

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE CIENCIAS



**TESIS**

**“ANÁLISIS DE COMPONENTES PRINCIPALES VS.  
MÉTODO DEL GRADIENTE PROYECTADO”**

PARA OBTENER EL TÍTULO PROFESIONAL DE LICENCIADO EN  
MATEMÁTICA

ELABORADO POR  
JUAN CARLOS ESPEJO DELZO

ASESOR  
WILFREDO SOSA SANDOVAL

LIMA-PERÚ

2017

*Pues, ¿quién soy yo para que pueda ofrecerte este trabajo?  
Porque todo viene de ti, y de tu mano te lo doy.*

# Agradecimientos

El autor desea agradecer al profesor Wilfredo Sosa por haberle brindado la magnífica oportunidad de trabajar bajo su orientación, por haberlo formado, por su apoyo y consejos, incluso encontrándose a 3169 kilómetros de distancia. También, desea agradecer al profesor William Eche garay por haber revisado este trabajo y por sus sugerencias y a la [National Oceanic and Atmospheric Administration](#) por brindar sus datos oceanográficos. Finalmente, el autor desea expresar la más profunda gratitud a sus padres por todo el amor recibido.

## Resumen

El objetivo de este trabajo es desarrollar el fundamento matemático que da lugar al [Análisis de Componentes Principales \(ACP\)](#) –técnica del análisis exploratorio de datos– que utilizaremos para estudiar [El Niño - Oscilación del Sur](#), fenómeno climático de suma importancia por ser de escala global.

Para conseguir nuestro objetivo, se sienta las bases matemáticas del álgebra lineal y análisis real, necesarios para desarrollar las herramientas de la optimización matemática que son empleadas en este trabajo: condiciones necesarias de optimalidad de primer y segundo orden para problemas sin restricciones y condiciones de optimalidad de [Karush-Kuhn-Tucker](#) para problemas con restricciones.

Estas herramientas de la optimización permiten comprender la concepción clásica de la técnica estadística del [ACP](#) como un método para consolidar  $n$  variables mutuamente correlacionadas de datos multidimensionales en nuevas variables mediante combinaciones lineales de las variables originales con mínima pérdida de la información en los datos observados. Y permiten también aseverar que el abordar el problema netamente geométrico de encontrar una recta o un plano que mejor se aproxime a un conjunto finito, no vacío, de vectores del espacio euclídeo de dimensión  $n$ , es la esencia matemática que desencadena de manera natural, en esta técnica estadística.

Finalmente, una vez más, estas herramientas de la optimización, nos permite desarrollar un esquema basado en las ideas del método del gradiente proyectado y el método de las potencias que puede ser implementado en un programa de computadora; y así, se aplicó el [ACP](#) para el estudio de base de datos de las temperaturas superficiales y de las profundidades de la isoterma de 20°C del océano pacífico ecuatorial.

# Índice general

<b>Prólogo</b>	<b>1</b>
<b>1. Introducción</b>	<b>2</b>
<b>2. Fundamento teórico</b>	<b>3</b>
2.1. Álgebra lineal . . . . .	3
2.1.1. Espacios vectoriales . . . . .	3
2.1.2. Transformaciones lineales . . . . .	7
2.1.3. Espacios producto interno . . . . .	8
2.1.4. El espacio euclídeo de dimensión finita . . . . .	12
2.2. Cálculo diferencial . . . . .	19
2.2.1. La regla de la cadena . . . . .	19
2.3. Optimización matemática . . . . .	21
2.3.1. Conjuntos convexos . . . . .	22
2.3.2. Funciones convexas . . . . .	23
2.3.3. Condiciones de optimalidad . . . . .	26
2.3.4. Método del gradiente proyectado . . . . .	29
2.4. Estadística . . . . .	30
2.4.1. Análisis de componentes principales . . . . .	30
2.5. Análisis numérico . . . . .	33
2.5.1. Método de las potencias . . . . .	33
2.5.2. Un esquema discreto . . . . .	35
<b>3. Metodología</b>	<b>37</b>
3.1. Planteamiento del problema . . . . .	37
3.2. Solución del problema . . . . .	38
3.3. Conexión con la técnica del ACP . . . . .	43

<b>4. Base de datos de la NOAA</b>	<b>45</b>
4.1. Lectura de los datos . . . . .	46
4.2. Preparación de los datos . . . . .	49
<b>5. Aplicación del ACP a la base de datos de la NOAA</b>	<b>51</b>
5.1. Temperatura superficial del mar . . . . .	51
5.2. Profundidad de la isoterma de 20 grados Celcius . . . . .	56
<b>6. Conclusión</b>	<b>62</b>
<b>Bibliografía</b>	<b>62</b>

# Índice de figuras

4.1. <i>Posición geográfica de la red de boyas de la NOAA.</i> . . . . .	45
5.1. <i>Representación diaria de la temperatura superficial del mar.</i> . . . . .	52
5.2. <i>Patrón mensual de la temperatura superficial del mar.</i> . . . . .	53
5.3. <i>Representación anual de la temperatura superficial del mar.</i> . . . . .	53
5.4. <i>Representación trimestral de la temperatura superficial del mar.</i> . . . . .	54
5.5. <i>Representación mensual de la temperatura superficial del mar.</i> . . . . .	54
5.6. <i>Representación mensual de la temperatura superficial del mar.</i> . . . . .	55
5.7. <i>Representación mensual de la temperatura superficial del mar.</i> . . . . .	55
5.8. <i>Representación mensual de la temperatura superficial del mar.</i> . . . . .	56
5.9. <i>Representación diaria de la profundidad de la isoterma de 20°C.</i> . . . . .	57
5.10. <i>Patrón mensual de la profundidad de la isoterma de 20°C.</i> . . . . .	58
5.11. <i>Representación anual de la profundidad de la isoterma de 20°C.</i> . . . . .	58
5.12. <i>Representación trimestral de la profundidad de la isoterma de 20°C.</i> . . . . .	59
5.13. <i>Representación mensual de la profundidad de la isoterma de 20°C.</i> . . . . .	59
5.14. <i>Representación mensual de la profundidad de la isoterma de 20°C.</i> . . . . .	60
5.15. <i>Representación mensual de la profundidad de la isoterma de 20°C.</i> . . . . .	60
5.16. <i>Representación mensual de la profundidad de la isoterma de 20°C.</i> . . . . .	61

# Índice de tablas

4.1. <i>Número de datos de temperatura superficial del mar recuperados en el periodo del 7 de marzo de 1980 al 15 de junio del 2011. Los años marcados con (*) indican que son bisiestos.</i> . . . . .	47
4.2. <i>Número de datos de profundidad de la isoterma de 20°C recuperados en el periodo del 7 de marzo de 1980 al 15 de junio del 2011. Los años marcados con (*) indican que son bisiestos.</i> . . . . .	48
4.3. <i>Arreglo de los datos de temperatura superficial del mar (°C).</i> . . . . .	49
4.4. <i>Arreglo de los datos de profundidad de la isoterma de 20°C (m).</i> . . . . .	50

# Índice de siglas

ACP	Análisis de Componentes Principales	<a href="#">i, 1, 30, 37, 43, 62</a>
ENOS	El Niño - Oscilación del Sur	<a href="#">i, 2</a>
KKT	Karush-Kuhn-Tucker	<a href="#">i, 27</a>
NOAA	National Oceanic and Atmospheric Administration	<a href="#">ii, 1, 45, 62</a>

# Índice de símbolos

$\text{spec}(A)$	El conjunto de todos los autovalores de $A$	18
$\mathbb{N}$	El conjunto de números naturales	3
$\mathbb{R}$	El conjunto de números reales	4
$T(C, a)$	El cono tangente de Bouligand de $C$ en $a$	26
$L(C, a)$	El cono tangente linealizado de $C$ en $a$	27
$d(x, S)$	Distancia del punto $x$ al conjunto $S$	10
$\mathbb{R}^n$	El espacio de las $n$ -uplas de números reales	12, 13
$\mathcal{M}(m, n)$	El espacio de las matrices de tamaño $m \times n$	14
$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$	El espacio de las transformaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$	16
$\mathcal{L}(E, F)$	El espacio de las transformaciones lineales de $E$ en $F$	8
$V$	Un espacio vectorial	4, 5
$\nabla f(a)$	El gradiente de $f$ en $a$	19
$\mathcal{H}(u, \alpha)$	Hiperplano con vector normal $u$	10
$\text{int}(C)$	El interior de $C$	26
$\nabla^2 f(a)$	La matriz hessiana de $f$ en $a$	19
$\  \cdot \ $	Norma	9
$\pi$	Relación entre la longitud de una circunferencia y su diámetro	4
$\mathcal{P}(z, u, v)$	Plano que pasa por $z$	6
$\langle \cdot, \cdot \rangle$	Producto interno	8
$\delta(B, \mathcal{L})$	La proximidad de $\mathcal{L}$ a $B$	37

$\text{proj}_C(a)$	La proyección de $a$ sobre $C$	23
$\mathcal{L}(z, u)$	Recta que pasa por $z$ con vector dirección $u$	6
$\text{span}(X)$	Subespacio generado por $X$	6

# Prólogo

En el presente trabajo, comenzamos colocando las bases matemáticas de nuestro estudio en el capítulo 2. En este capítulo definimos los espacios sobre los que trabajaremos como espacios vectoriales: el espacio euclídeo de dimensión finita y el espacio de las matrices simétricas. También se define un tipo especial de problema de optimización matemática: el problema de programación diferenciable y se describe el método del gradiente proyectado que es sobre el cual se dará un método del análisis numérico para encontrar el autovector asociado al mayor autovalor de una matriz simétrica.

En el capítulo 3 se plantea el problema geométrico de encontrar un plano que mejor se aproxime a un conjunto finito de puntos del espacio euclídeo de dimensión  $n$  como un problema de programación diferenciable. Con las herramientas de la optimización matemática desarrolladas en el capítulo 2 se da una solución a este problema y se establece esta metodología desarrollada no es otra cosa que la esencia matemática de la técnica estadística del [Análisis de Componentes Principales \(ACP\)](#).

Luego, en el capítulo 4 se muestra que los datos para aplicar la metodología desarrollada en el capítulo 3 –utilizando el esquema basado en el método del gradiente proyectado descrito en el capítulo 2– fueron descargados del sitio web de la [National Oceanic and Atmospheric Administration](#). También se explica como fueron tratados estos datos para poder aplicar sobre ellos la metodología desarrollada.

Finalmente, en el capítulo 5 se muestran los resultados de aplicar el [ACP](#) sobre dos bases de datos: la temperaturas superficiales del mar y las profundidades de la isoterma de 20 °C.

Juan Espejo

# Capítulo 1

## Introducción

La motivación del presente informe radica en el interés de estudiar el fenómeno de [El Niño - Oscilación del Sur \(ENOS\)](#), ya que, es el causante de condiciones climáticas extremas (como inundaciones y sequías) en muchas regiones del mundo, y el Perú no es la excepción.

Este fenómeno es el resultado de un calentamiento y enfriamiento cíclico de la superficie del Océano Pacífico central y oriental. *El Niño* se caracteriza por temperaturas oceánicas inusualmente cálidas en el Pacífico ecuatorial, frente a *La Niña*, que se caracteriza por temperaturas inusualmente frías del Océano Pacífico ecuatorial.

Entre las consecuencias de *El Niño*, tenemos el aumento de las precipitaciones a lo largo de la mayor parte del oeste americano, extendiéndose desde el sur de EE.UU. hasta el norte de Chile, que ha causado inundaciones destructivas, y las sequías en la región del Pacífico Occidental, a veces asociada con devastadores incendios en Australia.

Para empezar, es necesario sentar nuestro estudio sobre una base matemática sólida, ya que, es esta la que nos garantizará el correcto funcionamiento de las técnicas utilizadas en nuestro estudio. En matemáticas, lo primero que se hace es establecer el espacio en el cual se trabajará, y es en este contexto que enunciaremos nuestras definiciones y teoremas –probando aquellos que no se ha encontrado en la literatura– que emplearemos.

En nuestro caso, este contexto es el espacio vectorial. En particular, nos enfocaremos en el espacio euclídeo  $n$ -dimensional y el espacio de matrices simétricas de tamaño  $n$ . Finalmente, amoldamos las herramientas necesarias de la teoría de la optimización matemática a nuestro problema en particular.

# Capítulo 2

## Fundamento teórico

Del vasto edificio de las matemáticas desarrollamos algunos de sus cimientos que necesitaremos para sentar nuestro trabajo sobre una base matemática sólida.

### 2.1. Álgebra lineal

El álgebra lineal es la rama de la matemática relativa a los espacios vectoriales y a las transformaciones lineales entre dichos espacios. Cuando los espacios tienen dimensiones finitas, las transformaciones pueden ser identificadas por matrices. Así, el álgebra lineal, además de vectores y transformaciones lineales, lida también con matrices.

#### 2.1.1. Espacios vectoriales

La noción de espacio vectorial es la base del estudio que haremos. Ella es el terreno donde se desarrolla todo el álgebra lineal y por ello comenzamos nuestra exposición con este concepto. No obstante, antes necesitamos introducir el concepto de producto cartesiano.

Dado  $n \in \mathbb{N}$ , la idea intuitiva de una  $n$ -upla, digamos  $(x_1, \dots, x_n)$ , es la de una “secuencia” de  $n$  objetos  $x_1, \dots, x_n$  cualesquiera; es decir, en ella se identifica un primer objeto  $x_1$ , un segundo objeto  $x_2$ , y así sucesivamente hasta un  $n$ -ésimo objeto  $x_n$ . De ahora en adelante, al  $i$ -ésimo objeto de una  $n$ -upla  $x = (x_1, \dots, x_n)$  lo llamaremos la  $i$ -ésima coordenada de  $x$ . Diremos que las  $n$ -uplas  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_n)$  son iguales si sus respectivas coordenadas lo son; es decir,  $x_i = y_i$  para cada  $i = 1, \dots, n$ .

**Definición 2.1** (Producto cartesiano). Dado  $n \in \mathbb{N}$ . Si  $A_1, \dots, A_n$  son  $n$  conjuntos no vacíos, el *producto cartesiano*  $A_1 \times \dots \times A_n$ , es el conjunto de todas las  $n$ -uplas  $(a_1, \dots, a_n)$  con  $a_i \in A_i$  para cada  $i = 1, \dots, n$ .

**Observación 2.2.** No siempre  $A \times B$  y  $B \times A$  son iguales. Por ejemplo, si  $A = \{1, 2\}$  y  $B = \{\pi\}$ , luego  $A \times B$  es el conjunto formado por los pares ordenados  $(1, \pi)$  y  $(2, \pi)$ , mientras que  $B \times A$  está formado por los pares ordenados  $(\pi, 1)$  y  $(\pi, 2)$ . Así,  $A \times B \neq B \times A$ .

**Definición 2.3** (Espacio vectorial). Un *espacio vectorial* consiste en un conjunto no vacío  $V$ , cuyos elementos son llamados *vectores*, dotado de dos operaciones:

$$\begin{array}{ccc} V \times V & \longrightarrow & V \\ (u, v) & \mapsto & u + v \end{array} \qquad \begin{array}{ccc} \mathbb{R} \times V & \longrightarrow & V \\ (\alpha, v) & \mapsto & \alpha v \end{array}$$

llamadas *adición*, y *multiplicación por un número real*, respectivamente. Donde los vectores  $u+v$  y  $\alpha v$  son llamados *suma* de  $u$  y  $v$ , y *producto* de  $\alpha$  por  $v$ , respectivamente. Estas operaciones deben satisfacer para cada  $\alpha, \beta \in \mathbb{R}$  y  $u, v, w \in V$ :

- (D1)  $(u + v) + w = u + (v + w)$ ,  $(\alpha\beta)u = \alpha(\beta u)$  ;
- (D2)  $u + v = v + u$  ;
- (D3) existe un vector  $0$  tal que  $0 + v = v = v + 0$  ;
- (D4) existe un vector  $-v$  tal que  $(-v) + v = 0 = v + (-v)$  ;
- (D5)  $(\alpha + \beta)v = \alpha v + \beta v$ ,  $\alpha(u + v) = \alpha u + \alpha v$  ;
- (D6)  $1v = v$ ,

llamadas *axiomas de espacio vectorial*.

**Observación 2.4.**

- (i) Se demuestra la unicidad del vector  $0$  de la propiedad (D3), por ello es llamado *vector nulo* o *cero*. A pesar de que con el mismo símbolo, también, representamos al número cero, esta nueva representación no genera ambigüedad, ya que, siempre podremos inferir del contexto a cuál de los dos objetos el símbolo  $0$  está representando. También, se prueba la unicidad del vector  $-v$  de la propiedad (D4), y por ello es llamado el *vector opuesto* de  $v$ . Finalmente, escribiremos  $u - v$  en lugar de  $u + (-v)$ .

(ii) De esta manera, un espacio vectorial consiste de los siguientes 5 ingredientes: un conjunto no vacío  $V$ ,  $\mathbb{R}$ , dos operaciones y un conjunto de propiedades que todos los ingredientes previos deben cumplir. No obstante, es usual denotar a un espacio vectorial por el símbolo que representa al conjunto de vectores. Nosotros adoptaremos esta convención y en lo que resta de esta sección,  $V$  representará un espacio vectorial.

A continuación, definiremos algunos conceptos básicos del álgebra lineal, de esta manera, también fijaremos la notación empleada. La combinación lineal es el primer concepto que definiremos, ella emplea a la adición vectorial y multiplicación por un número real, las dos operaciones que caracterizan a los espacios vectoriales.

**Definición 2.5** (Combinación lineal). Dado  $X$  un subconjunto no vacío de  $V$ , una *combinación lineal* de vectores en  $X$  es una expresión de la forma

$$\alpha_1 x^1 + \cdots + \alpha_m x^m,$$

donde  $n \in \mathbb{N}$ ,  $x^1, \dots, x^m \in X$ ,  $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ , llamados los *coeficientes* de la combinación lineal. Cuando esta combinación lineal es igual al vector cero, se le llama *combinación lineal nula* y cuando estos coeficientes son todos iguales a cero, esta combinación lineal nula es llamada *combinación lineal trivial*.

**Definición 2.6** (Conjunto linealmente independiente). Decimos que un subconjunto no vacío  $X$  de  $V$  es *linealmente independiente* si, la única combinación lineal nula de vectores en él es la trivial. Caso contrario, se dice que  $X$  es *linealmente dependiente*.

**Observación 2.7.** Un conjunto no vacío  $X$  de vectores es linealmente independiente si, y sólo si, para todo  $x^1, \dots, x^m \in X$

$$\alpha_1 x^1 + \cdots + \alpha_m x^m = 0 \quad \text{implica que} \quad \alpha_1 = \cdots = \alpha_m = 0.$$

Así,  $X$  es linealmente dependiente si, y sólo si, existen  $m \in \mathbb{N}$ ,  $\alpha_1, \dots, \alpha_m \in \mathbb{R}$  no todos nulos y  $x^1, \dots, x^m \in X$ , tales que,  $\alpha_1 x^1 + \cdots + \alpha_m x^m = 0$ .

**Definición 2.8** (Subespacio vectorial). Decimos que un subconjunto no vacío  $X$  de  $V$  es un *subespacio vectorial* de  $V$  si

$$\alpha x + \beta y \in X,$$

para todo  $x, y \in X$  y todo  $\alpha, \beta \in \mathbb{R}$ .

**Observación 2.9.** Los conjuntos  $\{0\}$  y  $V$  mismo, son ejemplos obvios de subespacios vectoriales.  $\{0\}$  es llamado *el subespacio trivial*. Es importante señalar que en todo momento se da por hecho que  $V$  es distinto del subespacio trivial.

Dados un vector  $a$  y un subconjunto no vacío  $X$  de  $V$ , denotamos como  $a + X$  al conjunto

$$a + X := \{a + x : x \in X\}.$$

Los subespacios vectoriales forman parte de una familia más amplia de subconjuntos de  $V$  que pasamos a definir.

**Definición 2.10** (Conjunto afín). Todo conjunto de la forma  $a + S$ , donde  $a \in V$  y  $S$  es un subespacio de  $V$ , es llamado *conjunto afín* de  $V$ .

