Universidad Nacional "Pedro Ruiz Gallo"

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO ACADÉMICO DE MATEMÁTICA

EXISTENCIA DE SOLUCIÓN EN UN PROBLEMA DE OPTIMIZACIÓN NO LINEAL CON RESTRICCIONES DE NO NEGATIVIDAD

TESIS

PARA OBTENER EL TÍTULO PROFESIONAL DE

LICENCIADO EN MATEMÁTICA

AUTORES:

BACH. MAT. JANET MERCEDES BANDA ISIQUE

BACH. MAT. JUAN JULIO RIOS VALLEJOS

ASESOR:

LIC. MAT. ROLANDO CÓRDOVA DESCALZI

Lambayeque - Perú

2015

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO ACADÉMICO DE MATEMÁTICA

EXISTENCIA DE SOLUCIÓN EN UN PROBLEMA DE OPTIMIZACIÓN NO LINEAL CON RESTRICCIONES DE NO NEGATIVIDAD

TESIS PRESENTADA POR:

BACH. MAT. JANET MERCEDES BANDA ISIQUE

BACH. MAT. JUAN JULIO RIOS VALLEJOS

PARA OBTENER EL TÍTULO PROFESIONAL DE

LICENCIADO EN MATEMÁTICA

ASESOR: LIC. MAT. ROLANDO CÓRDOVA DESCALZI

LAMBAYEQUE - PERÚ

2015

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

ESCUELA PROFESIONAL DE MATEMÁTICA

Los firmantes, por la presente certifican que han leído y recomiendan a la Facultad de Ciencias Físicas y Matemáticas la aceptación de la tesis titulada "Existencia de solución en un problema de optimización no lineal con restricciones de no negatividad" Presentada por la Bach. Mat. Janet Mercedes Banda Isique y el Bach.Mat. Juan Julio Rios Vallejos, en el cumplimiento parcial de los requisitos necesarios para la obtención del Título Profesional de Licenciado en Matemática.

....

M. Sc. Enrique W. Cárpena Velásquez

Presidente del Jurado

....

Lic. Mat. Amado Malca Villalobos

Secretario del Jurado

....

Lic. Mat. Juan A. Cornetero Capitán

Vocal del Jurado

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

Título de la Tesis:

"Existencia de solución en un problema de optimización no lineal con restricciones de no negatividad"

Escuela Profesional:

Matemática

Autores:

Bach. Mat. JANET MERCEDES BANDA ISIQUE

Bach. Mat. JUAN JULIO RIOS VALLEJOS

Asesor:

Lic. Mat. CÓRDOVA DESCALZI, Rolando

Banda I.

....

Bach. Mat. JANET MERCEDES BANDA ISIQUE

Autor

Juan Rios

....

Bach. Mat. JUAN JULIO RIOS VALLEJOS

Autor

....

Lic. Mat. Rolando Córdova Descalzi

Asesor

DEDCATORIA

A Dios quien me dio la vida y me ha llenado de bendiciones todo este tiempo. A mis queridos padres **Luis Ernesto Banda Palacios** y **Isolina Isique Chavesta** quienes me han brindado el apoyo y aliento en el logro de mi formación profesional.

A **Fernando Pavel Díaz Chero** por su confianza y apoyo necesario para la culminación de mi meta programada.

Janet

A Dios que me ha dado la vida y fortaleza para poder terminar esta tesis.

Todas las aquellas personas que me impulsaron a seguir, nombrarlos sería agotador y además me olvidaría algunos. Para ustedes.

A mis padres Carlos Licandro Rios Quiliche y María Jesús vallejos Linares por estar ahí cuando más lo necesite.

A mis maestros que con su apoyo incondicional y sus sabios conocimientos me permitieron lograr mis metas trazadas.

Juan

AGRADECIMIENTO

Agradecemos a Dios por habernos acompañado y guiado a lo largo de nuestra carrera, a él por brindarnos una vida llena de aprendizajes, experiencias y sobre todo felicidad.

Debemos agradecer de manera especial a nuestros padres por apoyarnos en todo momento, por los valores que nos han inculcado, y por habernos dado la oportunidad de tener una excelente educación en el transcurso de nuestra vida.

Gracias a todas aquellas personas que han formado parte de nuestra vida profesional a las que nos encantaría agradecerles su amistad, consejos, apoyo, ánimos y compañía en los momentos más difíciles de nuestra vida.

Agradecemos también a nuestro asesor de tesis, por su valiosa guía y asesoramiento a la realización de la misma.

Para ellos: Muchas gracias y que Dios los bendiga.

