variables, son muy similares, lo que nos indica que las varianzas son muy similares, solamente resaltan los vectores correspondientes a las variables  $x_3$  y  $x_5$ , los cuales tienen una longitud menor a la de los cuatro vectores restantes, es decir,  $x_3$  y  $x_5$  son las variables que tienen las varianzas más pequeñas.

En la gráfica biplot (fig. 1.9.4), se puede ver que los zancudos que fueron medidos no forman algún grupo en especial, solamente se alcanza a distinguir cómo se forma un pequeño grupo cuyos individuos tienen scores ligeramente menores a cero para ambas componentes y el resto de los individuos se encuentran esparcidos en la gráfica, asimismo se puede distinguir al individuo 13 cómo se aleja del resto, este individuo tiene la característica de que el valor en la primera componente principal es el más pequeño de todos, esto es, en las tres etapas en las que fue medido, siempre fue el más pequeño con respecto al resto. Por otra parte, se observa que el individuo 27 es el que tiene el puntaje más alto en la primera componente, es decir, este insecto, durante las tres etapas, fue el más grande con respecto a los 87 insectos restantes.





## Capítulo 2

# CONCEPTOS EN OTRAS ÁREAS RELACIONADOS CON ESTADÍSTICA ESFÉRICA

### 2.1 Introducción

En el análisis de direcciones y de eventos periódicos, el sistema coordenado debe ser cambiado frecuentemente. Algunas veces las coordenadas rectangulares son muy apropiadas, pero para los propósitos de este trabajo, las coordenadas polares y esféricas brindan herramientas más apropiadas. El cambio de un sistema coordenado a otro requiere una aplicación cuidadosa de las funciones trigonométricas.

Asimismo los vectores y algunos conceptos básicos de física, como lo son centro de masa y momento de inercia, resultan una herramienta importante, cuando se trabaja con coordenadas rectangulares o en coordenadas polares, por lo que después de dar algunos conceptos básicos de trigonometría daremos la definición de un vector y hablaremos a grandes rasgos de los conceptos de centro de masa y momento de inercia.

## 2.2 Nociones básicas de Geometría

### 2.2.1 Ángulos y algunos conceptos básicos de trigonometría

Dadas  $\ell_1$  y  $\ell_2$  dos líneas rectas no coincidentes, si rotamos alrededor del punto de intersección (el vértice) a  $\ell_1$  en dirección contraria a las manecillas del reloj hasta que coincida con  $\ell_2$ , (como se muestra en la Fig. 2.2.1) la magnitud de la rotación que tuvimos que realizar para lograrlo es llamada *ángulo*, (en la figura denotado por  $\theta$ ). Los ángulos se miden en grados (denotados por  $^\circ$ ) o en radianes (*rad*).

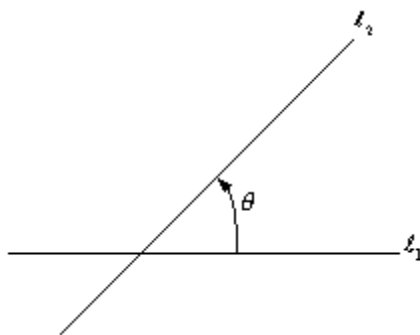


Fig. 2.2.1 Ángulo entre dos rectas.

Si  $\ell_1$  y  $\ell_2$  son dos rectas que coinciden, al rotar  $\ell_1$  hasta alcanzar a  $\ell_2$  obtendremos una circunferencia, su magnitud será de  $360^\circ$ , entonces un *grado* ( $1^\circ$ ) es  $\frac{1}{360}$  de una circunferencia y un *radián* (1 rad) será la medida de un ángulo que abarca un arco de circunferencia de longitud igual al radio por lo que habrá  $2\pi$  radianes en una circunferencia.

### Razones trigonométricas

A partir de los ángulos en un triángulo rectángulo se definen las razones trigonométricas, las que al generalizarse dan lugar a las funciones trigonométricas, por lo que es necesario, antes de hablar de las funciones trigonométricas, dar la definición de las razones trigonométricas.

Dado un triángulo rectángulo, a los dos lados que forman el ángulo recto los llamamos *catetos*, mientras que al lado opuesto al ángulo recto lo llamamos *hipotenusa* ( $H$ ). Por otro lado, si fijamos uno de los ángulos agudos, ( $\alpha$  en la figura 2.2.2) al

segmento BC que delimita a  $\alpha$  lo llamamos *cateto adyacente* (CA) y al otro cateto se le llama *cateto opuesto* (CO). Al definir las razones trigonométricas, supondremos que cateto opuesto, cateto adyacente o hipotenusa se refieren a las longitudes de dichos lados del triángulo.

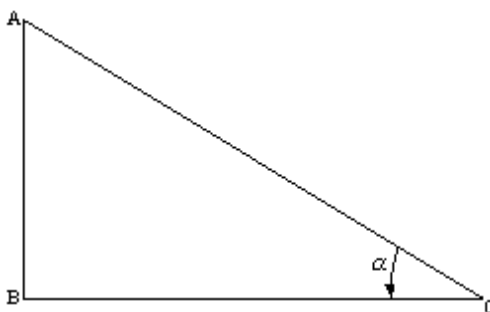


Fig. 2.2.2 Elementos de un triángulo rectángulo.

Entonces las razones trigonométricas quedan definidas de la siguiente manera:

$$\begin{aligned}\operatorname{sen} \alpha &= \frac{\text{cateto opuesto}}{\text{hipotenusa}} = \frac{BA}{CA} \\ \operatorname{cos} \alpha &= \frac{\text{cateto adyacente}}{\text{hipotenusa}} = \frac{CB}{CA} \\ \operatorname{tan} \alpha &= \frac{\text{cateto opuesto}}{\text{cateto adyacente}} = \frac{BA}{CB}\end{aligned}$$

donde  $\operatorname{sen} \alpha$ ,  $\operatorname{cos} \alpha$  y  $\operatorname{tan} \alpha$  son las abreviaturas de los términos *seno*, *coseno* y *tangente* del ángulo  $\alpha$ , respectivamente. Las razones recíprocas se definen a continuación:

$$\begin{aligned}\operatorname{csc} \alpha &= \frac{\text{hipotenusa}}{\text{cateto opuesto}} = \frac{CA}{BA} \\ \operatorname{sec} \alpha &= \frac{\text{hipotenusa}}{\text{cateto adyacente}} = \frac{CA}{CB} \\ \operatorname{cot} \alpha &= \frac{\text{cateto adyacente}}{\text{cateto opuesto}} = \frac{CB}{BA}\end{aligned}$$

## Funciones trigonométricas

Las *funciones trigonométricas* se definen a partir del *círculo trigonométrico* (fig. 2.2.3), el cual consta de una circunferencia de radio 1, con centro en el origen y de

un rayo que parte del origen y forma un ángulo  $\theta$  con la parte positiva del eje de las  $x$ . El rayo corta al círculo unitario en un punto  $P$ , cuya proyección  $Q$  en el eje  $x$  determina el triángulo rectángulo  $OQP$ .

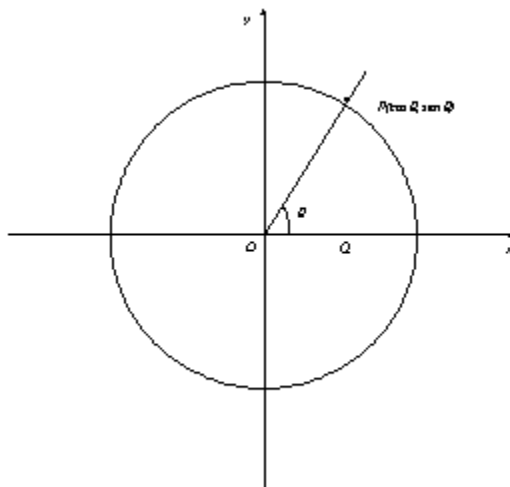


Fig. 2.2.3 Círculo trigonométrico.

La hipotenusa del triángulo  $OQP$  mide 1, puesto que es el radio del círculo, y por ello la abscisa y la ordenada de  $P$  son  $\cos \theta$  y  $\sin \theta$ , respectivamente, según es inmediato de las definiciones de las razones trigonométricas, ya que el denominador es 1. Por otra parte, se sigue del Teorema de Pitágoras que:

$$\cos^2 \theta + \sin^2 \theta = 1 \quad (2.1)$$

Dado lo anterior, ahora sí podemos definir las funciones trigonométricas.

**Definición 2.1** Si  $\theta$  es la medida en radianes de un ángulo, la función **seno de**  $\theta$  ( $\sin \theta$ ) es la ordenada del punto  $P$  en el círculo trigonométrico correspondiente al radio que forma un ángulo de medida  $\theta$  con la parte positiva del eje de las abscisas.

**Definición 2.2** Si  $\theta$  es la medida en radianes de un ángulo, la función **coseno de**  $\theta$  ( $\cos \theta$ ) es la abscisa del punto  $P$  en el círculo trigonométrico correspondiente al radio que forma un ángulo de medida  $\theta$  con la parte positiva del eje de las abscisas.

Si el dominio de  $\theta$  consiste de todos los números reales, las funciones  $\sin \theta$  y  $\cos \theta$  son *funciones periódicas* de  $\theta$ , esto es:

$$\sin(\theta + 2\pi) = \sin \theta \quad \text{y} \quad \cos(\theta + 2\pi) = \cos \theta$$

puesto que cualquier nueva rotación completa alrededor del círculo trigonométrico genera los mismos valores de  $\sen \theta$  y  $\cos \theta$ , entonces el periodo de estas funciones es de  $2\pi$ .

Las funciones seno y coseno pueden tomar valores negativos, pero la relación 2.1 implica que los valores no pueden ser menores que -1 ni mayores que 1, es decir,

$$-1 \leq \sen \theta \leq 1 \quad \text{y} \quad -1 \leq \cos \theta \leq 1.$$

**Definición 2.3** Si  $\theta$  es la medida en radianes de un ángulo, la función **tangente** de  $\theta$  es el cociente de la función seno entre la función coseno, es decir,  $\tan \theta = \frac{\sen \theta}{\cos \theta}$ .

Podemos ver que si  $\cos \theta = 0$ , la función tangente no está definida para los valores de  $\theta$  tales que  $\theta = \frac{(2k+1)\pi}{2}$ ,  $\forall k \in \mathbb{Z}$ .

La función  $\tan \theta$  es periódica de periodo  $\pi$ :

$$\tan(\theta + \pi) = \frac{\sen(\theta + \pi)}{\cos(\theta + \pi)} = \frac{\sen \theta \cos \pi + \cos \theta \sen \pi}{\cos \theta \cos \pi - \sen \theta \sen \pi}$$

ya que  $\sen(\theta + \pi) = \sen \theta \cos \pi + \sen \pi \cos \theta$  y  $\cos(\theta + \pi) = \cos \theta \cos \pi - \sen \theta \sen \pi$ , además  $\sen \pi = 0$  y  $\cos \pi = -1$ , entonces

$$\tan(\theta + \pi) = \frac{(\sen \theta)(-1) + (\cos \theta)(0)}{\cos \theta(-1) + (\sen \theta)(0)} = \frac{-\sen \theta}{-\cos \theta} = \tan \theta.$$

**Definición 2.4** Si  $\theta$  es la medida en radianes de un ángulo, la función **cosecante** de  $\theta$  es el cociente de la función constante 1 dividida entre la función seno, es decir,  $\csc \theta$  es el inverso multiplicativo de  $\sen \theta$ .

La función cosecante no está definida para los valores de  $\theta$  en los que el seno se anula, es decir, para  $\theta = k\pi$ ,  $\forall k \in \mathbb{Z}$ .

**Definición 2.5** Si  $\theta$  es la medida en radianes de un ángulo, la función **secante** de  $\theta$  es el cociente de la función constante 1 dividida entre la función coseno, es decir,  $\sec \theta$  es el inverso multiplicativo de  $\cos \theta$ .

Así, si  $\cos \theta = 0$ , la función secante no está definida para los valores de  $\theta$  tales que,  $\frac{(2k+1)\pi}{2}$ ,  $\forall k \in \mathbb{Z}$ .

**Definición 2.6** Si  $\theta$  es la medida en radianes de un ángulo, la función **cotangente** de  $\theta$  es el cociente de la función constante 1 dividida entre la función tangente, es decir,  $\cot \theta$  es el inverso multiplicativo de  $\tan \theta$ .

La función cotangente no está definida para los valores de  $\theta$  en los que el seno se anula, es decir, para  $\theta = k\pi$ ,  $\forall k \in \mathbb{Z}$ .

Las relaciones siguientes, las cuales pueden ser derivadas fácilmente de la figura 2.2.3, son muy útiles al momento de trabajar con distribuciones circulares y esféricas:

$$\begin{aligned}\operatorname{sen}(-\alpha) &= \operatorname{sen}(2\pi - \alpha) = -\operatorname{sen} \alpha \\ \operatorname{cos}(-\alpha) &= \operatorname{cos}(2\pi - \alpha) = \operatorname{cos} \alpha \\ \operatorname{sen}(\pi - \alpha) &= \operatorname{sen} \alpha \\ \operatorname{cos}(\pi - \alpha) &= -\operatorname{cos} \alpha \\ \operatorname{sen}(\pi + \alpha) &= -\operatorname{sen} \alpha \\ \operatorname{cos}(\pi + \alpha) &= -\operatorname{cos} \alpha\end{aligned}$$

Otras relaciones importantes, que pueden ser demostradas fácilmente son:

$$\begin{aligned}\operatorname{sen}(\alpha \pm \beta) &= \operatorname{sen} \alpha \operatorname{cos} \beta \pm \operatorname{cos} \alpha \operatorname{sen} \beta \\ \operatorname{cos}(\alpha \pm \beta) &= \operatorname{cos} \alpha \operatorname{cos} \beta \mp \operatorname{sen} \alpha \operatorname{sen} \beta\end{aligned}$$

### Funciones trigonométricas inversas

Ahora definiremos las funciones inversas, para hacerlo, es importante definir el intervalo en el que va a quedar el ángulo.

Si el ángulo  $\theta$  crece de  $-\frac{\pi}{2}$  a  $\frac{\pi}{2}$ , entonces la función  $\operatorname{sen} \theta$  crece de  $-1$  a  $1$ , por lo que  $\theta$  queda únicamente determinado en este intervalo. Entonces,  $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$  es llamada la *función inversa* de  $y = \operatorname{sen} \theta$ , denominada *arcoseno de  $y$* , la que escribimos de la siguiente manera:

$$\begin{aligned}\theta &= \operatorname{arcsen} y && \left( y \in [-1, 1], \quad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \right) \\ \text{o} \quad \theta &= \operatorname{sen}^{-1} y\end{aligned}$$

Si el ángulo  $\theta$  crece de  $0$  a  $\pi$ ,  $\operatorname{cos} \theta$  decrece de  $1$  a  $-1$ . Entonces la *función inversa* de  $y = \operatorname{cos} \theta$ , denominada *arcoseno de  $y$* , está dada por:

$$\begin{aligned}\theta &= \operatorname{arccos} y && (y \in [-1, 1], \quad \theta \in [0, \pi]) \\ \text{o} \quad \theta &= \operatorname{cos}^{-1} y.\end{aligned}$$

Por otro lado, la función  $y = \operatorname{tan} \theta$  está definida únicamente para los valores de  $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ . De esta manera la *función inversa* de  $y = \operatorname{tan} \theta$ , llamada *arcotangente*

de  $y$ , es:

$$\begin{aligned} \theta &= \arctan y && \left( y \in (-\infty, \infty), \quad \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \right) \\ \text{o} \quad \theta &= \tan^{-1} y \end{aligned}$$

Asimismo, la función  $y = \csc \theta$  está definida únicamente para los valores de  $\theta \in (0, \pi)$ . Así, la *función inversa* de  $y = \csc \theta$ , denominada *arcocosecante de  $y$* , es:

$$\begin{aligned} \theta &= \operatorname{arccsc} y && (y \in \{(-\infty, -1] \cup [1, \infty)\}, \quad \theta \in (0, \pi)) \\ \text{o} \quad \theta &= \csc^{-1} y \end{aligned}$$

Ahora, la función  $y = \sec \theta$  está definida únicamente para los valores de  $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ . Así, la *función inversa* de  $y = \sec \theta$ , llamada *arcosecante de  $y$* , está dada por:

$$\begin{aligned} \theta &= \operatorname{arcsec} y && \left( y \in \{(-\infty, -1] \cup [1, \infty)\}, \quad \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \right) \\ \text{o} \quad \theta &= \sec^{-1} y \end{aligned}$$

Por último, la función  $y = \cot \theta$  está definida únicamente para los valores de  $\theta \in (0, \pi)$ . por lo que la *función inversa* de  $y = \cot \theta$ , denominada *arcocotangente de  $y$* , es:

$$\begin{aligned} \theta &= \operatorname{arccot} y && (y \in (-\infty, \infty), \quad \theta \in (0, \pi)) \\ \text{o} \quad \theta &= \cot^{-1} y \end{aligned}$$

## 2.2.2 Vectores

Los vectores fueron introducidos por los físicos para estudiar conceptos como fuerza y velocidad en la cual no sólo la magnitud sino también la dirección son importantes. Hoy en día, los vectores no sólo son utilizados por los físicos, también por los matemáticos y cualquier otro profesionista que requiera utilizarlos en el análisis de direcciones.

A continuación enumeramos algunas definiciones y propiedades básicas en el estudio de los vectores.

**Definición 2.7** *Un **escalar** es un número puro, es decir, se puede ver a los escalares, desde el punto de vista de la física, como cantidades que pueden especificarse en forma completa por un número y una unidad y que, por lo tanto, tienen solamente magnitud.*



**Definición 2.8** *Un vector es un segmento rectilíneo dirigido.*

En la siguiente figura, los puntos  $P$  y  $Q$  forman el vector  $\mathbf{v}$ , el cual también lo podemos denotar como el vector  $\overrightarrow{QP}$ . El punto  $Q$  es el punto inicial del vector y  $P$  es el punto final. Su longitud se escribe como  $|\overrightarrow{QP}|$  o  $|\mathbf{v}|$ .

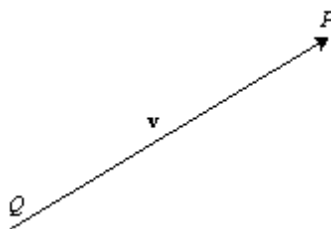


Fig. 2.2.4 Vector  $\mathbf{v}$ .

**Definición 2.9** *Dos vectores son iguales si y sólo si tienen la misma longitud y dirección.*

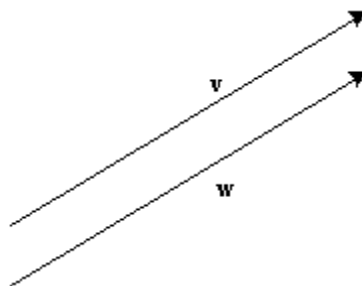


Fig. 2.2.5 Dos vectores iguales.

Así, los vectores que se muestran en la figura anterior son iguales aunque no tengan el mismo punto inicial, pero se puede observar que tienen la misma longitud y dirección.

Dos o más vectores se suman por la regla del polígono: desde el punto extremo de la flecha de un vector, se traza el segundo vector (con su longitud y dirección correctos); del extremo de éste se traza el tercer vector; y así sucesivamente, como se

indica en la figura:

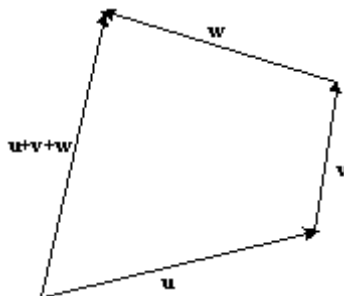


Fig. 2.2.6 Suma de los vectores  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{w}$ .

**Definición 2.10** *La suma vectorial o vector resultante es el vector que va desde el origen del primer vector hasta el extremo final del último, es decir, el vector que completa el polígono.*

Notése que éste podría ser un polígono alabeado en el espacio tridimensional, es decir, que no está en un plano.

Por otro lado, si sumamos  $k$  veces el mismo vector:  $\mathbf{v} + \mathbf{v} + \dots + \mathbf{v} = k\mathbf{v}$ , como se muestra en la figura:

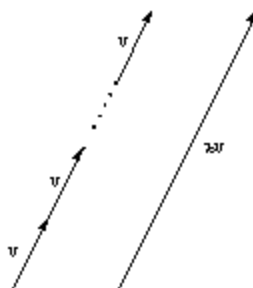


Fig. 2.2.7  $k$  veces el vector  $\mathbf{v}$ .

podemos ver que  $k\mathbf{v}$  es un vector que tiene la misma dirección, pero su longitud es  $k$  veces mayor que la longitud del vector  $\mathbf{v}$ . Sin embargo, si  $k < 0$ ,  $k\mathbf{v}$  tiene dirección contraria al de  $\mathbf{v}$  y es  $|k|$  veces mayor. Por lo anterior, si dos vectores son paralelos, uno de ellos será el producto escalar por el otro, e inversamente, si  $n\mathbf{u} = m\mathbf{v}$  ( $m$  y  $n$  escalares distintos de cero).

**Definición 2.11** El *vector unitario*  $\mathbf{i}' = \frac{\mathbf{u}}{|\mathbf{u}|}$ , es el vector de magnitud 1 en la misma dirección del vector  $\mathbf{u}$ .

De la definición anterior se puede escribir al vector  $\mathbf{u}$  de la siguiente manera:

$$\mathbf{u} = |\mathbf{u}| \mathbf{i}'$$

**Definición 2.12** La *resta vectorial* es el vector que va desde el extremo final del segundo vector hasta el extremo final del primer vector.

La definición anterior se puede ver de la siguiente manera: dados dos vectores  $\mathbf{u}$  y  $\mathbf{v}$ , el vector  $\mathbf{u} - \mathbf{v}$  se define como el vector que sumado al vector  $\mathbf{v}$  da como resultado el  $\mathbf{u}$ , como se puede ver en la figura 2.2.8.