**Observación 2.11.** Está claro que todo subespacio  $S$  de  $V$  es un conjunto afín por ser de la forma  $0 + S$ .

**Definición 2.12** (Subespacio generado). Al conjunto de todas las combinaciones lineales de un subconjunto no vacío  $X$  de  $V$  llamamos de *subespacio generado* por  $X$  y lo denotamos como  $\text{span}(X)$  o  $\text{span}(x^1, \dots, x^m)$  si  $X = \{x^1, \dots, x^m\}$ . Diremos que  $X$  genera a un subespacio  $A$  de  $V$ , si  $\text{span}(X) = A$ .

**Observación 2.13.** El nombre de subespacio generado no es casualidad, de hecho, se demuestra que  $\text{span}(X)$  es siempre un subespacio vectorial.

De ahora en adelante, utilizaremos indistintamente los términos *punto* y vector.

**Ejemplo 2.14** (Recta). Sea  $z$  un punto y  $u$  un vector no nulo. Se define al conjunto

$$z + \text{span}(u)$$

como la recta que pasa por el punto  $z$  con vector dirección  $u$  y se le denota  $\mathcal{L}(z, u)$ .

**Ejemplo 2.15** (Plano). Sean  $z$  un punto y  $u, v$  vectores linealmente independientes. El siguiente conjunto afín:

$$\mathcal{P}(z, u, v) = z + \text{span}(u, v)$$

es llamado de plano.

**Definición 2.16** (Base). Decimos que un subconjunto no vacío de  $V$  es una *base* de  $V$  si es linealmente independiente y genera  $V$ .

**Propiedad 2.17.** Sea  $n \in \mathbb{N}$ . Si el espacio vectorial  $V$  admite una base de  $n$  elementos, entonces cualquier otra base de  $V$  posee también  $n$  elementos.

*Demostración.* Ver prueba del corolario 2 de [Lima, 2011a]. □

**Definición 2.18** (Dimensión de una base). Se dice que un espacio vectorial tiene *dimensión finita* si admite una base con un número finito de elementos. En virtud de la propiedad 2.17, este número es el mismo para todas las bases de dicho espacio y es llamado de *dimensión* de dicho espacio. Por extensión, se dice que el subespacio trivial tiene dimensión cero.

### 2.1.2. Transformaciones lineales

**Definición 2.19** (Transformación lineal). Sean  $E$  y  $F$  espacios vectoriales. Una *transformación lineal*  $A : E \rightarrow F$  es una correspondencia que asocia a cada vector  $u \in E$  un vector  $Au := A(u) \in F$  de modo que para cada  $u, v \in E$  y cada  $\alpha \in \mathbb{R}$  se tenga:

$$\begin{aligned}A(u + v) &= Au + Av, \\A(\alpha u) &= \alpha Au.\end{aligned}$$

**Observación 2.20.** Si  $A : E \rightarrow F$  es una transformación lineal, entonces para todo  $u, v, v^1, \dots, v^m \in E$  y todo  $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ :

- (i)  $A(0) = 0$ ,
- (ii)  $A(\alpha_1 v^1 + \dots + \alpha_m v^m) = \alpha_1 Av^1 + \dots + \alpha_m Av^m$ ,
- (iii)  $A(-v) = -Av$ ,
- (iv)  $A(u - v) = Au - Av$ .

**Observación 2.21.**

- (i) La *suma* de dos transformaciones lineales  $A, B : E \rightarrow F$  y el *producto* de un número  $\alpha \in \mathbb{R}$  por una transformación lineal  $A : E \rightarrow F$  son las transformaciones lineales  $A + B : E \rightarrow F$  y  $\alpha A : E \rightarrow F$ , definidas respectivamente por

$$(A + B)v = Av \quad \text{y} \quad (\alpha A)v = \alpha Av,$$

para todo  $v \in E$ .

(ii) El símbolo  $0$  indica la transformación lineal nula  $0 : E \rightarrow F$ , definida por  $0v = 0$ . A pesar de que con el mismo símbolo representamos al número cero, al vector cero y ahora a la transformación lineal nula, esta nueva representación no genera ambigüedad, ya que, siempre podremos inferir del contexto a cuál de estos tres objetos el símbolo  $0$  está representando.

(iii) Finalmente, definimos  $-A : E \rightarrow F$  por  $(-A)v = -Av$ . Se ve que

$$(-A) + A = 0 = A + (-A).$$

(iv) Sean  $E$  y  $F$  espacios vectoriales. Denotamos por  $\mathcal{L}(E, F)$  al espacio vectorial provisto del conjunto de las transformaciones lineales de  $E$  en  $F$  con la suma y producto definidos arriba.

**Definición 2.22** (isomorfismo). Sean  $E$  y  $F$  espacios vectoriales. Llamamos de *isomorfismo* a una transformación lineal biyectiva  $A : E \rightarrow F$  y decimos que los espacios vectoriales  $E$  y  $F$  son *isomorfos*.

**Observación 2.23.** Desde el punto de vista del álgebra lineal que dos espacios vectoriales sean isomorfos quiere decir que ambos espacios son en esencia el mismo espacio vectorial.

**Propiedad 2.24.** Dos espacios vectoriales de dimensión finita son isomorfos si y solo si tienen la misma dimensión.

*Demostración.* Ver las páginas 66 y 67 de [Lima, 2011a]. □

### 2.1.3. Espacios producto interno

El concepto de producto interno en un espacio vectorial enriquece su estructura vectorial, permitiendo abordar ciertas nociones geométricas como perpendicularidad, longitud y distancia.

**Definición 2.25** (Producto interno). Un *producto interno* en  $V$  es una función

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$$

tal que a cada  $(u, v) \in V \times V$  le asigna un número real  $\langle u, v \rangle$ , llamado producto interno de  $u$  por  $v$ , de modo que para cada  $u, v, w \in V$  y  $\alpha, \beta \in \mathbb{R}$ , se cumple:

(i)  $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$ ,

(ii)  $\langle u, v \rangle = \langle v, u \rangle$ ,

(iii)  $\langle v, v \rangle \geq 0$ ,

(iv)  $\langle v, v \rangle = 0$  si, y sólo si,  $v = 0$ .

**Observación 2.26.** En lo sucesivo, asumiremos que nuestro espacio vectorial  $V$  está equipado con un producto interno  $\langle \cdot, \cdot \rangle$ .

**Definición 2.27** (Ortogonalidad entre dos vectores). Dados  $u, v \in V$ . Decimos que  $u$  y  $v$  son *ortogonales* si  $\langle u, v \rangle = 0$ .

**Definición 2.28** (Norma o longitud). Dado  $v \in V$ . La *norma*, o *longitud* de  $v$  está definido por

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

Se dice que  $v$  es un *vector unitario* si  $\|v\| = 1$ .

**Observación 2.29.** La función *norma*  $\|\cdot\| : V \rightarrow \mathbb{R}$ ,  $\|v\| = \sqrt{\langle v, v \rangle}$ , satisface que para cada  $\alpha \in \mathbb{R}$  y  $v \in V$ :

(i)  $\|\alpha v\| = |\alpha| \|v\|$ ,

(ii)  $\|v\| \geq 0$ ,

(iii)  $\|v\| = 0 \Leftrightarrow v = 0$ ,

**Definición 2.30** (Distancia entre dos vectores). Dados  $u, v \in V$ . La *distancia* entre  $u$  y  $v$  está definida por  $\|u - v\|$ .

**Propiedad 2.31** (Teorema de Pitágoras). Si  $u$  y  $v$  son vectores ortogonales, entonces

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

*Demostración.*

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2. \end{aligned}$$

□

**Observación 2.32.** Dado un vector  $v$  no nulo. Podemos obtener un vector unitario  $u$  haciendo

$$u = \frac{1}{\|v\|} v.$$

Este proceso es conocido como *normalización* del vector  $v$ .

**Ejemplo 2.33** (Hiperplano). Sean  $u, z \in V$  con  $u \neq 0$  y sea  $\alpha = \langle u, z \rangle$ . El conjunto

$$\mathcal{H}(u, \alpha) := \{x \in \mathbb{R}^n ; \langle x, u \rangle = \alpha\}$$

es llamado de *hiperplano* que pasa por  $z$  con vector normal  $u$ .

**Observación 2.34.**

- (i) El hiperplano  $\mathcal{H} = \mathcal{H}(u, \alpha)$  que pasa por  $z$  es un conjunto afín. En efecto,  $\mathcal{H} = z + S$ , donde  $S = \{x \in \mathbb{R}^n ; \langle u, x \rangle = 0\}$  es un subespacio vectorial.
- (ii) Todo hiperplano de una espacio vectorial de dimensión  $n \in \mathbb{N}$  surge de trasladar un subespacio de dimensión  $n - 1$ .

**Definición 2.35** (Conjunto ortonormal). Decimos que un subconjunto  $X$  de  $V$  es un *conjunto ortonormal* si para cada  $u, v \in X$ :

$$\langle u, v \rangle = \begin{cases} 0 & \text{si } u \neq v \\ 1 & \text{si } u = v. \end{cases}$$

**Definición 2.36** (Base ortonormal). Decimos que un subconjunto de  $V$  es una *base ortonormal* de  $V$  si, es una base de  $V$  y es un conjunto ortonormal.

**Definición 2.37** (Distancia de una punto a un conjunto). Dados un punto  $x$  y un subconjunto no vacío  $S$  de  $V$ . Al ínfimo del conjunto

$$\{\|x - y\| ; y \in S\},$$

se le llama *distancia* de  $x$  a  $S$  y se le denota como  $d(x, S)$ .

**Observación 2.38.**

- (i)  $d(x, S)$  está bien definida, pues el conjunto

$$\{\|x - y\| : y \in S\}$$

es no vacío y está acotado inferiormente por el cero.

- (ii) Sean  $x, y \in V$ . Como  $d(x, \{y\}) = \|x - y\|$ , la distancia de un punto a un conjunto generaliza de manera natural la noción de distancia entre dos puntos.
- (iii) Hallar la distancia de un punto a un conjunto dado, puede resultar una tarea difícil; sin embargo, cuando dicho conjunto es una recta, se puede dar una fórmula para ello.

**Propiedad 2.39.** Sea  $x \in V$  y sea  $\mathcal{L} = \mathcal{L}(z, u)$  una recta de  $V$ . Entonces, existe un único  $\hat{x} \in \mathcal{L}$  tal que

$$\|x - \hat{x}\| \leq \|x - y\|$$

para todo  $y \in \mathcal{L}$ . Además, se tiene que  $\hat{x} = z + \langle x - z, v \rangle v$ .

*Demostración.* Sea  $\hat{x} = z + \langle x - z, v \rangle v$ , está claro que  $\hat{x}$  está en  $\mathcal{L}$ . Dado  $y$  en  $\mathcal{L}$ , entonces  $\hat{x} - y = \alpha v$ , para algún  $\alpha$  en  $\mathbb{R}$ . Por otro lado,

$$\begin{aligned} \langle x - \hat{x}, \hat{x} - y \rangle &= \langle x - \hat{x}, \alpha v \rangle \\ &= \alpha \langle x - \hat{x}, v \rangle \\ &= \alpha \langle x - z - \langle x - z, v \rangle v, v \rangle \\ &= \alpha (\langle x - z, v \rangle - \langle x - z, v \rangle) \\ &= 0, \end{aligned}$$

lo cual, por el Teorema de Pitágoras, se tiene que

$$\begin{aligned} \|x - y\|^2 &= \|(x - \hat{x}) + (\hat{x} - y)\|^2 \\ &= \|x - \hat{x}\|^2 + \|\hat{x} - y\|^2. \end{aligned}$$

De esta última igualdad, queda claro que para todo  $y \neq \hat{x}$  en  $\mathcal{L}$  se cumple que

$$\|x - \hat{x}\| < \|x - y\|,$$

lo cual implica la unicidad de  $\hat{x}$ . □

**Observación 2.40.** Este teorema nos dice que hallar la distancia de un punto  $x$  a una recta  $\mathcal{L} = \mathcal{L}(z, u)$  se reduce a calcular la distancia entre los vectores  $x$  y  $\hat{x}$ , es decir,

$$d(x, \mathcal{L}) = \|x - z - \langle x - z, u \rangle u\|. \quad (2.1)$$

Con el mismo espíritu de la prueba de este resultado, se demuestra el siguiente resultado.

**Propiedad 2.41.** Sea  $x \in V$  y sea  $\mathcal{P} = \mathcal{P}(z, u, v)$  un plano de  $V$ . Entonces, existe un único  $\hat{x} \in \mathcal{P}$  tal que

$$\|x - \hat{x}\| \leq \|x - y\|$$

para todo  $y \in \mathcal{P}$ . Además, se tiene que  $\hat{x} = z + \langle x - z, u \rangle u + \langle x - z, v \rangle v$ .

**Propiedad 2.42.** Sea  $x \in V$  y sea  $\mathcal{H} = \mathcal{H}(u, \alpha)$  el hiperplano de  $V$  que pasa por  $z \in V$ . Entonces, existe un único  $\hat{x} \in \mathcal{H}$  tal que

$$\|x - \hat{x}\| \leq \|x - y\|$$

para todo  $y \in \mathcal{H}$ . Además, se tiene que  $\hat{x} = z + \langle x - z, u \rangle u$ .

*Demostración.* Sea  $\hat{x} = x - \langle x - z, u \rangle u$ , luego

$$\begin{aligned}\langle \hat{x}, u \rangle &= \langle x - \langle x - z, u \rangle u, u \rangle \\ &= \langle x, u \rangle - \langle x - z, u \rangle \\ &= \langle x, u \rangle - \langle x, u \rangle + \langle z, u \rangle \\ &= \langle z, u \rangle.\end{aligned}$$

Además, como el hiperplano  $\mathcal{H}$  pasa por  $z$ ,  $\langle z, u \rangle = \alpha$ . Entonces,  $\langle \hat{x}, u \rangle = \alpha$ , lo cual implica que  $\hat{x}$  está en  $\mathcal{H}$ . Dado  $y$  en  $\mathcal{H}$ , entonces  $\langle y, u \rangle = \alpha$ . Por otro lado,

$$\begin{aligned}\langle x - \hat{x}, \hat{x} - y \rangle &= \langle x - (x - \langle x - z, u \rangle u), \hat{x} - y \rangle \\ &= \langle \langle x - z, u \rangle u, \hat{x} - y \rangle \\ &= \langle x - z, u \rangle \langle u, \hat{x} - y \rangle \\ &= \langle x - z, u \rangle (\langle u, \hat{x} \rangle - \langle u, y \rangle) \\ &= \langle x - z, u \rangle (\alpha - \alpha) \\ &= 0,\end{aligned}$$

lo cual, por el Teorema de Pitágoras, se tiene que

$$\begin{aligned}\|x - y\|^2 &= \|(x - \hat{x}) + (\hat{x} - y)\|^2 \\ &= \|x - \hat{x}\|^2 + \|\hat{x} - y\|^2.\end{aligned}$$

De esta última igualdad, queda claro que para todo  $y \neq \hat{x}$  en  $\mathcal{L}$  se cumple que

$$\|x - \hat{x}\| < \|x - y\|,$$

lo cual implica la unicidad de  $\hat{x}$ . □

#### 2.1.4. El espacio euclídeo de dimensión finita

En virtud de la propiedad 2.24, viendo a través de los lentes del álgebra lineal, existe un sólo espacio euclídeo de dimensión  $n \in \mathbb{N}$ . Sin embargo, daremos tres de los ejemplos más importantes de espacios vectoriales de dimensión finita:  $\mathbb{R}^n$ ,  $\mathcal{M}(m, n)$  y  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ .

**Definición 2.43** (El conjunto de las  $n$ -uplas de números reales). Dado  $n \in \mathbb{N}$ , el conjunto de las  $n$ -uplas de números reales, denotado por  $\mathbb{R}^n$ , es el producto cartesiano de los  $n$  conjuntos todos iguales a  $\mathbb{R}$ .

**Observación 2.44.**

- (i) Para cada  $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$  y  $\alpha \in \mathbb{R}$ , definimos la *suma* de  $x$  e  $y$  como la  $n$ -upla

$$x + y = (x_1 + y_1, \dots, x_n + y_n)$$

y el *producto* de  $\alpha$  por  $x$  como la  $n$ -upla

$$\alpha x = (\alpha x_1, \dots, \alpha x_n).$$

- (ii) Se verifica que estas operaciones satisfacen los axiomas de espacio vectorial, donde el vector nulo o cero es  $(0, \dots, 0)$  y el vector opuesto de  $x = (x_1, \dots, x_n)$  es  $-x = (-x_1, \dots, -x_n)$ . En lo sucesivo, denotaremos con  $\mathbb{R}^n$  al espacio vectorial de las  $n$ -uplas de números reales provisto de la suma y producto definidos arriba.

**Ejemplo 2.45.** En  $\mathbb{R}^n$ , para cada  $i = 1, \dots, n$  denotemos  $e^i$  al vector cuya  $i$ -ésima coordenada es igual a 1 y las demás son todas iguales a 0. Luego, se verifica que el conjunto  $\{e^1, \dots, e^n\}$  es linealmente independiente.

**Ejemplo 2.46** (Base canónica de  $\mathbb{R}^n$ ). El conjunto  $\{e^1, \dots, e^n\}$  definido en el ejemplo 2.45, genera  $\mathbb{R}^n$ . En efecto, cada vector  $x = (x_1, \dots, x_n)$  se puede expresar como  $x_1 e^1 + \dots + x_n e^n$ . Por lo tanto, ya que es además un conjunto linealmente independiente, es una base y recibe el nombre de *base canónica*, llamada también *base natural*, de  $\mathbb{R}^n$ .

**Definición 2.47** (Producto interno canónico). Sean  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$  en  $\mathbb{R}^n$ . Definimos el *producto interno canónico* de  $x$  e  $y$  como  $x_1 y_1 + \dots + x_n y_n$ .

**Observación 2.48.** Se verifica que el producto interno canónico es un producto interno en  $\mathbb{R}^n$ . De esta manera, las nociones de ortogonalidad entre dos vectores, longitud de un vector, vector unitario, distancia entre dos vectores y base ortonormal, se hacen presentes en  $\mathbb{R}^n$ . De ahora en adelante denotaremos con  $\mathbb{R}^n$  a nuestro espacio vectorial usual dotado con el producto interno canónico.

**Definición 2.49** (El conjunto de las matrices de tamaño  $m \times n$ ). Sean  $m, n \in \mathbb{N}$ . Una *matriz* de tamaño  $m \times n$ ,  $A$ , es un arreglo rectangular de números reales dispuestos en  $m$  filas y  $n$  columnas, donde el número que yace en la  $i$ -ésima fila y  $j$ -ésima columna, denotado como  $A_j^i$ , es llamado la  $(i, j)$ -ésima *entrada* de  $A$ . Así, esta

matriz se representa como:

$$A = \begin{bmatrix} A_1^1 & A_2^1 & \cdots & A_n^1 \\ A_1^2 & A_2^2 & \cdots & A_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ A_1^m & A_2^m & \cdots & A_n^m \end{bmatrix}.$$

- (i) Cuando  $n = 1$ ,  $A$  es llamada *matriz columna* y su  $(i, 1)$ -ésima entrada es llamada simplemente  $i$ -ésima entrada.
- (ii) Cuando  $m = 1$ ,  $A$  es llamada *matriz fila* y su  $(1, j)$ -ésima entrada es llamada simplemente  $j$ -ésima entrada.
- (iii) Cuando  $m = n$ ,  $A$  es llamada *matriz cuadrada* de tamaño  $n$ .
- (iv) Denotaremos con  $\mathcal{M}(m, n)$  al conjunto de todas las matrices de tamaño  $m \times n$ .

**Observación 2.50.**

- (i) Decimos que  $A, B \in \mathcal{M}(m, n)$  son *iguales* si sus respectivas entradas son iguales, es decir, la  $(i, j)$ -ésima entrada de  $A$  es igual a la  $(i, j)$ -ésima entrada de  $B$  para cada  $i = 1, \dots, m$  y cada  $j = 1, \dots, n$ .
- (ii) Para cada  $A, B \in \mathcal{M}(m, n)$  y  $\alpha \in \mathbb{R}$ , definimos la *suma* de  $A$  y  $B$  como la matriz  $A + B \in \mathcal{M}(m, n)$  cuya  $(i, j)$ -ésima entrada es la suma de las  $(i, j)$ -ésimas entradas de  $A$  y  $B$ . Y el *producto* de  $\alpha$  por  $A$  como la matriz  $\alpha A \in \mathcal{M}(m, n)$  cuya  $(i, j)$ -ésima entrada es el producto de  $\alpha$  por la  $(i, j)$ -ésima entradas de  $A$ .
- (iii) Se verifica que estas operaciones satisfacen los axiomas de espacio vectorial. En lo sucesivo, denotaremos con  $\mathcal{M}(m, n)$  al espacio de las matrices de tamaño  $m \times n$  provisto de la suma y producto definidos arriba.