RESUMEN

En esta tesis se presenta un modelo matemático para la creación de carteras de inversión óptimas en un mercado financiero, cuya solución queda garantizada mediante la aplicación del teorema de Karush – Kuhn – Tucker. Primeramente se usan las definiciones de valor esperado, varianza y covarianzas para obtener el rendimiento esperado, varianza y covarianza de cada activo que forma parte de la cartera, con esta información se determina tanto el valor esperado como la varianza y covarianza de la cartera de inversión, luego se construye la matriz de varianzas y covarianzas de la cartera de inversión, y de este modo se obtiene un modelo de programación cuadrática con restricciones el cual es el modelo matemático para la creación de carteras de inversión óptimas en un mercado financiero , puesto que se quiere optimizar el valor esperado y el riesgo el problema es de minimización. Finalmente para poner de manifiesto la fuerza de la técnica descrita en este trabajo, se aplica en el caso particular de un inversionista que desea invertir en una cartera formada por acciones de las empresas peruanas: Alicorp S.A.A, Cervecerías unidas Backus& Johnston S.A.A y Gloria S.A., cuya información fue obtenida de la página Web de la Bolsa de Valores de Lima, el modelo implementado para este caso particular fue resuelto utilizando la herramienta de Solver de la planilla de cálculo de Excel.

ABSTRACT

In this thesis a mathematical model for creating optimal investment portfolios in a financial market, the solution is ensured by applying the Karush theorem is presented. First the definitions of expected value, variance and covariance are used to obtain the expected return, variance and covariance of each asset that is part of the portfolio, then this information is determined by both the expected value and variance and covariance portfolio investment, then the matrix of variances and covariances of the investment portfolio is constructed, and thus a model of quadratic constraint programming which is the mathematical model for creating portfolios optimal investment in a financial market is obtained, since you want to optimize the expected value and risk minimization is the problem. Finally, to demonstrate the strength of the technique described in this paper applies in the case of an investor who wants to invest in a portfolio consisting of shares of Peruvian companies: Alicorp SAA, together Breweries Backus & Johnston SAA and Gloria SA, whose information was obtained from the Web site of the Lima Stock Exchange, the model implemented in this particular case was solved using the Solver in Excel spreadsheet.

ÍNDICE GENERAL

PÁG.

CAPÍTULO 1

BASE TEÓRICA

En este capítulo se dan las definiciones y resultados básicos útiles para el desarrollo de la parte central de este trabajo.

1.1. ESPACIOS VECTORIALES

DEFINICION: Sea $V \neq \emptyset$ un conjunto, k un campo sobre el cual se definen las siguientes dos operaciones, la llamadas adición, que se denota por (+) y producto, que se denota por (.)

$$+ : V \times V \to V$$
$$(x,y) \to +(x,y) = x + y$$

$$\cdot : V \times V \to V$$
$$(\alpha, x) \to .(\alpha, x) = \alpha . x$$

Si estas operaciones satisfacen las siguientes:

A_1.- $x + y = y + x,\ \forall\, x,y\ \in V$ (propiedad conmutativa).

A_2.- $x + (y + z) = (x + y) + z\ \forall\, x,y,z\ \in V$ (propiedad asociativa para la adición).

A_3.- $\forall\, x \in V$, existe $0 \in V$, tal que $x + 0 = 0 + x = x$ donde 0 se denomina elemento neutro aditivo o cero para la adición.

A_4.-$\forall\, x \in V$, existe $-x \in V$, tal que $x + (-x) = (-x) + x = 0$, donde $-x$ se denomina elemento opuesto de x para la adición.

B_1.- $\alpha(\beta x) = (\alpha\beta)x,\ \forall\, x, \alpha, \beta\ \in k$ (propiedad conmutativa para el producto interno)

B_2.-$(\alpha+\beta)x = \alpha x + \beta x,\ \forall\, x \in V, \alpha, \beta \epsilon\, k$ (propiedad distributiva respecto a los escalares)

B_3.- $\alpha(x+y) = \alpha x + \alpha y,\ \forall\, x, y \in V, \alpha \epsilon\, k$ (propiedad distributiva respecto a los elementos del conjunto V)

B_4.- $\forall\ x \in V$, existe un único $1 \in k$ llamado elemento identidad multiplicativo tal que $1.x = x$

En este caso se dice que $(V, +, .)$ es un espacio vectorial sobre K.

Observación:

Los elementos de V se llaman vectores y los elementos de K se llaman escalares.

Como V esta definido sobre los elementos de K, se dice que V es un K espacio vectorial.