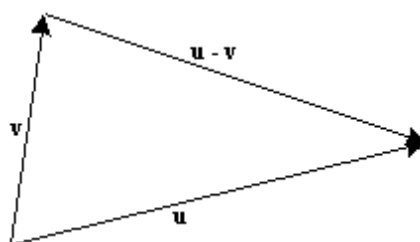


Fig. 2.2.8 Resta de los vectores  $\mathbf{u}$  y  $\mathbf{v}$ .

Las operaciones de suma y resta de vectores conducen a un vector especial importante: el *vector nulo*. En el caso de que la suma de los vectores conduzca a un polígono cerrado, el extremo del último vector coincidiendo con el origen del primer vector, se define la suma por un vector de magnitud cero, llamado *vector nulo*. A continuación daremos la definición de vector nulo.

**Definición 2.13** Se llama *vector nulo* al vector de magnitud cero, el cual se puede indicar como  $\mathbf{0}$ , el cual no tiene dirección, es indefinido.

A partir de la definición de suma vectorial y de vector nulo podemos ver que:

$$\mathbf{u} + (-\mathbf{u}) = \mathbf{0} \quad \text{y} \quad \mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$$

y si

$$m\mathbf{u} + n\mathbf{v} = \mathbf{0} \quad \implies \quad n\mathbf{v} = \mathbf{0} - m\mathbf{u} = -m\mathbf{u}$$

(el vector que sumado a  $m\mathbf{u}$  da  $\mathbf{0}$ ) por la definición de resta. Todo lo anterior justifica la transposición de un vector de un miembro de la ecuación al otro, como en álgebra.

Después de haber definido la suma y resta de vectores podemos enunciar las propiedades de la suma de vectores.

1.  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$     *ley conmutativa de la suma.*

Se sigue del hecho de que  $\mathbf{u} + \mathbf{v}$  y  $\mathbf{v} + \mathbf{u}$ , partiendo de un punto común, forman un paralelogramo; y la diagonal, por tanto, es simultáneamente el vector  $\mathbf{u} + \mathbf{v}$  y el vector  $\mathbf{v} + \mathbf{u}$ , como se puede observar en la siguiente figura.

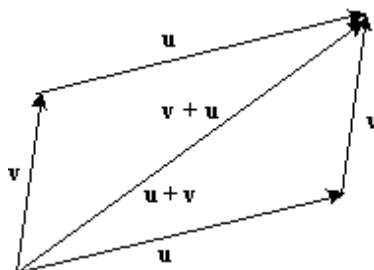


Fig. 2.2.9 Conmutativad  
de la suma.

2.  $\mathbf{u} + \mathbf{v} + \mathbf{w} = (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v} + \mathbf{w})$     *ley asociativa de la suma.*

La ley anterior se sigue del hecho de que cuando se suman vectores, el mismo vector cierra los tres polígonos (Fig. 2.2.10).

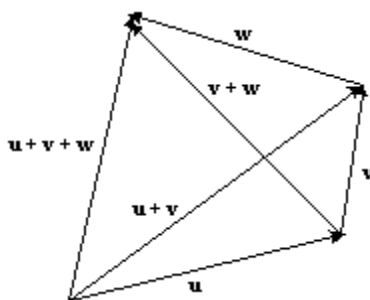


Fig. 2.2.10 Asociatividad  
de la suma.

3.  $\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \dots + \mathbf{u}_n$  *ley conmutativa generalizada.*

El método de la adición de vectores nos permite decir que este vector es el mismo, cualquiera que sea el número de vectores agrupados, es decir, podemos escribir

$$(\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{u}_3 + \mathbf{u}_4 + \mathbf{u}_5) + (\mathbf{u}_6 + \dots + \mathbf{u}_n)$$

es decir, tenemos una ley asociativa generalizada, ya que el vector necesario para cerrar este polígono siempre es el mismo. Por tanto, al colocar paréntesis abarcando pares y aplicar la ley conmutativa en forma repetida, podemos poner a los vectores  $\mathbf{u}_i$  en cualquier orden que deseemos.

4.  $m(n\mathbf{u}) = (mn)\mathbf{u}$  *ley asociativa para la multiplicación de un vector por escalares.*

Cualesquiera que sean los signos de  $m$  y  $n$ , los dos vectores  $m(n\mathbf{u})$  y  $(mn)\mathbf{u}$  deben tener la misma dirección (igual que  $\mathbf{u}$  si  $m, n$  tienen el mismo signo, dirección opuesta al de  $\mathbf{u}$  en el caso contrario). Las magnitudes son las mismas porque

$$|m||n\mathbf{u}| = |m|(|n||\mathbf{u}|) = |(mn)|\mathbf{u}$$

siendo  $m$  y  $n$  números reales.

5.  $(m+n)\mathbf{u} = m\mathbf{u} + n\mathbf{u}$  *una forma de la ley distributiva.*

Otra vez, las direcciones son las mismas. Esto es evidente si  $m$  y  $n$  tienen el mismo signo, o si  $m+n=0$ . En caso contrario, la dirección se determina por el signo del escalar que tenga un mayor valor numérico,  $|m|$  o  $|n|$ . Con respecto a las magnitudes,

$$|m+n||\mathbf{u}| = |m||\mathbf{u}| + |n||\mathbf{u}|$$

si  $m$  y  $n$  tienen el mismo signo; mientras que

$$|m+n||\mathbf{u}| = (|m| - |n|)|\mathbf{u}|$$

si  $m$  y  $n$  tienen signos contrarios, ya que éstos son números reales.

6.  $m(\mathbf{u} + \mathbf{v}) = m\mathbf{u} + m\mathbf{v}$  *ley distributiva.*

Para demostrar esta propiedad necesitamos tomar dos casos  $m \geq 0$  y  $m < 0$ :

- (a) i.  $m \geq 0$  Para demostrar que la dirección de ambos vectores es la misma, primero formemos dos triángulos: uno al sumar los vectores  $\mathbf{u}$  y  $\mathbf{v}$  y el otro al sumar los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ , como se muestra en la figura 2.2.11

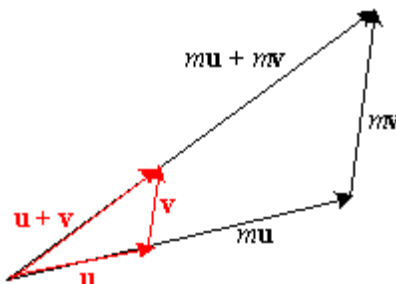


Fig. 2.2.11  $\mathbf{u} + \mathbf{v}$  y  $m\mathbf{u} + m\mathbf{v}$ .

Por la definición de vectores paralelos,  $\mathbf{u} \parallel m\mathbf{u}$  y  $\mathbf{v} \parallel m\mathbf{v}$ , por lo que podemos escribir:

$$\frac{|m\mathbf{u}|}{|\mathbf{u}|} = \frac{|m\mathbf{v}|}{|\mathbf{v}|}$$

ó

$$\frac{|m| |\mathbf{u}|}{|\mathbf{u}|} = \frac{|m| |\mathbf{v}|}{|\mathbf{v}|}$$

así que el ángulo comprendido entre el vector  $\mathbf{u}$  y el vector  $\mathbf{v}$  es igual al ángulo formado por los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ , por lo que el triángulo formado al sumar los vectores  $\mathbf{u}$  y  $\mathbf{v}$  es semejante al triángulo que se obtiene al sumar a los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ . De esta manera, al ser estos dos triángulos semejantes, se sigue que los ángulos interiores de ambos triángulos son iguales. Así, por todo lo anterior, se puede afirmar que la dirección del vector  $m\mathbf{u} + m\mathbf{v}$  es igual a la del vector  $\mathbf{u} + \mathbf{v}$  y, por ende, a la del vector que resulta al sumar al vector  $\mathbf{u} + \mathbf{v}$ ,  $m$  veces:  $m(\mathbf{u} + \mathbf{v})$ .

Con respecto a la igualdad de las magnitudes de ambos vectores, como los lados del triángulo formado por los vectores  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{u} + \mathbf{v}$  son proporcionales a los lados del triángulo formado por los vectores  $m\mathbf{u}$ ,  $m\mathbf{v}$  y  $m\mathbf{u} + m\mathbf{v}$ , es decir,

$$\frac{|m\mathbf{u}|}{|\mathbf{u}|} = \frac{|m\mathbf{v}|}{|\mathbf{v}|} = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

de otra manera,

$$\frac{|m||\mathbf{u}|}{|\mathbf{u}|} = \frac{|m||\mathbf{v}|}{|\mathbf{v}|} = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

y al ser las magnitudes de los vectores  $\mathbf{u}$  y  $\mathbf{v}$  escalares y  $m \geq 0$  podemos ver que:

$$m = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

lo que implica que

$$\begin{aligned} m|\mathbf{u} + \mathbf{v}| &= |m\mathbf{u} + m\mathbf{v}| \\ \iff |m||\mathbf{u} + \mathbf{v}| &= |m\mathbf{u} + m\mathbf{v}| \\ \therefore |m(\mathbf{u} + \mathbf{v})| &= |m\mathbf{u} + m\mathbf{v}| \end{aligned}$$

ii.  $m < 0$

Igual que para el caso anterior, para demostrar la igualdad de las direcciones de ambos vectores, formemos dos triángulos: uno al sumar los vectores  $\mathbf{u}$  y  $\mathbf{v}$  y el otro al sumar los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ , como se muestra en la figura 2.2.12

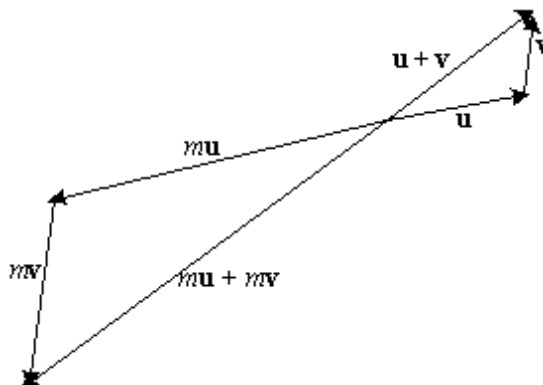


Fig. 2.2.12  $\mathbf{u} + \mathbf{v}$  y  $m\mathbf{u} + m\mathbf{v}$  cuando  $m < 0$ .

De nuevo, por la definición de vectores paralelos,  $\mathbf{u} \parallel m\mathbf{u}$  y  $\mathbf{v} \parallel m\mathbf{v}$ , entonces podemos escribir:

$$\frac{|m\mathbf{u}|}{|\mathbf{u}|} = \frac{|m\mathbf{v}|}{|\mathbf{v}|}$$

además, como el ángulo comprendido entre  $\mathbf{u}$  y  $\mathbf{v}$  es igual al ángulo formado por los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ , ya que son ángulos opuestos por el vértice, se sigue que el triángulo formado al sumar los vectores  $\mathbf{u}$  y  $\mathbf{v}$  es semejante al triángulo que se obtiene al sumar a los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$ . De nuevo, al ser estos dos triángulos semejantes se sigue que los ángulos interiores del primer triángulo son iguales a los ángulos del segundo triángulo pero, al ser la dirección de los vectores  $m\mathbf{u}$  y  $m\mathbf{v}$  opuesta a la de los vectores  $\mathbf{u}$  y  $\mathbf{v}$ , la dirección del vector  $\mathbf{u} + \mathbf{v}$  es opuesta a la del vector  $m\mathbf{u} + m\mathbf{v}$ ; ahora, como  $m < 0$ , el vector  $m(\mathbf{u} + \mathbf{v})$  tiene dirección opuesta a la del vector  $\mathbf{u} + \mathbf{v}$ , lo que implica que su dirección es la misma que la del vector  $m\mathbf{u} + m\mathbf{v}$ .

Por otra parte, para demostrar la igualdad de magnitudes, tenemos que los lados del triángulo formado por los vectores  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{u} + \mathbf{v}$  son proporcionales a los lados del triángulo formado por los vectores  $m\mathbf{u}$ ,  $m\mathbf{v}$  y  $m\mathbf{u} + m\mathbf{v}$ , por lo que tenemos que

$$\frac{|m\mathbf{u}|}{|\mathbf{u}|} = \frac{|m\mathbf{v}|}{|\mathbf{v}|} = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

de otra forma,

$$\frac{|m| |\mathbf{u}|}{|\mathbf{u}|} = \frac{|m| |\mathbf{v}|}{|\mathbf{v}|} = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

y al ser las magnitudes de los vectores  $\mathbf{u}$  y  $\mathbf{v}$  escalares y  $m < 0$  tenemos que:

$$|m| = \frac{|m\mathbf{u} + m\mathbf{v}|}{|\mathbf{u} + \mathbf{v}|}$$

entonces

$$\begin{aligned} |m| |\mathbf{u} + \mathbf{v}| &= |m\mathbf{u} + m\mathbf{v}| \\ \therefore |m(\mathbf{u} + \mathbf{v})| &= |m\mathbf{u} + m\mathbf{v}| \end{aligned}$$

Con lo que queda demostrada la propiedad 6.

Antes de llevar los vectores a un sistema coordenado vale la pena enunciar el *teorema base*.

**Teorema 2.14 (Teorema base, parte a)** *Un vector  $\mathbf{w}$  se puede expresar como una combinación lineal única de dos vectores dados no paralelos y no nulos,  $\mathbf{u}$  y  $\mathbf{v}$ , en el mismo plano:  $\mathbf{w} = m\mathbf{u} + n\mathbf{v}$ , siendo  $m$  y  $n$  coeficientes escalares únicos.*



**Demostración.** Por el extremo  $P$  de  $\mathbf{w} = \overrightarrow{OP}$ , tracemos una recta paralela a  $\mathbf{u}$ , y por el otro extremo  $O$  de  $\mathbf{w}$ , una paralela a  $\mathbf{v}$ , como se puede apreciar en la figura 2.2.13. Estas rectas se cortan en un punto  $R$ . Entonces  $\mathbf{w} = \overrightarrow{OR} + \overrightarrow{RP}$ . Como  $\overrightarrow{RP}$  es paralelo a  $\mathbf{u}$ , es un escalar multiplicado por  $\mathbf{u}$ , es decir,  $m\mathbf{u}$ ; y análogamente  $\overrightarrow{OR}$  es un escalar multiplicado por  $\mathbf{v}$ , es decir,  $n\mathbf{v}$ . Por consiguiente  $\mathbf{w}$  es una combinación lineal de  $\mathbf{u}$  y  $\mathbf{v}$ .

Ahora falta demostrar que esta combinación obtenida es única, por lo que ahora supongamos que podemos escribir a  $\mathbf{w}$  como dos combinaciones lineales distintas:  $\mathbf{w} = m_1\mathbf{u} + n_1\mathbf{v}$  y como  $\mathbf{w} = m_2\mathbf{u} + n_2\mathbf{v}$ . Esto significa que:

$$\begin{aligned} m_1\mathbf{u} + n_1\mathbf{v} &= m_2\mathbf{u} + n_2\mathbf{v} \\ \iff m_1\mathbf{u} - m_2\mathbf{u} &= n_2\mathbf{v} - n_1\mathbf{v} \\ (m_1 - m_2)\mathbf{u} &= (n_2 - n_1)\mathbf{v} \end{aligned}$$

lo que implicaría que los vectores  $\mathbf{u}$  y  $\mathbf{v}$  son paralelos lo que es una contradicción, o bien que cada miembro es el vector nulo, en tal caso  $m_1 = m_2$  y  $n_1 = n_2$ , ya que  $\mathbf{u} \neq \mathbf{0}$  y  $\mathbf{v} \neq \mathbf{0}$ , lo que demuestra que  $\mathbf{w}$  se puede expresar sólo en una forma como una combinación lineal de  $\mathbf{u}$  y  $\mathbf{v}$ . ■

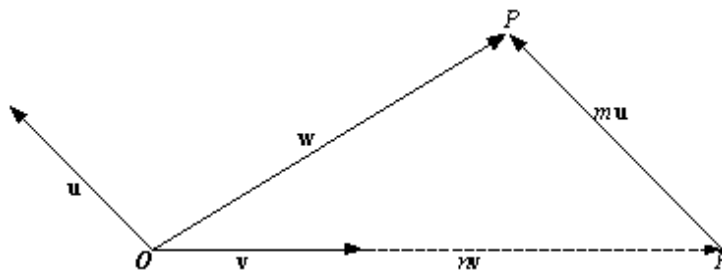


Fig. 2.2.13 Representación gráfica del teorema base parte a.

**Teorema 2.15 (Teorema base, parte b)** *En el espacio tridimensional, considérense  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{w}$  como tres vectores no nulos, no paralelos al mismo plano, sin que sean paralelos dos de ellos. Entonces cualquier vector  $\mathbf{a}$  se puede expresar como una combinación lineal única de  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{w}$ :  $\mathbf{a} = l\mathbf{u} + m\mathbf{v} + n\mathbf{w}$ .*

**Demostración.** Si los vectores parten del mismo punto  $O$ , (figura 2.2.14) una recta paralela a  $\mathbf{w}$  que pase por el extremo  $P$  de  $\mathbf{a} = \overrightarrow{OP}$  cortará al plano de  $\mathbf{u}$ ,  $\mathbf{v}$  en

el punto  $R$ . Entonces  $\overrightarrow{OR}$ , del mismo plano de  $\mathbf{u}$ ,  $\mathbf{v}$ , es una combinación lineal única de  $\mathbf{u}$ ,  $\mathbf{v}$  (por la parte a);  $\overrightarrow{OR}$ , siendo paralelo a  $\mathbf{w}$ , es un escalar multiplicado por  $\mathbf{u}$ ; y, por supuesto,  $\mathbf{a} = \overrightarrow{OR} + \overrightarrow{RP}$ . Por consiguiente,  $\mathbf{a}$  es una combinación lineal de  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{w}$ : Con respecto a la unicidad, supongamos nuevamente que:  $\mathbf{a} = l_1\mathbf{u} + m_1\mathbf{v} + n_1\mathbf{w} = l_2\mathbf{u} + m_2\mathbf{v} + n_2\mathbf{w}$ . Por lo tanto:

$$(l_1 - l_2)\mathbf{u} + (m_1 - m_2)\mathbf{v} = (n_2 - n_1)\mathbf{w}$$

si  $n_1 \neq n_2$ , se tendrá a  $\mathbf{w}$  como una combinación lineal de  $\mathbf{u}$  y  $\mathbf{v}$ , y por tanto paralelo al plano formado por  $\mathbf{u}$  y  $\mathbf{v}$ , lo que es una contradicción. En consecuencia,  $n_1 = n_2$ ; y podemos ver ahora que  $l_1 = l_2$  y  $m_1 = m_2$  por el mismo razonamiento que en la demostración de la parte a. ■

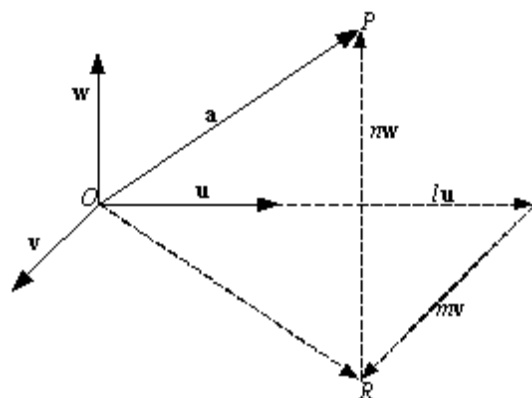


Fig. 2.2.14 Representación gráfica del teorema base parte b.

Los vectores  $\mathbf{u}$  y  $\mathbf{v}$  de la parte a y los vectores  $\mathbf{u}$ ,  $\mathbf{v}$  y  $\mathbf{w}$  de la parte b forman lo que se conoce como una *base* del correspondiente sistema vectorial o *espacio vectorial*. El número de vectores en una base es el mismo que el número de dimensiones del espacio que interviene. Los escalares  $l$ ,  $m$  y  $n$  son las *componentes* del vector  $l\mathbf{u} + m\mathbf{v} + n\mathbf{w}$ . Ningún vector base puede ser el vector nulo.

Aunque la visión geométrica de los espacios vectoriales bidimensionales y tridimensionales son suficientes para los propósitos de este trabajo, los espacios vectoriales se pueden generalizar a cualquier número de dimensiones y, de hecho, a un número infinito de las mismas (llamados espacios de Hilbert).

## 2.2.3 Coordenadas rectangulares

### Coordenadas rectangulares

Ahora procederemos a llevar el concepto de vector a un sistema coordenado, para ello definiremos los vectores base.

Empezaremos en dos dimensiones: tomaremos como base dos vectores perpendiculares, cada uno con longitud igual a uno: uno horizontal, llamado  $\mathbf{i}$ , el otro vertical llamado  $\mathbf{j}$  como se puede ver en la figura 2.3.1. Entonces cualquier vector  $\mathbf{r}$  que va desde el punto fijo  $O$  al punto  $P$  en el plano, se puede expresar en forma única como una combinación lineal de  $\mathbf{i}$  y de  $\mathbf{j}$  :

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j}$$

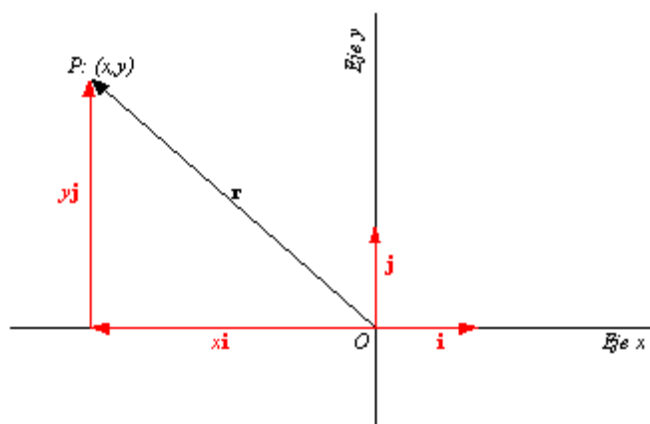


Fig. 2.3.1 Los vectores base para el plano cartesiano.