**Definición 2.51** (Producto matricial). Para cada  $A \in \mathcal{M}(m, n)$  y cada  $B \in \mathcal{M}(n, p)$  definimos el *producto* de  $A$  por  $B$  como la matriz  $AB \in \mathcal{M}(m \times p)$  cuya  $(i, j)$ -ésima entrada está dada por  $A_1^i B_j^1 + \cdots + A_n^i B_j^n$ .

**Observación 2.52.**

- (i) Para cada  $A \in \mathcal{M}(m, n)$  definimos la *transpuesta* de  $A$  como la matriz  $A^T \in \mathcal{M}(n, m)$  cuya  $(j, i)$ -ésima entrada es igual a la  $(i, j)$ -ésima entrada de  $A$ .

(ii) Al espacio de las matrices columnas  $\mathcal{M}(n, 1)$  podemos dotarlo de un producto interno si para cada  $x, y \in \mathcal{M}(n, 1)$  definimos el producto interno de  $x$  e  $y$  como

$$x^\top y = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + \cdots + x_n y_n.$$

De esta manera, las nociones de ortogonalidad entre dos vectores, longitud de un vector, vector unitario y base ortonormal, se hacen presentes en  $\mathcal{M}(n, 1)$ . De ahora en adelante denotaremos con  $\mathcal{M}(n, 1)$  a nuestro espacio vectorial usual dotado con el producto interno que se acaba de definir.

Las tres propiedades a continuación son propiedades bastante universales y por ello sólo las enunciamos.

**Propiedad 2.53.** Sean  $A \in \mathcal{M}(m, n)$  y  $B \in \mathcal{M}(n, p)$ . Entonces,

$$AB = \sum_{i=1}^n \begin{bmatrix} A_i^1 \\ \vdots \\ A_i^m \end{bmatrix} \begin{bmatrix} B_1^i & \cdots & B_p^i \end{bmatrix}.$$

**Propiedad 2.54.** Sean  $A, B, C \in \mathcal{M}(m, n)$ , sean  $D, E \in \mathcal{M}(n \times p)$  y sea  $F \in \mathcal{M}(p \times q)$ . Se cumple:

- (i)  $(A + B) + C = A + (B + C)$ .
- (ii)  $(AD)F = A(DF)$ .
- (iii)  $(A + B)D = AD + BD$ .
- (iv)  $A(D + E) = AD + AE$ .

**Propiedad 2.55.** Sean  $A \in \mathcal{M}(m, n)$  y  $B \in \mathcal{M}(n, p)$ . Entonces,

$$(AB)^\top = B^\top A^\top \quad \text{y} \quad (A^\top)^\top = A.$$

**Definición 2.56** (Matriz simétrica y semidefinida positiva). Sea  $A$  una matriz cuadrada de tamaño  $n$ . Se dice que  $A$  es una *matriz simétrica* si  $A = A^\top$ . Si una matriz simétrica  $A$  satisface

$$x^\top Ax \geq 0$$

para todo  $x \in \mathcal{M}(n, 1)$ , entonces se le llama *matriz semidefinida positiva*.

**Ejemplo 2.57** (Matriz identidad). Aquella matriz cuadrada de tamaño  $n$  cuya  $(i, j)$ -ésima entrada es igual a 0 para todo  $i \neq j$  e igual a 1 caso contrario:

$$I_n := \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix},$$

es llamada *matriz identidad* y simplemente denotada como  $I$  cuando  $n$  se desprende del contexto.

**Propiedad 2.58.** Sea  $A \in \mathcal{M}(m, n)$ . Entonces,  $A^\top A$  es semidefinida positiva.

*Demostración.* En virtud de la propiedad 2.55, tenemos que  $A^\top A$  es una matriz simétrica. Sean  $x \in \mathcal{M}(n, 1)$ . Luego, otra vez por la misma propiedad:

$$x^\top (A^\top A)x = (x^\top A^\top)(Ax) = (Ax)^\top (Ax) \geq 0.$$

□

**Definición 2.59** (El espacio de las transformaciones lineales de  $\mathbb{R}^n$  en  $\mathbb{R}^m$ ). En virtud de la observación 2.21, al espacio vectorial de las transformaciones lineales de  $\mathbb{R}^n$  en  $\mathbb{R}^m$  lo denotamos con  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ .

**Observación 2.60.** De ahora en adelante, nos mantendremos con los lentes del álgebra lineal puestos y en virtud de la propiedad 2.24:

- (i) No haremos distinción entre  $\mathbb{R}^n$  y  $\mathcal{M}(n, 1)$  pues ambos tienen dimensión  $n$ .
- (ii) No haremos distinción entre  $\mathcal{M}(m, n)$  y  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  pues ambos tienen dimensión  $mn$ .

**Propiedad 2.61.** Sea  $u \in \mathbb{R}^n$  unitario. Entonces,  $A := I - uu^\top$  es semidefinida positiva.

*Demostración.* Es claro que  $A$  es una matriz simétrica. Luego, como  $A = AA$ , entonces  $A = A^\top A$  y por la propiedad 2.58, concluimos la prueba. □

**Propiedad 2.62.** Para cada  $x, u \in \mathbb{R}^n$  con  $\|u\| = 1$  se cumple que la expresión

$$\|x - \langle x, u \rangle u\|^2$$

es equivalente a las siguientes expresiones:

- (i)  $x^\top (I - uu^\top) x$ .

$$(ii) \quad x^\top x - u^\top (xx^\top) u.$$

*Demostración.* Por un lado tenemos:

$$\begin{aligned} \|x - \langle x, u \rangle u\|^2 &= \|x\|^2 - 2\langle x, \langle x, u \rangle u \rangle + \|\langle x, u \rangle u\|^2 \\ &= \|x\|^2 - 2\langle x, u \rangle^2 + \langle x, u \rangle^2 \\ &= \|x\|^2 - \langle x, u \rangle \langle x, u \rangle \\ &= \langle x, x \rangle - \langle x, \langle x, u \rangle u \rangle \\ &= x^\top x - x^\top ((x^\top u) u) \\ &= x^\top x - x^\top ((u^\top x) u) \\ &= x^\top x - x^\top (u(u^\top x)) \\ &= x^\top Ix - x^\top (uu^\top) x \\ &= x^\top (I - uu^\top) x. \end{aligned}$$

Por otro lado:

$$\begin{aligned} x^\top x - x^\top (uu^\top) x &= x^\top x - (x^\top u) (u^\top x) \\ &= x^\top x - (u^\top x) (x^\top u) \\ &= x^\top x - u^\top (xx^\top) u. \end{aligned}$$

□

La siguiente propiedad se sigue de la propiedad 2.62 por un argumento inductivo.

**Propiedad 2.63.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$  y sea  $u \in \mathbb{R}^n$  unitario. Entonces, la expresión

$$\sum_{i=1}^m \|x^i - \langle x^i, u \rangle u\|^2$$

es equivalente a las siguientes expresiones:

- (i)  $\sum_{i=1}^m \langle x^i, (I - uu^\top) x^i \rangle.$
- (ii)  $\sum_{i=1}^m \langle x^i, x^i \rangle - \langle u, (\sum_{i=1}^m x^i (x^i)^\top) u \rangle.$

Como el objeto de estudio del presente trabajo son las matrices simétricas, entonces restringiremos nuestro análisis a este tipo de matrices. En este contexto, definimos a continuación dos conceptos que se relacionan uno con el otro.

**Definición 2.64** (Autovalor y autovector). Un *autovector* de una matriz simétrica  $A$  de tamaño  $n$  es un vector  $x \in \mathbb{R}^n$  no nulo tal que

$$Ax = \lambda x$$

para algún  $\lambda \in \mathbb{R}$ . El número  $\lambda$  es llamado *autovalor* de  $A$  asociado al autovector  $x$ . Denotaremos como  $\text{spec}(A)$  al conjunto de todos los autovalores de  $A$ .

**Propiedad 2.65.** Sea  $A$  una matriz semidefinida positiva. Todo autovalor de  $A$  es un número real no negativo.

*Demostración.* Sea  $\lambda \in \text{spec}(A)$  y sea  $x$  un autovector asociado a  $\lambda$ . Como

$$\lambda \langle x, x \rangle = \langle x, \lambda x \rangle = \langle x, Ax \rangle \geq 0,$$

y  $x \neq 0$ , entonces  $\lambda \geq 0$ . □

**Teorema 2.66.** Sea  $A$  una matriz semidefinida positiva de tamaño  $n$ . Entonces, existe una base ortonormal  $\{u^1, \dots, u^n\}$  de  $\mathbb{R}^n$  formada por autovectores de  $A$  donde sus respectivos autovalores asociados satisfacen:

$$\lambda_1 \geq \dots \geq \lambda_n \geq 0 \quad \text{y} \quad \text{spec}(A) = \{\lambda_1, \dots, \lambda_n\}.$$

*Demostración.* En virtud del teorema espectral, teorema 13.6 de [Lima, 2011a], sea  $\mathcal{B} := \{u^1, \dots, u^n\}$  una base ortonormal de  $\mathbb{R}^n$  formada por autovectores de  $A$  y sean  $\lambda_1, \dots, \lambda_n$  sus respectivos autovalores asociados. Luego, por la propiedad 2.65, podemos reindexar dichos autovectores y autovalores de tal manera que

$$\lambda_1 \geq \dots \geq \lambda_n \geq 0.$$

Sea  $\lambda \in \text{spec}(A)$ , entonces  $Ax = \lambda x$  para algún  $x \in \mathbb{R}^n$  no nulo. Así, por ser  $\mathcal{B}$  una base de  $\mathbb{R}^n$ , tenemos

$$x = \alpha_1 u^1 + \dots + \alpha_n u^n$$

para  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$  no todos iguales a cero, ya que  $x$  es no nulo. Para fijar ideas, asumamos  $\alpha_1 \neq 0$ . Luego, como

$$\begin{aligned} \alpha_1 \lambda_1 u^1 + \dots + \alpha_n \lambda_n u^n &= \alpha_1 A u^1 + \dots + \alpha_n A u^n \\ &= A(\alpha_1 u^1 + \dots + \alpha_n u^n) \\ &= \lambda(\alpha_1 u^1 + \dots + \alpha_n u^n) \\ &= \alpha_1 \lambda u^1 + \dots + \alpha_n \lambda u^n, \end{aligned}$$

entonces

$$\alpha_1(\lambda_1 - \lambda)u^1 + \dots + \alpha_n(\lambda_n - \lambda)u^n = 0,$$

lo cual, por ser  $\mathcal{B}$  linealmente independiente, implica que

$$\alpha_1(\lambda_1 - \lambda) = \dots = \alpha_n(\lambda_n - \lambda) = 0.$$

Entonces, como  $\alpha_1(\lambda_1 - \lambda) = 0$  y  $\alpha_1 \neq 0$ ,  $\lambda = \lambda_1$ , lo cual implica que  $\text{spec}(A) \subset \{\lambda_1, \dots, \lambda_n\}$ . □

## 2.2. Cálculo diferencial

El cálculo diferencial es una parte del análisis matemático que consiste en el estudio de cómo cambian las funciones cuando sus variables cambian. El principal objeto de estudio en el cálculo diferencial es la derivada.

### 2.2.1. La regla de la cadena

A continuación, presentaremos la regla de la cadena cuya prueba puede verse en la página 256 de [Lima, 2011b] y mostraremos algunas aplicaciones de este resultado del cálculo diferencial.

**Teorema 2.67** (Regla de la cadena). Sea  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  una función diferenciable en  $a \in \mathbb{R}^n$  y sea  $G : \mathbb{R}^m \rightarrow \mathbb{R}^p$  una función diferenciable en  $b = F(a)$ . Entonces,  $H : \mathbb{R}^n \rightarrow \mathbb{R}^p$  definida mediante  $H(x) = G(F(x))$  para cada  $x \in \mathbb{R}^n$ , es diferenciable en  $a$  y se tiene que  $H'(a) = G'(b)F'(a)$ .

**Definición 2.68** (Gradiente y Hessiana). Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función dos veces diferenciable. Para cada  $a \in \mathbb{R}^n$ , se define el *gradiente* de  $f$  en  $a$  como la matriz columna  $\nabla f(a)$  de tamaño  $n \times 1$  cuya  $i$ -ésima entrada está dada por

$$\frac{\partial f}{\partial x_i}(a)$$

y la *matriz hessiana* de  $f$  en  $a$  como la matriz cuadrada  $\nabla^2 f(a)$  de tamaño  $n$  cuya  $(i, j)$ -ésima entrada está dada por

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a).$$

**Ejemplo 2.69.** Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función dos veces diferenciable y  $a, d \in \mathbb{R}^n$ . Sea la función  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  definida como sigue

$$\theta(t) = f(H(t)) \quad \text{con} \quad H(t) = a + td.$$

Entonces, para cada  $t \in \mathbb{R}$  se cumple:

- (i)  $\theta'(t) = \langle \nabla f(p + td), d \rangle$ .
- (ii)  $\theta''(t) = \langle d, \nabla^2 f(p + td)d \rangle$ .

**Ejemplo 2.70.** Sea  $A$  una matriz simétrica de tamaño  $n$  y sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  definida mediante

$$f(x) = \langle x, Ax \rangle.$$

Entonces, ya que  $f$  es dos veces diferenciable por ser una función polinomial, para cada  $z \in \mathbb{R}^n$  se cumple:

(i)  $\nabla f(z) = 2Az.$

(ii)  $\nabla^2 f(z) = 2A.$

De la definición del concepto de derivada se tiene el siguiente resultado.

**Lema 2.71.** Sean  $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  funciones diferenciables en  $a \in \mathbb{R}^n$ . Entonces, la función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  definida mediante

$$f(x) = f_1(x) + \dots + f_m(x)$$

es diferenciable en  $a$  y se tiene que  $f'(a) = f'_1(a) + \dots + f'_m(a)$ .

**Teorema 2.72.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$ , sea  $A$  una matriz simétrica de tamaño  $n$  y sean  $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  funciones diferenciables definidas por

$$f_i(y) = \langle x^i - y, A(x^i - y) \rangle$$

para cada  $i = 1, \dots, m$ . Luego, la función  $f := f_1 + \dots + f_m$  cumple que para cada  $z \in \mathbb{R}^n$ :

(i)  $\nabla f(z) = -2A(x^1 + \dots + x^m - mz).$

(ii)  $\nabla^2 f(z) = 2mA.$

*Demostración.* Sea  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g(z) = \langle z, Az \rangle$ . Luego,  $f_i = g \circ h_i$  donde  $h_i(z) = x^i - z$  para cada  $i = 1, \dots, m$ . Sea  $z \in \mathbb{R}^n$ . En virtud de la regla de la cadena y el ejemplo 2.70:

$$\nabla f_i(z) = \nabla g(h_i(z))(-I) = 2A(x^i - z)(-I) = -2A(x^i - z)$$

para cada  $i = 1, \dots, m$ . Luego, del lema 2.71,

$$\nabla f(z) = \sum_{i=1}^m -2A(x^i - z) = -2A \sum_{i=1}^m (x^i - z) = -2A(x^1 + \dots + x^m - mz).$$

Finalmente, (ii) se sigue de (i). □

## 2.3. Optimización matemática

En un problema típico de optimización matemática el objetivo es minimizar (o maximizar) una determinada función, denominada función de costos (beneficios, producción, etc.), sujeto a restricciones en la variable disponibilidad (tiempo, presupuesto, etc.) Este problema se dice que es un *problema de programación matemática*, siguiendo la terminología militar americana de la Segunda Guerra Mundial.

La formulación general de un problema de minimización tiene la siguiente forma:

$$\text{mín } \{ f(x) ; x \in C \}, \quad (2.2)$$

donde

- (i)  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es llamada *función objetivo*,
- (ii)  $\emptyset \neq C \subset \mathbb{R}^n$  es llamado *conjunto de las soluciones admisibles o factibles o viables*,
- (iii) el conjunto de aquellos  $x \in C$  donde  $f(x)$  es igual a la expresión (2.2) es llamado *conjunto de las soluciones optimales* y
- (iv) el valor que una solución optimal toma bajo  $f$  es llamado de *valor optimal*.

Un problema de optimización consiste en encontrar el conjunto de soluciones optimales y el valor optimal. Si  $C = \mathbb{R}^n$ , se dice que el problema es un *problema sin restricciones*.

Sea  $a \in C$ . Se dice que  $f$  tiene un *mínimo global* sobre  $C$  en  $a$  si este vector es una solución optimal del problema (2.2). Se dice que  $f$  tiene un *mínimo local* sobre  $C$  en  $a$  si existe una bola abierta  $B := B(a, r)$  tal que

$$f(a) \leq f(x), \quad \forall x \in C \cap B.$$

Además, nos enfocaremos en un tipo especial de problema de minimización:

$$\begin{aligned} &\text{mín } \{ f(x) ; x \in C \} \\ &C = \{ x \in \mathbb{R}^n ; h_1(x) = 0, \dots, h_q(x) = 0 \}. \end{aligned} \quad (2.3)$$

En este planteamiento,  $f, h_1, \dots, h_q : \mathbb{R}^n \rightarrow \mathbb{R}$  son diferenciables y (2.3) es llamado *problema de programación diferenciable*. A continuación, resumiremos propiedades topológicas y geométricas de conjuntos convexos y funciones convexas que serán empleados para desarrollar condiciones de optimalidad para estos problemas.

### 2.3.1. Conjuntos convexos

**Definición 2.73** (Conjunto convexo). Sea  $X \subset \mathbb{R}^n$ . Se dice que  $X$  es un *conjunto convexo* si

$$\alpha x + (1 - \alpha)y \in X,$$

para todo  $x, y \in X$  y todo  $\alpha \in [0, 1]$ .

**Ejemplo 2.74.**

(i) Todo conjunto afín de  $\mathbb{R}^n$  es un conjunto convexo: conjuntos unitarios, rectas, planos, etc.

(ii) Sea  $v \in \mathbb{R}^n$  y sea  $c \in \mathbb{R}$ . El conjunto

$$\{x \in \mathbb{R}^n ; \langle x, v \rangle \leq c\}$$

es un conjunto convexo.

**Propiedad 2.75.** Intersección de conjuntos convexos es un conjunto convexo.

**Ejemplo 2.76.** Sea  $A$  una matriz de tamaño  $m \times n$  y sea  $b \in \mathbb{R}^m$ . El conjunto

$$\{x \in \mathbb{R}^n ; Ax \leq b\}$$

es un conjunto convexo.

**Definición 2.77** (Cono). Sea  $X \subset \mathbb{R}^n$ . Se dice que  $X$  es un *cono* si

$$\lambda x \in X,$$

para todo  $x \in X$  y todo  $\lambda > 0$ .

**Ejemplo 2.78.**

(i) Todo subespacio vectorial es un cono.

(ii) El conjunto

$$\{(x, y) \in \mathbb{R}^2 ; x \geq 0, y \geq 0\}$$

es un cono.

**Observación 2.79.** El siguiente teorema es un caso particular del teorema 1.5 de [Crouzeix et al., 2011]. Y los dos corolarios que se siguen son consecuencias directas del teorema a continuación.

**Teorema 2.80.** Sea  $C \subset \mathbb{R}^n$  un convexo cerrado no vacío y sea  $a \in \mathbb{R}^n$ . Entonces, existe un único  $b \in \mathbb{R}^n$  tal que

$$b \in C, \quad \forall x \in C : \|b - a\| \leq \|x - a\|. \quad (2.4)$$

Se dice que  $b$  es la *proyección* de  $a$  sobre  $C$  y se denota por  $\text{proj}_C(a)$ . Más aún, la proposición (2.4) es equivalente a la siguiente proposición:

$$b \in C, \quad \forall x \in C : \langle a - b, x - b \rangle \leq 0. \quad (2.5)$$

**Corolario 2.81.** Sea  $C \subset \mathbb{R}^n$  un convexo cerrado no vacío y sea  $a \in \mathbb{R}^n$ . Luego,

$$a \in C \quad \Leftrightarrow \quad a = \text{proj}_C(a).$$

**Corolario 2.82.** Sea  $C \subset \mathbb{R}^n$  un convexo cerrado no vacío, sea  $a \in \mathbb{R}^n$  y sea  $f(x) = \|x - a\|$ . El problema de minimización

$$\text{mín} \{ f(x) ; x \in C \},$$

tiene como única solución optimal al vector  $\text{proj}_C(a)$ .

**Ejemplo 2.83** (Proyección sobre una recta, un plano y un hiperplano). Sean  $\mathcal{L} = \mathcal{L}(z, u)$ ,  $\mathcal{P} = \mathcal{P}(z, u, v)$  y  $\mathcal{H} = \mathcal{H}(w, \alpha)$  una recta, un plano y un hiperplano que pasa por el punto  $z$ , respectivamente. Entonces, en virtud de las propiedades 2.39, 2.41 y 2.42, para cada  $x \in \mathbb{R}^n$ :

$$\text{proj}_{\mathcal{L}}(x) = z + \langle x - z, v \rangle v.$$

$$\text{proj}_{\mathcal{P}}(x) = z + \langle x - z, u \rangle u + \langle x - z, v \rangle v.$$

$$\text{proj}_{\mathcal{H}}(x) = x - \langle x - z, w \rangle w.$$

### 2.3.2. Funciones convexas

Comenzaremos definiendo el concepto de función convexa. Luego, caracterizaremos a las funciones convexas que son dos veces diferenciables.

**Definición 2.84** (Función convexa). Se dice que una función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es una *función convexa* si

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y),$$

para todo  $x, y \in \mathbb{R}^n$  y todo  $\alpha \in [0, 1]$ .

**Ejemplo 2.85.** Las siguientes funciones  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  son convexas:

- (i)  $f(x) = x$ ,
- (ii)  $f(x) = 2x + 1$ ,
- (iii)  $f(x) = x^2$ .

**Lema 2.86.** Sea  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  una función dos veces diferenciable. Las siguientes proposiciones son equivalentes:

- (i)  $\theta$  es una función convexa.
- (ii)  $\theta(t) \geq \theta(s) + (t - s)\theta'(s)$  para todo  $s, t \in \mathbb{R}$ .
- (iii)  $\theta''(t) \geq 0$  para todo  $t \in \mathbb{R}$ .

*Demostración.* Ver pruebas de la proposición 2.9 de [Crouzeix et al., 2011] y del teorema 11 del capítulo 8 de [Lima, 2011c].  $\square$

**Observación 2.87.** El siguiente resultado nos muestra un camino natural para estudiar a las funciones convexas de varias variables a partir de las funciones de una variable.

**Lema 2.88.** Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Para cada  $p, d \in \mathbb{R}^n$  se define la función  $\theta_{p,d} : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\theta_{p,d}(t) = f(p + td)$ . Las siguientes proposiciones son equivalentes:

- (i)  $f$  es una función convexa.
- (ii)  $\theta_{p,d}$  es una función convexa para todo  $p, d \in \mathbb{R}^n$ .

*Demostración.* Primero, asumamos la proposición (i). Dados  $p, d \in \mathbb{R}^n$ , afirmamos que  $\theta_{p,d}$  es una función convexa. En efecto, dados  $s, t \in \mathbb{R}$  y  $\alpha \in [0, 1]$ ,

$$\begin{aligned}
 \theta_{p,d}(\alpha s + (1 - \alpha)t) &= f(p + \alpha sd + (1 - \alpha)td) \\
 &= f(\alpha(p + sd) + (1 - \alpha)(p + td)) \\
 &\leq \alpha f(p + sd) + (1 - \alpha)f(p + td) \\
 &\leq \alpha \theta_{p,d}(s) + (1 - \alpha)\theta_{p,d}(t).
 \end{aligned}$$

Ahora, asumamos la proposición (ii). Dados  $x, y \in \mathbb{R}^n$  y  $\alpha \in [0, 1]$ . Por hipótesis,  $\theta_{y,x-y}$  es una función convexa, entonces

$$\begin{aligned}
 f(\alpha x + (1 - \alpha)y) &= f(y + \alpha(x - y)) \\
 &= \theta_{y,x-y}(\alpha 1 + (1 - \alpha)0) \\
 &\leq \alpha \theta_{y,x-y}(1) + (1 - \alpha)\theta_{y,x-y}(0) \\
 &\leq \alpha f(x) + (1 - \alpha)f(y).
 \end{aligned}$$

$\square$

**Teorema 2.89.** Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función dos veces diferenciable. Las siguientes proposiciones son equivalentes:

- (i)  $f$  es una función convexa.
- (ii)  $f(y) \geq f(x) + \langle y - x, \nabla f(x) \rangle$  para todo  $x, y \in \mathbb{R}^n$ .
- (iii)  $\nabla^2 f(x)$  es una matriz semidefinida positiva para todo  $x \in \mathbb{R}^n$ .

*Demostración.* Primero, vamos a demostrar que las proposiciones (i) y (ii) son equivalentes. En virtud del lema 2.88,  $f$  es una función convexa si, y sólo si, para cada  $p, d \in \mathbb{R}^n$ ,

$$\theta_{p,d}(t) \text{ es una función convexa.} \quad (2.6)$$

Como  $\theta_{p,d}$  es una función diferenciable, en base al ejemplo 2.69, entonces

$$\theta'_{p,d}(s) = \langle \nabla f(p + sd), d \rangle$$

por el lema 2.86, la proposición (2.6) es equivalente a

$$f(p + td) \geq f(p + sd) + (t - s)\langle \nabla f(p + sd), d \rangle, \text{ para todo } s, t \text{ en } \mathbb{R}. \quad (2.7)$$

Por un lado, si se cumple la proposición (i), entonces para todo  $p, d \in \mathbb{R}^n$  se cumple la proposición (2.7). Fijando  $x, y \in \mathbb{R}^n$ , luego reemplazando  $p = x$ ,  $d = y - x$ ,  $s = 0$  y  $t = 1$  en la proposición (2.7), tenemos que

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle. \quad (2.8)$$

Por otro lado, si se cumple la proposición (ii), entonces para todo  $x, y \in \mathbb{R}^n$  se cumple la desigualdad (2.8). Fijando  $p, d \in \mathbb{R}^n$ , para cada  $s, t \in \mathbb{R}$  reemplazamos  $x = p + sd$  e  $y = p + td$  en la desigualdad (2.8), luego tenemos que se cumple la proposición (2.7), y se tiene la proposición (i).

Ahora, demostraremos que las proposiciones (i) y (iii) son equivalentes. Una vez más, por el lema 2.88, es necesario y suficiente probar que todo  $p, d \in \mathbb{R}^n$  se cumple la proposición (2.6). Como  $\theta_{p,d}$  es una función dos veces diferenciable, en base al ejemplo 2.69, entonces

$$\theta''_{p,d}(t) = \langle d, \nabla^2 f(p + td)d \rangle,$$

por el lema 2.86, la proposición (2.6) es equivalente a

$$\langle d, \nabla^2 f(p + td)d \rangle \geq 0, \text{ para todo } t \text{ en } \mathbb{R}. \quad (2.9)$$

Por un lado, si se cumple la proposición (i), entonces para todo  $p, d \in \mathbb{R}^n$  se satisface la proposición (2.9). Fijando  $x \in \mathbb{R}^n$ , luego reemplazando  $p = x$  y  $t = 0$  en la proposición (2.9), tenemos que

$$\langle d, \nabla^2 f(x)d \rangle \geq 0, \text{ para todo } d \text{ en } \mathbb{R}^n, \quad (2.10)$$

y tenemos la proposición (iii).

Por otro lado, si se cumple la proposición (iii), entonces todo  $x \in \mathbb{R}^n$  se satisface la proposición (2.10). Fijando  $p, d \in \mathbb{R}^n$ , para cada  $t \in \mathbb{R}$  reemplazamos  $x = p + td$  en la proposición (2.10), luego tenemos que se cumple la proposición (2.9), y tenemos la proposición (i).  $\square$

### 2.3.3. Condiciones de optimalidad

**Teorema 2.90** (Condición suficiente de optimalidad). Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función dos veces diferenciable donde  $\nabla^2 f(x)$  es una matriz semidefinida positiva para cada  $x \in \mathbb{R}^n$ . Si  $\nabla f(a) = 0$  para algún  $a \in \mathbb{R}^n$ , entonces  $a$  es un mínimo global de  $f$ .

*Demostración.* Como  $\nabla^2 f(x)$  es una matriz semidefinida positiva para cada  $x \in \mathbb{R}^n$ , en virtud del teorema 2.89, se sigue que

$$f(x) \geq f(a) + \langle x - a, \nabla f(a) \rangle,$$

para todo  $x \in \mathbb{R}^n$ . Finalmente, como  $\nabla f(a) = 0$ , la conclusión se sigue.  $\square$

Se ha visto una condición suficiente de optimalidad para problemas sin restricciones: teorema 2.90. Ahora, veamos una condición necesaria de optimalidad para el caso con restricciones, teorema: 2.96. Para ello, introduciremos dos nociones de cono tangente.

**Definición 2.91** (Cono tangente de Bouligand). Sea  $C \subset \mathbb{R}^n$  y sea  $a \in C$ . El *cono tangente de Bouligand* de  $C$  en  $a$ , denotado por  $\mathbf{T}(C, a)$ , es el conjunto de vectores  $d \in \mathbb{R}^n$  para los que

- (i) existe una sucesión  $(d_k) \subset \mathbb{R}^n$  que converge a  $d$ ,
- (ii) existe una sucesión  $(t_k) \subset (0, \infty)$  que converge a 0 y
- (iii)  $a + t_k d_k \in C$  para todo  $k$ .

**Observación 2.92.**  $\mathbf{T}(C, a)$  es un cono que contiene al origen. Si  $a \in \text{int}(C)$ , entonces  $\mathbf{T}(C, a) = \mathbb{R}^n$ .

**Definición 2.93** (Cono tangente linealizado). Sea

$$C = \{x \in \mathbb{R}^n ; h_1(x) = 0, \dots, h_q(x) = 0\}, \quad (2.11)$$

donde  $h_1, \dots, h_q : \mathbb{R}^n \rightarrow \mathbb{R}$  son diferenciables y sea  $a \in C$ . El conjunto

$$L(C, a) = \{d \in \mathbb{R}^n ; \langle \nabla h_1(a), d \rangle = 0, \dots, \langle \nabla h_q(a), d \rangle = 0\}$$

es llamado de *cono tangente linealizado* de  $C$  en  $a$ .

**Observación 2.94.**  $L(C, a)$  es un cono que contiene al origen; además, siempre que  $C$  esté dado por (2.11), se tiene que

$$T(C, a) \subset L(C, a).$$

No obstante, ambos conos en general no son iguales. Las condiciones que implican la igualdad de ambos conos se llaman *condiciones de calificación de las restricciones*.

**Teorema 2.95** (Condición de calificación de las restricciones). Sea  $C$  dado por (2.11) y sea  $a \in C$ . Si  $\nabla h_1(a), \dots, \nabla h_q(a)$  son linealmente independientes, entonces

$$T(C, a) = L(C, a).$$

*Demostración.* Ver prueba del teorema 3.5 de [Crouzeix et al., 2011].  $\square$

**Teorema 2.96** (Condiciones KKT). Dado un problema de programación diferenciable (2.3) con  $f, h_1, \dots, h_q : \mathbb{R}^m \rightarrow \mathbb{R}$  de clase  $C^1$  y sea  $a \in C$  con  $\nabla h_1(a), \dots, \nabla h_q(a)$  linealmente independientes. Si  $f$  tiene un mínimo local sobre  $C$  en  $a$ , entonces existe  $(\alpha_1, \dots, \alpha_q) \in \mathbb{R}^q$  tal que

$$\nabla f(a) + \alpha_1 \nabla h_1(a) + \dots + \alpha_q \nabla h_q(a) = 0. \quad (2.12)$$

*Demostración.* Se sigue de ensamblar el teorema 2.95 con el teorema 3.7 de [Crouzeix et al., 2011].  $\square$

**Corolario 2.97.** Sea  $S$  una matriz simétrica de tamaño  $n$ . En virtud del teorema 2.66, sea

$$\{u^1, \dots, u^n\}$$

una base ortonormal de  $\mathbb{R}^n$  formada por autovectores de  $S$  y sean

$$\lambda_1 \geq \dots \geq \lambda_n$$

sus respectivos autovalores. Para cada  $q = 1, \dots, n$ :  $u^q$  es una solución optimal y  $\lambda_q$  el valor optimal del problema:

$$\text{máx} \{ \langle x, Sx \rangle ; \langle x, u^1 \rangle = 0, \dots, \langle x, u^{q-1} \rangle = 0, \|x\| = 1 \}. \quad (2.13)$$

*Demostración.* Sea  $q \in \{1, \dots, n\}$  fijo. En primer lugar, como el conjunto de las soluciones admisibles es compacto y la función objetivo es continua, por el teorema de Weierstrass, el problema (2.13) tiene al menos una solución optimal  $a$ . Luego,  $a$  es solución optimal de un problema del tipo (2.3), esto lo logramos haciendo:

$$\begin{aligned} f(x) &= -\langle x, Sx \rangle, \\ h_1(x) &= \langle x, u^1 \rangle, \\ &\vdots \\ h_{q-1}(x) &= \langle x, u^{q-1} \rangle, \\ h_q(x) &= \|x\| - 1. \end{aligned}$$

Así, como

$$\begin{aligned} \nabla h_1(a) &= u^1, \dots, \nabla h_{q-1}(a) = u^{q-1}, \nabla h_q(a) = 2a, \\ \langle a, u^1 \rangle &= 0, \dots, \langle a, u^{q-1} \rangle = 0, \end{aligned}$$

entonces  $\nabla h_1(a), \dots, \nabla h_q(a)$  son linealmente independientes, y por el teorema 2.96, existe  $(\alpha_1, \dots, \alpha_q) \in \mathbb{R}^q$  que satisface la ecuación (2.12). Es decir, se debe cumplir:

$$-2Sa + \alpha_1 u^1 + \dots + \alpha_{q-1} u^{q-1} + 2\alpha_q a = 0. \quad (2.14)$$

Pero, como  $a$  es ortogonal a  $u^1, \dots, u^{q-1}$ , entonces

$$a = \gamma_q u^q + \dots + \gamma_n u^n$$

para algún  $\gamma_q, \dots, \gamma_n \in \mathbb{R}$ . Así,

$$-2Sa = (-2\gamma_q \lambda_q) u^q + \dots + (-2\gamma_n \lambda_n) u^n. \quad (2.15)$$

Luego, reemplazando (2.15) en la ecuación (2.14), tenemos:

$$\alpha_1 u^1 + \dots + \alpha_{q-1} u^{q-1} + (-2\gamma_q \lambda_q) u^q + \dots + (-2\gamma_n \lambda_n) u^n + 2\alpha_q a = 0,$$

de donde se infiere que

$$\alpha_1 = \dots = \alpha_{q-1} = 0.$$

Finalmente, de la ecuación (2.14), tenemos que  $a$  debe ser un autovector de  $S$ , y así, la conclusión se sigue.  $\square$

**Observación 2.98.** El corolario que se acaba de probar generaliza el ejemplo 3.6.2 de [Crouzeix et al., 2011].

### 2.3.4. Método del gradiente proyectado

El *método del gradiente proyectado* fue introducido por Goldstein en [Goldstein, 1964] para resolver el siguiente problema de optimización:

$$\text{máx } \{ f(x) ; x \in C \}.$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es diferenciable y  $C \subset \mathbb{R}^n$  es cerrado. La esencia de este método se basa en dos hechos:

(i) Condición necesaria de primer orden: si  $a$  es un máximo local de  $f$ , entonces

$$\langle \nabla f(a), d \rangle \leq 0 \quad \forall d \in T(C, a). \quad (2.16)$$

(ii) Proyección sobre un convexo cerrado no vacío: ver teorema 2.80.

Sea  $S$  una matriz simétrica de tamaño  $n$ . En virtud del teorema 2.66, sea

$$\{ u^1, \dots, u^n \}$$

una base ortonormal de  $\mathbb{R}^n$  formada por autovectores de  $S$  y sean

$$\lambda_1 \geq \dots \geq \lambda_n$$

sus respectivos autovalores. Para cada  $q = 1, \dots, n$ , en virtud del corolario 2.97,  $u^q$  es una solución optimal y  $\lambda_q$  el valor optimal del problema (2.13).

Luego, aplicando el método del gradiente proyectado, si  $a$  es una solución optimal del problema (2.13), entonces se debe cumplir la proposición (2.16). Además, por el teorema 2.95,

$$T(C, a) = L(C, a) = \{ d \in \mathbb{R}^n ; \langle u^1, d \rangle = 0, \dots, \langle u^{q-1}, d \rangle = 0, \langle a, d \rangle = 0 \}.$$

Así,

$$T(C, a) = \mathcal{H}(a, 1) - a,$$

donde

$$\mathcal{H}(a, 1) = \{ d \in \mathbb{R}^n ; \langle u^1, d \rangle = 0, \dots, \langle u^{q-1}, d \rangle = 0, \langle a, d \rangle = 1 \}.$$

Así, la proposición (2.16) es equivalente a la siguiente proposición:

$$\langle \nabla f(a), x - a \rangle \leq 0 \quad \forall x \in \mathcal{H}(a, 1).$$

Y esta a su vez es equivalente a

$$\langle (a + \alpha Sa) - a, x - a \rangle \leq 0 \quad \forall x \in \mathcal{H}(a, 1), \forall \alpha > 0.$$

Finalmente, en virtud de la proposición (2.5), si  $a$  es solución optimal del problema (2.13), entonces

$$a = \text{proj}_{\mathcal{H}(a, 1)}((I + \alpha S)a) \quad \forall \alpha > 0. \quad (2.17)$$

## 2.4. Estadística

A continuación, sintetizamos la descripción sobre la estadística, el análisis multivariante y el [Análisis de Componentes Principales \(ACP\)](#) dada por [\[Konishi, 2014\]](#) en las páginas xxiii y 225-233.

El objetivo de las ciencias estadísticas es desarrollar la metodología y la teoría para extraer información útil de los datos y para razonablemente inferir y elucidar fenómenos con incertidumbre en varios campos de las ciencias sociales y naturales. Los datos contienen información sobre el fenómeno considerado y el objetivo del análisis estadístico es expresar esta información en una forma comprensible usando procedimientos estadísticos.

El análisis multivariante se refiere a las técnicas utilizadas para analizar datos que surgen de múltiples variables entre las cuales hay algunas relaciones. Estas técnicas se han utilizado ampliamente para extraer información y patrones útiles a partir de datos multivariados y para comprender la estructura de fenómenos aleatorios. Estas técnicas incluyen la regresión, el análisis discriminante, el análisis cluster, el [ACP](#), etc., y se basan principalmente en la linealidad de las variables observadas. A continuación, pasamos a describir una de estas técnicas:

### 2.4.1. Análisis de componentes principales

El [ACP](#) es un método para consolidar las variables mutuamente correlacionadas de datos multidimensionales en nuevas variables mediante combinaciones lineales de las variables originales con mínima pérdida de la información en los datos observados. La información de interés se puede extraer de los datos a través de esta fusión de las múltiples variables que caracterizan a los individuos.

El [ACP](#) también puede ser utilizado como una técnica para la aprehensión visual de las estructuras de datos mediante la consolidación de datos de alta dimensión en un menor número de variables, la reducción de la dimensión (reducción de la dimensionalidad, también conocida como compresión dimensional), y la proyección de los resultados sobre una recta unidimensional, un plano bidimensional o un espacio tridimensional.

Para el uso del [ACP](#), se requiere que la base de datos contenga  $m$  observaciones de  $n$  variables. Toda esa información puede ser representada de manera compacta en una matriz de tamaño  $m \times n$ . Así, denotemos por  $A$  la matriz que representa la base de datos y denotemos por  $x^i$  la  $i$ -ésima línea de la matriz  $A$ . Luego, cada línea de la matriz  $A$  puede ser considerada como un vector de  $\mathbb{R}^n$ .