Las rectas indefinidas que pasan por  $O$  (el punto inicial común a ambos vectores) sobre  $\mathbf{i}$  y  $\mathbf{j}$  son los ejes coordenados, el eje  $x$  y el eje  $y$  respectivamente, o un sistema de referencia llamado plano cartesiano (debido a Descartes). Los coeficientes escalares  $(x, y)$  se denominan coordenadas rectangulares o coordenadas cartesianas del punto  $P$  con relación a  $O$  como origen.

El vector  $\mathbf{r} = \overrightarrow{OP}$  que parte del origen se denomina *vector de posición* de  $P$ , ya que sus componentes son las coordenadas rectangulares  $(x, y)$  de su punto terminal  $P$ .

De manera semejante, en tres dimensiones, tres vectores mutuamente perpendiculares de longitud uno, llamados  $\mathbf{i}$ ,  $\mathbf{j}$  y  $\mathbf{k}$ , se toman como base. Entonces el vector

de posición

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

tiene por componentes las coordenadas rectangulares  $(x, y, z)$  del punto terminal  $P$ . Los vectores unidad  $\mathbf{i}$ ,  $\mathbf{j}$  y  $\mathbf{k}$  se llaman también *terna fundamental*.

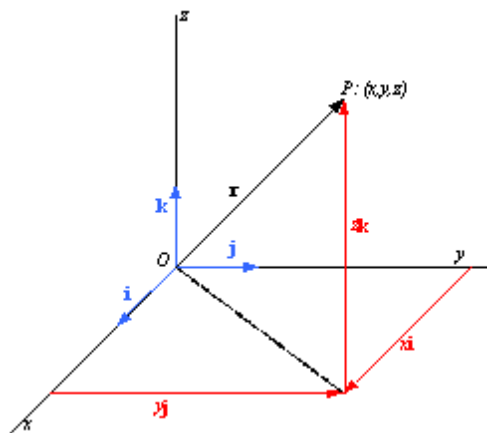


Fig. 2.3.2 Los vectores base  $\mathbf{i}$ ,  $\mathbf{j}$  y  $\mathbf{k}$ .

Es evidente, de la figura 2.3.1, que por el teorema de Pitágoras

$$|\mathbf{r}| = \sqrt{x^2 + y^2}$$

en dos dimensiones aun cuando  $x$  o  $y$  o ambas sean negativas. Similarmente, en tres dimensiones, como se ve en la figura 2.3.2:

$$|\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$$

y en general, el vector  $\mathbf{v} = l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$ , en cualquier lugar del espacio, tiene por magnitud  $\sqrt{l^2 + m^2 + n^2}$ .  $l$ ,  $m$  y  $n$  son las *componentes direccionales* del vector  $\mathbf{v}$ . A veces se les suele llamar *números directores*.

### Suma de vectores en un sistema coordenado

Para tratar el problema de la suma de vectores en un sistema coordenado analizaremos el caso de sumar dos vectores en el espacio tridimensional, tomando en cuenta que para el caso de dos vectores en el plano cartesiano no tendríamos coordenada  $z$ . Primero, sean  $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$  y  $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$  dos vectores no paralelos en

el espacio tridimensional y sea  $\mathbf{r} = r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k}$  el vector resultante al sumar a los vectores  $\mathbf{a}$  y  $\mathbf{b}$ , lo que significa que

$$\mathbf{r} = \mathbf{a} + \mathbf{b}$$

$$\iff r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} + (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k})$$

en el sistema coordenado dado, aplicando la propiedad asociativa:

$$r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k} = (a_1\mathbf{i} + b_1\mathbf{i}) + (a_2\mathbf{j} + b_2\mathbf{j}) + (a_3\mathbf{k} + b_3\mathbf{k})$$

utilizando la propiedad 5:

$$r_1\mathbf{i} + r_2\mathbf{j} + r_3\mathbf{k} = (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j} + (a_3 + b_3)\mathbf{k}$$

por lo que los vectores  $\mathbf{r}$  y  $\mathbf{a} + \mathbf{b}$ , sólo pueden ser iguales si lo son sus componentes correspondientes (por el teorema base); es decir, si

$$r_1 = a_1 + b_1$$

$$r_2 = a_2 + b_2$$

$$r_3 = a_3 + b_3$$

De esta manera, tenemos la siguiente regla para sumar vectores: descomponer cada vector en sus componentes en un sistema de coordenadas dado; la suma algebraica de las componentes individuales sobre cada eje particular es la componente de la suma vectorial sobre dicho eje; la suma vectorial puede ser reconstruida cuando se conocen sus componentes.

## Vectores de desplazamiento

Hasta ahora hemos tratado el problema de caracterizar la posición de un punto en el plano cartesiano con un vector que parte del origen (vector de posición), por lo que continuaremos con el problema de encontrar el vector formado por dos puntos fijos, o visto otra manera, representar la trayectoria de una partícula que va de un punto  $P$  a un punto  $Q$  a través de un vector (*vector de desplazamiento*) ya sea en el plano o en el espacio tridimensional, es decir, queremos formar el vector  $\overrightarrow{PQ}$ .

Para hacerlo tomaremos el caso en que ambos puntos estén en el espacio tridimensional, tomando en cuenta que para el caso en que ambos puntos estuvieran en el plano no tendríamos coordenadas en  $z$ . Como tenemos dos puntos fijos  $P$  y  $Q$  con

coordenadas  $(x_1, y_1, z_1)$  y  $(x_2, y_2, z_2)$ , respectivamente, podemos formar sus correspondientes vectores de posición:

$$\overrightarrow{OP} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k} \quad \text{y} \quad \overrightarrow{OQ} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}$$

En la figura siguiente, se puede ver que el vector buscado,  $\overrightarrow{PQ}$  se puede formar al restar el vector  $\overrightarrow{OP}$  del vector  $\overrightarrow{OQ}$  :

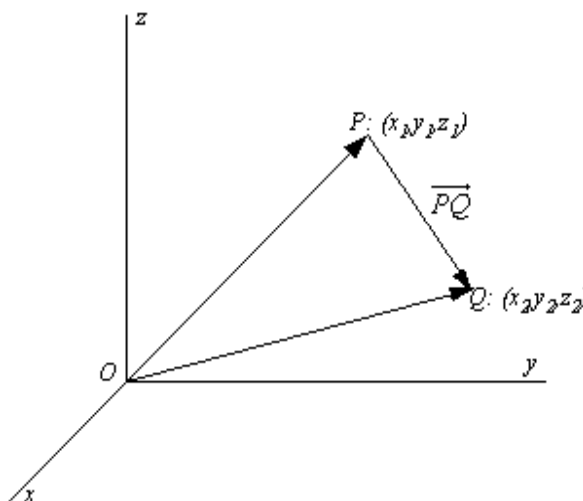


Fig. 2.3.3 Vector  $\overrightarrow{PQ}$

esto es

$$\overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k} - (x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k})$$

Aplicando la propiedad distributiva de la suma de vectores

$$\overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k} - x_1\mathbf{i} - y_1\mathbf{j} - z_1\mathbf{k}$$

utilizando la propiedad 5.

$$\overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}$$

Así que para obtener un vector, basta con restar a las componentes del vector de posición del punto donde se localiza el extremo final del vector las componentes del vector de posición del punto inicial.

### Ángulos y cosenos directores

Los ángulos  $\alpha$ ,  $\beta$  y  $\gamma$  que forman el vector  $\mathbf{v} = l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$  con los tres ejes coordenados (Fig. 2.3.4), se llaman *ángulos directores* del vector.

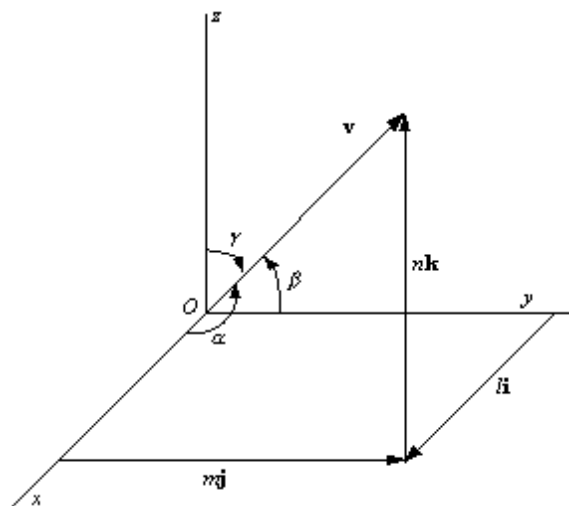


Fig. 2.3.4 Cosenos directores.

Entonces, los cosenos de los ángulos directores estarán dados de la siguiente manera:

$$\cos \alpha = \frac{l}{\sqrt{l^2 + m^2 + n^2}}, \quad \cos \beta = \frac{m}{\sqrt{l^2 + m^2 + n^2}} \quad \text{y} \quad \cos \gamma = \frac{n}{\sqrt{l^2 + m^2 + n^2}}$$

por lo que se conocen como *cosenos directores* de  $\mathbf{v}$ . Elevando al cuadrado cada uno y sumando, tenemos que

$$\begin{aligned} \cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma &= \left( \frac{l}{\sqrt{l^2 + m^2 + n^2}} \right)^2 + \left( \frac{m}{\sqrt{l^2 + m^2 + n^2}} \right)^2 + \left( \frac{n}{\sqrt{l^2 + m^2 + n^2}} \right)^2 \\ &= \frac{l^2 + m^2 + n^2}{l^2 + m^2 + n^2} = 1 \end{aligned}$$

Asimismo, es posible hablar de ángulos directores y de cosenos directores de una recta, pero una recta tiene dos sentidos, por lo que también tiene dos conjuntos de ángulos directores ( $\alpha$ ,  $\beta$  y  $\gamma$  y  $180^\circ - \alpha$ ,  $180^\circ - \beta$  y  $180^\circ - \gamma$ ) y de cosenos directores ( $\cos \alpha$ ,  $\cos \beta$  y  $\cos \gamma$  y  $-\cos \alpha$ ,  $-\cos \beta$  y  $-\cos \gamma$ ).

## 2.3 Conceptos importantes de física

La estadística al igual que todas las ciencias, necesita apoyarse en conceptos de otras áreas. Así la estadística esférica se apoya en conceptos importantes de física, como lo *son centro de masa y momento de inercia*, por lo que esta sección estará centrada en explicarlos a grandes rasgos.

### 2.3.1 Centro de masa

#### Definición y obtención

Supongamos que tenemos un sistema que consta de  $n$  partículas de masas  $m_i$ , cuyos vectores de posición son  $\mathbf{r}_i = x_i\mathbf{i} + y_i\mathbf{j} + z_i\mathbf{k}$ ,  $i = 1, 2, \dots, n$ . Definimos el *vector de posición del centro de masa* del sistema  $\mathbf{R}$  como

$$\mathbf{R} = \sum_{i=1}^n \frac{m_i \mathbf{r}_i}{M} \quad (2.2)$$

Aquí  $M$  es la masa total del sistema

$$M = \sum_{i=1}^n m_i.$$

Entonces para obtener el centro de masa de un sistema de  $n$  partículas, reescribimos la ecuación 2.2:

$$\mathbf{R} = \frac{\sum_{i=1}^n m_i \mathbf{r}_i}{\sum_{i=1}^n m_i} = \frac{\sum_{i=1}^n m_i (x_i \mathbf{i} + y_i \mathbf{j} + z_i \mathbf{k})}{\sum_{i=1}^n m_i} = \frac{\sum_{i=1}^n m_i x_i \mathbf{i}}{\sum_{i=1}^n m_i} + \frac{\sum_{i=1}^n m_i y_i \mathbf{j}}{\sum_{i=1}^n m_i} + \frac{\sum_{i=1}^n m_i z_i \mathbf{k}}{\sum_{i=1}^n m_i} \quad (2.3)$$

Ahora, sean:

$$X = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}, \quad Y = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i} \quad \text{y} \quad Z = \frac{\sum_{i=1}^n m_i z_i}{\sum_{i=1}^n m_i}$$

por lo que podemos escribir a 2.3 de la siguiente manera:

$$\mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$$



### 2.3.2 Momento de inercia

De nuevo, supongamos que tenemos un sistema que consta de  $n$  partículas de masas  $m_1, m_2, \dots, m_n$ . Definimos el *momento de inercia*  $I$  del sistema con respecto a un eje  $JJ'$  como:

$$I = \sum_{i=1}^n m_i r_i^2$$

donde  $r_i$  es la distancia de la  $i$  - *ésima* partícula al eje  $JJ'$ .

## Capítulo 3

# CONCEPTOS BÁSICOS DE ESTADÍSTICA ESFÉRICA

### 3.1 Introducción

Las observaciones direccionales son tan antiguas como la Estadística. La teoría del error tal como fue desarrollada por Gauss y Laplace tiene su primera importante aplicación en el análisis de medidas astronómicas. Estas consisten principalmente de medidas dirección de cuerpos celestes con respecto a la tierra, las cuales eran representadas como puntos sobre la bóveda celeste. Dado que los errores en las observaciones eran inevitables, algunas medidas repetidas de la misma dirección no coincidían pero tendían a esparcirse alrededor de su "verdadera" localización, generalmente en un pequeñísimo radio del orden de segundos de arco. Para enfrentar la variación anterior fue necesario el análisis estadístico. Como la curvatura en la superficie de la bóveda celeste es insignificante sobre un área de este tamaño, la superficie puede ser aproximada por un número infinito de planos tangentes, con prácticamente toda la probabilidad concentrada en el área inmediata adyacente al punto de contacto. Esta sustitución llevó al desarrollo de distribuciones de errores lineales y planares las cuales pudieron ser también utilizadas para datos no-direccionales.

Sin embargo, existen áreas de la investigación como lo son la Geología, Biología y Meteorología, existen medidas direccionales que no pueden ser aproximadas de manera lineal. Los métodos para analizar datos direccionales aparecieron mucho después. La Estadística circular tiene sus inicios en el siglo antepasado cuando Lord Rayleigh desarrolló su famosa prueba para una muestra. Sin embargo, esta área de la

estadística no ha recibido la atención suficiente, pero cabe destacar los importantes trabajos de K. V. Mardia, Edward Batschelet, Christopher Bingham, Geoffrey S. Watson, entre otros.

En general, existen dos tipos de datos direccionales: mediciones que especifican una línea que tiene sentido y mediciones que sólo especifican una línea recta sin hacer distinción entre sus puntas.

En este capítulo abordaremos los conceptos más básicos para hacer un análisis muy general de datos direccionales que se encuentran sobre una esfera, con frecuencia llamados datos esféricos; aunque cabe destacar que estos mismos pueden ser generalizados cuando se tienen datos sobre una circunferencia o sobre una hiperesfera (una esfera en más de tres dimensiones).

## 3.2 Datos direccionales

Los datos direccionales en tres dimensiones surgen en muchas áreas de la experimentación científica y observacional. Como ejemplos de datos direccionales de varios campos, podemos enumerar muchos, entre ellos tenemos: en Astrofísica, las direcciones de llegada de los baños de rayos cósmicos; en Cristalografía las direcciones de los ejes ópticos en cristales de cuarzo; en Meteorología podemos hablar de las direcciones del viento observadas en un lugar dado; en Oceanografía podemos mencionar las mediciones de las direcciones de las corrientes oceánicas; etc. Además, observaciones que no son en algún sentido orientaciones, algunas veces pueden ser reexpresadas en la forma de orientaciones y analizadas como datos direccionales.

Por observaciones direccionales nos referimos a la orientación de una línea recta en el espacio. En algunos contextos resulta útil referirnos a la línea recta como dirigida, la cual recibe el nombre de *vector*, y en otros casos, como no dirigida, a la que llamamos *eje*. Cualquier vector puede ser representado por un punto  $P$  sobre la superficie de una esfera unitaria con centro en el origen, siendo  $\overrightarrow{OP}$  la línea dirigida; cualquier eje puede ser representado por un par de puntos  $Q, P$  en los extremos de un diámetro  $QOP$  de una esfera, o equivalentemente por un punto  $P$  sobre un hemisferio unitario (la mitad de una esfera de radio 1 dividida por un plano que pasa por el centro). Dos cantidades son requeridas para definir un punto  $P$  sobre la esfera o hemisferio, y por tanto, una medida esférica, ya sea vectorial o axial.

### 3.3 Sistemas coordenados en estadística circular

#### 3.3.1 Coordenadas polares

Para definir las coordenadas polares, primero, en un plano fijaremos un punto  $O$  al que llamaremos *polo u origen*, y a partir de él tracemos un rayo  $\mathcal{L}$  al que llamaremos *eje polar*, cualquier punto  $P$  del plano pertenece a una única circunferencia con centro en el polo y cuyo radio sea igual a la distancia  $d(P, O)$ .

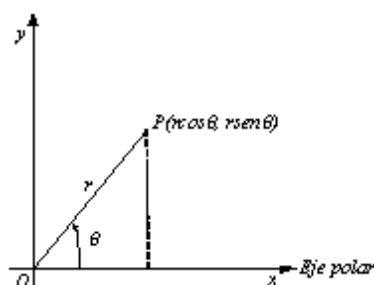
Entonces, las *coordenadas polares* del punto  $P$  serán los números  $(r, \theta)$ , llamados *radio* y *ángulo*, respectivamente, donde

$$r = d(O, P) \quad \text{y} \quad \theta = \text{ángulo entre } \mathcal{L} \text{ y } OP$$

Después de haber definido lo anterior, podemos establecer fórmulas para obtener las coordenadas polares del punto  $P$  con coordenadas cartesianas  $(x, y)$  con el origen como polo:

$$x = r \cos \theta \quad \text{y} \quad y = r \operatorname{sen} \theta$$

como se puede apreciar en la siguiente figura:



3.3.1. Transformación de coordenadas rectangulares a polares.

De esta manera podemos plantear las relaciones entre las coordenadas rectangulares y las coordenadas polares, ya que resultan de gran ayuda para hacer transformaciones de un sistema de coordenadas a otro:

$$x = r \cos \theta \quad y = r \operatorname{sen} \theta \quad r = \sqrt{x^2 + y^2}$$

$$\theta = \arctan \frac{y}{x} = \operatorname{arcsen} \frac{y}{\sqrt{x^2 + y^2}} = \operatorname{arccos} \frac{x}{\sqrt{x^2 + y^2}}$$

siempre que  $x$  y  $y$  no sean ambas cero. Se puede observar que  $\theta$  puede tener muchos valores, difiriendo cada valor de otro en un múltiplo de  $2\pi$  *radianes* o  $360^\circ$ , pero su cuadrante es único si recordamos que  $x$  y  $y$  son números que tienen signos, además, de que  $r$ , siendo una distancia, siempre es positiva para cualquier punto (salvo para el origen).

## 3.4 Sistemas coordenados en estadística esférica

### 3.4.1 Coordenadas esféricas

Existen diversas formas de representar un vector en un espacio tridimensional, algunas de las cuales resultan más útiles, en el sentido de que facilitan los cálculos que se realicen con los vectores que representan ciertos tipos de datos. Así, para el análisis de componentes principales para datos direccionales, las coordenadas esféricas resultan de gran ayuda.

Tomemos una esfera de radio 1 ( $\rho = 1$ ) con centro en el origen (en el sistema coordenado rectangular), cuya superficie es tocada por un vector  $\overrightarrow{OP}$  en el punto  $P$  con coordenadas  $(x, y, z)$ , como se muestra en la figura 3.4.1. Supongamos también que  $\theta$  es el ángulo entre el eje  $z$  y el vector  $\overrightarrow{OP}$  medido en el sentido de las manecillas del reloj y que  $\phi$  es el ángulo entre el eje  $x$  y la línea  $ON$  medido en el sentido contrario al de las manecillas del reloj, donde  $N$  es el pie de la perpendicular que baja desde el punto  $P$  al plano  $XY$ .

Para hacer la transformación de coordenadas rectangulares a coordenadas esféricas, necesitamos la longitud del segmento  $ON$  a la que llamaremos  $d$ . Así, por el teorema de Pitágoras:

$$\begin{aligned} d^2 &= x^2 + y^2 & \text{y} & \quad \rho^2 = d^2 + z^2, & \text{en este caso } \rho = 1 \\ \implies & 1 = x^2 + y^2 + z^2 \end{aligned}$$

y utilizando razones trigonométricas, se puede ver que

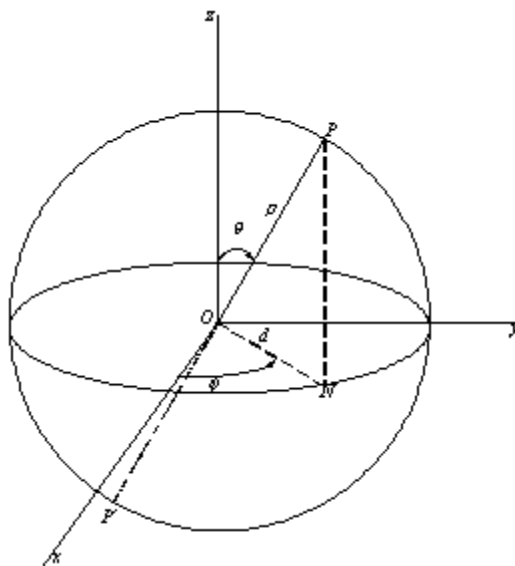
$$x = d \cos \phi, \quad y = d \sin \phi, \quad z = \rho \cos \theta \quad \text{y} \quad d = \rho \sin \theta$$

por lo que

$$x = \rho \cos \phi \sin \theta, \quad y = \rho \sin \phi \sin \theta, \quad z = \rho \cos \theta$$

pero  $\rho = 1$ ,

$$x = \cos \phi \sin \theta, \quad y = \sin \phi \sin \theta, \quad z = \cos \theta \tag{3.1}$$



3.4.1 Coordenadas esféricas  $\theta$  y  $\phi$  del punto  $P$ .

Los posibles valores de los ángulos son:

$$0^\circ \leq \theta \leq 180^\circ \quad \text{y} \quad 0^\circ \leq \phi \leq 360^\circ$$

El ángulo  $\theta$  puede ser descrito como la colatitud del punto  $P$  mientras que  $\phi$  puede ser descrito como la longitud de  $P$ , el origen es el polo de referencia y el eje  $z$  el eje polar. El eje  $x$  se encuentra en  $\theta = \frac{\pi}{2}$ ,  $\phi = 0$ , mientras que el eje  $y$  está en  $\theta = \phi = \frac{\pi}{2}$ .

Para una observación axial, el vector  $\overrightarrow{OP}$  es reemplazado por el diámetro que va del punto  $P$  al punto  $P'$ , pasando por el origen. Como los datos axiales suelen ser indicados en el hemisferio inferior, las coordenadas de  $P'$  serían, en términos de  $\theta$  y  $\phi$ ,  $(180^\circ - \theta, \phi + 180^\circ)$ .

### 3.4.2 Coordenadas geográficas

Si colocamos a la Tierra en un sistema coordenado rectangular, con su centro en el origen  $O$ , el punto de intersección del eje  $z$  con la superficie terrestre es el Polo Norte, en coordenadas esféricas el Polo Norte se encuentra en  $\theta = 0^\circ$ . La variación del punto  $P$  a lo largo del eje  $z$  de Norte a Sur es medida por el ángulo  $\theta'$ , denominada latitud, el cual es medido entre  $OP$  y el plano  $xy$  (el plano ecuatorial con coordenadas esféricas  $\theta = 90^\circ$ ), con los ángulos por arriba del él siendo positivos y por debajo,

negativos. La longitud  $\phi'$  es medida como el ángulo entre el meridiano que contiene al punto  $P$  y el meridiano "cero", o de otra manera, el ángulo entre la proyección  $ON$  de  $OP$  sobre el plano  $xy$  y el eje  $x$ , en sentido opuesto al de las manecillas del reloj. Así, podemos expresar a las coordenadas geográficas en términos de las coordenadas esféricas de la siguiente manera:

$$\theta' = 90^\circ - \theta, \quad \phi' = \phi$$

por lo que 3.1 puede simplificarse de la siguiente manera:

$$x = \cos \phi' \sin (90^\circ - \theta'), \quad y = \sin \phi' \sin (90^\circ - \theta'), \quad z = \cos (90^\circ - \theta')$$

como  $\sin (90^\circ - \theta') = \cos \theta'$  y  $\cos (90^\circ - \theta') = \sin \theta'$ , entonces

$$x = \cos \phi' \cos \theta', \quad y = \sin \phi' \cos \theta', \quad z = \sin \theta'$$

y los valores de  $\theta'$  y  $\phi'$  quedan definidos de la siguiente manera:

$$-90^\circ \leq \theta \leq 90^\circ \quad \text{y} \quad 0^\circ \leq \phi' \leq 360^\circ$$

En la figura siguiente se puede observar la manera en que son medidas las coordenadas geográficas.

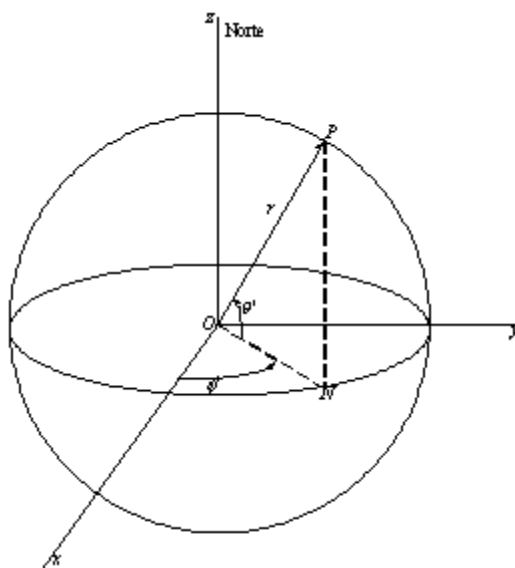


Fig. 3.4.2. Coordenadas geograficas.

En el hemisferio norte (la mitad superior de la esfera cortada a través del ecuador)  $\theta = 90^\circ - \theta'$  es la latitud norte. En el hemisferio sur,  $\theta = 90^\circ + \theta'$ , siendo  $\theta'$  la latitud sur. Los círculos grandes a través de los polos son los meridianos y el meridiano principal  $\phi = 0^\circ$ .

### 3.5 Proyecciones más utilizadas en estadística esférica

Para facilitar la visualización y por consiguiente el análisis de datos esféricos, sería muy útil poder representarlos sobre un plano, es decir, realizar una proyección. Una proyección de datos esféricos es una representación de la observaciones esféricas sobre un plano. En muchos casos, los datos se encuentran confinados en un hemisferio. Si éste no es el caso, las proyecciones separadas pueden ser hechas para los datos en cada hemisferio o representar a todos los datos sobre un mismo plano pero indicando a los datos de un hemisferio con círculos llenos ( $\bullet$ ) y a los correspondientes al otro hemisferio con círculos vacíos ( $\circ$ ), por ejemplo.

Existen muchos tipos de proyecciones creadas a partir de las diferentes necesidades de investigadores como astrónomos, cartógrafos, geólogos y otros, por lo que en esta sección sólo daremos descripción de las más utilizadas.<sup>1</sup> La elección de la proyección a utilizar dependerá de las propiedades de los datos que uno quiera conservar después de haberla realizado. Si los datos están esparcidos sobre ambos hemisferios, lo más deseable es poder mostrar ambos hemisferios al mismo tiempo.

Consideremos el caso en que las observaciones sean de tipo axial las cuales pueden ser representadas sobre un hemisferio.

Sea  $P$  un punto sobre el hemisferio superior de una esfera unitaria, con colatitud  $\theta$  y longitud  $\phi$ . Denotemos por  $P'$  su proyección sobre el plano  $xy$ , con coordenadas  $(x', y')$ . Algunas veces es conveniente trabajar con coordenadas polares  $(r, \psi)$  en el plano, como ya hemos descrito en la sección 3.3, así, tenemos que

$$x' = r \cos \psi, \quad y' = r \operatorname{sen} \psi$$

y de manera inversa:

$$r = \sqrt{x'^2 + y'^2}, \quad \psi = \tan^{-1} \left( \frac{y'}{x'} \right)$$

---

<sup>1</sup>Todas las proyecciones enumeradas en [8].



A continuación enumeramos las proyecciones más comunes:

### 1. Proyección Wulff o estereográfica

*Definición*

$$r = \left| \tan \left( \frac{\theta}{2} \right) \right|, \quad \psi = \phi, \quad 0 \leq r \leq 1, \quad 0 \leq \psi < 2\pi$$

*Propiedades*

Esta proyección preserva los ángulos: círculos máximos y mínimos son proyectados en áreas circulares. La densidad de los puntos dibujados en diferentes áreas puede ser fácilmente analizada a través de esta proyección. Es muy usada en trabajos de ingeniería, morfología de cristales, paleomagnetismo y geología estructural.

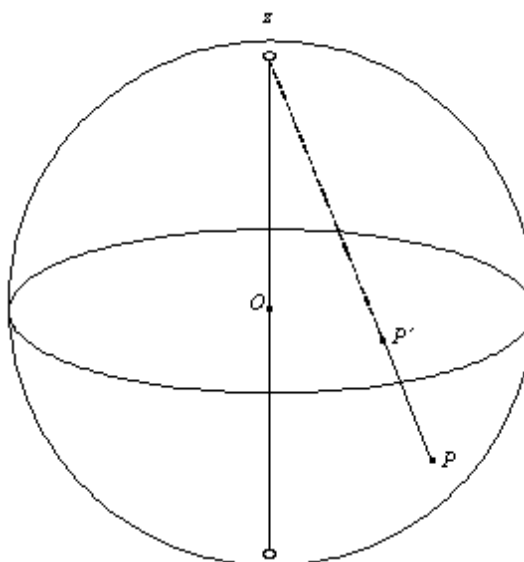


Fig. 3.5.1. Proyección Wulff del punto  $P$  en  $P'$ .

### 2. Proyección Lambert o Schmidt

*Definición*

$$r = 2 \left| \sin \left( \frac{\theta}{2} \right) \right|, \quad \psi = \phi, \quad 0 \leq r \leq \sqrt{2}, \quad 0 \leq \psi < 2\pi$$

### *Propiedades*

Es una proyección que preserva el área del objeto que se está proyectando: áreas sobre la superficie de la esfera son mapeadas sobre áreas iguales sobre el plano de proyección. Así, la densidad de los puntos es preservada.

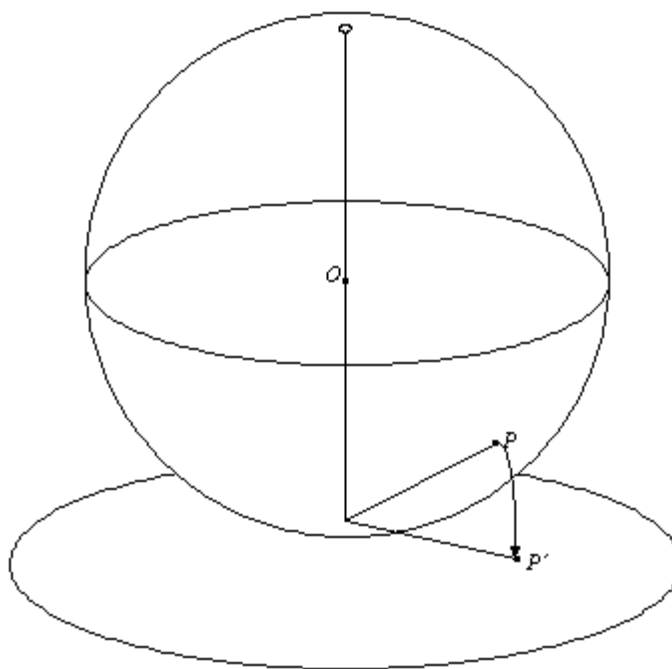


Fig. 3.5.2 Proyección Lambert del punto  $P$  en  $P'$ .

### 3. Proyección ortográfica

#### *Definición*

$$x' = \sin \phi, \quad y' = \cos \theta.$$

#### *Propiedades*

Esta proyección primero fue de interés estético para representar los datos en dos dimensiones. Sin embargo, los astrónomos la utilizan mucho.

### 4. Proyección central o gnómica

#### *Definición*

$$\text{Para } \frac{\pi}{2} < \theta \leq \pi, \quad r = -\tan \theta, \quad \psi = \phi$$

Esta proyección puede ser utilizada únicamente para datos que se encuentran en un hemisferio, sin embargo, tiene la propiedad de que todos los

círculos máximos son proyectados como líneas rectas.

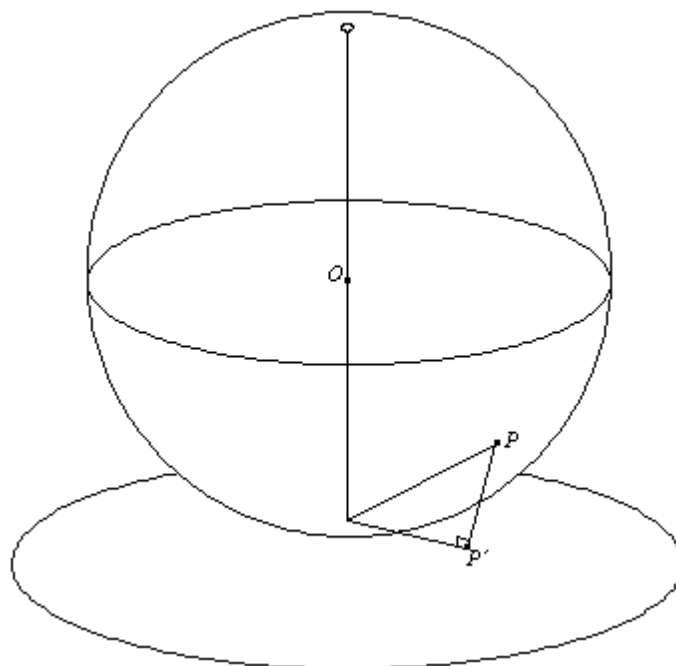


Fig. 3.5.3. Proyeccion ortografica del punto  $P$  en  $P'$ .

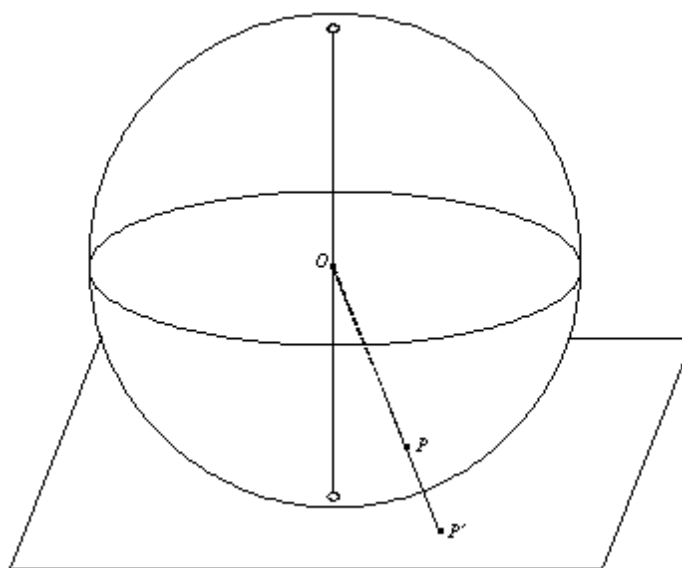


Fig. 3.5.4 Proyeccion central del punto  $P$  en el punto  $P'$ .

### 3.6 Rotación de vectores unitarios y de ejes

Si las observaciones se encuentran sobre una esfera, los puntos de los hemisferios superior e inferior pueden ser proyectados utilizando los métodos de proyección descritos en la sección anterior.

Antes de llevar a cabo una proyección sería deseable rotar la esfera en el caso, por ejemplo, de que los puntos estén agrupados alrededor del ecuador o alrededor de alguna dirección en especial.

Sean  $(\theta, \phi)$  las coordenadas esféricas de un vector unitario medido en relación a un polo de referencia que tiene coordenadas  $(0^\circ, 0^\circ)$ . Para encontrar las coordenadas  $(\theta', \phi')$  relativas a otra dirección, digamos  $(\theta_0, \phi_0)$  como nuevo polo. Para lograrlo es necesario realizar la siguiente rotación. Sea  $\psi_0$  un ángulo arbitrario y definamos a  $a$  como  $\cos \psi_0$  y  $b$  como  $\sin \psi_0$ . La forma general de una matriz que da una rotación de los ejes  $x, y$  y  $z$  con  $(\theta_0, \phi_0)$  como la nueva dirección  $z$  es:

$$= \mathbf{A}(\theta_0, \phi_0, \psi_0) \begin{pmatrix} a \cos \theta_0 \cos \phi_0 - b \sin \phi_0 & a \cos \theta_0 \sin \phi_0 + b \cos \phi_0 & -a \sin \theta_0 \\ -b \cos \theta_0 \cos \phi_0 - a \sin \phi_0 & -b \cos \theta_0 \sin \phi_0 + a \cos \phi_0 & b \sin \theta_0 \\ \sin \theta_0 \cos \phi_0 & \sin \theta_0 \sin \phi_0 & \cos \theta_0 \end{pmatrix}$$

Así, los cosenos directores de  $(\theta', \phi')$  están dados por:

$$\begin{pmatrix} \sin \theta' \cos \phi' \\ \sin \theta' \sin \phi' \\ \cos \theta' \end{pmatrix} = \mathbf{A}(\theta_0, \phi_0, \psi_0) \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}$$

El ángulo  $\psi_0$  es una rotación alrededor del eje polar a través de  $(\theta_0, \phi_0)$ , y no tiene gran importancia en muchos casos. Si éste es fijado en cero, podemos usar la matriz de rotación simplificada:

$$\mathbf{A}(\theta_0, \phi_0, 0) = \begin{pmatrix} \cos \theta_0 \cos \phi_0 & \cos \theta_0 \sin \phi_0 & -\sin \theta_0 \\ -\sin \phi_0 & \cos \phi_0 & 0 \\ \sin \theta_0 \cos \phi_0 & \sin \theta_0 \sin \phi_0 & \cos \theta_0 \end{pmatrix} \quad (3.2)$$

Note que  $\theta'$  es el ángulo entre  $(\theta, \phi)$  y  $(\theta_0, \phi_0)$ .

Para datos axiales, suponga que el eje  $(x, y, z)$  es medido en relación al eje polar que tiene como vector director al vector canónico  $\mathbf{k}$  y queremos encontrar sus coordenadas  $(x', y', z')$  relativas a  $(x_0, y_0, z_0)$ . Por conveniencia, primero se determina uno

de los finales del eje  $(x_0, y_0, z_0)$  en coordenadas esféricas, es decir, se obtienen las coordenadas  $(\theta_0, \phi_0)$ . Después, como para el caso de los vectores, se obtiene  $\mathbf{A}(\theta_0, \phi_0, \psi_0)$  o  $\mathbf{A}(\theta_0, \phi_0, 0)$ , y así tenemos que

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{A}(\theta_0, \phi_0, \psi_0) \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$z'$  es el coseno del más pequeño de los ángulos que forman los ejes  $(x, y, z)$  y  $(x_0, y_0, z_0)$

## 3.7 Medidas descriptivas

### 3.7.1 Dirección media

Antes de obtener la dirección media para datos direccionales sobre una esfera, estudiaremos la dirección media en estadística circular. Primero, suponga que tenemos 5 observaciones sobre un círculo de radio 1, con sus correspondientes ángulos  $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ ,

$$\theta_1 = 90^\circ, \quad \theta_2 = 0^\circ, \quad \theta_3 = 60^\circ, \quad \theta_4 = 330^\circ \quad y \quad \theta_5 = 120^\circ$$

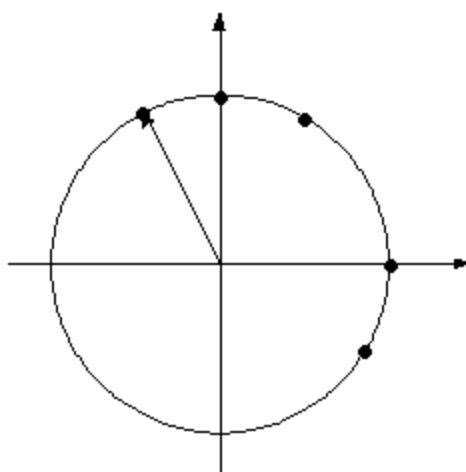


Fig. 3.7.1 Dirección media.

Como puede observarse en la figura anterior, se esperaría que la dirección media,  $\bar{\theta}$  fuera un ángulo entre  $0^\circ$  y  $60^\circ$ , sin embargo, al calcular la media aritmética directa-

mente con los ángulos vemos que

$$\begin{aligned}\bar{\theta} &= \frac{1}{5} (\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5) \\ &= \frac{1}{5} (90^\circ + 0^\circ + 60^\circ + 330^\circ + 120^\circ) \\ &= \frac{1}{5} (600^\circ) = 120^\circ\end{aligned}$$

claramente, este valor para la observación media no es adecuado. Si representamos a  $\theta_4$  como  $-30^\circ$ , tenemos que la dirección media ahora será

$$\begin{aligned}\bar{\theta} &= \frac{1}{5} (\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5) \\ &= \frac{1}{5} (90^\circ + 0^\circ + 60^\circ - 30^\circ + 120^\circ) \\ &= \frac{1}{5} (240^\circ) = 48^\circ\end{aligned}$$

que resulta un valor para la dirección media más aceptable, sin embargo si estuviéramos trabajando con una cantidad grande de observaciones el hacer estos cambios en los ángulos que pudieran estar influyendo en el resultado, sería una tarea complicada.

Una manera más apropiada para obtener la dirección media está basada en un procedimiento diferente. Dado un conjunto de  $n$  observaciones, las podemos colocar en un círculo unitario (círculo de radio igual a 1), asignando a cada una, una masa de tamaño  $M$ , entonces, para encontrar la dirección media basta con encontrar el centro de masa a partir de los vectores de posición de cada una de las  $n$  observaciones y sus respectivas masas  $M$ .