En primer lugar, se plantea encontrar el eje de proyección, i.e., el vector unitario sobre el que proyectaremos nuestros datos, conocido como el primer componente principal, que maximiza la varianza total. Veamos esto en términos precisos; estos  $m$  datos  $n$ -dimensionales son expresados como

$$y_i = \omega_1 x_1^i + \cdots + \omega_n x_n^i = \omega^\top x^i, \quad i = 1, \dots, m,$$

donde  $\omega = (\omega_1, \dots, \omega_n)$  representa el vector de coeficientes. Así, la media de los datos  $y^1, \dots, y^m$  está dada por

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y^i = \omega^\top \bar{x},$$

donde

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x^i.$$

Luego, la varianza de los datos  $y^1, \dots, y^m$  está dada por

$$\sigma^2 = \omega^\top S \omega, \tag{2.18}$$

donde  $S$  es la matriz de covarianza de  $A$  definida por

$$S = \frac{1}{m} B^\top B \tag{2.19}$$

en la que  $B$  es la matriz de tamaño  $m \times n$  cuya  $i$ -ésima línea está dada por  $x^i - \bar{x}$ . En efecto, la ecuación (2.18) se sigue del siguiente sistema:

$$\begin{aligned} \sigma^2 &= \frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2 \\ &= \frac{1}{m} \sum_{i=1}^m (\omega^\top x^i - \omega^\top \bar{x})^2 \\ &= \frac{1}{m} \sum_{i=1}^m \omega^\top (x^i - \bar{x})(x^i - \bar{x})^\top \omega \\ &= \omega^\top \left( \frac{1}{m} \sum_{i=1}^m (x^i - \bar{x})(x^i - \bar{x})^\top \right) \omega. \end{aligned}$$

Así, el problema de encontrar el vector de coeficientes  $\omega = (\omega_1, \dots, \omega_n)$  que corresponda a la máxima varianza de los datos  $y^1, \dots, y^m$  tiene el formato:

$$\text{máx} \{ \langle \omega, S \omega \rangle ; \|\omega\| = 1 \}. \tag{2.20}$$

Esta restricción se aplica porque  $\|\omega\|$  sería infinitamente grande sin ella y así la varianza divergiría. En virtud del teorema 2.66, sea

$$\{u^1, \dots, u^n\}$$

una base ortonormal de  $\mathbb{R}^n$  formada por autovectores de  $S$  y sean

$$\lambda_1 \geq \dots \geq \lambda_n$$

sus respectivos autovalores. Luego, gracias al corolario 2.97,  $u^1$  es una solución óptima del problema (2.20) y  $\lambda_1$  es el valor óptimo del mismo.

A continuación, se plantea encontrar el eje de proyección que maximiza la varianza total bajo la restricción de la ortogonalidad al primer componente principal. Este eje se conoce como el segundo componente principal. En términos precisos, siguiendo el mismo razonamiento de arriba, el nuevo problema planteado es del siguiente formato:

$$\text{máx } \{ \langle \omega, S\omega \rangle ; \langle \omega, u^1 \rangle, \|\omega\| = 1 \}. \quad (2.21)$$

De nuevo, por el corolario 2.97,  $u^2$  es una solución óptima del problema (2.21) y  $\lambda_2$  es el valor óptimo del mismo.

Continuando de la misma manera, el tercer componente principal se define como el eje que maximiza la varianza de los  $m$   $n$ -dimensionales datos proyectados bajo el requisito de ortogonalidad al primer y segundo componente principal. Y, de nuevo, por el corolario 2.97,  $u^3$  es dicho eje (salvo el sentido de la dirección) y  $\lambda_3$  es la máxima varianza de los datos proyectados. Por sucesivas repeticiones de este proceso, podemos obtener las  $n$  componentes principales de nuestra base de datos.

Sea  $U$  la matriz cuadrada de tamaño  $n$  cuya  $i$ -ésima columna está dada por  $u^i$  y sea  $\Lambda$  la matriz diagonal de tamaño  $n$  cuyo  $i$ -ésimo elemento de la diagonal principal es  $\lambda_i$ . Sintetizando lo obtenido hasta ahora tenemos

$$SU = U\Lambda \quad \text{y} \quad U^T U = I,$$

de donde se siguen las siguientes ecuaciones:

$$U^T S U = \Lambda. \quad (2.22)$$

$$S = U\Lambda U^T = \lambda_1 u^1 (u^1)^T + \dots + \lambda_n u^n (u^n)^T. \quad (2.23)$$

$$\text{tr}(S) = \lambda_1 + \dots + \lambda_n = \text{tr}(\Lambda). \quad (2.24)$$

Así, la ecuación (2.22) muestra que la matriz simétrica  $S$ , dada por la ecuación (2.19), puede ser diagonalizada por la matriz ortogonal  $U$ ; la ecuación (2.23) es conocida

como la descomposición espectral de la matriz de covarianza  $S$ ; y la ecuación (2.24) nos dice que la suma de las varianzas de las variables originales es igual a la suma de las varianzas de las componentes principales. Como resultado, el ACP se convierte así en el problema de encontrar autovectores de la matriz de covarianza de nuestra base de datos.

## 2.5. Análisis numérico

El análisis numérico es el estudio de algoritmos que utilizan la aproximación numérica (a diferencia de las manipulaciones simbólicas generales) para los problemas del análisis matemático. Los complejos problemas matemáticos que enfrentan los científicos e ingenieros no siempre pueden ser resueltos por métodos analíticos, por lo que los métodos numéricos son a menudo necesarios.

Uno de los problemas más importantes de esta rama de las matemáticas es el diseño de algoritmos eficientes para encontrar los autovalores y autovectores de una matriz, ya que, estos dos conceptos juegan un papel importante en las aplicaciones del álgebra lineal y la tarea de hallar las raíces de un polinomio no es viable cuando el grado de este es grande.

### 2.5.1. Método de las potencias

El método de las potencias es un algoritmo que calcula el autovector asociado al mayor (en norma) autovalor de una matriz. Aunque este método fue introducido por Von Mises en [von Mises and Pollaczek-Geiringer, 1929] hace casi 90 años, en la actualidad sigue siendo muy útil para ciertos problemas computacionales. Por ejemplo, Google lo utiliza para calcular el PageRank<sup>1</sup> de los documentos en su motor de búsqueda, ver [Fernández, 2007], y Twitter lo utiliza para mostrar a los usuarios las recomendaciones de a quién seguir, ver [Gupta et al., 2013].

A continuación, describiremos el método de las potencias. Sea  $A$  una matriz cuadrada invertible de tamaño  $n$ . En primer lugar, consideramos la función *cociente de Rayleigh* que se define para cada  $x \neq 0$  como sigue:

$$r(x) = \frac{\langle x, Ax \rangle}{\langle x, x \rangle}.$$

---

<sup>1</sup>PageRank es una marca registrada y patentada por Google el 9 de enero de 1999 que ampara una familia de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos (o páginas web) indexados por un motor de búsqueda.

Si  $x$  es un autovector, entonces  $Ax = r(x)x$ ; es decir,  $r(x)$  es su autovalor asociado. Luego, supongamos que  $\{v_1, \dots, v_n\}$  es una base de  $\mathbb{R}^n$  formada por autovectores de  $A$  y  $\lambda_1, \dots, \lambda_n$  sus respectivos autovalores asociados tales que

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|;$$

o sea,  $A$  es una matriz con *autovalor dominante*.

Ahora, tomamos  $x_0 \in \mathbb{R}^n$  unitario tal que

$$x_0 = \alpha_1 v_1 + \dots + \alpha_n v_n,$$

para ciertos coeficientes  $\alpha_1, \dots, \alpha_n$  con  $\alpha_1 \neq 0$ . Luego, por un argumento inductivo,

$$A^k x_0 = \lambda_1^k \left( \alpha_1 v_1 + \alpha_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k v_2 + \dots + \alpha_n \left( \frac{\lambda_n}{\lambda_1} \right)^k v_n \right) \quad \forall k = 0, 1, \dots \quad (2.25)$$

Así,

$$\lim_{k \rightarrow \infty} \frac{A^k x_0}{\|A^k x_0\|} = v_1 \quad (2.26)$$

porque

$$\lim_{k \rightarrow \infty} \left( \frac{\lambda_i}{\lambda_1} \right)^k = 0 \quad \forall i = 2, 3, \dots$$

Resumiendo: si  $A$  es una matriz cuadrada de tamaño  $n$  tal que

- (i)  $A$  es invertible,
- (ii)  $A$  tiene autovalor dominante  $\lambda_1$  y
- (iii) el vector unitario inicial  $x_0$  tiene coeficiente no nulo en la dirección de un autovector  $v_1$  asociado a  $\lambda_1$ ;

entonces, se tiene (2.26) donde

- (i) la sucesión  $(Ax_k)$  está dada por (2.25) y
- (ii)  $\{v_1, \dots, v_n\}$  es una base de  $\mathbb{R}^n$  formada por autovectores de  $A$  y  $\lambda_1, \dots, \lambda_n$  son sus respectivos autovalores.

Se verifica que para

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

(matriz invertible y simétrica), la sucesión generada por el método de las potencias diverge para cualquier

$$x_0 \neq \frac{1}{2}(\sqrt{2}, \sqrt{2})$$

porque  $A$  no tiene autovalor dominante.

## 2.5.2. Un esquema discreto

Nosotros necesitamos encontrar los autovalores y autovectores de una matriz simétrica, a saber una matriz de covarianza de una base de datos, para llevar a cabo el análisis de componentes principales de dicha base de datos. En el corolario 2.97 se obtuvo que para encontrar cada autovector se necesita resolver un problema de optimización que se sabe teóricamente que dicho autovector es su solución, pero no se sabe cómo hallarlo.

Para ello, aplicaremos ideas del método del gradiente proyectado para obtener un método iterativo que construya una matriz definida positiva en cada iteración y aplique el método de las potencias para encontrar dicho autovector que es solución del problema de optimización planteado. El poder de este esquema discreto es que funciona para cualquier matriz simétrica.

Con el fin de describir este esquema, enunciamos algunos resultados teóricos que serán sobre los que se fundamenta dicho esquema.

**Observación 2.99.** Si  $S$  es una matriz simétrica no nula y  $\alpha^{-1} = \sum_{i,j} |s_{ij}| + 1$ , entonces

$$1 - \alpha|\lambda| > 0 \quad \forall \lambda \in \text{spec}(S). \quad (2.27)$$

**Observación 2.100.** Sea  $S$  una matriz simétrica y sea  $\alpha > 0$  tal que satisface la proposición (2.27). Entonces:

- (i)  $P = I + \alpha S$  es una matriz definida positiva.
- (ii)  $\text{spec}(S) = \alpha^{-1} \text{spec}(P) - 1$ .
- (iii)  $P$  y  $S$  tienen los mismos autovectores.

**Observación 2.101.** Sea  $S$  una matriz simétrica, sea  $\alpha > 0$  tal que satisface la proposición (2.27), sea  $P = I + \alpha S$  y sea  $F : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ ,

$$F(x) = \frac{Px}{\|Px\|}. \quad (2.28)$$

Entonces, son equivalentes:

- (i)  $a$  es un punto fijo de  $F$ .
- (ii)  $a$  es un autovector unitario de  $P$ .
- (iii)  $a$  es un autovector unitario de  $S$ .

**Observación 2.102.** Sea  $S$  una matriz simétrica, sea  $\alpha^{-1} = \sum_{i,j} |s_{ij}| + 1$  y sea  $x_0 \in \mathbb{R}^n$  unitario. Entonces, la sucesión  $(x_k)$  definida por

$$x_{k+1} = \frac{(I + \alpha S)x_k}{\|(I + \alpha S)x_k\|} \quad \forall k = 0, 1, \dots \quad (2.29)$$

converge a un autovector unitario de  $S$ .

A continuación, pasamos a describir el esquema. Sea  $\mathcal{B}$  la base canónica de  $\mathbb{R}^n$  y  $S$  una matriz simétrica de tamaño  $n$ . Hallaremos una base ortonormal  $\{u^1, \dots, u^n\}$  de  $\mathbb{R}^n$  formada por autovectores de  $S$  y sus respectivos autovalores a través de  $n$  iteraciones:

(i) Antes de la primera iteración hacemos

$$U_1 = I, \quad S_1 = S, \quad V_1 = \mathbb{R}^n,$$

donde  $I$  es la matriz identidad de tamaño  $n$ .

(ii) En la  $i$ -ésima iteración encontraremos el autovector  $u^i$  y su respectivo autovalor:

- a) Caso  $S_i$  sea no nulo: encontramos un autovector unitario  $u^i$  de  $S_i$  en  $V_i$  siguiendo el método iterativo de la observación 2.102.
- b) Caso  $S_i$  sea nulo: encontramos un  $u^i \in \mathcal{B}$  tal que

$$\|U_i u^i\| \geq \|U_i v\| \quad \forall v \in \mathcal{B}.$$

Luego, hacemos

$$U_{i+1} = (I - u^i (u^i)^\top) U_i, \quad S_{i+1} = U_{i+1} S_i U_{i+1}, \quad V_{i+1} = (\text{span}(u^1, \dots, u^i))^\top.$$

# Capítulo 3

## Metodología

Utilizaremos, en lo sucesivo, la expresión *nube de puntos* para referirnos a un conjunto finito no vacío de puntos de  $\mathbb{R}^n$ . Y diremos que la nube de puntos  $y^1, \dots, y^m$  es proyección de la nube de puntos  $x^1, \dots, x^m$  sobre el hiperplano  $\mathcal{H} \subset \mathbb{R}^n$  si  $y^i = \text{proj}_{\mathcal{H}}(x^i)$  para cada  $i = 1, \dots, m$ .

Dado que la metodología del [Análisis de Componentes Principales \(ACP\)](#) consiste en buscar ejes sobre los cuales la varianza de las proyecciones de una nube de puntos es “máxima”, surge una pregunta natural: ¿qué relación hay entre dichos ejes y la recta o el plano que “mejor se aproxima” en algún sentido a dicha nube de puntos?

En este capítulo describiremos la metodología que utilizaremos para aplicar la técnica del [ACP](#).

### 3.1. Planteamiento del problema

En primer lugar, estamos interesados en:

“dada una nube de puntos, encontrar una recta que mejor se aproxime a dicha nube de puntos.”

A continuación, plantearemos este problema en términos más precisos. En primer lugar, dada una nube de puntos  $B = \{x^1, \dots, x^m\}$  y una recta  $\mathcal{L} = \mathcal{L}(z, u)$  de  $\mathbb{R}^n$  definimos la *proximidad* de  $\mathcal{L}$  a  $B$  como

$$\delta(B, \mathcal{L}) = \sum_{i=1}^m d(x^i, \mathcal{L})^2.$$

Esta definición es clave, ya que, utilizar  $d(x^k, \mathcal{L})^2$  en esta definición, nos llevará a que el problema planteado sea un problema de programación diferenciable y, de esta manera, podremos usar las herramientas disponibles para este tipo de problemas.

En segundo lugar, dada una nube de puntos  $B$ , la idea de recta que mejor se aproxime a  $B$ , es la de una recta cuya proximidad a  $B$  sea mínima. De esta manera, ya que, una recta queda caracterizada por un punto y un vector unitario, este problema replanteado en un formato matemático es como sigue:

“dada una nube de puntos  $B$ , hallar un punto  $\bar{z} \in \mathbb{R}^n$  y un vector unitario  $\bar{u} \in \mathbb{R}^n$  tales que

$$\delta(B, \mathcal{L}(\bar{z}, \bar{u})) \leq \delta(B, \mathcal{L}(z, u)),$$

para todo punto  $z \in \mathbb{R}^n$  y todo vector unitario  $u \in \mathbb{R}^n$ .”

Habiendo replanteado este problema en términos más precisos, mostraremos que este problema es un problema de programación diferenciable. En efecto, sean  $\varphi, h : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  definidas para cada  $z, u \in \mathbb{R}^n$  por

$$\varphi(z, u) = \sum_{i=1}^m \|x^i - z - \langle x^i - z, u \rangle u\|^2 \quad \text{y} \quad h(z, u) = \|u\|^2 - 1. \quad (3.1)$$

Así,  $\varphi$  y  $h$  son funciones diferenciables por ser funciones polinomiales. Luego, en virtud de la ecuación 2.1, este problema consiste en resolver el siguiente problema de optimización:

$$\begin{aligned} & \text{mín} \{ \varphi(z, u) ; (z, u) \in C \} \\ & C = \{ (z, u) \in \mathbb{R}^{2n} ; h(z, u) = 0 \}. \end{aligned} \quad (3.2)$$

Finalmente, planteamos el problema a resolver:

**Problema 3.1.** Dada una nube de puntos  $x^1, \dots, x^m$  encontrar un plano  $\mathcal{P}(\bar{z}, \bar{u}, \bar{v})$  tal que

- (i)  $\mathcal{L}(\bar{z}, \bar{u})$  sea una recta que mejor se aproxime a la nube  $x^1, \dots, x^m$  y
- (ii)  $\mathcal{L}(\bar{z}, \bar{v})$  sea una recta que mejor se aproxime, en el hiperplano  $\mathcal{H} = \mathcal{H}(\bar{u}, \alpha)$  que pasa por  $\bar{z}$ , a la nube  $y^1, \dots, y^m$  que surge de proyectar la nube  $x^1, \dots, x^m$  sobre  $\mathcal{H}$ .

## 3.2. Solución del problema

En esta sección encontraremos una solución para el problema 3.1.

**Lema 3.2.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$ , sea  $u \in \mathbb{R}^n$  unitario y sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$f(z) = \sum_{k=1}^m \|x^k - z - \langle x^k - z, u \rangle u\|^2.$$

Entonces,

$$\bar{z} := \frac{1}{m} \sum_{i=1}^m x^i$$

es una solución óptima del problema:

$$\text{mín} \{ f(z) ; z \in \mathbb{R}^n \}.$$

*Demostración.* En efecto, por la propiedad 2.63 (i), tenemos que para cada  $z \in \mathbb{R}^n$ :

$$f(z) = \sum_{i=1}^m \langle x^i - z, (I - uu^\top)(x^i - z) \rangle.$$

Luego, por la propiedad 2.61 y el teorema 2.72,

$$(i) \quad \nabla f(z) = -2(I - uu^\top)(x^1 + \dots + x^m - mz).$$

$$(ii) \quad \nabla^2 f(z) = 2m(I - uu^\top).$$

Finalmente, en virtud del teorema 2.90, la conclusión se sigue. □

**Lema 3.3.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$ , sea

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i$$

y sea  $A$  la matriz de tamaño  $m \times n$  cuya  $i$ -ésima fila es el vector  $x^i - \bar{z}$  para cada  $i = 1, \dots, m$ . Sea  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$g(x) = \sum_{i=1}^m \|x^i - \bar{z} - \langle x^i - \bar{z}, x \rangle x\|^2.$$

Entonces, el autovector unitario asociado al mayor autovalor de la matriz  $A^\top A$  es una solución óptima del siguiente problema:

$$\begin{aligned} & \text{mín} \{ g(x) ; x \in C \} \\ & C = \{ u \in \mathbb{R}^n ; \|u\| = 1 \}. \end{aligned}$$

*Demostración.* En primer lugar, por la propiedad 2.63 (ii), tenemos que para cada  $u \in C$ :

$$g(u) = c - \langle u, \Lambda u \rangle,$$

donde

$$c = \sum_{i=1}^m \langle x^i - \bar{z}, x^i - \bar{z} \rangle \quad \text{y} \quad \Lambda = \sum_{i=1}^m (x^i - \bar{z})(x^i - \bar{z})^\top.$$

Luego, por la propiedad 2.53,  $\Lambda = A^\top A$ , y por la propiedad 2.58,  $\Lambda$  es una matriz semidefinida positiva. Entonces, por el teorema 2.66, sea  $\{u^1, \dots, u^n\}$  una base ortonormal de  $\mathbb{R}^n$  formada por autovectores de  $\Lambda$  y sean

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

sus respectivos autovalores asociados. Sea  $v \in C$ . Como  $\mathcal{B}$  es una base ortonormal, existen  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$  tales que

$$u = \alpha_1 u^1 + \dots + \alpha_n u^n \quad \text{y} \quad \alpha_1^2 + \dots + \alpha_n^2 = 1.$$