### Obtención de la dirección media

Cada una de las  $n$  observaciones con coordenadas  $(x_i, y_i)$ , representada sobre el círculo unitario tiene un vector de posición de longitud 1,  $\mathbf{e}_i$ ,  $i = 1, 2, \dots, n$ , es decir,

$$\mathbf{e}_i = x_i \mathbf{i} + y_i \mathbf{j}, \quad |\mathbf{e}_i| = 1 \quad i = 1, 2, \dots, n$$

asignando a cada observación una masa  $M$ , entonces, el vector de posición del centro de masa o dicho de otra manera, el vector de la dirección media estará dado por:

$$\mathbf{m} = \frac{\sum_{i=1}^n M \mathbf{e}_i}{\sum_{i=1}^n M} = \frac{M \sum_{i=1}^n \mathbf{e}_i}{nM} = \frac{\sum_{i=1}^n \mathbf{e}_i}{n} = \frac{\sum_{i=1}^n (x_i \mathbf{i} + y_i \mathbf{j})}{n} = \frac{\sum_{i=1}^n x_i}{n} \mathbf{i} + \frac{\sum_{i=1}^n y_i}{n} \mathbf{j}$$

Así, obtenemos el vector resultante  $\sum_{i=1}^n \mathbf{e}_i$  cuya longitud está dividida por  $n$ . De esta manera llamamos a  $\mathbf{m}$  *dirección media o vector medio* de la muestra. Entonces este vector tendrá componentes  $\bar{x}$  y  $\bar{y}$  (que son las coordenadas del centro de masa), respectivamente:

$$\mathbf{m} = \frac{\sum_{i=1}^n \mathbf{e}_i}{n} = \frac{\sum_{i=1}^n (x_i \mathbf{i} + y_i \mathbf{j})}{n} = \frac{\sum_{i=1}^n x_i \mathbf{i}}{n} + \frac{\sum_{i=1}^n y_i \mathbf{j}}{n} = \bar{x} \mathbf{i} + \bar{y} \mathbf{j}$$

Ahora, si definimos los vectores de posición de las observaciones de la muestra con sus coordenadas polares tenemos que cada observación tendrá un vector de posición

$$\mathbf{e}_i = \cos \theta_i \mathbf{i} + \text{sen } \theta_i \mathbf{j} \quad i = 1, 2, \dots, n$$

ya que las coordenadas polares de cada una de las observaciones estarán dadas por:

$$x_i = \cos \theta_i \quad y_i = \text{sen } \theta_i$$

pues las observaciones se encuentran sobre una circunferencia de radio igual a 1. Por otro lado, si  $\bar{x}$  y  $\bar{y}$  son las coordenadas del centro de masa, tenemos que

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n \cos \theta_i \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n \text{sen } \theta_i \end{aligned}$$

Por otra parte, sean  $R$  y  $r$  las longitudes de los vectores  $\sum_{i=1}^n \mathbf{e}_i$  y  $\mathbf{m}$ , respectivamente, es decir,

$$R = \left| \sum_{i=1}^n \mathbf{e}_i \right| = \sqrt{\left( \sum_{i=1}^n x_i \right)^2 + \left( \sum_{i=1}^n y_i \right)^2}, \quad r = |\mathbf{m}| = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (3.3)$$

entonces

$$r = \frac{R}{n} \quad (3.4)$$

Luego, es fácil deducir a partir de las ecuaciones 3.3 y 3.4 que  $r$  será igual a 1 y el centro de masa estará en la circunferencia en el caso excepcional de que todas las observaciones tengan la misma dirección, mientras que si las observaciones están muy dispersas alrededor de la circunferencia, el centro de masa se encontrará en el interior del círculo, lo anterior nos lleva a afirmar que

$$0 \leq R \leq n \quad \text{y} \quad 0 \leq r \leq 1$$

Así, tenemos un vector de posición del centro de masa, que también nos está indicando la dirección media, por lo que ahora sería interesante encontrar el ángulo que forma este vector con el eje  $x$ , primero, escribamos la longitud del vector  $\mathbf{m}$  en términos de las componentes polares del vector medio

$$r = \frac{1}{n} \sqrt{\left(\sum_{i=1}^n \cos \theta_i\right)^2 + \left(\sum_{i=1}^n \operatorname{sen} \theta_i\right)^2}$$

de la ecuación podemos ver que tendremos un caso especial cuando las componentes del vector medio sean 0, es decir,  $\bar{x} = \bar{y} = 0$  lo que implica que  $r = 0$ . En este caso el vector medio será el vector cero, si no consideramos este caso tendremos que el vector medio formará un ángulo muy bien definido con el eje  $x$ , al cual se le llama *ángulo medio muestral* y lo denotamos por  $\bar{\theta}$ . Para calcularlo tomemos en cuenta que

$$\tan \bar{\theta} = \frac{\operatorname{sen} \bar{\theta}}{\cos \bar{\theta}} = \frac{\bar{y}}{\bar{x}}$$

así que

$$\bar{\theta} = \begin{cases} \arctan\left(\frac{\bar{y}}{\bar{x}}\right) & \text{si } \bar{x} > 0 \\ 180^\circ + \arctan\left(\frac{\bar{y}}{\bar{x}}\right) & \text{si } \bar{x} < 0 \end{cases}$$

tomando en cuenta los casos excepcionales

$$\bar{\theta} = \begin{cases} 90^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} > 0 \\ 270^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} < 0 \\ \text{indeterminado} & \text{si } \bar{x} = 0 \text{ y } \bar{y} = 0 \end{cases}$$

Una vez que hemos revisado la obtención de la dirección media en estadística circular nos será mucho más sencillo encontrar la dirección media en el caso en que contemos con datos que puedan ser caracterizados sobre una esfera. Así, podemos definir a la dirección media en estadística esférica de manera similar a la que utilizamos en estadística circular.

Sean  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  observaciones. Para encontrar la dirección media coloquemos a las  $n$  observaciones sobre una esfera unitaria ( $\rho = 1$ ), con coordenadas  $(x_i, y_i, z_i)$   $i = 1, 2, \dots, n$ ; si cada uno de estos puntos tiene vector de dirección  $\mathbf{e}_i$  y a cada observación le asignamos una masa de tamaño  $M$ , entonces la dirección media de estas  $n$  observaciones será el vector resultante del centro de masa de estas  $n$  observaciones y el centro de masa tendrá coordenadas  $(\bar{x}, \bar{y}, \bar{z})$ :

$$\mathbf{m} = \frac{\sum_{i=1}^n \mathbf{e}_i}{n} = \frac{\sum_{i=1}^n (x_i \mathbf{i} + y_i \mathbf{j} + z_i \mathbf{k})}{n} = \frac{\sum_{i=1}^n x_i}{n} \mathbf{i} + \frac{\sum_{i=1}^n y_i}{n} \mathbf{j} + \frac{\sum_{i=1}^n z_i}{n} \mathbf{k} = \bar{x} \mathbf{i} + \bar{y} \mathbf{j} + \bar{z} \mathbf{k}$$



Entonces este vector tendrá componentes  $\bar{x}$ ,  $\bar{y}$  y  $\bar{z}$  (que son las coordenadas del centro de masa), respectivamente.

Ahora si definimos a los vectores de posición de las observaciones de la muestra con sus coordenadas esféricas tenemos que cada observación tendrá un vector de posición

$$\mathbf{e}_i = \cos \phi_i \operatorname{sen} \theta_i \mathbf{i} + \operatorname{sen} \phi_i \operatorname{sen} \theta_i \mathbf{j} + \cos \theta_i \mathbf{k} \quad i = 1, 2, \dots, n$$

ya que las coordenadas esféricas de cada una de las observaciones estarán dadas por:

$$x_i = \cos \phi_i \operatorname{sen} \theta_i \quad y_i = \operatorname{sen} \phi_i \operatorname{sen} \theta_i \quad z_i = \cos \theta_i$$

pues las observaciones se encuentran sobre una esfera de radio igual a 1. Por otro lado, si  $\bar{x}$ ,  $\bar{y}$  y  $\bar{z}$  son las coordenadas del centro de masa, tenemos que

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n \cos \phi_i \operatorname{sen} \theta_i \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n \operatorname{sen} \phi_i \operatorname{sen} \theta_i \\ \bar{z} &= \frac{1}{n} \sum_{i=1}^n \cos \theta_i \end{aligned}$$

De igual manera, sean  $R$  y  $r$  las longitudes de los vectores  $\sum_{i=1}^n \mathbf{e}_i$  y  $\mathbf{m}$ , respectivamente, es decir,

$$\begin{aligned} R &= \left| \sum_{i=1}^n \mathbf{e}_i \right| = \sqrt{\left( \sum_{i=1}^n x_i \right)^2 + \left( \sum_{i=1}^n y_i \right)^2 + \left( \sum_{i=1}^n z_i \right)^2} \\ r &= |\mathbf{m}| = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2} \end{aligned}$$

entonces

$$r = \frac{R}{n} \tag{3.5}$$

De igual manera, podemos afirmar que:

$$0 \leq R \leq n \quad \text{y} \quad 0 \leq r \leq 1$$

Así, tenemos un vector de posición del centro de masa, que también nos está indicando la dirección media, por lo que ahora sería interesante escribir este vector en coordenadas esféricas, para ello, sean  $\bar{\theta}$  y  $\bar{\phi}$  los ángulos medios que forman a las coordenadas

esféricas del centro de masa, para obtenerlas, podemos escribir:

$$\begin{aligned}\bar{x} &= \cos \bar{\phi} \sen \bar{\theta} \\ \bar{y} &= \sen \bar{\phi} \sen \bar{\theta} \\ \bar{z} &= \cos \bar{\theta} \\ \tan \bar{\phi} &= \frac{\sen \bar{\phi}}{\cos \bar{\phi}} = \frac{\bar{y}}{\bar{x}}\end{aligned}$$

así que

$$\bar{\phi} = \begin{cases} \arctan\left(\frac{\bar{y}}{\bar{x}}\right) & \text{si } \bar{x} > 0 \\ 180^\circ + \arctan\left(\frac{\bar{y}}{\bar{x}}\right) & \text{si } \bar{x} < 0 \end{cases}$$

con los casos excepcionales

$$\bar{\phi} = \begin{cases} 90^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} > 0 \\ 270^\circ & \text{si } \bar{x} = 0 \text{ y } \bar{y} < 0 \\ \text{indeterminado} & \text{si } \bar{x} = 0 \text{ y } \bar{y} = 0 \end{cases}$$

y  $\bar{\theta} = \arccos(\bar{z})$ , tomando en cuenta que  $0^\circ \leq \bar{\theta} \leq 180^\circ$ .

### Propiedades

Sea  $\mathbf{L}_i = (x_i, y_i, z_i)'$ ,  $i = 1, 2, \dots, n$ . Considere la rotación de ejes definida por

$$\mathbf{L}_{i*} = \mathbf{A}\mathbf{L}_i$$

donde  $\mathbf{A}$  es una matriz ortogonal ( $\mathbf{A}' = \mathbf{A}^{-1}$ ) y  $\mathbf{L}_{i*}$  es el vector  $\mathbf{L}_i$  transformado. Es fácilmente demostrable que  $R$  es invariante bajo rotaciones y los cosenos directores de la nueva dirección media  $\mathbf{m}_* = (\bar{x}_*, \bar{y}_*, \bar{z}_*)'$  están dados por

$$\mathbf{m}_* = \mathbf{A}\mathbf{m}$$

Además, si  $(\theta_0, \phi_0)$  es el nuevo polo bajo la transformación, la matriz  $\mathbf{A}$  es la matriz de coeficientes en 3.2. Si aplicamos la transformación anterior a cada una de las  $n$  observaciones, tenemos que:

$$\begin{aligned}\begin{pmatrix} x_{i*} \\ y_{i*} \\ z_{i*} \end{pmatrix} &= \begin{pmatrix} \cos \theta_0 \cos \phi_0 & \cos \theta_0 \sen \phi_0 & -\sen \theta_0 \\ -\sen \phi_0 & \cos \phi_0 & 0 \\ \sen \theta_0 \cos \phi_0 & \sen \theta_0 \sen \phi_0 & \cos \theta_0 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \\ \begin{pmatrix} x_{i*} \\ y_{i*} \\ z_{i*} \end{pmatrix} &= \begin{pmatrix} (\cos \theta_0 \cos \phi_0) x_i + (\cos \theta_0 \sen \phi_0) y_i + (-\sen \theta_0) z_i \\ (-\sen \phi_0) x_i + (\cos \phi_0) y_i \\ (\sen \theta_0 \cos \phi_0) x_i + (\sen \theta_0 \sen \phi_0) y_i + (\cos \theta_0) z_i \end{pmatrix}\end{aligned}$$

y al obtener las componentes del vector  $\mathbf{m}_0$  tenemos que:

$$\begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_{i^*} \\ \frac{1}{n} \sum_{i=1}^n y_{i^*} \\ \frac{1}{n} \sum_{i=1}^n z_{i^*} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n [(\cos \theta_0 \cos \phi_0) x_i + (\cos \theta_0 \sin \phi_0) y_i + (-\sin \theta_0) z_i] \\ \frac{1}{n} \sum_{i=1}^n [(-\sin \phi_0) x_i + (\cos \phi_0) y_i] \\ \frac{1}{n} \sum_{i=1}^n [(\sin \theta_0 \cos \phi_0) x_i + (\sin \theta_0 \sin \phi_0) y_i + (\cos \theta_0) z_i] \end{pmatrix}$$

a partir de la ecuación anterior es posible observar que la nueva dirección media con respecto al nuevo polo es precisamente la dirección media con respecto al polo original, pero rotada.

$$\begin{pmatrix} \bar{x}_* \\ \bar{y}_* \\ \bar{z}_* \end{pmatrix} = \begin{pmatrix} (\cos \theta_0 \cos \phi_0) \bar{x} + (\cos \theta_0 \sin \phi_0) \bar{y} + (-\sin \theta_0) \bar{z} \\ (-\sin \phi_0) \bar{x} + (\cos \phi_0) \bar{y} \\ (\sin \theta_0 \cos \phi_0) \bar{x} + (\sin \theta_0 \sin \phi_0) \bar{y} + (\cos \theta_0) \bar{z} \end{pmatrix}$$

### 3.7.2 Varianza

Como mencionamos anteriormente,  $r$  tomará el valor 1 en el caso en que todas las observaciones tengan la misma dirección, mientras que si las observaciones están muy dispersas alrededor de la circunferencia, el valor de  $r$  será muy cercano a cero. Así, es natural pensar en  $r$  como una medida de dispersión de los datos.

Por otro lado es fácil comprobar lo anterior al demostrar que  $\mathbf{m}$  tiene la siguiente propiedad de minimización: sea  $S(\mathbf{a})$  la media aritmética de las distancias cordales cuadradas entre los vectores  $\mathbf{x}_i$  y el vector  $\mathbf{a}$ , donde  $|\mathbf{a}| = 1$ . Entonces tenemos que

$$\text{Min} \quad S(\mathbf{a}) = \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i - \mathbf{a}|^2 \quad \text{s.a.} \quad \mathbf{a}'\mathbf{a} = 1$$

Pero

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i - \mathbf{a}|^2 &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{a})' (\mathbf{x}_i - \mathbf{a}) \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i' \mathbf{x}_i - \mathbf{x}_i' \mathbf{a} - \mathbf{a}' \mathbf{x}_i + \mathbf{a}' \mathbf{a}) \\ &= \frac{1}{n} \sum_{i=1}^n (1 - 2\mathbf{x}_i' \mathbf{a} + 1) \end{aligned}$$

ya que  $|\mathbf{x}_i| = 1$  y  $|\mathbf{a}| = 1$ . Entonces

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i - \mathbf{a}|^2 &= \frac{1}{n} \sum_{i=1}^n (2 - 2\mathbf{x}'_i \mathbf{a}) \\ &= \frac{2}{n} \left( n - \mathbf{a}' \sum_{i=1}^n \mathbf{x}_i \right) \\ &= 2 - 2\mathbf{m}'\mathbf{a} \end{aligned}$$

Por lo que el problema queda planteado ahora de la siguiente manera:

$$\text{Min } S(\mathbf{a}) = 2 - 2\mathbf{m}'\mathbf{a} \quad \text{s.a. } \mathbf{a}'\mathbf{a} = 1$$

Para resolver este problema, el método apropiado es el de multiplicadores de Lagrange. Este método consiste en plantear la siguiente función:

$$\varphi = 2 - 2\mathbf{m}'\mathbf{a} - \lambda(\mathbf{a}'\mathbf{a} - 1)$$

que lleva al problema de encontrar el valor mínimo de la función anterior, por lo que derivando  $\varphi$  con respecto a  $\mathbf{a}$ , tenemos:

$$\frac{d\varphi}{d\mathbf{a}} = 2\mathbf{m} - 2\lambda\mathbf{a}$$

e igualando al vector cero esta derivada simbólica tenemos:

$$\begin{aligned} 2\mathbf{m} - 2\lambda\mathbf{a} &= \mathbf{0} \Leftrightarrow \\ 2(\mathbf{m} - \lambda\mathbf{a}) &= \mathbf{0} \Leftrightarrow \\ \mathbf{m} - \lambda\mathbf{a} &= \mathbf{0} \Leftrightarrow \\ \mathbf{m} &= \lambda\mathbf{a} \Leftrightarrow \\ \mathbf{a} &= \frac{1}{\lambda}\mathbf{m} \end{aligned}$$

pero  $\mathbf{a}'\mathbf{a} = 1$

$$\Rightarrow \mathbf{a}'\mathbf{a} = \left( \frac{1}{\lambda}\bar{\mathbf{x}} \right)' \left( \frac{1}{\lambda}\bar{\mathbf{x}} \right) = \frac{1}{\lambda^2} \mathbf{m}'\mathbf{m} = 1$$

entonces

$$\begin{aligned} \lambda^2 &= |\mathbf{m}|^2 \\ \therefore \lambda &= |\mathbf{m}| = r \\ \therefore \mathbf{a} &= \frac{1}{r}\mathbf{m} \end{aligned}$$

Entonces el mínimo de  $S(\mathbf{a})$  se da cuando  $\mathbf{a} = \frac{1}{r}\mathbf{m}$  y éste es:

$$\begin{aligned} \min S(\mathbf{a}) &= 2 \left( 1 - \mathbf{m}' \left( \frac{1}{r} \mathbf{m} \right) \right) \\ &= 2 \left( 1 - \frac{1}{r} \mathbf{m}' \mathbf{m} \right) \\ &= 2 \left( 1 - \frac{1}{r} r^2 \right) \\ &= 2(1 - r) \end{aligned}$$

A la cantidad  $2(1 - r)$  se le llama *varianza esférica*, sin embargo, es más usual trabajar, con datos direccionales, con  $r$  en lugar de  $2(1 - r)$ , y ésta mide el agrupamiento alrededor de la dirección media. De nuevo, es fácil notar que  $r \approx 0$  cuando las observaciones están uniformemente distribuidas y  $r \approx 1$  cuando están muy concentradas en algún punto.

## Capítulo 4

# COMPONENTES PRINCIPALES PARA DATOS DIRECCIONALES

### 4.1 Introducción

Este capítulo se centrará en ver que la información que se obtiene de las componentes principales cuando se trabaja con datos direccionales al igual que cuando se trabaja con datos lineales, será básicamente descriptiva; sin embargo, tendrá un significado distinto a la que se obtiene cuando se utiliza un análisis de componentes principales para datos lineales.

Al igual que en el caso de datos lineales, las observaciones direccionales pueden seguir alguna de las formas de distribuciones esféricas: uniforme, unimodal, bimodal, etc. El problema de identificar la posible distribución que sigue la población de la que fueron tomadas las observaciones direccionales puede ser resuelto utilizando las proyecciones que mencionamos en el capítulo anterior, pero el utilizar el momento de inercia de los puntos de la muestra puede ser un método más efectivo, ya que como veremos más adelante, proporciona información importante sobre la dispersión de los mismos.

La interpretación que en este capítulo se da del momento de inercia se debe a Geoffrey S. Watson, quien por primera vez le da este significado en su artículo publicado en 1966 titulado: "*The statistics of orientation data*".

## 4.2 Formas de distribuciones esféricas

Las formas básicas de distribuciones esféricas son las siguientes:

### Distribución uniforme

Se dice que los datos direccionales siguen una distribución uniforme si los puntos están uniformemente distribuidos sobre la esfera. En este caso, se puede ver que no existe alguna dirección en la que se acumule un número importante de observaciones. Las observaciones de la figura 4.2.1 siguen una distribución uniforme<sup>1</sup>.

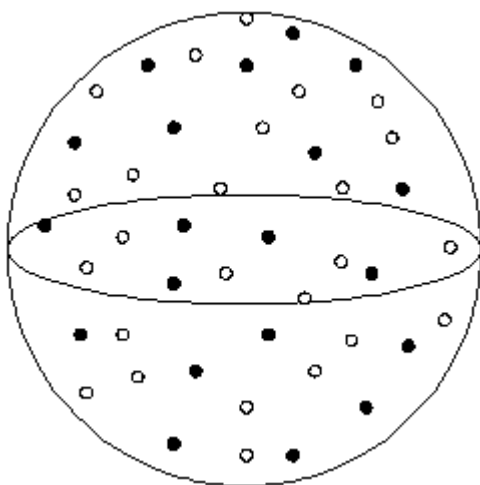


Fig. 4.2.1 Distribucion uniforme.

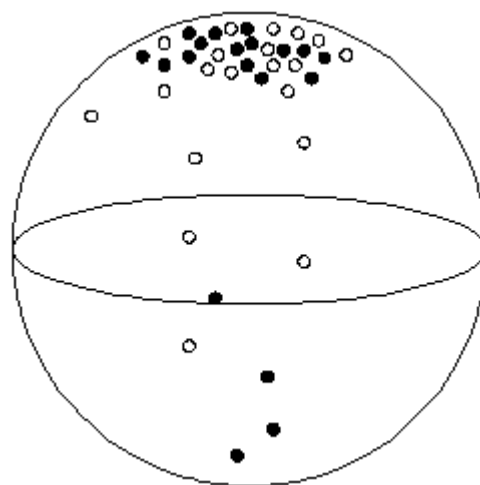


Fig. 4.2.2 Distribucion unimodal.

### Distribución unimodal

En este caso, las observaciones tienen una dirección modal, es decir, existe una dirección en la que se acumula un buen número de datos. Si hay simetría rotacional alrededor de la dirección modal, llamamos a la distribución una *distribución unipolar*, un ejemplo de este tipo de distribuciones es cuando los puntos están distribuidos sobre la esfera en la forma de un trompo, como se puede ver en la figura 4.2.2.

<sup>1</sup>En las figuras que ilustran las distintas formas de distribuciones esféricas, los círculos abiertos denotan observaciones que se encuentran por detrás de la esfera, mientras que los círculos cerrados son las observaciones que son visibles frontalmente.

### Distribución bimodal

En este tipo de distribuciones existen dos direcciones modales. Si la distribución es simétrica rotacionalmente alrededor del eje que une ambas modas o alrededor del eje que une las dos antimodas, se dice que es una *distribución bipolar*. Distribuciones bipolares suceden también en el caso axial cuando los puntos muestrales están concentrados en las puntas de un eje. Podemos ver un ejemplo de una distribución bimodal en la figura 4.2.3.

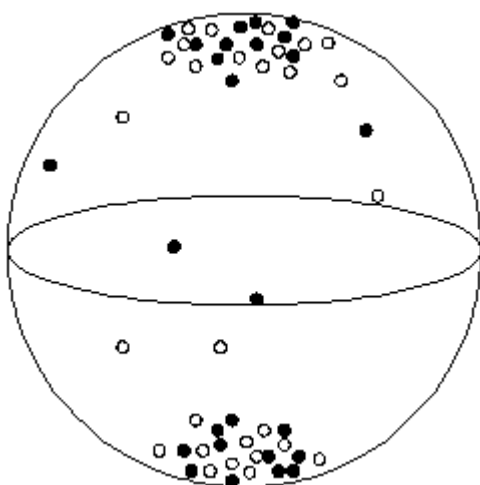


Fig. 4.2.3 Distribucion bimodal.