Luego,

$$\begin{aligned} \langle u, \Lambda u \rangle &= \langle u, \Lambda(\alpha_1 u^1 + \dots + \alpha_n u^n) \rangle \\ &= \langle u, (\alpha_1 u^1 + \dots + \alpha_n u^n) \rangle \\ &= \langle u, (\alpha_1 \Lambda u^1 + \dots + \alpha_n \Lambda u^n) \rangle \\ &= \langle (\alpha_1 u^1 + \dots + \alpha_n u^n), (\alpha_1 \lambda_1 u^1 + \dots + \alpha_n \lambda_n u^n) \rangle \\ &= \alpha_1^2 \lambda_1 + \dots + \alpha_n^2 \lambda_n. \\ &\leq \alpha_1^2 \lambda_1 + \dots + \alpha_n^2 \lambda_1 \\ &\leq (\alpha_1^2 + \dots + \alpha_n^2) \lambda_1 \\ &\leq \lambda_1 \langle u^1, u^1 \rangle \\ &\leq \langle u^1, \lambda_1 u^1 \rangle \\ &\leq \langle u^1, \Lambda u^1 \rangle, \end{aligned}$$

y así,  $g(u^1) \leq g(u)$ . □

**Lema 3.4.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$ , sea

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i$$

y sea  $A$  la matriz de tamaño  $m \times n$  cuya  $i$ -ésima fila es el vector  $x^i - \bar{z}$  para cada  $i = 1, \dots, m$ . Sean  $\bar{u}$  y  $\bar{v}$  los autovectores de  $A^\top A$  asociados a su mayor y segundo

mayor autovalor, respectivamente. Sea  $\mathcal{H} = \mathcal{H}(\bar{u}, \alpha)$  el hiperplano que pasa por  $\bar{z}$ , sean  $y^1, \dots, y^m$  las proyecciones de  $x^1, \dots, x^m$  sobre  $\mathcal{H}$  y sea  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$g(x) = \sum_{i=1}^m \|y^i - \bar{z} - \langle y^i - \bar{z}, x \rangle x\|^2.$$

Entonces,  $\bar{v}$  es una solución optimal del problema:

$$\begin{aligned} & \text{mín } \{ g(x) ; x \in C \} \\ C &= \{ v \in \mathbb{R}^n ; \langle v, \bar{u} \rangle, \|v\| = 1 \}. \end{aligned}$$

*Demostración.* En primer lugar, para cada  $v \in C$ , por la propiedad 2.63 (ii) y el hecho de ser  $v$  ortogonal a  $\bar{u}$ ,

$$\begin{aligned} g(v) &= \sum_{i=1}^m \|(x^i - \alpha_i \bar{u}) - \bar{z} - \langle (x^i - \alpha_i \bar{u}) - \bar{z}, v \rangle v\|^2 \\ &= \sum_{i=1}^m \|x^i - \bar{z} - \alpha_i \bar{u} - \langle x^i - \bar{z} - \alpha_i \bar{u}, v \rangle v\|^2 \\ &= c - \left\langle v, \left( \sum_{i=1}^m (x^i - \bar{z} - \alpha_i \bar{u})(x^i - \bar{z} - \alpha_i \bar{u})^\top \right) v \right\rangle \\ &= c - \left\langle v, \left( \sum_{i=1}^m (x^i - \bar{z})(x^i - \bar{z})^\top \right) v \right\rangle - \left\langle v, \left( \sum_{i=1}^m \alpha_i (x^i - \bar{z})(\bar{u})^\top \right) v \right\rangle \\ &\quad - \left\langle v, \left( \sum_{i=1}^m \alpha_i \bar{u} (x^i - \bar{z})^\top \right) v \right\rangle + \left\langle v, \left( \sum_{i=1}^m \alpha_i^2 \bar{u} (\bar{u})^\top \right) v \right\rangle \\ &= c - \langle v, \Lambda v \rangle, \end{aligned}$$

donde

$$c = \sum_{i=1}^m \|x^i - \bar{z} - \alpha_i \bar{u}\|^2 \quad \text{y} \quad \Lambda = \sum_{i=1}^m (x^i - \bar{z})(x^i - \bar{z})^\top.$$

Finalmente, por el mismo razonamiento que se siguió en la demostración del lema 3.3, se concluye que  $g(\bar{v}) \leq g(v)$ .  $\square$

**Teorema 3.5.** Sean  $x^1, \dots, x^m \in \mathbb{R}^n$ , sea

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i$$

y sea  $A$  la matriz de tamaño  $m \times n$  cuya  $i$ -ésima fila es el vector  $x^i - \bar{z}$  para cada  $i = 1, \dots, m$ . Sean  $\bar{u}$  y  $\bar{v}$  los autovectores de  $A^\top A$  asociados a su mayor y segundo

mayor autovalor, respectivamente. Sea  $\mathcal{H} = \mathcal{H}(\bar{u}, \alpha)$  el hiperplano que pasa por  $\bar{z}$  y sean  $y^1, \dots, y^m$  las proyecciones de  $x^1, \dots, x^m$  sobre  $\mathcal{H}$ . Entonces,  $\mathcal{P}(\bar{z}, \bar{u}, \bar{v})$  es un plano tal que

- (i)  $\mathcal{L}(\bar{z}, \bar{u})$  es una recta que mejor se aproxima a la nube  $x^1, \dots, x^m$  y
- (ii)  $\mathcal{L}(\bar{z}, \bar{v})$  es una recta que mejor se aproxima, en el hiperplano  $\mathcal{H}$ , a la nube  $y^1, \dots, y^m$ .

*Demostración.*

- (i) Sean  $\varphi, h : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  definidas por las correspondencias (3.1) y sea  $C \subset \mathbb{R}^n$  el conjunto del problema (3.2). Para cada vector  $u \in \mathbb{R}^n$  unitario definimos  $f_u : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(z) = \varphi(z, u)$ . Luego, por el lema 3.2, para todo  $(z, u) \in C$ :

$$\varphi(\bar{z}, u) \leq \varphi(z, u). \quad (3.3)$$

Ahora, definamos  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g(x) = \varphi(\bar{z}, x)$ . Finalmente, por el lema 3.3 y la desigualdad (3.3), para todo  $(z, u) \in C$ :

$$\varphi(\bar{z}, \bar{u}) \leq \varphi(\bar{z}, u) \leq \varphi(z, u).$$

- (ii) Denotemos por  $\mathcal{B}$  a la nube  $y^1, \dots, y^m$ . Toda recta  $\mathcal{L}(z, v)$  en  $\mathcal{H}$  cumple que  $v$  es ortogonal a  $\bar{u}$  y la media de  $y^1, \dots, y^m$  coincide con la media de  $x^1, \dots, x^m$ . Luego, sea  $v \in \mathbb{R}^n$  unitario y ortogonal a  $\bar{u}$ . Entonces, para todo  $z \in \mathbb{R}^n$ , por el lema 3.2, tenemos:

$$\sum_{i=1}^m \|y^i - \bar{z} - \langle y^i - \bar{z}, v \rangle v\|^2 \leq \delta(\mathcal{B}, \mathcal{L}(z, v)). \quad (3.4)$$

Finalmente, por el lema 3.4 y la desigualdad (3.4), tenemos:

$$\delta(\mathcal{B}, \mathcal{L}(\bar{z}, \bar{v})) \leq \sum_{i=1}^m \|y^i - \bar{z} - \langle y^i - \bar{z}, v \rangle v\|^2 \leq \delta(\mathcal{B}, \mathcal{L}(z, v)).$$

□

**Definición 3.6** (Eje principal y plano principal). Con la notación del teorema 3.5:

- (i) La recta  $\mathcal{L}(\bar{z}, \bar{u})$  es llamada *eje principal* de la nube de puntos  $x^1, \dots, x^m$ .
- (ii) El plano  $\mathcal{P}(\bar{z}, \bar{u}, \bar{v})$  es llamado *plano principal* de la nube de puntos  $x^1, \dots, x^m$ .

**Observación 3.7.** Con la notación del teorema 3.5, el plano principal de la nube de puntos  $x^1, \dots, x^m$  es un plano que resuelve el problema 3.1.

### 3.3. Conexión con la técnica del ACP

En esta sección, estableceremos la conexión entre la técnica estadística del análisis de componentes principales y el concepto de plano principal introducido en la sección anterior:

**Teorema 3.8.** Con la notación del teorema 3.5, encontrar el plano principal  $\mathcal{P}(\bar{z}, \bar{u}, \bar{v})$  de la nube de puntos  $x^1, \dots, x^m$  es precisamente aplicar el ACP a dicha nube y establecer:

- (i)  $\bar{u}$  como la primera componente principal y
- (ii)  $\bar{v}$  como la segunda componente principal.

*Demostración.* Sea  $A$  la matriz definida en el teorema 3.5, sea  $\Lambda = A^\top A$  y sea  $S$  la matriz de covarianza de  $A$  definida en la ecuación (2.19). Sean  $u \in \mathbb{R}^n$  y la  $\lambda \in \mathbb{R}$ . La conclusión se sigue de verificar que  $u$  es autovector de  $S$  con autovalor asociado  $\lambda$  si y solamente si  $u$  es autovector de  $\Lambda$  con autovalor asociado  $m\lambda$ .  $\square$

**Observación 3.9.** En virtud del teorema 3.8, podemos afirmar que hemos arribado a la técnica estadística del ACP conducidos por el motor de encontrar una recta que mejor se aproxima a cierta nube de puntos.

Finalmente, como aplicación del teorema 3.8, dada una nube de puntos, estamos interesados en representar esta nube en el plano principal  $\mathcal{P} = \mathcal{P}(\bar{z}, \bar{u}, \bar{v})$  de dicha nube de puntos. Como para cada  $x \in \mathbb{R}^n$ :

$$\text{proj}_{\mathcal{P}}(x) = \bar{z} + \langle x - \bar{z}, \bar{u} \rangle \bar{u} + \langle x - \bar{z}, \bar{v} \rangle \bar{v}$$

y las coordenadas de  $x$  en el plano  $\mathcal{P}$  están dadas por

$$(\langle x - \bar{z}, \bar{u} \rangle, \langle x - \bar{z}, \bar{v} \rangle),$$

esta tarea puede ser sintetizada en el siguiente algoritmo:

---

**Algoritmo 1** Representación de una nube de puntos en su plano principal

---

**Entrada:** una nube de puntos  $x^1, \dots, x^m \in \mathbb{R}^n$  con  $n > 2$

**Salida:**  $w^1, \dots, w^m$ , representación de  $x^1, \dots, x^m$  en el plano principal de esta

$$\bar{z} \leftarrow \frac{1}{m} \sum_{i=1}^m x^i$$
$$A \leftarrow \begin{bmatrix} x^1 - \bar{z} \\ x^2 - \bar{z} \\ \vdots \\ x^m - \bar{z} \end{bmatrix}$$

$$\Lambda \leftarrow A^T A$$

$\bar{u}$   $\leftarrow$  autovector unitario de  $\Lambda$  asociado a su mayor autovalor

$\bar{v}$   $\leftarrow$  autovector unitario de  $\Lambda$  asociado a su segundo mayor autovalor

**para**  $i = 1$  hasta  $m$  **hacer**

$$w^i \leftarrow (\langle x^i - \bar{z}, \bar{u} \rangle, \langle x^i - \bar{z}, \bar{v} \rangle)$$

**fin para**

---

## Capítulo 4

### Base de datos de la NOAA

La [National Oceanic and Atmospheric Administration \(NOAA\)](#) es una agencia científica del Departamento de Comercio de los Estados Unidos cuyas actividades se centran en las condiciones de los océanos y la atmósfera. Esta agencia opera una red de 88 boyas, distribuidas de manera casi regular, sobre un arreglo de de  $8 \times 11$  nodos a lo largo y ancho del Océano Pacífico ecuatorial (ver figura 4.1).

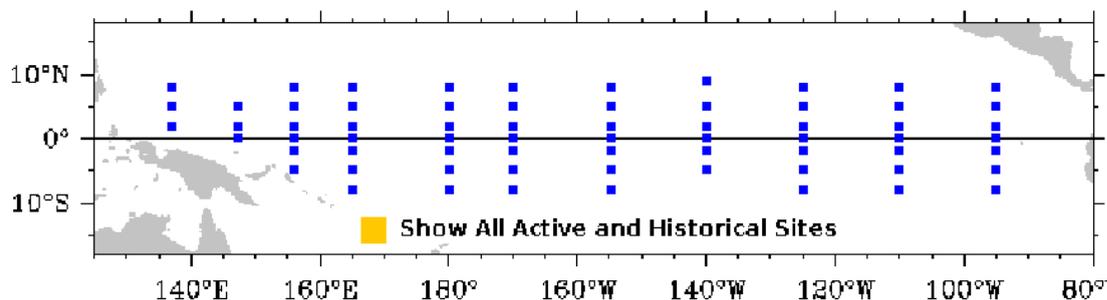


Figura 4.1: Posición geográfica de la red de boyas de la NOAA.

Estas boyas miden parámetros oceanográficos y ambientales como: precipitación, temperatura del aire, presión a nivel del mar, temperatura superficial del mar, salinidad y profundidad de la isoterma de 20 °C. Según la información de la propia agencia, la profundidad de la isoterma de 20 °C es calculada a partir de perfiles de temperatura mediante la interpolación lineal de profundidad marina versus temperatura.

NOAA transmite diariamente los datos recogidos, cuya calidad ha sido probada y ajustada cuando fue necesaria, poniendolos a disposición de investigadores de todo el mundo. Nosotros analizaremos los datos diarios de temperatura superficial del mar y profundidad de la isoterma de 20 °C que descargamos de la página web de la NOAA ([http://www.pmel.noaa.gov/tao/data\\_deliv/deliv.html](http://www.pmel.noaa.gov/tao/data_deliv/deliv.html).)

## 4.1. Lectura de los datos

Descargamos todos los datos disponibles, de estos dos parámetros, entre el 7 de marzo de 1980 y el 15 de junio del 2011. Como análisis preliminar, se construyó una matriz por cada parámetro en la cual se brinda el número de datos que se han recuperado por cada año, para cada boya. En las tablas [4.1](#) y [4.2](#) apenas mostramos la información del número de datos encontrados por año, sólo para 7 boyas. Cabe señalar que, por ejemplo, 2S170W etiqueta a la boya que tiene latitud 2 grados sur, y longitud 170 grados oeste.

Año	0N095W	2S137E	2S147E	2S156E	2S165E	2S180E	2S170W
*1980	0	0	0	0	0	0	0
1981	180	0	0	0	0	0	0
1982	365	0	0	0	0	0	0
1983	108	0	0	0	0	0	0
*1984	0	0	0	0	0	0	0
1985	0	0	0	0	25	0	0
1986	0	0	0	0	176	0	0
1987	0	0	0	0	209	0	0
*1988	0	0	0	0	296	0	0
1989	0	0	0	0	269	0	0
1990	0	0	0	0	143	0	0
1991	0	0	0	125	275	0	0
*1992	47	0	0	273	362	0	296
1993	363	0	0	335	361	280	345
1994	209	0	0	288	363	364	333
1995	317	0	0	364	363	363	260
*1996	364	0	0	361	364	363	364
1997	239	0	0	364	313	364	362
1998	232	0	0	362	167	364	363
1999	284	0	0	365	335	364	363
*2000	362	0	0	366	366	366	363
2001	224	0	0	365	365	365	365
2002	364	0	0	365	329	365	365
2003	365	0	0	365	364	365	364
*2004	163	0	0	366	293	366	366
2005	365	0	0	250	365	365	365
2006	364	0	0	312	365	365	365
2007	189	0	0	365	365	365	208
*2008	266	0	0	366	361	366	311
2009	61	0	0	365	337	248	364
2010	193	0	0	365	364	349	362
2011	164	0	0	166	166	164	166

*Tabla 4.1: Número de datos de temperatura superficial del mar recuperados en el periodo del 7 de marzo de 1980 al 15 de junio del 2011. Los años marcados con (\*) indican que son bisiestos.*

Año	2S110W	2S095W	5S137E	5S147E	5S156E	5S165E	5S180E
*1980	0	0	0	0	0	0	0
1981	0	0	0	0	0	0	0
1982	0	0	0	0	0	0	0
1983	0	0	0	0	0	0	0
*1984	0	0	0	0	0	0	0
1985	124	0	0	0	0	0	0
1986	190	0	0	0	0	0	0
1987	224	0	0	0	0	230	0
*1988	116	0	0	0	0	54	0
1989	117	0	0	0	0	0	0
1990	329	0	0	0	0	0	0
1991	301	0	0	0	0	144	0
*1992	0	0	0	0	0	290	0
1993	114	127	0	0	16	283	282
1994	363	177	0	0	363	326	362
1995	364	95	0	0	365	364	159
*1996	364	289	0	0	363	215	96
1997	222	344	0	0	361	331	171
1998	362	228	0	0	149	361	58
1999	349	179	0	0	365	305	363
*2000	365	337	0	0	366	278	365
2001	365	281	0	0	313	365	358
2002	365	365	0	0	365	365	365
2003	352	342	0	0	365	364	365
*2004	305	192	0	0	366	366	366
2005	62	254	0	0	365	357	364
2006	157	282	0	0	365	230	365
2007	248	130	0	0	365	364	305
*2008	263	201	0	0	366	312	131
2009	263	360	0	0	365	201	364
2010	201	258	0	0	365	309	363
2011	156	165	0	0	166	165	164

Tabla 4.2: Número de datos de profundidad de la isoterma de 20°C recuperados en el periodo del 7 de marzo de 1980 al 15 de junio del 2011. Los años marcados con (\*) indican que son bisiestos.

## 4.2. Preparación de los datos

Luego de revisar estas dos matrices, se encontraron que 20 boyas no presentan información alguna y que las 68 restantes las presenta de manera casi regular desde 1994. Por ello, decidimos limitar nuestro análisis a las 68 boyas restantes en el periodo del 1 de enero de 1994 al 15 de junio del 2011.

Por otro lado, como se advierte en las dos tablas de arriba, aún tenemos datos faltantes. Si, por ejemplo, en la boya 2S110W no tenemos dato para el día 15 de junio de 1999; entonces, se llena con el promedio aritmético de los datos disponibles de todos los 15 de junio entre 1994 y 2011 de esta boya. De este modo, llenamos los datos faltantes y remediamos este inconveniente.

Así, tenemos una serie temporal de temperaturas superficiales del mar y otra de profundidades de la isoterma de 20°C con un total de 6375 observaciones diarias, listas para ser analizadas. En las tablas 4.3 y 4.4 apenas se muestra la información de dichas series para 6 boyas en un periodo de quince días.

Año	Mes	Día	9N140W	8N137E	8N156E	8N165E	8N180E	8N170W
2011	6	1	26.56	29.47	29.32	28.85	28.28	27.75
2011	6	2	26.59	29.85	29.25	28.87	28.22	27.74
2011	6	3	26.53	29.49	29.22	28.88	28.20	27.78
2011	6	4	26.52	29.33	29.26	28.83	28.14	27.77
2011	6	5	26.58	29.38	29.31	28.86	28.18	27.74
2011	6	6	26.63	29.40	29.28	28.87	28.22	27.74
2011	6	7	26.56	29.54	29.26	28.86	28.18	27.78
2011	6	8	26.58	29.51	29.20	28.87	28.20	27.71
2011	6	9	26.65	29.49	29.19	28.89	28.19	27.72
2011	6	10	26.63	29.46	29.28	28.82	28.15	27.76
2011	6	11	26.62	29.69	29.20	28.81	28.15	27.77
2011	6	12	26.68	29.55	29.23	28.79	28.17	27.74
2011	6	13	26.60	29.56	29.29	28.81	28.18	27.75
2011	6	14	26.54	29.51	29.26	28.89	28.17	27.77
2011	6	15	26.55	29.41	29.17	28.81	28.17	27.73

Tabla 4.3: Arreglo de los datos de temperatura superficial del mar (°C).

Año	Mes	Día	0N137E	0N147E	0N156E	0N165E	0N180E	0N170W
1994	1	1	168.525	170.079	132.489	164.965	167.610	161.382
1994	1	2	173.471	168.741	133.221	165.567	168.572	163.693
1994	1	3	171.779	167.842	134.756	166.438	167.863	165.302
1994	1	4	167.712	165.359	134.294	167.713	168.576	164.837
1994	1	5	173.550	169.660	133.750	164.785	169.456	166.106
1994	1	6	178.096	169.487	139.040	164.732	169.446	164.686
1994	1	7	175.292	167.743	133.942	166.188	167.205	164.514
1994	1	8	175.466	168.167	145.875	165.280	169.342	159.947
1994	1	9	173.233	166.240	146.172	162.528	168.795	160.106
1994	1	10	175.713	167.131	139.314	161.112	169.577	165.467
1994	1	11	172.646	169.210	143.836	162.205	169.238	164.884
1994	1	12	176.338	170.391	144.583	163.505	170.190	162.369
1994	1	13	181.803	168.763	132.668	168.470	170.662	160.739
1994	1	14	179.215	167.847	127.735	168.789	167.870	161.996
1994	1	15	176.452	169.122	129.767	167.100	167.893	161.206

*Tabla 4.4: Arreglo de los datos de profundidad de la isoterma de 20°C (m).*

## Capítulo 5

# Aplicación del ACP a la base de datos de la NOAA

En este capítulo, se muestran los resultados que se obtienen al someter las series temporales de temperaturas superficiales del mar y de profundidades de la isoterma de 20 °C a la técnica del análisis de componentes principales descrita en el algoritmo 1. Así, nuestras dos nubes de puntos a procesar serán las observaciones diarias de los datos de las 68 boyas seleccionadas; es decir, se trata de dos nubes de 6375 puntos de  $\mathbb{R}^{68}$ . Es importante resaltar que en todas las gráficas que siguen, el eje de las abscisas del sistema coordenado mostrado representa al eje principal.

### 5.1. Temperatura superficial del mar

En la figura 5.1 se representan las observaciones diarias de la temperatura superficial del mar sobre su plano principal. Debido a la gran cantidad de puntos no se aprecia con claridad algún patrón de comportamiento de este parámetro, aunque se destaca el pulgar de la mano que forma esta representación como un pequeño grupo de puntos que se ubican lejos del origen y cercanos al semieje positivo del eje principal.

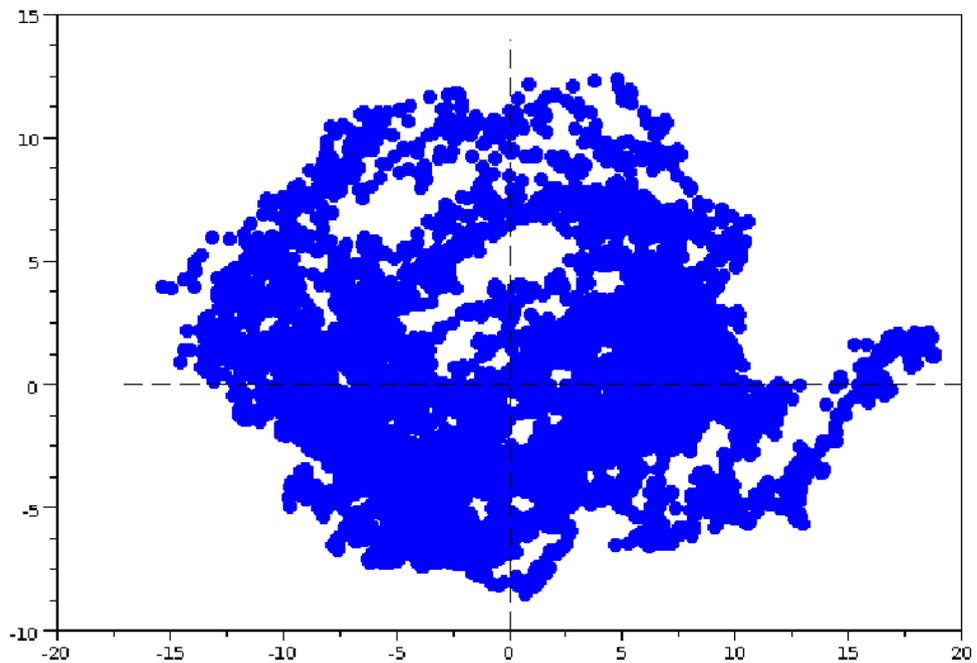
Con la finalidad de apreciar algún patrón en los meses, calculamos la media de todos los puntos de la figura 5.1 que pertenecen a un mismo mes y obtenemos así 12 puntos, donde cada uno representa un mes. Así, en la figura 5.2 se aprecia que de agosto a noviembre los puntos se encuentran muy cercanos, confinados en esta pequeña región; sin embargo, en el resto de puntos hay una gran variación de uno a otro, sobre todo de enero a febrero.

Por otro lado, calculando la media de todos los puntos que pertenecen a un mismo año obtenemos 17 puntos. De esta forma, en la figura 5.3 se observa que hay un cambio

drástico en los puntos que representan a los años 1997 y 1998, años en los que ocurrió el último fenómeno *El Niño* catalogado de muy fuerte.

Luego, calculamos la media de todos los puntos que pertenecen a un mismo trimestre de cada año, donde el primer trimestre es el periodo de enero a marzo y el último de octubre a diciembre. A pesar de que en la figura 5.4 algunos años se solapan, lo más importante es que se observan claramente 4 regiones. Sin embargo, sólo del tercer trimestre de 1997 al segundo trimestre de 1998 se aprecia un comportamiento anómalo.

Finalmente, con el fin de tener un mejor monitoreo y precisar el mes en el que comenzó la anomalía presentada en la figura 5.4, calculamos la media de todos los puntos que pertenecen a un mismo mes de cada año. Así, en la figura 5.7 se detecta que dicha anomalía comenzó en el mes de julio de 1997.



*Figura 5.1: Representación diaria de la temperatura superficial del mar.*

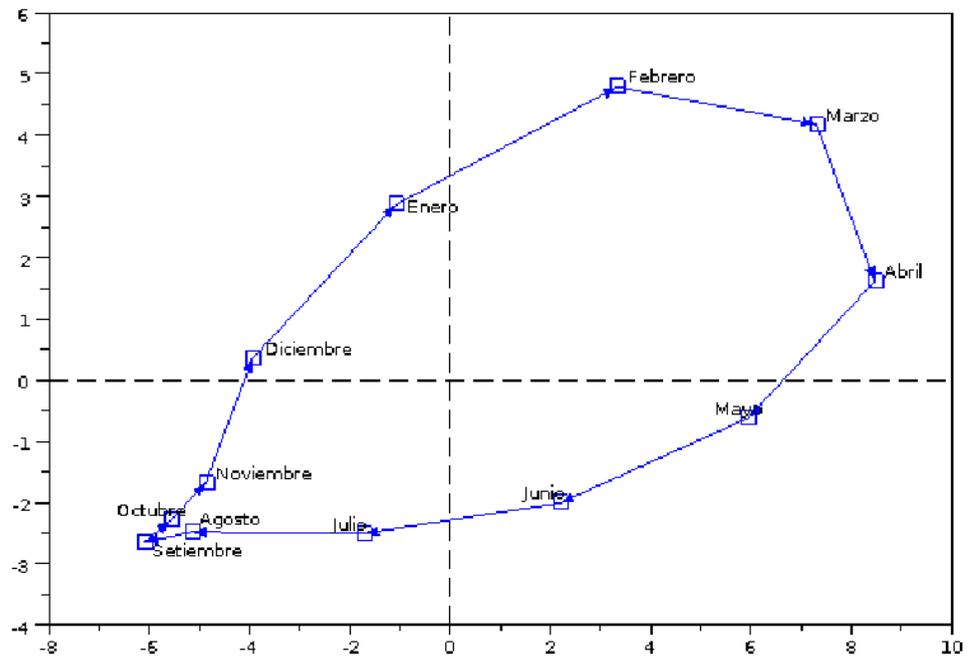


Figura 5.2: Patrón mensual de la temperatura superficial del mar.

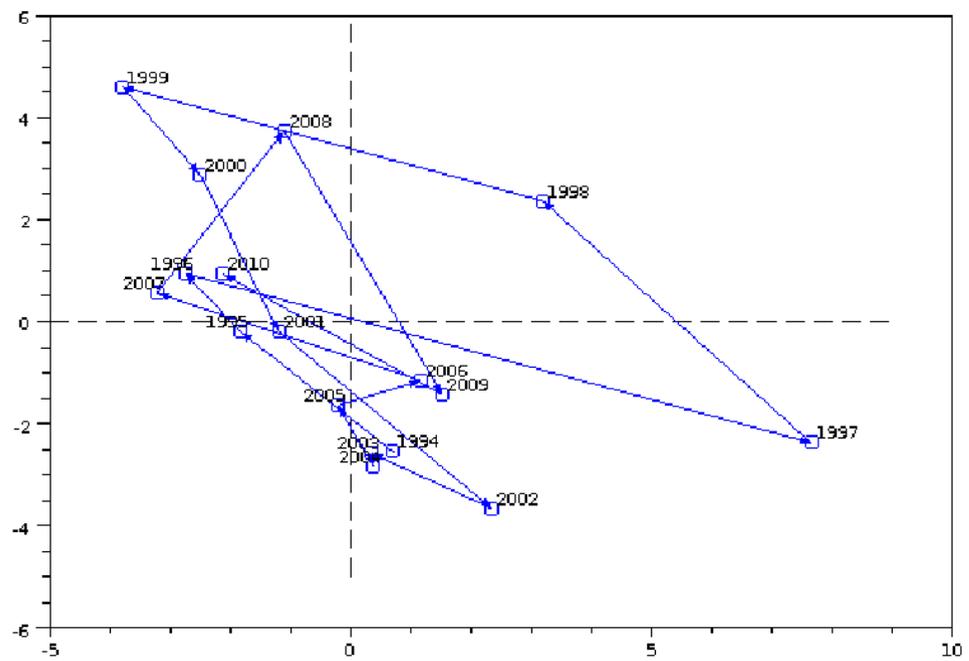


Figura 5.3: Representación anual de la temperatura superficial del mar.

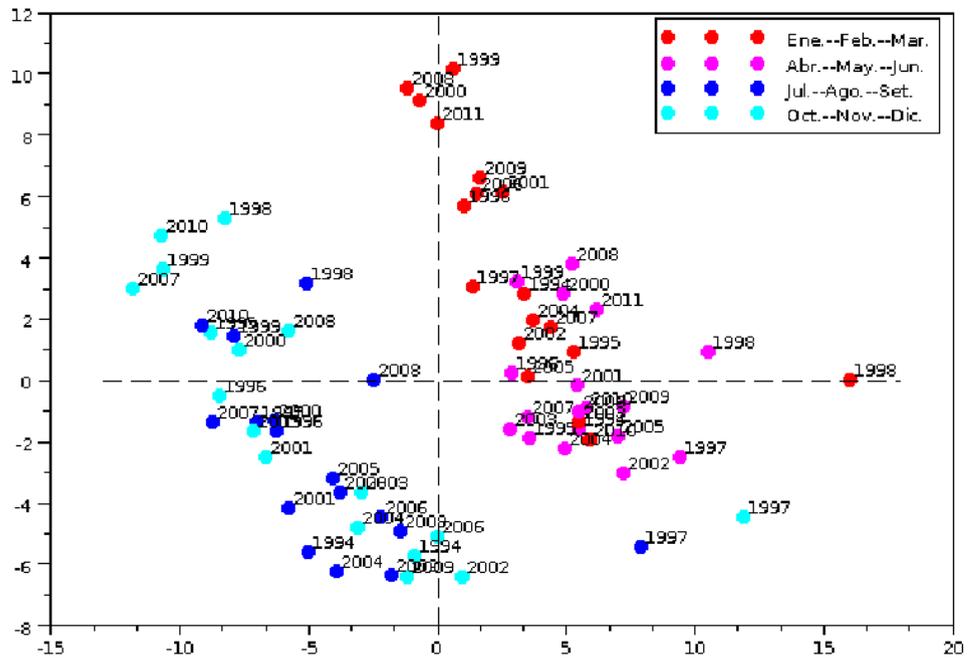


Figura 5.4: Representación trimestral de la temperatura superficial del mar.

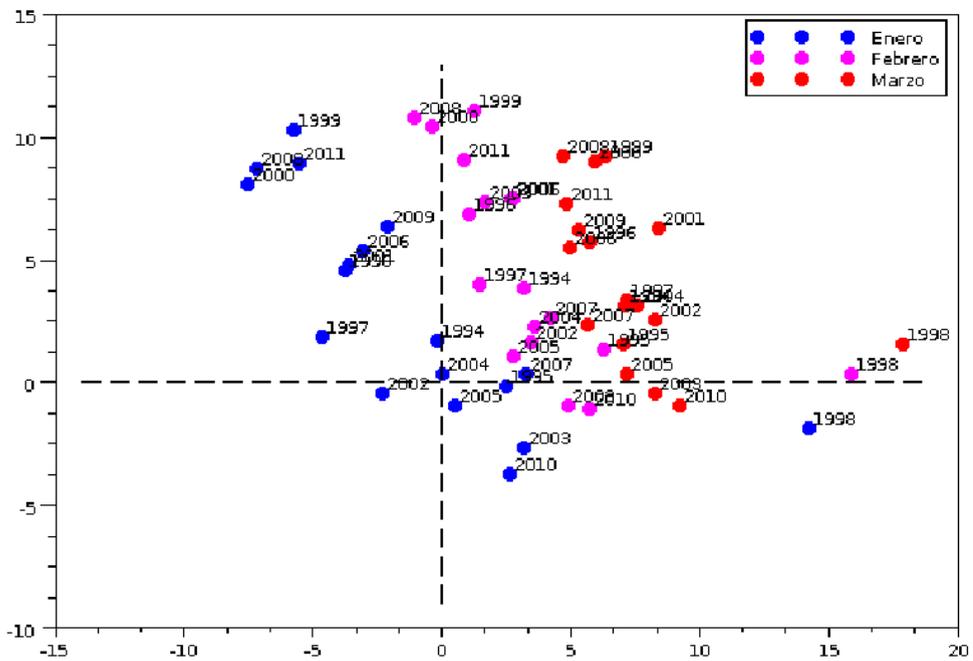


Figura 5.5: Representación mensual de la temperatura superficial del mar.

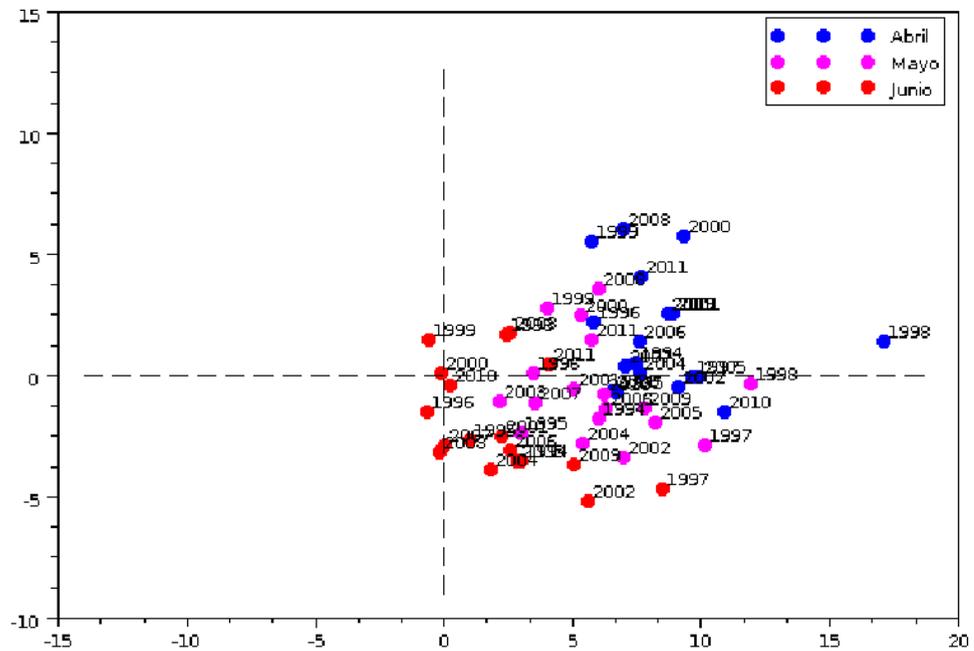


Figura 5.6: Representación mensual de la temperatura superficial del mar.

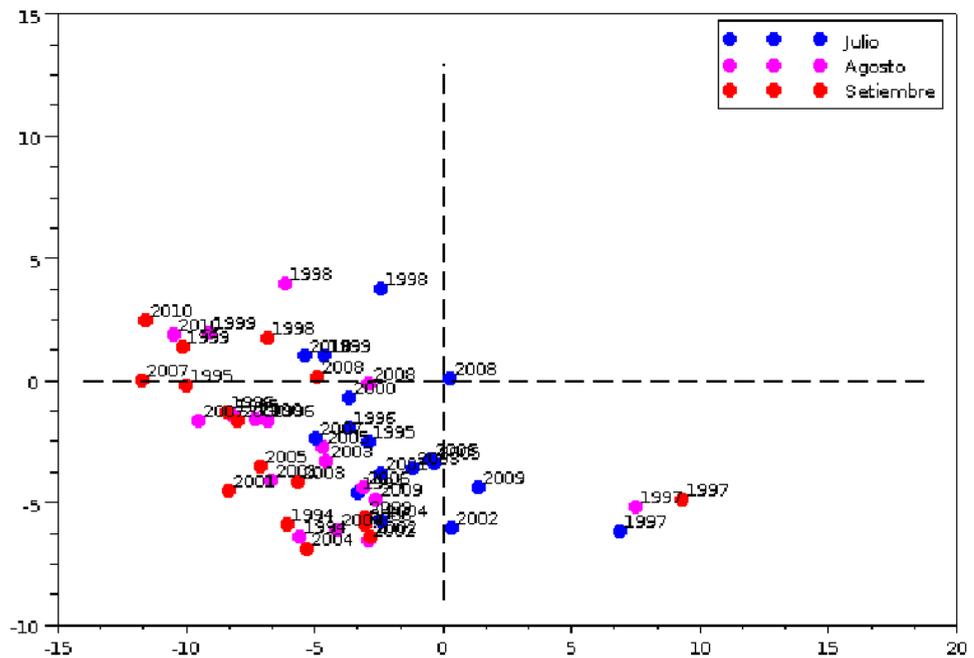


Figura 5.7: Representación mensual de la temperatura superficial del mar.

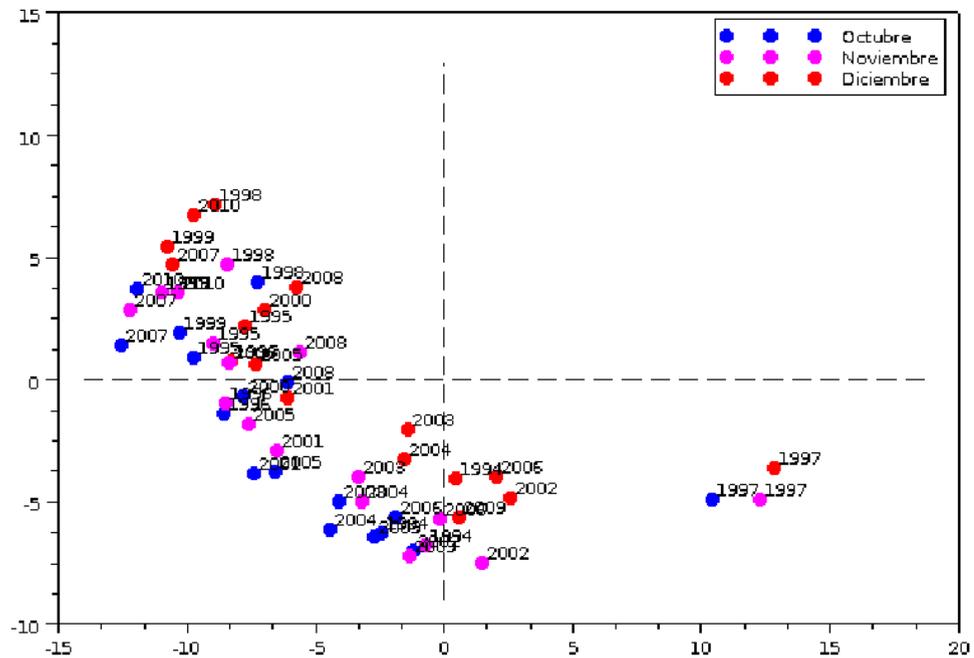


Figura 5.8: Representación mensual de la temperatura superficial del mar.

## 5.2. Profundidad de la isoterma de 20 grados Celcius

En la figura 5.9 se representan las observaciones diarias de las profundidades de la isoterma de 20°C sobre su plano principal. A diferencia de las temperaturas superficiales del mar, aquí es bastante notoria la peculiar distribución de un pequeño grupo de puntos en el cuarto cuadrante.

Con la finalidad de apreciar algún patrón en los meses, calculamos la media de todos los puntos de la figura 5.9 que pertenecen a un mismo mes y obtenemos así 12 puntos, donde cada uno representa un mes. De esta manera, en la figura 5.10 se aprecia, a diferencia de la temperatura superficial del mar, que de noviembre a enero los puntos se encuentran muy cercanos, confinados en esta pequeña región, mientras que en el resto de puntos hay una mayor variación de uno a otro.