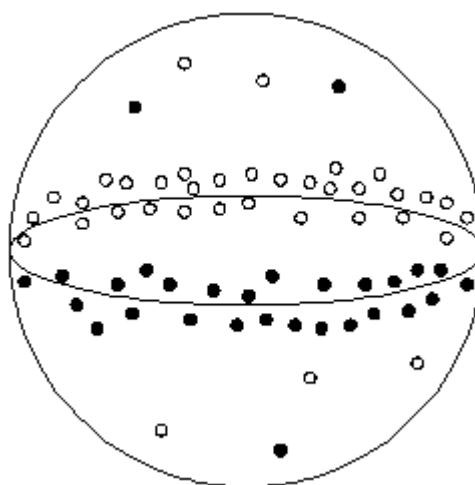


Fig. 4.2.2 Distribucion cinturón.

### Distribución cinturón

Si una distribución axial tiene su máximo en un plano alrededor de un círculo máximo en lugar de los puntos finales de un eje, se dice que la distribución que siguen los datos es una distribución cinturón, es decir, cuando las observaciones están distribuidas en un patrón elíptico, como un cinturón que rodea al círculo máximo. Si existe una simetría rotacional alrededor del eje perpendicular al plano que contiene al círculo máximo, se dice que es una *distribución cinturón simétrica*. En la figura 4.2.4 se puede ver que las observaciones rodean al ecuador.



### 4.3 Momento de inercia y ejes principales

Retomando lo visto en el capítulo 2, para obtener el momento de inercia,  $I$ , alrededor de cualquier eje tenemos que

$$I = \sum_{i=1}^n m_i r_i^2$$

donde  $r_i$  es la distancia perpendicular de la partícula  $i$  – *sima* al eje de referencia y  $m_i$  es la masa de la partícula  $i$  – *ésima*. Obviamente, el valor que tome el momento de inercia dependerá del eje de referencia que se considere.

Ahora, si tomamos  $n$  partículas y a cada una le asignamos una masa de tamaño 1, tenemos que el momento de inercia alrededor de un eje cualquiera, será:

$$I = \sum_{i=1}^n r_i^2$$

es decir, estará dado únicamente por las distancias cuadradas de las partículas al eje.

Por otro lado, sean  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  los  $n$  vectores de dirección, normalizados, de  $n$  observaciones, a las que asignaremos, una masa de tamaño 1 a cada una. Supongamos que  $\mathbf{u} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$  es una dirección fija, ( $|\mathbf{u}| = 1$ ) y sea  $\theta_i$  el ángulo entre la observación  $i$  – *ésima* y la dirección fija  $\mathbf{u}$ . Finalmente, sea  $r_i$  la distancia de la observación a  $\mathbf{u}$ , como se puede ver en la figura 4.3.1.

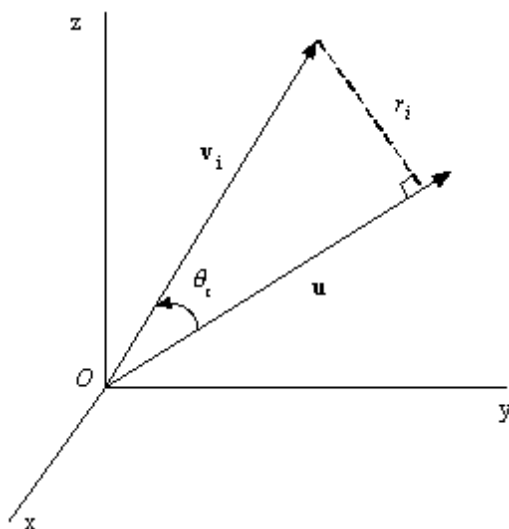


Fig. 4.3.1. Distancia de la observación  $i$  a  $\mathbf{u}$ .

Es posible notar que

$$\begin{aligned}\operatorname{sen} \theta_i &= \frac{r_i}{|\mathbf{v}_i|} = r_i \\ \implies r_i &= \operatorname{sen} \theta_i\end{aligned}$$

donde  $\mathbf{v}_i = l_i \mathbf{i} + m_i \mathbf{j} + n_i \mathbf{k}$ , además,

$$\operatorname{sen} \theta_i = \sqrt{1 - \cos^2 \theta_i}$$

y

$$\begin{aligned}\cos \theta_i &= \frac{\mathbf{v}_i \cdot \mathbf{u}}{|\mathbf{v}_i| |\mathbf{u}|} = \mathbf{v}_i \cdot \mathbf{u} \\ \implies \cos^2 \theta_i &= (\mathbf{v}_i \cdot \mathbf{u})^2 \\ \implies \operatorname{sen} \theta_i &= \sqrt{1 - (\mathbf{v}_i \cdot \mathbf{u})^2}\end{aligned}$$

así que

$$r_i = \operatorname{sen} \theta_i = \sqrt{1 - (\mathbf{v}_i \cdot \mathbf{u})^2}$$

De lo anterior podemos ver que el momento de inercia de las  $n$  observaciones alrededor de la dirección  $\mathbf{u}$  será:

$$I = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n \operatorname{sen}^2 \theta_i = \sum_{i=1}^n (1 - \cos^2 \theta_i) \quad (4.1)$$

$$\begin{aligned}&= \sum_{i=1}^n [1 - (\mathbf{v}_i \cdot \mathbf{u})^2] \\ &= \sum_{i=1}^n [1 - (l_i u + m_i v + n_i w)^2] \\ &= \sum_{i=1}^n 1 - \sum_{i=1}^n (l_i u + m_i v + n_i w)^2 \\ &= n - \sum_{i=1}^n (l_i u + m_i v + n_i w)^2 \quad (4.2)\end{aligned}$$

Como  $|\mathbf{u}| = |\mathbf{u}|^2 = 1$ , podemos expresar al momento de inercia de la siguiente manera:

$$\begin{aligned}I &= |\mathbf{u}|^2 n - \sum_{i=1}^n (l_i u + m_i v + n_i w)^2 \\ &= (u^2 + v^2 + w^2) n - \sum_{i=1}^n (l_i u + m_i v + n_i w)^2\end{aligned}$$

$$\begin{aligned}
&= n(u^2 + v^2 + w^2) - \sum_{i=1}^n (u^2 l_i^2 + 2uwl_i m_i + 2uwl_i n_i + 2vwm_i n_i + v^2 m_i^2 + w^2 n_i^2) \\
&= u^2 n - u^2 \sum_{i=1}^n l_i^2 - uv \sum_{i=1}^n l_i m_i - uw \sum_{i=1}^n l_i n_i - uv \sum_{i=1}^n l_i m_i + v^2 n - v^2 \sum_{i=1}^n m_i^2 \\
&\quad - vw \sum_{i=1}^n m_i n_i - uw \sum_{i=1}^n l_i n_i - vw \sum_{i=1}^n m_i n_i + w^2 n - w^2 \sum_{i=1}^n n_i^2 \\
&= \left( un - u \sum_{i=1}^n l_i^2 - v \sum_{i=1}^n l_i m_i - w \sum_{i=1}^n l_i n_i \right) u \\
&\quad + \left( -u \sum_{i=1}^n l_i m_i + vn - v \sum_{i=1}^n m_i^2 - w \sum_{i=1}^n m_i n_i \right) v \\
&\quad + \left( -u \sum_{i=1}^n l_i n_i - v \sum_{i=1}^n m_i n_i + wn - w \sum_{i=1}^n n_i^2 \right) w \\
&= \begin{pmatrix} un - u \sum_{i=1}^n l_i^2 - v \sum_{i=1}^n l_i m_i - w \sum_{i=1}^n l_i n_i \\ -u \sum_{i=1}^n l_i m_i + vn - v \sum_{i=1}^n m_i^2 - w \sum_{i=1}^n m_i n_i \\ -u \sum_{i=1}^n l_i n_i - v \sum_{i=1}^n m_i n_i + wn - w \sum_{i=1}^n n_i^2 \end{pmatrix}' \begin{pmatrix} u \\ v \\ w \end{pmatrix} \\
&= \begin{pmatrix} u(n - \sum_{i=1}^n l_i^2) - v \sum_{i=1}^n l_i m_i - w \sum_{i=1}^n l_i n_i \\ -u \sum_{i=1}^n l_i m_i + v(n - \sum_{i=1}^n m_i^2) - w \sum_{i=1}^n m_i n_i \\ -u \sum_{i=1}^n l_i n_i - v \sum_{i=1}^n m_i n_i + w(n - \sum_{i=1}^n n_i^2) \end{pmatrix}' \begin{pmatrix} u \\ v \\ w \end{pmatrix} \\
&= (u, v, w) \begin{pmatrix} n - \sum_{i=1}^n l_i^2 & -\sum_{i=1}^n l_i m_i & -\sum_{i=1}^n l_i n_i \\ -\sum_{i=1}^n l_i m_i & n - \sum_{i=1}^n m_i^2 & -\sum_{i=1}^n m_i n_i \\ -\sum_{i=1}^n l_i n_i & -\sum_{i=1}^n m_i n_i & n - \sum_{i=1}^n n_i^2 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix}
\end{aligned}$$

sea

$$\mathbf{B} = \begin{pmatrix} n - \sum_{i=1}^n l_i^2 & -\sum_{i=1}^n l_i m_i & -\sum_{i=1}^n l_i n_i \\ -\sum_{i=1}^n l_i m_i & n - \sum_{i=1}^n m_i^2 & -\sum_{i=1}^n m_i n_i \\ -\sum_{i=1}^n l_i n_i & -\sum_{i=1}^n m_i n_i & n - \sum_{i=1}^n n_i^2 \end{pmatrix}$$

$$\therefore I = \mathbf{u}' \mathbf{B} \mathbf{u} \tag{4.3}$$

por otro lado, sea

$$\mathbf{T} = \begin{pmatrix} \sum_{i=1}^n l_i^2 & \sum_{i=1}^n l_i m_i & \sum_{i=1}^n l_i n_i \\ \sum_{i=1}^n l_i m_i & \sum_{i=1}^n m_i^2 & \sum_{i=1}^n m_i n_i \\ \sum_{i=1}^n l_i n_i & \sum_{i=1}^n m_i n_i & \sum_{i=1}^n n_i^2 \end{pmatrix}$$

entonces la matriz  $\mathbf{B}$  puede ser escrita de la siguiente manera:

$$\begin{aligned}\mathbf{B} &= \begin{pmatrix} n & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & n \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^n l_i^2 & \sum_{i=1}^n l_i m_i & \sum_{i=1}^n l_i n_i \\ \sum_{i=1}^n l_i m_i & \sum_{i=1}^n m_i^2 & \sum_{i=1}^n m_i n_i \\ \sum_{i=1}^n l_i n_i & \sum_{i=1}^n m_i n_i & \sum_{i=1}^n n_i^2 \end{pmatrix} \\ &= n\mathbf{I} - \mathbf{T}\end{aligned}$$

La matriz  $\mathbf{T}$  es llamada *matriz de orientación* [8]. Si colocamos las  $n$  observaciones en una matriz de datos,  $\mathbf{L}$ :

$$\mathbf{L} = \begin{pmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ \vdots & \vdots & \vdots \\ l_n & m_n & n_n \end{pmatrix}$$

podemos ver que

$$\mathbf{T} = \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i' = \mathbf{L}'\mathbf{L}$$

que es la matriz de sumas de cuadrados y productos cruzados de los cosenos directores de las  $n$  observaciones  $(l_i, m_i, n_i)$ ,  $i = 1, 2, \dots, n$ .

Cabe destacar, por un lado, que el cuadrado del coseno del ángulo en 4.1 garantiza el hecho de que estemos hablando de ángulos entre dos líneas  $\mathbf{u}$  y  $\mathbf{v}_i$  que pasan por el origen, por esta razón, a los vectores  $\mathbf{v}_i$ ,  $i = 1, 2, \dots, n$  se les consideran ejes en vez de direcciones.

Es posible observar en la ecuación 4.3 que la variación del momento de inercia estará dada de acuerdo a la elección de la dirección fija, de esta manera el momento de inercia resulta muy útil para la obtención de información sobre la dispersión de las observaciones. De acuerdo a lo anterior podemos plantearnos el problema de encontrar la dirección para la cual el momento de inercia sea mínimo, dicho de otra manera, encontrar la dirección para la cual, el total de las distancias cuadradas de todas las observaciones a esa dirección sea mínima, así, definimos a la *primera dirección principal* o *primer eje principal* como la dirección  $\mathbf{b}_1$  la cual minimiza los ángulos  $\theta_i$ , entre  $\mathbf{b}_1$  y  $\mathbf{v}_i$ ,  $i = 1, 2, \dots, n$ , esto es, cuando

$$I = \sum_{i=1}^n [1 - \cos^2 \theta_i] = \mathbf{b}_1' \mathbf{B} \mathbf{b}_1$$

es mínimo.

La *segunda dirección principal* o *segundo eje principal*  $\mathbf{b}_2$  es la dirección, alrededor de la cual, se produce el momento de inercia segundo más pequeño, es decir, la dirección que minimiza los ángulos entre  $\mathbf{b}_2$  y  $\mathbf{v}_i$ ,  $i = 1, 2, \dots, n$ , pero que además, sea perpendicular a  $\mathbf{b}_1$ .

Por último, podemos encontrar la *tercera dirección principal* o *tercer eje principal*  $\mathbf{b}_3$  que es la dirección alrededor de la cual las observaciones tienen el momento de inercia más grande, en comparación con las otras dos direcciones, pero que además sea perpendicular a  $\mathbf{b}_1$  y a  $\mathbf{b}_2$ .

## 4.4 Obtención de las direcciones principales a partir de la matriz $\mathbf{B}$

Para obtener las direcciones principales debemos plantearnos el problema de encontrar la dirección  $\mathbf{b}_1$  tal que  $I$  sea mínimo y además  $\mathbf{b}'_1 \mathbf{b}_1 = 1$ , como en el caso de componentes principales para datos lineales, tenemos que

$$\text{Min } \mathbf{b}'_1 \mathbf{B} \mathbf{b}_1 \quad \text{sujeto a } \mathbf{b}'_1 \mathbf{b}_1 = 1$$

Para resolver este problema, de nueva cuenta utilizaremos el método de multiplicadores de Lagrange. Entonces planteamos la siguiente función:

$$\varphi = \mathbf{b}'_1 \mathbf{B} \mathbf{b}_1 - \beta (\mathbf{b}'_1 \mathbf{b}_1 - 1)$$

que lleva al problema de encontrar el valor mínimo de la función anterior, por lo que derivando  $\varphi$  con respecto a  $\mathbf{b}_1$ , tenemos:

$$\frac{d\varphi}{d\mathbf{b}_1} = 2\mathbf{B}\mathbf{b}_1 - 2\beta\mathbf{b}_1$$

e igualando esta derivada simbólica al vector cero, tenemos:

$$\begin{aligned} 2\mathbf{B}\mathbf{b}_1 - 2\beta\mathbf{b}_1 &= \mathbf{0} \Leftrightarrow \\ 2(\mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}_1) &= \mathbf{0} \Leftrightarrow \\ \mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}_1 &= \mathbf{0} \Leftrightarrow \\ (\mathbf{B} - \beta\mathbf{I})\mathbf{b}_1 &= \mathbf{0} \end{aligned} \tag{4.4}$$

la última ecuación tiene solución no trivial si  $\mathbf{B} - \beta\mathbf{I}$  es una matriz singular, por lo que su determinante debe ser igual a cero:

$$|\mathbf{B} - \beta\mathbf{I}| = 0 \tag{4.5}$$

que como sabemos es el polinomio característico de  $\mathbf{B}$  y su grado es 3, dado que éste es el orden de  $\mathbf{B}$ . Como vimos

$$\begin{aligned}\mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}_1 &= \mathbf{0} \Leftrightarrow \\ \mathbf{B}\mathbf{b}_1 &= \beta\mathbf{b}_1\end{aligned}$$

por lo que para resolver el problema basta con encontrar los valores y vectores propios de  $\mathbf{B}$ .

Ahora el momento de inercia alrededor de  $\mathbf{b}_1$  estará dado por:

$$I = \mathbf{b}'_1\mathbf{B}\mathbf{b}_1$$

por otro lado,

$$\begin{aligned}\mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}_1 &= \mathbf{0} \Leftrightarrow \\ \mathbf{B}\mathbf{b}_1 &= \beta\mathbf{b}_1 \Leftrightarrow \\ \mathbf{b}'_1\mathbf{B}\mathbf{b}_1 &= \beta\mathbf{b}'_1\mathbf{b}_1 \\ \implies \mathbf{b}'_1\mathbf{B}\mathbf{b}_1 &= \beta\end{aligned}$$

ya que  $\mathbf{b}'_1\mathbf{b}_1 = 1$

$$\therefore I = \beta$$

entonces podemos observar que el momento de inercia alrededor del primer eje principal es  $\beta$ , una raíz del polinomio 4.5, por lo que nuestra primera dirección principal es el vector propio asociado al valor propio más pequeño de  $\mathbf{B}$ ,  $\beta_1$ .

Ahora, podríamos encontrar una segunda dirección, digamos  $\mathbf{b}_2$ , tal que, alrededor de ella, se obtenga el momento de inercia más pequeño pero que sea mayor al que se obtiene alrededor del primer eje principal, y además, que sea perpendicular a éste, por lo que ahora tenemos que plantear un nuevo problema:

$$\text{Min } \mathbf{b}'_2\mathbf{B}\mathbf{b}_2 \quad \text{sujeto a } \mathbf{b}'_2\mathbf{b}_2 = 1 \quad \text{y} \quad \mathbf{b}'_2\mathbf{b}_1 = 0$$

nuevamente, utilizando multiplicadores de Lagrange definiremos la siguiente función:

$$\psi = \mathbf{b}'_2\mathbf{B}\mathbf{b}_2 - \beta(\mathbf{b}'_2\mathbf{b}_2 - 1) - \gamma\mathbf{b}'_2\mathbf{b}_1$$

para resolver el problema basta con encontrar el mínimo de  $\psi$ . Derivando  $\psi$  con respecto a  $\mathbf{b}_2$  e igualando al vector cero:

$$\begin{aligned}\frac{d\psi}{d\mathbf{b}_2} &= 2\mathbf{B}\mathbf{b}_2 - 2\beta\mathbf{b}_2 - \gamma\mathbf{b}_1 \\ 2\mathbf{B}\mathbf{b}_2 - 2\beta\mathbf{b}_2 - \gamma\mathbf{b}_1 &= \mathbf{0}\end{aligned}\tag{4.6}$$

multiplicando por  $\mathbf{b}'_1$  por la izquierda tenemos:

$$2\mathbf{b}'_1\mathbf{B}\mathbf{b}_2 - 2\beta\mathbf{b}'_1\mathbf{b}_2 - \gamma\mathbf{b}'_1\mathbf{b}_1 = 0$$

luego si multiplicamos por la izquierda a 4.3 por  $\mathbf{b}'_2$ , tenemos

$$\begin{aligned}\mathbf{b}'_2(\mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}_1) &= \mathbf{b}'_2\mathbf{0} \implies \\ \mathbf{b}'_2\mathbf{B}\mathbf{b}_1 - \beta\mathbf{b}'_2\mathbf{b}_1 &= 0 \implies \\ \mathbf{b}'_2\mathbf{B}\mathbf{b}_1 &= \beta\mathbf{b}'_2\mathbf{b}_1\end{aligned}$$

pero queremos que  $\mathbf{b}'_2\mathbf{b}_1 = \mathbf{b}'_1\mathbf{b}_2 = 0$ , por lo que tenemos que  $\gamma = 0$  ya que

$$\mathbf{b}'_2\mathbf{B}\mathbf{b}_1 = \mathbf{b}'_2\mathbf{b}_1 = 0 \tag{4.7}$$

y  $\mathbf{b}'_1\mathbf{b}_1 = 1$ , por lo que la ecuación 4.5 puede ser reducida a:

$$\begin{aligned}\frac{d\psi}{d\mathbf{b}_2} &= 2\mathbf{B}\mathbf{b}_2 - 2\beta\mathbf{b}_2 \\ 2\mathbf{B}\mathbf{b}_2 - 2\beta\mathbf{b}_2 &= \mathbf{0} \quad \Leftrightarrow \\ 2(\mathbf{B}\mathbf{b}_2 - \beta\mathbf{b}_2) &= \mathbf{0} \quad \Leftrightarrow \\ (\mathbf{B} - \beta\mathbf{I})\mathbf{b}_2 &= \mathbf{0}\end{aligned}$$

al igual que para obtener  $\mathbf{b}_1$ , solamente basta con tomar el segundo valor propio más pequeño  $\beta_2$  y  $\mathbf{b}_2$  su vector propio correspondiente. Así, la segunda dirección principal será, precisamente  $\mathbf{b}_2$ . A partir de este resultado es posible ver que la tercera dirección principal será el vector propio,  $\mathbf{b}_3$  asociado al valor propio más grande  $\beta_3$ .

Así, tenemos que los valores propios de la matriz  $\mathbf{B}$  tienen la siguiente relación de orden:

$$\beta_1 \leq \beta_2 \leq \beta_3$$

y el momento de inercia alrededor de cada eje principal, será

$$I_{\mathbf{b}_i} = \mathbf{b}'_i\mathbf{B}\mathbf{b}_i \quad i = 1, 2, 3$$

pero sabemos que cada  $\mathbf{b}_i$  es un vector propio de la matriz  $\mathbf{B}$ , entonces cumplen lo siguiente:

$$\begin{aligned}\mathbf{B}\mathbf{b}_i - \beta_i\mathbf{b}_i &= \mathbf{0} \Leftrightarrow \\ \mathbf{B}\mathbf{b}_i &= \beta_i\mathbf{b}_i \Leftrightarrow \\ \mathbf{b}'_i\mathbf{B}\mathbf{b}_i &= \beta_i\mathbf{b}'_i\mathbf{b}_i \\ \implies \mathbf{b}'_i\mathbf{B}\mathbf{b}_i &= \beta_i\end{aligned}$$

$$\therefore I_{\mathbf{b}_i} = \mathbf{b}'_i \mathbf{B} \mathbf{b}_i = \beta_i \quad i = 1, 2, 3 \quad (4.8)$$

Así, si los tres valores propios de  $\mathbf{B}$  son distintos, los vectores propios  $\mathbf{b}_1$ ,  $\mathbf{b}_2$  y  $\mathbf{b}_3$  pueden ser vistos como los cosenos directores de un conjunto de tres ejes coordenados. Cuando algunos de los valores propios de  $\mathbf{B}$  sean iguales, no existe problema, pues los vectores propios se escogen de tal manera que sean ortogonales.

Al tener todos los vectores propios podemos formar a la matriz  $\mathbf{V}$  de  $3 \times 3$ , donde:  $\mathbf{V} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$  y  $\mathbf{b}'_j = (b_{1j}, b_{2j}, b_{3j})$   $j = 1, 2, 3$ , y al tomar los vectores  $\mathbf{b}_j$  de tal manera que  $\mathbf{b}'_j \mathbf{b}_j = 1$ , la matriz  $\mathbf{V}$  es ortogonal, es decir  $\mathbf{V}\mathbf{V}' = \mathbf{I}$ . Así, al multiplicar los cosenos directores  $(l_i, m_i, n_i)$   $i = 1, 2, \dots, n$  por la matriz  $\mathbf{V}'$ , obtendremos los nuevos cosenos directores  $(l_i^*, m_i^*, n_i^*)$   $i = 1, 2, \dots, n$  de las observaciones pero con respecto a los nuevos ejes:

$$\begin{aligned} \mathbf{v}_i^* &= \mathbf{V}' \mathbf{v}_i \\ \begin{pmatrix} l_i^* \\ m_i^* \\ n_i^* \end{pmatrix} &= \begin{pmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{pmatrix} \begin{pmatrix} l_i \\ m_i \\ n_i \end{pmatrix} \end{aligned}$$

Si hacemos el producto  $\mathbf{V}'\mathbf{B}\mathbf{V}$  formaremos una matriz de  $3 \times 3$  a la que denotaremos por  $\mathbf{\Lambda}^*$ , cuyos elementos sobre la diagonal serán los momentos de inercia con respecto a cada uno de los tres ejes principales. Esta matriz es diagonal ya que es posible ver que si hacemos el producto de matrices  $\mathbf{V}'\mathbf{B}\mathbf{V}$  tendremos:

$$\begin{aligned} \mathbf{V}'\mathbf{B}\mathbf{V} &= \begin{pmatrix} \mathbf{b}'_{(1)} \\ \mathbf{b}'_{(2)} \\ \mathbf{b}'_{(3)} \end{pmatrix} (\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3) (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \\ &= \begin{pmatrix} \mathbf{b}'_{(1)} (\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3) \\ \mathbf{b}'_{(2)} (\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3) \\ \mathbf{b}'_{(3)} (\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3) \end{pmatrix} (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \\ &= \begin{pmatrix} \mathbf{b}'_{(1)} \mathbf{B} \\ \mathbf{b}'_{(2)} \mathbf{B} \\ \mathbf{b}'_{(3)} \mathbf{B} \end{pmatrix} (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \\ &= \begin{pmatrix} \mathbf{b}'_{(1)} \mathbf{B} \mathbf{b}_1 & \mathbf{b}'_{(1)} \mathbf{B} \mathbf{b}_2 & \mathbf{b}'_{(1)} \mathbf{B} \mathbf{b}_3 \\ \mathbf{b}'_{(2)} \mathbf{B} \mathbf{b}_1 & \mathbf{b}'_{(2)} \mathbf{B} \mathbf{b}_2 & \mathbf{b}'_{(2)} \mathbf{B} \mathbf{b}_3 \\ \mathbf{b}'_{(3)} \mathbf{B} \mathbf{b}_1 & \mathbf{b}'_{(3)} \mathbf{B} \mathbf{b}_2 & \mathbf{b}'_{(3)} \mathbf{B} \mathbf{b}_3 \end{pmatrix} \end{aligned}$$



donde  $\mathbf{B}_i$   $i = 1, 2, 3$  es el  $i$ -ésimo vector columna de la matriz  $\mathbf{B}$ . Tenemos por 4.8 y 4.7, que

$$\mathbf{b}'_{(i)} \mathbf{B} \mathbf{b}_j = \begin{cases} \beta_i & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

entonces tenemos que

$$\mathbf{\Lambda}^* = \mathbf{V}' \mathbf{B} \mathbf{V} = \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_2 & 0 \\ 0 & 0 & \beta_3 \end{pmatrix} \quad (4.9)$$

Despejando  $\mathbf{B}$  de la ecuación 4.9:

$$\begin{aligned} \mathbf{\Lambda}^* &= \mathbf{V}' \mathbf{B} \mathbf{V} \iff \\ \mathbf{V} \mathbf{\Lambda}^* &= \mathbf{V} \mathbf{V}' \mathbf{B} \mathbf{V} \iff \\ \mathbf{V} \mathbf{\Lambda}^* \mathbf{V}' &= \mathbf{B} \mathbf{V} \mathbf{V}' \\ \therefore \mathbf{V} \mathbf{\Lambda}^* \mathbf{V}' &= \mathbf{B} \end{aligned}$$

Por otra parte, podemos formar una nueva matriz de datos  $\mathbf{L}^*$  con los nuevos cosenos directores de las  $n$  observaciones :

$$\mathbf{L}^* = \begin{pmatrix} l_1^* & m_1^* & n_1^* \\ l_2^* & m_2^* & n_2^* \\ \vdots & \vdots & \vdots \\ l_n^* & m_n^* & n_n^* \end{pmatrix} = \mathbf{L} \mathbf{V}$$

y la matriz de orientación  $\mathbf{T}^*$  de las observaciones cuyos cosenos directores estarán dados de acuerdo a los ejes principales, será:

$$\mathbf{T}^* = \mathbf{L}^* \mathbf{L}^* = (\mathbf{L} \mathbf{V})' \mathbf{L} \mathbf{V} = \mathbf{V}' \mathbf{L}' \mathbf{L} \mathbf{V} = \mathbf{V}' \mathbf{T} \mathbf{V}$$

y

$$\mathbf{B}^* = n\mathbf{I} - \mathbf{T}^* = n\mathbf{I} - \mathbf{V}' \mathbf{T} \mathbf{V} \quad (4.10)$$

Despejando  $\mathbf{V}' \mathbf{T} \mathbf{V}$  de la ecuación 4.10, tenemos:

$$\mathbf{V}' \mathbf{T} \mathbf{V} = n\mathbf{I} - \mathbf{B}^* \quad (4.11)$$

como la matriz  $\mathbf{V}$  es ortogonal, podemos multiplicar 4.11 por la izquierda por  $\mathbf{V}$  y por la derecha por  $\mathbf{V}'$  :

$$\begin{aligned} \mathbf{V} \mathbf{V}' \mathbf{T} \mathbf{V} \mathbf{V}' &= \mathbf{V} (n\mathbf{I} - \mathbf{B}^*) \mathbf{V}' \\ \implies \mathbf{T} &= \mathbf{V} (n\mathbf{I} - \mathbf{B}^*) \mathbf{V}' \\ \implies \mathbf{T} &= n\mathbf{V} \mathbf{V}' - \mathbf{V} \mathbf{B}^* \mathbf{V}' \\ \implies \mathbf{T} &= n\mathbf{I} - \mathbf{V} \mathbf{B}^* \mathbf{V}' \end{aligned}$$

multiplicando por  $-1$  ambos lados de la igualdad anterior:

$$-\mathbf{T} = \mathbf{V}\mathbf{B}^*\mathbf{V}' - n\mathbf{I}$$

sumando  $n\mathbf{I}$  a ambos lados de la igualdad:

$$\begin{aligned} n\mathbf{I} - \mathbf{T} &= \mathbf{V}\mathbf{B}^*\mathbf{V}' \\ \therefore \mathbf{B} &= \mathbf{V}\mathbf{B}^*\mathbf{V}' \end{aligned} \quad (4.12)$$

Ahora, si despejamos a  $\mathbf{B}^*$  de 4.11 tendremos:

$$\begin{aligned} \mathbf{V}\mathbf{B}^*\mathbf{V}' &= \mathbf{B} \iff \\ \mathbf{V}'\mathbf{V}\mathbf{B}^*\mathbf{V}' &= \mathbf{V}'\mathbf{B} \iff \\ \mathbf{B}^*\mathbf{V}'\mathbf{V} &= \mathbf{V}'\mathbf{B}\mathbf{V} \iff \\ \mathbf{B}^* &= \mathbf{V}'\mathbf{B}\mathbf{V} \\ \therefore \mathbf{B}^* &= \mathbf{\Lambda}^* \end{aligned}$$

Al calcular la traza de la matriz  $\mathbf{B}$ , podemos ver que:

$$\begin{aligned} tr(\mathbf{B}) &= tr(n\mathbf{I} - \mathbf{T}) = \left(n - \sum_{i=1}^n l_i^2\right) + \left(n - \sum_{i=1}^n m_i^2\right) + \left(n - \sum_{i=1}^n n_i^2\right) \\ &= n - \sum_{i=1}^n l_i^2 + n - \sum_{i=1}^n m_i^2 + n - \sum_{i=1}^n n_i^2 \\ &= 3n - \left(\sum_{i=1}^n l_i^2 + \sum_{i=1}^n m_i^2 + \sum_{i=1}^n n_i^2\right) \\ &= 3n - \sum_{i=1}^n (l_i^2 + m_i^2 + n_i^2) \\ &= 3n - n = 2n \end{aligned}$$

y como  $\mathbf{V}\mathbf{\Lambda}^*\mathbf{V}' = \mathbf{B}$

$$\begin{aligned} tr(\mathbf{B}) &= tr(\mathbf{V}\mathbf{\Lambda}^*\mathbf{V}') \\ &= tr(\mathbf{V}'\mathbf{V}\mathbf{\Lambda}^*) \\ &= tr(\mathbf{\Lambda}^*) \\ &= tr(\mathbf{B}^*) \\ &= \sum_{i=1}^3 \beta_i = 2n \end{aligned} \quad (4.13)$$

que nos dice que la traza de la matriz  $\mathbf{B}$  es igual a la suma de los tres momentos de inercia alrededor de los tres ejes principales.

## 4.5 Obtención de las direcciones principales a partir de la matriz $\mathbf{T}$

Para propósitos estadísticos es más conveniente trabajar con la matriz de orientación  $\mathbf{T}$  en vez de la matriz  $\mathbf{B}$ . Ahora, al trabajar con la matriz  $\mathbf{T}$  tendremos que la primera dirección principal  $\mathbf{t}$ , es la dirección que minimiza los ángulos  $\theta_i$  entre  $\mathbf{t}$  y  $\mathbf{v}_i$ ,  $i = 1, 2, \dots, n$ , esto es, cuando

$$\begin{aligned} \sum_{i=1}^n \cos^2 \theta_i &= \sum_{i=1}^n (\mathbf{t}'\mathbf{v}_i)^2 \\ &= \sum_{i=1}^n \mathbf{t}'\mathbf{v}_i\mathbf{v}_i'\mathbf{t} \\ &= \mathbf{t}' \left( \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i' \right) \mathbf{t} \\ &= \mathbf{t}'\mathbf{T}\mathbf{t} \end{aligned}$$

es máximo. Entonces el problema de encontrar la primera dirección principal,  $\mathbf{t}_1$ , queda planteado de la siguiente manera:

$$\text{Max } \mathbf{t}'_1\mathbf{T}\mathbf{t}_1 \quad \text{sujeto a } \mathbf{t}'_1\mathbf{t}_1 = 1$$

y al igual que en el caso de encontrar los ejes principales a partir de la matriz  $\mathbf{B}$  es posible ver que la dirección  $\mathbf{t}_1$  que maximiza  $\mathbf{t}'\mathbf{T}\mathbf{t}$  es el vector propio asociado al valor propio  $\tau$  más grande de la matriz  $\mathbf{T}$ . Para encontrar la segunda dirección seguimos el mismo procedimiento y vemos que la segunda dirección principal será el vector propio asociado al segundo valor propio más grande y la tercera dirección principal será la dirección asociada al valor propio más pequeño  $\tau_3$ .

Por otro lado, podemos establecer la siguiente relación de orden entre los valores propios de  $\mathbf{T}$ :

$$\tau_1 \geq \tau_2 \geq \tau_3$$

Tenemos que los valores y vectores propios de  $\mathbf{T}$  cumplen la siguiente ecuación:

$$\mathbf{T}\mathbf{t} = \tau\mathbf{t} \tag{4.14}$$

Por otra parte, los valores y vectores propios de  $\mathbf{B} = n\mathbf{I} - \mathbf{T}$  cumplen lo siguiente:

$$\begin{aligned}
 (n\mathbf{I} - \mathbf{T}) \mathbf{b} - \beta \mathbf{b} &= \mathbf{0} \iff \\
 n\mathbf{Ib} - \mathbf{Tb} - \beta \mathbf{b} &= \mathbf{0} \iff \\
 \mathbf{Tb} &= n\mathbf{Ib} - \beta \mathbf{b} \iff \\
 \mathbf{Tb} &= (n - \beta) \mathbf{b}
 \end{aligned} \tag{4.15}$$

Así, de las ecuaciones 4.14 y 4.14 podemos afirmar lo siguiente:

$$\mathbf{t}_i = \mathbf{b}_i \quad i = 1, 2, 3 \tag{4.16}$$

y

$$\tau_i = n - \beta_i \quad i = 1, 2, 3$$

De esta manera, por la ecuación 4.1 el momento de inercia alrededor de cada uno de los ejes principales será:

$$I_{\mathbf{t}_i} = n - \mathbf{t}_i' \mathbf{T} \mathbf{t}_i \quad i = 1, 2, 3$$

de otra manera:

$$I_{\mathbf{t}_i} = n - \tau_i \quad i = 1, 2, 3 \tag{4.17}$$

además podemos ver que

$$\begin{aligned}
 tr(\mathbf{T}) &= \sum_{i=1}^n l_i^2 + \sum_{i=1}^n m_i^2 + \sum_{i=1}^n n_i^2 \\
 &= \sum_{i=1}^n (l_i^2 + m_i^2 + n_i^2) \\
 &= n
 \end{aligned}$$

y que

$$\begin{aligned}
 \sum_{i=1}^3 \tau_i &= \sum_{i=1}^3 (n - \beta_i) \\
 &= \sum_{i=1}^3 n - \sum_{i=1}^3 \beta_i \\
 &= 3n - 2n = n
 \end{aligned} \tag{4.18}$$

## 4.6 Interpretación de las direcciones principales obtenidas a partir de la matriz $\mathbf{B}$

Como la distancias perpendiculares del punto  $(l_i, m_i, n_i)$  sobre la esfera a la dirección  $\mathbf{u}$  es la misma que la distancia perpendicular del punto diametralmente opuesto  $(-l_i, -m_i, -n_i)$  a  $\mathbf{u}$ , el momento de inercia no distingue entre los datos direccionales y los datos axiales. Sin embargo, el valor de  $r$ , que como vimos en el capítulo 3 también nos da información sobre la dispersión de los datos, puede ser utilizada para distinguir entre estos dos casos.

Ahora ya tenemos las herramientas suficientes para detectar la forma de la distribución de los puntos  $(l_i, m_i, n_i)$ , para esto, plantearemos casos basándonos en los valores de  $\beta_1, \beta_2, \beta_3$  y de  $r$ .

### Caso 1

**$\beta_3$  grande** De la ecuación 4.12 se sigue que  $\beta_1$  y  $\beta_2$  son pequeños. Este hecho nos indica que la suma de las distancias cuadradas de los puntos a la dirección de  $\mathbf{b}_3$  es grande pero pequeña a las direcciones de  $\mathbf{b}_1$  y  $\mathbf{b}_2$ . Como  $\mathbf{b}_1, \mathbf{b}_2$  y  $\mathbf{b}_3$  son ortogonales entre sí, entonces se deduce que la mayoría de las observaciones están concentradas en el plano que contiene a  $\mathbf{b}_1$  y  $\mathbf{b}_2$  cuya normal es  $\mathbf{b}_3$ . Entonces, valores grandes de  $\beta_3$  implican que los datos siguen una distribución cinturón.

### Caso 2

**$\beta_2$  y  $\beta_3$  grandes** En este caso  $\beta_1$  es pequeño. Al tener  $\beta_1$  un valor pequeño nos indica que la suma de las distancias cuadradas de las observaciones a esta dirección es pequeña, lo que significa que la mayoría de las observaciones están concentradas sobre la esfera alrededor de las puntas del diámetro generado por el vector  $\mathbf{b}_1$ . Como consecuencia de 4.12 y 4.8, la distribución es unimodal para  $r$  grande, de otra manera es bimodal.

### Caso 3

**$\beta_1 \approx \beta_2$**  De 4.8 y 4.12, el momento de inercia será casi constante alrededor de cualquier diámetro del círculo que contiene a  $\mathbf{b}_1$  y a  $\mathbf{b}_2$ . Así, existe una aproximación a una simetría rotacional alrededor de  $\mathbf{b}_3$ . Además, si  $\beta_3$  es grande entonces la

distribución será una distribución cinturón simétrica con simetría rotacional alrededor de  $\mathbf{b}_3$ . El caso en que  $\beta_2 \approx \beta_3$  es interpretado de manera similar.

#### Caso 4

$\beta_1 \approx \beta_2 \approx \beta_3$  De igual manera, a partir de 4.8 y 4.12, el momento de inercia será el mismo cualquiera que sea  $\mathbf{b}_i$ . Entonces, en este caso, la distribución es uniforme. Como el momento de inercia no hace distinción entre datos direccionales y axiales, esto implica que los ejes no tienen una orientación preferida.

Los casos anteriores son también aplicables al caso en que las observaciones sean circulares. En este caso, los valores propios de  $\mathbf{B}$  nos indicarán si se trata de distribuciones uniformes, unimodales o bimodales dependiendo de  $r$ .

## 4.7 Interpretación de las direcciones principales obtenidas a partir de la matriz $\mathbf{T}$

Los resultados anteriores son aplicables de igual forma cuando se encuentran las direcciones principales a partir de la matriz  $\mathbf{T}$ , pues como vimos en la ecuación 4.16,  $\mathbf{t}_i = \mathbf{b}_i$ ,  $i = 1, 2, 3$ . Entonces ahora tenemos los 4 casos siguientes:

#### Caso 1

$\tau_3$  **pequeño** Las ecuaciones 4.17 y 4.17 nos indican que entonces  $\tau_1$  y  $\tau_2$  son grandes y como consecuencia, la suma de las distancias cuadradas de los puntos a la dirección de  $\mathbf{t}_3$  es grande pero pequeñas a las direcciones  $\mathbf{t}_1$  y  $\mathbf{t}_2$ . Como  $\mathbf{t}_1$ ,  $\mathbf{t}_2$  y  $\mathbf{t}_3$  son ortogonales entre sí, entonces se deduce que la mayoría de las observaciones están concentradas en el plano que contiene a  $\mathbf{t}_1$  y  $\mathbf{t}_2$  cuya normal es  $\mathbf{t}_3$ . Entonces, valores pequeños de  $\tau_3$  implican que los datos siguen una distribución cinturón.

#### Caso 2

$\tau_2$  y  $\tau_3$  **pequeños** En este caso tenemos que  $\tau_1$  es grande lo que nos indica que la suma de las distancias cuadradas de las observaciones a esta dirección es pequeña, entonces tendremos observaciones que están concentradas sobre la esfera alrededor de las puntas del diámetro generado por el vector  $\mathbf{t}_1$ . Como consecuencia de 4.17 y 4.17, la distribución es unimodal para  $r$  grande y en caso contrario, es bimodal.

### Caso 3

$\tau_1 \approx \tau_2$  De 4.17 y 4.17, el momento de inercia será casi constante alrededor de cualquier diámetro del círculo que contiene a  $\mathbf{t}_1$  y a  $\mathbf{t}_2$ , entonces tendremos que va a existir una aproximación a una simetría rotacional alrededor de  $\mathbf{t}_3$ . Además, si  $\tau_3$  es pequeño entonces la distribución será una distribución cinturón simétrica con simetría rotacional alrededor de  $\mathbf{t}_3$ . El caso en que  $\tau_2 \approx \tau_3$  es interpretado de manera similar.

### Caso 4

$\tau_1 \approx \tau_2 \approx \tau_3$  Es posible ver, en las ecuaciones 4.17 y 4.17 que ahora tenemos el caso en que el momento de inercia será el mismo cualquiera que sea  $\mathbf{t}_i$ , por lo que tendremos datos que siguen una distribución uniforme. Como el momento de inercia no hace distinción entre datos direccionales y axiales, esto implica que los ejes no tienen una orientación preferida.