Por otro lado, calculando la media de todos los puntos que pertenecen a un mismo año (a excepción del año 2011, por no estar completo) obtenemos 17 puntos, donde cada uno representa un año. Así, en la figura 5.11 se observa, como en la temperatura superficial del mar, que hay un cambio drástico en los puntos que representan a los años 1997 y 1998. Más aún, la variación de uno al otro se realiza en el tercer cuadrante.

Continuando con el análisis de la evolución de este parámetro, ahora calculamos la media de todos los puntos que pertenecen a un mismo trimestre de cada año. A pesar de

que en la figura 5.12 algunos años se solapan, lo más resaltante es que hay 4 regiones bien marcadas. Sin embargo, sólo del primer trimestre de 1997 al último trimestre de 1998 se aprecia un comportamiento anómalo. De nuevo, la tendencia es desplazarse, lo más alejado posible del origen, al tercer cuadrante.

Finalmente, con el fin de tener un mejor monitoreo y precisar el mes en el que comenzó la anomalía presentada en la figura 5.12, calculamos la media de todos los puntos que pertenecen a un mismo mes de cada año. Así, en la figura 5.14 se detecta que este comportamiento distinto comenzó en el mes de abril de 1997.

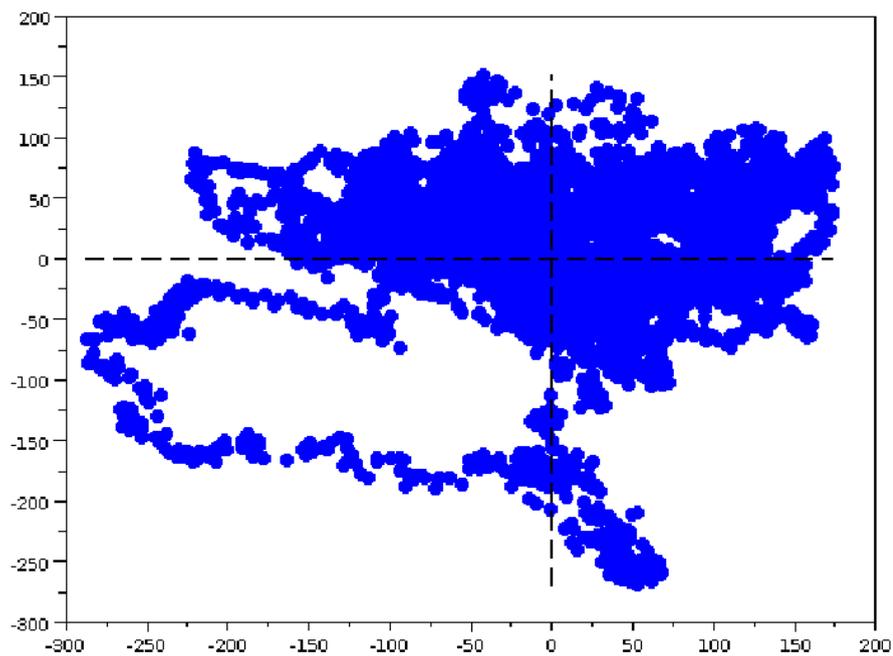


Figura 5.9: Representación diaria de la profundidad de la isoterma de 20°C.

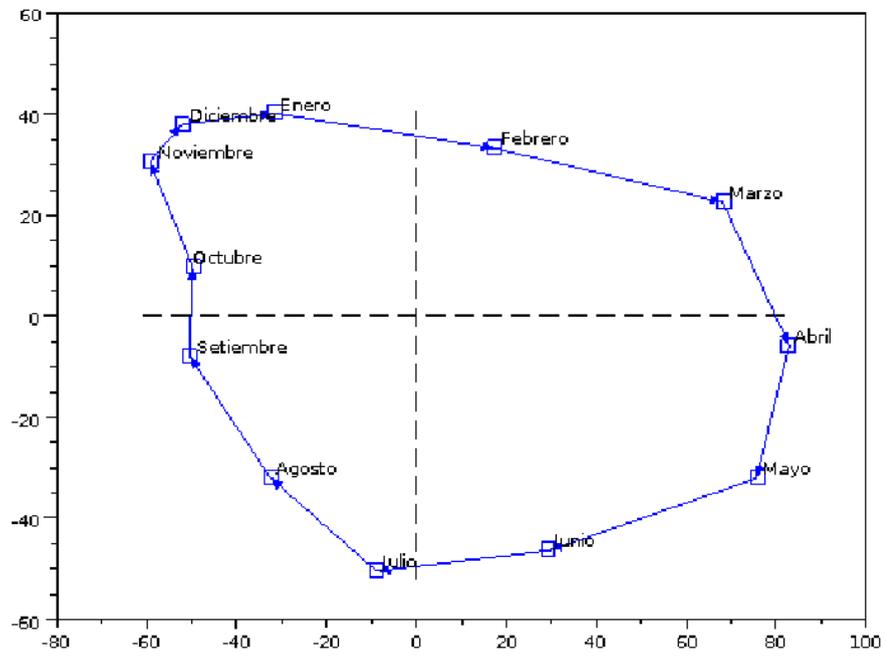


Figura 5.10: Patrón mensual de la profundidad de la isoterma de 20°C.

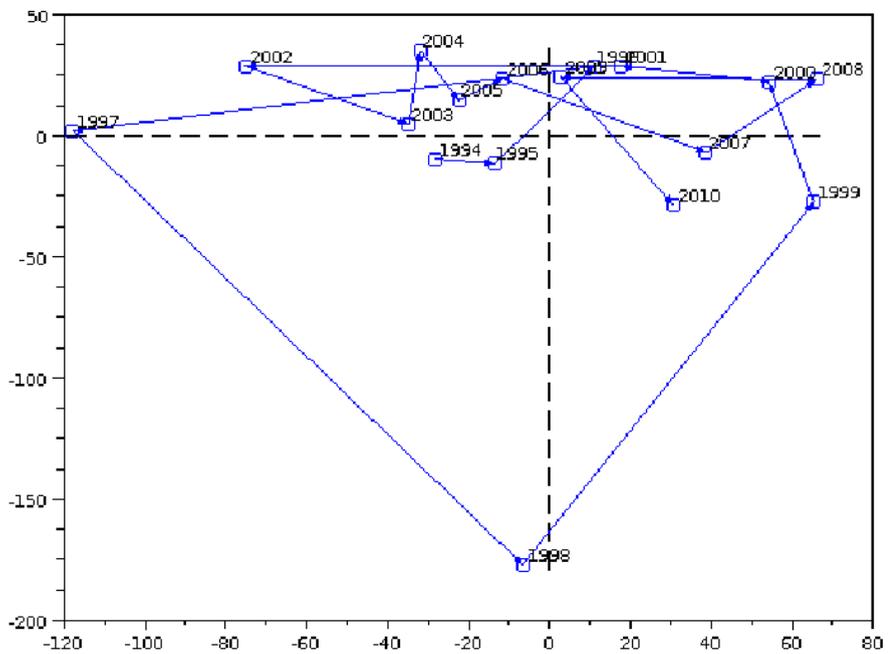


Figura 5.11: Representación anual de la profundidad de la isoterma de 20°C.

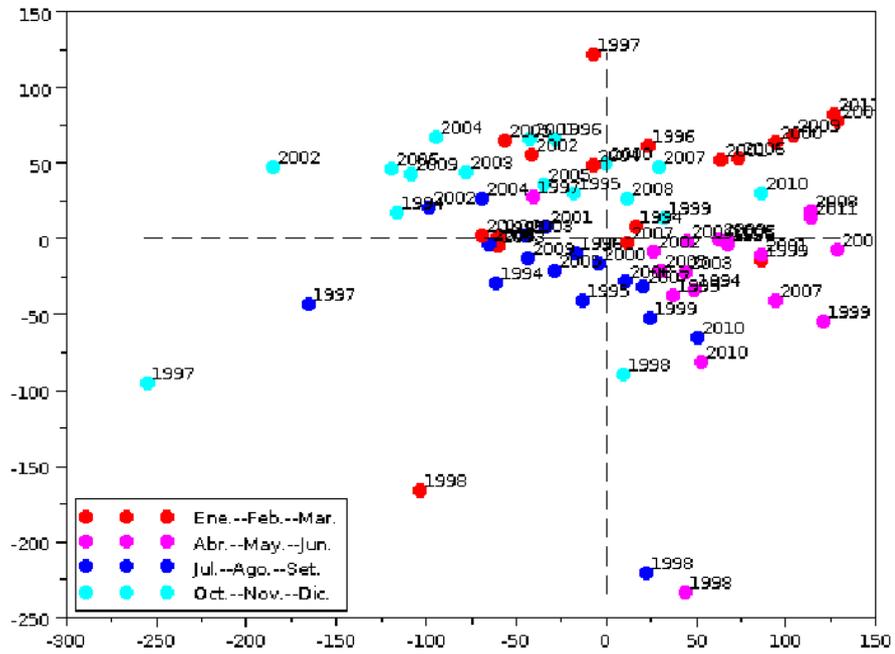


Figura 5.12: Representación trimestral de la profundidad de la isoterma de 20°C.

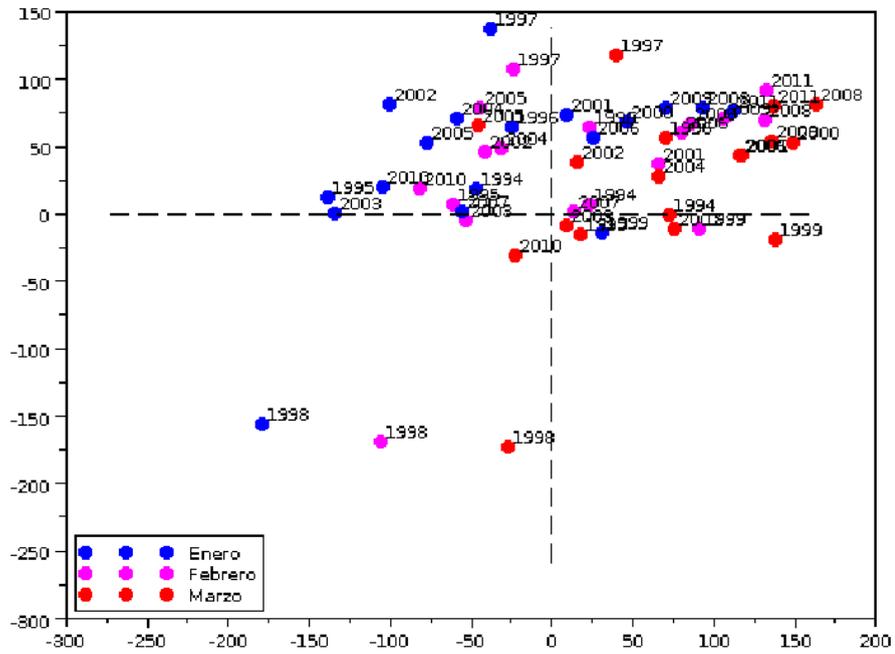


Figura 5.13: Representación mensual de la profundidad de la isoterma de 20°C.

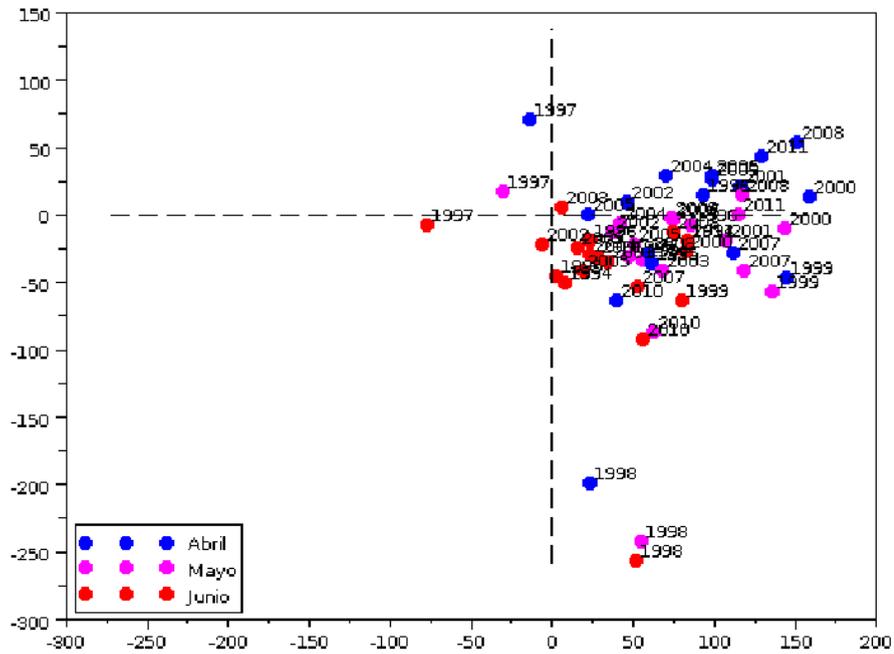


Figura 5.14: Representación mensual de la profundidad de la isoterma de 20°C.

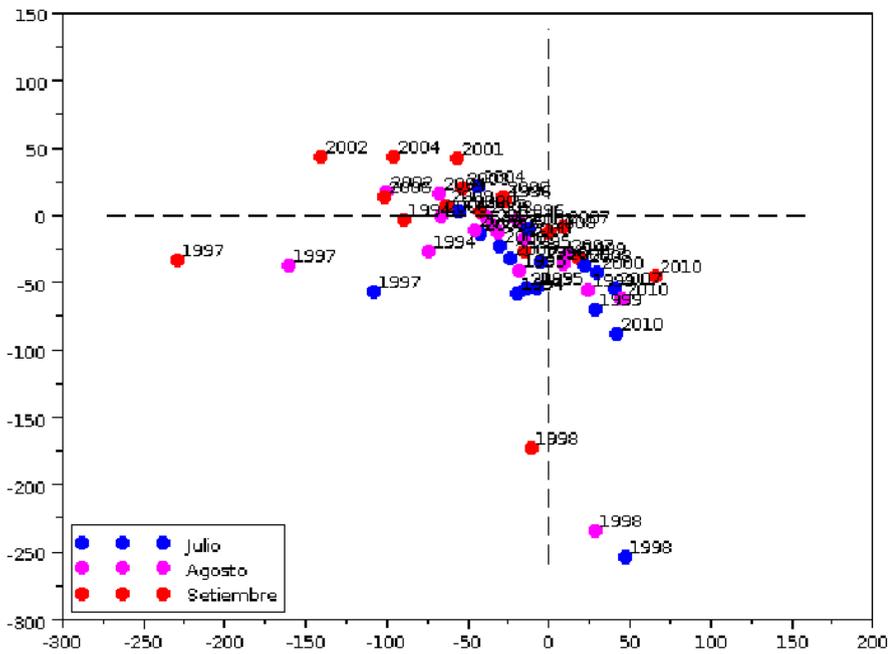


Figura 5.15: Representación mensual de la profundidad de la isoterma de 20°C.

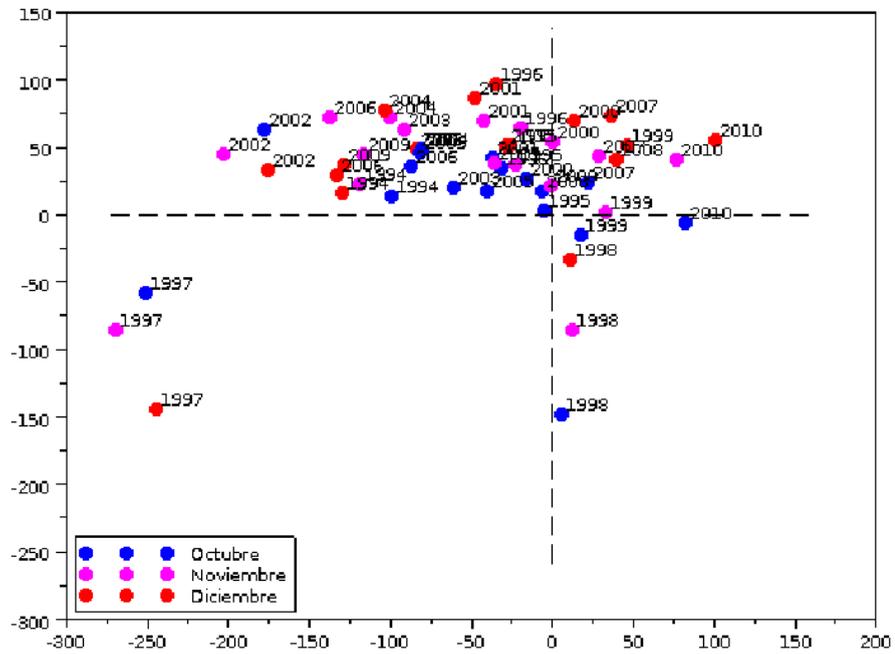


Figura 5.16: Representación mensual de la profundidad de la isoterma de 20°C.

# Capítulo 6

## Conclusión

En virtud del teorema 3.8, a la noción clásica de la técnica estadística del ACP como un método para consolidar variables mutuamente correlacionadas de datos multidimensionales en nuevas variables, se ha acrecentado la perspectiva geométrica de representar dichos datos multidimensionales (nube de puntos) en un plano que mejor se aproxime a dichos datos.

Al aplicar el algoritmo 1 en las bases de datos del capítulo 5 para analizar algunas bases de datos oceanográficos de la NOAA, se utilizó la técnica del ACP para la aprehensión visual de estas estructuras de datos de alta dimensión mediante la consolidación de los mismos en una nueva base de datos sobre un plano.

Continuando con un razonamiento inductivo, el teorema 3.8 puede ser extendido para incluir más componentes principales. Así, por ejemplo, podríamos representar estas mismas bases de datos del capítulo 4 en un espacio tridimensional. En efecto, siguiendo la notación del teorema 3.8, sea  $\bar{w}$  el autovector de  $A^T A$  asociado a su tercer mayor autovalor. Luego,

- (i)  $\mathcal{L}(\bar{z}, \bar{u})$  es una recta que mejor se aproxima a la nube  $x^1, \dots, x^m$ ,
- (ii)  $\mathcal{L}(\bar{z}, \bar{v})$  es una recta que mejor se aproxima, en el hiperplano  $\mathcal{H}$ , a la nube  $y^1, \dots, y^m$  y
- (iii)  $\mathcal{L}(\bar{z}, \bar{w})$  es una recta que mejor se aproxima, en el hiperplano  $\mathcal{H}' = \mathcal{H}(\bar{w}, \beta)$  de  $\mathcal{H}$  que pasa por  $\bar{z}$ , a la nube que surge de proyectar  $y^1, \dots, y^m$  sobre  $\mathcal{H}'$ .

Finalmente, cada  $x^i$  queda representado por

$$(\langle x^i - \bar{z}, \bar{u} \rangle, \langle x^i - \bar{z}, \bar{v} \rangle, \langle x^i - \bar{z}, \bar{w} \rangle).$$

# Bibliografía

- [Crouzeix et al., 2011] Crouzeix, J. P., Keraghel, A., and Sosa, W. (2011). *Programación Matemática Diferenciable*. Universidad Nacional de Ingeniería.
- [Fernández, 2007] Fernández, P. (2007). Google’s secret and linear algebra. *Newsletter of The European Mathematical Society*, 63:10–15.
- [Goldstein, 1964] Goldstein, A. A. (1964). Convex programmin in hilbert space. *Bulletin of The American Mathematical Society*, 70(5):709–710.
- [Gupta et al., 2013] Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., and Zadeh, R. (2013). Wtf: the who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*, pages 505–514.
- [Konishi, 2014] Konishi, S. (2014). *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling*. Texts in statistical science. CRC Press.
- [Lima, 2011a] Lima, E. L. (2011a). *Álgebra Linear*. IMPA. 8 edição.
- [Lima, 2011b] Lima, E. L. (2011b). *Curso de Análise*, volume 2. IMPA. 11 edição.
- [Lima, 2011c] Lima, E. L. (2011c). *Curso de Análise*, volume 1. IMPA. 13 edição.
- [von Mises and Pollaczek-Geiringer, 1929] von Mises, R. and Pollaczek-Geiringer, H. (1929). Praktische verfahren der gleichungsauflösung. (german) [métodos prácticos de resolución de ecuaciones]. *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik*, 9:152–164.