## 4.8 Ejemplo

En la tabla siguiente se muestran 75 medidas de las direcciones sobre el eje  $c$  de granos de calcita de las montañas Taconic de Nueva York [2]. Las coordenadas esféricas y los cosenos directores de estas mediciones:

$\theta$	$\phi$	$\cos \phi$ <i>I</i>	$\sin \phi$ <i>II</i>	$\sin \theta$ <i>III</i>	$l_i$ <i>I</i> $\times$ <i>III</i>	$m_i$ <i>II</i> $\times$ <i>III</i>	$n_i$ $\cos \theta$
62	315	0.7071	-0.7071	0.8829	0.6243	-0.6243	0.4695
46	18	0.9511	0.3090	0.7193	0.6841	0.2223	0.6947
57	51	0.6293	0.7771	0.8387	0.5278	0.6518	0.5446
74	212	-0.8480	-0.5299	0.9613	-0.8152	-0.5094	0.2756
71	203	-0.9205	-0.3907	0.9455	-0.8704	-0.3694	0.3256
65	44	0.7193	0.6947	0.9063	0.6519	0.6296	0.4226
79	173	-0.9925	0.1219	0.9816	-0.9743	0.1196	0.1908
74	5	0.9962	0.0872	0.9613	0.9576	0.0838	0.2756
51	28	0.8829	0.4695	0.7771	0.6862	0.3648	0.6293
66	184	-0.9976	-0.0698	0.9135	-0.9113	-0.0637	0.4067
87	53	0.6018	0.7986	0.9986	0.6010	0.7975	0.0523

$\theta$	$\phi$	$\cos \phi$ <i>I</i>	$\text{sen } \phi$ <i>II</i>	$\text{sen } \theta$ <i>III</i>	$l_i$ <i>I</i> $\times$ <i>III</i>	$m_i$ <i>II</i> $\times$ <i>III</i>	$n_i$ $\cos \theta$
68	236	-0.5592	-0.8290	0.9272	-0.5185	-0.7687	0.3746
48	73	0.2924	0.9563	0.7431	0.2173	0.7107	0.6691
60	9	0.9877	0.1564	0.8660	0.8554	0.1355	0.5000
69	95	-0.0872	0.9962	0.9336	-0.0814	0.9300	0.3584
68	28	0.8829	0.4695	0.9272	0.8187	0.4353	0.3746
62	37	0.7986	0.6018	0.8829	0.7052	0.5314	0.4695
88	208	-0.8829	-0.4695	0.9994	-0.8824	-0.4692	0.0349
16	170	-0.9848	0.1736	0.2756	-0.2714	0.0479	0.9613
83	232	-0.6157	-0.7880	0.9925	-0.6111	-0.7821	0.1219
72	52	0.6157	0.7880	0.9511	0.5855	0.7494	0.3090
82	233	-0.6018	-0.7986	0.9903	-0.5960	-0.7909	0.1392
66	350	0.9848	-0.1736	0.9135	0.8997	-0.1586	0.4067
34	313	0.6820	-0.7314	0.5592	0.3814	-0.4090	0.8290
77	332	0.8829	-0.4695	0.9744	0.8603	-0.4574	0.2250
78	202	-0.9272	-0.3746	0.9781	-0.9069	-0.3664	0.2079
88	46	0.6947	0.7193	0.9994	0.6942	0.7189	0.0349
80	297	0.4540	-0.8910	0.9848	0.4471	-0.8775	0.1736
87	339	0.9336	-0.3584	0.9986	0.9323	-0.3579	0.0523
59	224	-0.7193	-0.6947	0.8572	-0.6166	-0.5954	0.5150
41	56	0.5592	0.8290	0.6561	0.3669	0.5439	0.7547
81	47	0.6820	0.7314	0.9877	0.6736	0.7223	0.1564
65	39	0.7771	0.6293	0.9063	0.7043	0.5704	0.4226
86	184	-0.9976	-0.0698	0.9976	-0.9951	-0.0696	0.0698
80	343	0.9563	-0.2924	0.9848	0.9418	-0.2879	0.1736
66	31	0.8572	0.5150	0.9135	0.7831	0.4705	0.4067
85	337	0.9205	-0.3907	0.9962	0.9170	-0.3892	0.0872
80	253	-0.2924	-0.9563	0.9848	-0.2879	-0.9418	0.1736
61	9	0.9877	0.1564	0.8746	0.8639	0.1368	0.4848
57	26	0.8988	0.4384	0.8387	0.7538	0.3676	0.5446
72	355	0.9962	-0.0872	0.9511	0.9474	-0.0829	0.3090
83	74	0.2756	0.9613	0.9925	0.2736	0.9541	0.1219
83	219	-0.7771	-0.6293	0.9925	-0.7714	-0.6246	0.1219
87	22	0.9272	0.3746	0.9986	0.9259	0.3741	0.0523
35	27	0.8910	0.4540	0.5736	0.5111	0.2604	0.8192
63	41	0.7547	0.6561	0.8910	0.6725	0.5846	0.4540
89	243	-0.4540	-0.8910	0.9998	-0.4539	-0.8909	0.0175
78	28	0.8829	0.4695	0.9781	0.8637	0.4592	0.2079
88	214	-0.8290	-0.5592	0.9994	-0.8285	-0.5589	0.0349
85	331	0.8746	-0.4848	0.9962	0.8713	-0.4830	0.0872
79	156	-0.9135	0.4067	0.9816	-0.8968	0.3993	0.1908
77	275	0.0872	-0.9962	0.9744	0.0849	-0.9707	0.2250



$\theta$	$\phi$	$\cos \phi$ <i>I</i>	$\text{sen } \phi$ <i>II</i>	$\text{sen } \theta$ <i>III</i>	$l_i$ <i>I</i> $\times$ <i>III</i>	$m_i$ <i>II</i> $\times$ <i>III</i>	$n_i$ $\cos \theta$
81	49	0.6561	0.7547	0.9877	0.6480	0.7454	0.1564
73	291	0.3584	-0.9336	0.9563	0.3427	-0.8928	0.2924
63	128	-0.6157	0.7880	0.8910	-0.5486	0.7021	0.4540
46	115	-0.4226	0.9063	0.7193	-0.3040	0.6519	0.6947
67	8	0.9903	0.1392	0.9205	0.9115	0.1281	0.3907
86	162	-0.9511	0.3090	0.9976	-0.9487	0.3083	0.0698
31	321	0.7771	-0.6293	0.5150	0.4003	-0.3241	0.8572
64	60	0.5000	0.8660	0.8988	0.4494	0.7784	0.4384
85	165	-0.9659	0.2588	0.9962	-0.9623	0.2578	0.0872
84	44	0.7193	0.6947	0.9945	0.7154	0.6909	0.1045
64	183	-0.9986	-0.0523	0.8988	-0.8976	-0.0470	0.4384
44	18	0.9511	0.3090	0.6947	0.6607	0.2147	0.7193
45	305	0.5736	-0.8192	0.7071	0.4056	-0.5792	0.7071
85	321	0.7771	-0.6293	0.9962	0.7742	-0.6269	0.0872
53	54	0.5878	0.8090	0.7986	0.4694	0.6461	0.6018
45	353	0.9925	-0.1219	0.7071	0.7018	-0.0862	0.7071
61	178	-0.9994	0.0349	0.8746	-0.8741	0.0305	0.4848
83	43	0.7314	0.6820	0.9925	0.7259	0.6769	0.1219
87	35	0.8192	0.5736	0.9986	0.8180	0.5728	0.0523
88	291	0.3584	-0.9336	0.9994	0.3581	-0.9330	0.0349
67	72	0.3090	0.9511	0.9205	0.2845	0.8755	0.3907
80	21	0.9336	0.3584	0.9848	0.9194	0.3529	0.1736
80	142	-0.7880	0.6157	0.9848	-0.7760	0.6063	0.1736

En la figura 4.8.1 podemos observar las observaciones graficadas sobre una esfera de radio 1.

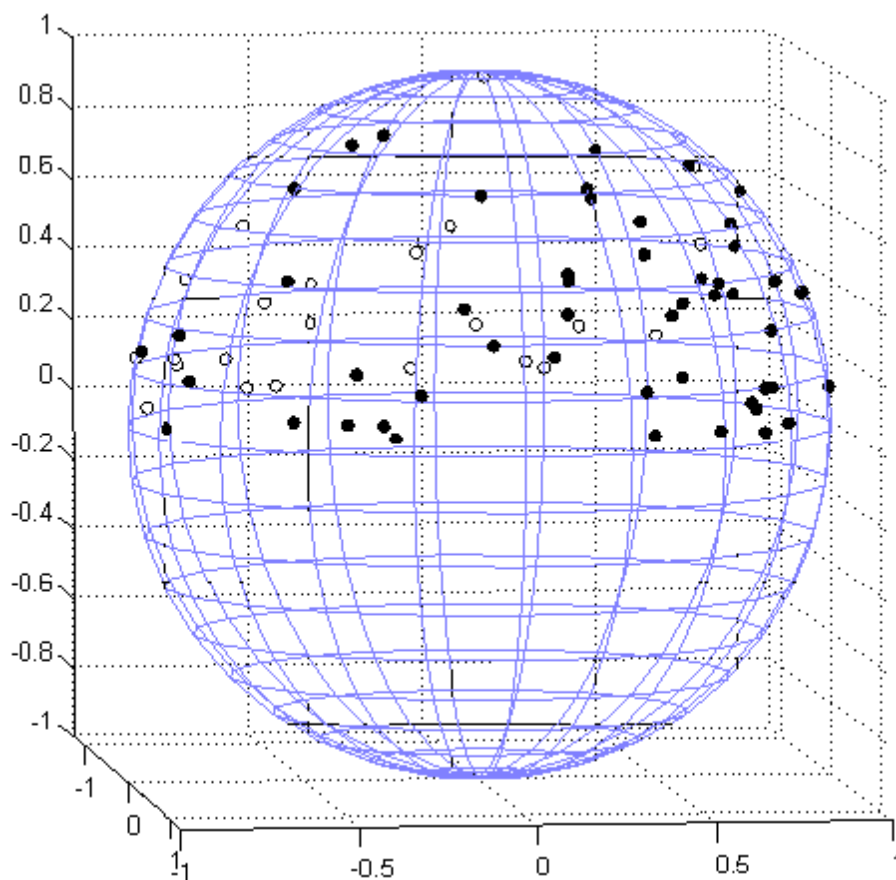


Fig. 4.8.1 Observaciones graficadas sobre una esfera de radio 1.

En esta gráfica no es posible observar fácilmente la forma de la distribución que siguen los datos, pero sí es posible observar que todas las observaciones se encuentran en el hemisferio superior. Ahora, haremos una proyección Lambert de los mismos, la cual se puede apreciar en la figura 4.8.2.

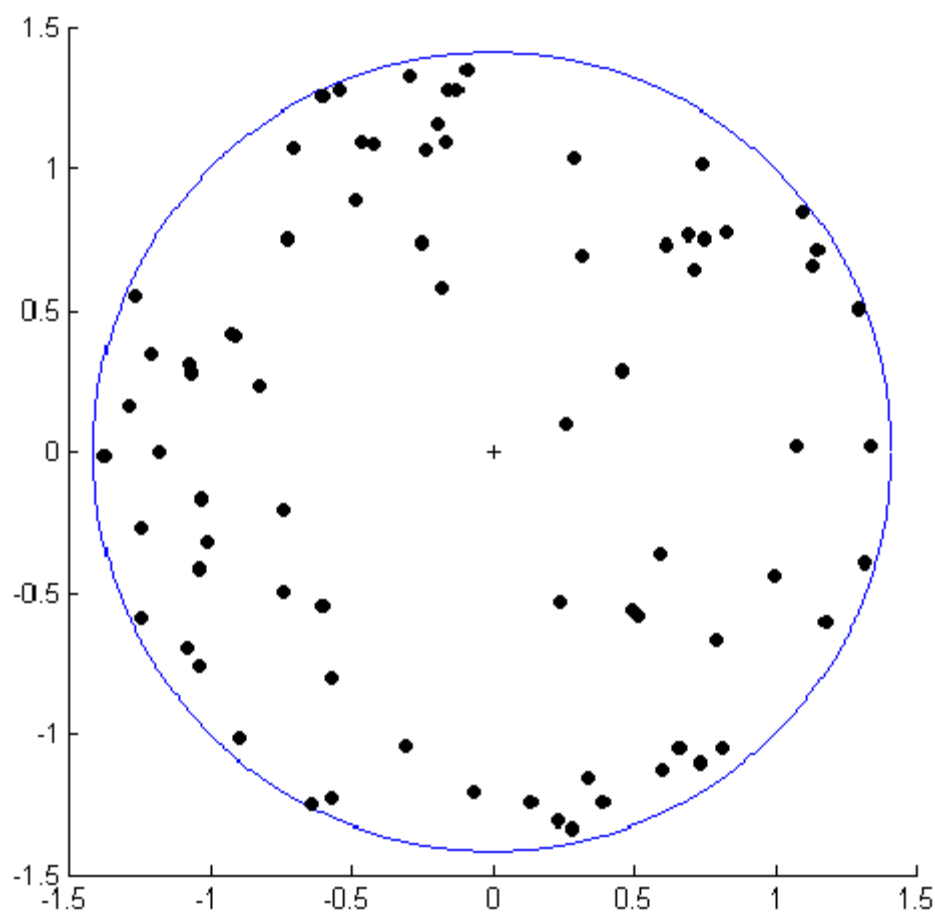


Fig. 4.8.2 Proyeccion Lambert de los granos de calcita.

En esta proyección no se puede observar claramente si las observaciones tienen una dirección preferida, la mayoría están dispersas. Para poder concluir algo sobre la forma de la distribución de las observaciones obtendremos las direcciones principales.

Primero, encontraremos las direcciones principales utilizando la matriz  $\mathbf{B}$ . Para ello, formaremos la matriz de orientación, la matriz  $\mathbf{T}$  :

$$\mathbf{T} = \begin{pmatrix} 38.0644 & 8.6952 & 6.8892 \\ 8.6952 & 24.1364 & 3.3513 \\ 6.8892 & 3.3513 & 12.7995 \end{pmatrix}$$

y la matriz  $\mathbf{B}$  será:

$$\mathbf{B} = \begin{pmatrix} 75 & 0 & 0 \\ 0 & 75 & 0 \\ 0 & 0 & 75 \end{pmatrix} - \begin{pmatrix} 38.0644 & 8.6952 & 6.8892 \\ 8.6952 & 24.1364 & 3.3513 \\ 6.8892 & 3.3513 & 12.7995 \end{pmatrix}$$

$$\therefore \mathbf{B} = \begin{pmatrix} 36.9356 & -8.6952 & -6.8892 \\ -8.6952 & 50.8636 & -3.3513 \\ -6.8892 & -3.3513 & 62.2005 \end{pmatrix}$$

los valores propios de  $\mathbf{B}$ , son:

$$\beta_1 = 30.8851, \quad \beta_2 = 55.0397 \quad \text{y} \quad \beta_3 = 64.0748$$

y los correspondientes vectores propios, que son los cosenos directores de las direcciones principales, son:

$$\begin{aligned} \mathbf{b}'_1 &= (0.8754, 0.4209, 0.2376) \\ \mathbf{b}'_2 &= (-0.4341, 0.9008, 0.0039) \\ \mathbf{b}'_3 &= (-0.2124, -0.1066, 0.9713) \end{aligned}$$

por lo que la matriz de vectores propios queda de la siguiente manera:

$$\mathbf{V} = \begin{pmatrix} 0.8754 & -0.4341 & -0.2124 \\ 0.4209 & 0.9008 & -0.1066 \\ 0.2376 & 0.0039 & 0.9713 \end{pmatrix}$$

Como es posible ver,  $\beta_2 \approx \beta_3$ , con valores grandes lo que nos indica que la mayoría de los datos están concentrados alrededor del punto final del vector  $\mathbf{b}_1$ . Es posible afirmar que la forma de la distribución que siguen estos datos es unimodal, ya que el hecho de que todas las observaciones se encuentren sobre el hemisferio superior impide la posibilidad de que  $\mathbf{b}_1$  genere un diámetro y por ende, que existan observaciones concentradas en la otra punta de ese diámetro. Ahora, calculemos  $r$ , entonces tenemos

que:

$$\begin{aligned}
 \sum_{i=1}^{75} l_i &= 15.2691 \quad \Rightarrow \quad \left( \sum_{i=1}^{75} l_i \right)^2 = 233.1466 \\
 \sum_{i=1}^{75} m_i &= 4.8216 \quad \Rightarrow \quad \left( \sum_{i=1}^{75} m_i \right)^2 = 23.2477 \\
 \sum_{i=1}^{75} n_i &= 25.1980 \quad \Rightarrow \quad \left( \sum_{i=1}^{75} n_i \right)^2 = 634.9375 \\
 R &= 29.8552 \\
 \Rightarrow r &= \frac{29.8552}{75} = 0.3981 \\
 \therefore r &= 0.3981
 \end{aligned}$$

Como  $r$  es pequeño, se puede afirmar que existe dispersión en los datos.

Al utilizar la matriz  $\mathbf{T}$  tenemos que los valores propios son:

$$\tau_1 = 44.1149, \quad \tau_2 = 19.9603 \quad \text{y} \quad \tau_3 = 10.9252$$

y sus correspondientes vectores propios son los mismos que los de la matriz  $\mathbf{B}$ :

$$\begin{aligned}
 \mathbf{t}'_1 &= (0.8754, 0.4209, 0.2376) \\
 \mathbf{t}'_2 &= (-0.4341, 0.9008, 0.0039) \\
 \mathbf{t}'_3 &= (-0.2124, -0.1066, 0.9713)
 \end{aligned}$$

Como es posible ver,  $\tau_1 \approx \tau_2$ , con valores pequeños lo que indica lo que habíamos obtenido de la matriz  $\mathbf{B}$ : que la mayoría de los datos están concentrados alrededor del punto final del vector  $\mathbf{t}_1$  y por tanto, la forma de la distribución de los datos es unimodal.

## CONCLUSIONES

El desarrollo de este trabajo se centró en encontrar la equivalencia entre las componentes principales y el tratamiento que de éstas se da en estadística esférica; para ello, fue necesario utilizar el concepto de momento de inercia.

El momento de inercia fue utilizado bajo el supuesto de que a cada una de las  $n$  observaciones direccionales en muestra, colocadas sobre una esfera de radio 1, se les asigna una masa de tamaño uno. Lo cual nos dio como resultado que el momento de inercia de  $n$  observaciones alrededor de un eje cualquiera, es igual a la suma de las distancias cuadradas al eje de referencia. Por lo anterior, el momento de inercia resulta una herramienta muy útil al obtener información sobre la dispersión que existe en los datos en muestra.

El desarrollo del momento de inercia, bajo este supuesto, dio lugar a dos matrices importantes: la matriz de orientación  $\mathbf{T}$  y la matriz  $\mathbf{B}$ . La matriz de orientación  $\mathbf{T}$  es la matriz de sumas de cuadrados y productos cruzados de los cosenos directores de las  $n$  observaciones, es decir, el producto de la matriz de observaciones transpuesta, por la matriz de observaciones. Así, la matriz  $\mathbf{B}$  es igual a la diferencia entre la matriz identidad por el tamaño de la muestra, y la matriz de orientación  $\mathbf{T}$ .

Para encontrar las componentes principales, en estadística esférica, se planteó como objetivo encontrar la dirección que minimizara el momento de inercia, es decir, encontrar la dirección para la cual, la suma de distancias cuadradas de las observaciones a esta dirección, fuera mínima. De esta manera, se encontró que las direcciones principales son las equivalentes a las componentes principales pero trabajándolas con datos direccionales. El planteamiento que dio lugar a las direcciones principales fue hecho en las matrices  $\mathbf{B}$  y  $\mathbf{T}$  en lugar de la matriz de varianzas y covarianzas o de correlaciones que dan lugar a la obtención de las componentes principales. También fue posible encontrar que las direcciones principales obtenidas a partir de la matriz  $\mathbf{B}$  son las mismas que las obtenidas a partir de la matriz de orientación  $\mathbf{T}$ .

Asimismo, se encontró que la información que ofrecen las direcciones o ejes prin-

cipales, difiere de aquélla que arrojan las componentes principales. Dado que las direcciones principales son obtenidas utilizando el momento de inercia, dan información sobre la dispersión de los datos y partir de ésta, es posible caracterizar la forma de la distribución que siguen los datos, ya sea unimodal, bimodal, uniforme o cinturón.

Finalmente, se da un ejemplo para ilustrar la manera en que se puede identificar la forma de la distribución que siguen los datos, a partir de la obtención de las direcciones principales.

## Bibliografía

- [1] Batschelet, Edward, *Circular Statistics in Biology*, London, Academic Press Inc., 1981, xv-371 pp. (Mathematics in biology).
- [2] Bingham Christopher, "*Distributions on the sphere and on the projective plane*", Ph. D. Thesis, Yale University, 1964.
- [3] Braun, Eliezer, *Mecánica*, México, Trillas, 1988, 367 pp.
- [4] Chatfield, Christopher y Alexander J. Collins, *Introduction to Multivariate Analysis*, London, Chapman and Hall, 1980, x-246 pp.
- [5] Dillon, William R y Matthew Goldstein, *Multivariate Analysis. Methods and Applications*, USA, John & Wiley Sons, 1984, 587 pp. (Wiley Series in probability and statistics).
- [6] Everitt, Brian Sidney y Graham Dunn, *Applied Multivariate Data Analysis*, London, Edward Arnold, 1991, 304 pp.
- [7] Feynman, Richard Phillips, *The feynman lectures on physics*, USA, Addison Wesley, 1963.
- [8] Fisher, N. I.; T. Lewis y B. J. J. Embleton, *Statistical analysis of spherical data*, New York, Cambridge University Press, 1987, xiv-329 pp.
- [9] Flury, Bernhard, *A First Course in Multivariate Statistics*, New York, Springer, 1997, xiii-713 pp.
- [10] Johnson, Richard Arnold y Dean W. Wichern, *Applied Multivariate Statistical Analysis*, 3a. ed., New Jersey, Prentice-Hall, 1992, xiv-642 pp.
- [11] Mardia, Kantilal Varichand, *Statistics of Directional Data*, London, Academic Press, 1972, 357 pp.



- [12] Mardia, Kantilal Varichand; J. T. Kent y J. M. Bibby, *Multivariate Analysis*, USA, Academic Press, Inc., 1982, xv-521 pp. (Probability and mathematical statistics).
- [13] Mardia, Kanti V. y Peter E. Jupp, *Directional Statistics*, London, John & Wiley Sons Ltd., 1999, xxi-429 pp. (Wiley Series in probability and statistics).
- [14] Ramírez Garlarza, Ana Irene, *Geometría Analítica: Una Introducción a la Geometría*, México, Facultad de Ciencias, UNAM, 1998, vii-337 pp.
- [15] Rencher, Alvin C, *Methods of multivariate analysis*, USA, John & Wiley Sons, 1995, xvi-627 pp. (Wiley Series in probability and mathematical statistics. Applied probability and statistics).
- [16] Seber, George Arthur Frederick, *Multivariate Observations*, USA, John & Wiley Sons, 1984, 686 pp. (Wiley Series in probability and statistics).
- [17] Watson, Geoffrey S., *Statistics on spheres*, [s. l. i.], John Wiley & Sons Inc., 1983, x-238 pp. (University of Arkansas Lecture Notes in the Mathematical Sciences, Vol. 6).
- [18] Watson, Geoffrey S., "*The Statistics of Orientation Data*", *Journal of Geology*, Chicago, 74, 5, 1966, pp. 786 - 797.
- [19] Wexler, Charles, *Geometría Analítica: Un Enfoque Vectorial*, Barcelona, Montaner y Simon, 1968, 303 pp.