De la definición anterior se sigue que para un conjunto $V \neq \emptyset$ para que sea un espacio vectorial sobre un campo K debe tener definidas do operaciones ´´ suma ´´ y ´´multiplicación por un escalar´´ y que cumple los ocho axiomas mencionados, en caso que no cumpla con alguno de dichos axiomas no es un espacio vectorial.

Ejemplo 1: Probar que el conjunto de todos los polinomios con coeficientes reales,

$$P[x] = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, n \in N, a_0, a_1, \dots, a_n \in \mathbb{R}\}$$

es un espacio vectorial, con las operaciones definidas como sigue:

$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$, donde $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$; $q(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$ y $\lambda p(x) = \lambda a_0 + \lambda a_1 x + \lambda a_2 x^2 + \cdots + \lambda a_n x^n$.

En efecto:

1.) Sea $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$,

$q(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$ dos elementos de $P[x]$

Probaremos que: $p(x) + q(x) = q(x) + p(x)$

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

$$= (b_0 + a) + (b_1 + a_1)x + \cdots + (b_n + a_n)x^n$$
$$= q(x) + p(x)$$

2.) Sea $p_1(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, $p_2(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n$, $p_3(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$ tres elementos de $P[x]$

Probaremos que: $p_1(x) + \big(p_2(x) + p_3(x)\big) = \big(p_1(x) + p_2(x)\big) + p_3(x)$

3.) El elemento neutro es el polinomio nulo $q(x) = 0$, puesto que para cualquier $p(x) \in P[x]$ se verifica que: $p(x) + q(x) = p(x) + 0 = p(x)$

4.) Dado cualquier polinomio $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ de $P[x]$ se verifica que el polinomio $\bar{p}(x) = -a_0 - a_1x - a_2x^2 - \cdots - a_nx^n$ es su elemento opuesto, puesto que $p(x) + \bar{p}(x) = q(x) = 0$

5.) Si $p_1(x), p_2(x) \in P[x]$ y $\alpha \in$ se debe probar que:

$$\alpha[p_1(x) + p_2(x)] = \alpha p_1(x) + \alpha p_2(x)$$

6.) Si : $\alpha, \beta \in \mathbb{R}$ y $p(x) \in P[x]$ se debe probar que:

$$(\alpha + \beta)p(x) = \alpha p(x) + \beta p(x)$$

$$(\alpha + \beta)p(x) = (\alpha + \beta)(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n)$$
$$= \alpha(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + \beta(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n)$$
$$= \alpha p(x) + \beta p(x)$$

7.) Si $\alpha, \beta \in \mathbb{R}$ y $p(x) \in P[x]$ se debe probar que: $(\alpha\beta)p(x) = \alpha(\beta p(x))$

$$(\alpha\beta)p(x) = (\alpha\beta)(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n)$$
$$= \alpha(\beta a_0 + \beta a_1x + \beta a_2x^2 + \cdots + \beta a_nx^n)$$
$$= \alpha(\beta p(x))$$

8.) $\exists\ 1 \in \mathbb{R}$, tal que $1.p(x) = p(x)$ para todo polinomio $p(x)$ de $P[x]$ con lo cual se ha probado que el conjunto $P[x]$ de polinomios de coeficientes reales es un espacio vectorial sobre$\mathbb{R}$.

Ejemplo 2: Probar que el conjunto de todas las matrices de orden $m\,x\,n$ sobre $K = \mathbb{R}$ denotado por $\mathbb{R}^{mxn}$ esta provisto de dos operaciones suma (+) y producto (.). Probaremos que $(\mathbb{R}^{mxn}, +, .)$ es un espacio vectorial sobre $K = \mathbb{R}$.

En efecto:

1.) Sea $f, g \in \mathbb{R}^{mxn} \Rightarrow f, g: X \to \mathbb{R}$ de donde

$$f = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix}, \qquad g = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{mn} \end{bmatrix}$$

$$f + g = \begin{bmatrix} f_{11} + g_{11} & \cdots & f_{1n} + g_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} + g_{m1} & \cdots & f_{mn} + g_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} g_{11} + f_{11} & \cdots & g_{1n} + f_{1n} \\ \vdots & \ddots & \vdots \\ g_{m1} + f_{m1} & \cdots & g_{mn} + f_{mn} \end{bmatrix} = g + f$$

Por lo tanto $f + g = g + f$

2.) Sea $f, g, h \in \mathbb{R}^{mxn} \Rightarrow f, g, h: X \to \mathbb{R}$ de donde

$$f + (g + h) = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix} + \begin{bmatrix} g_{11} + h_{11} & \cdots & g_{1n} + h_{1n} \\ \vdots & \ddots & \vdots \\ g_{m1} + h_{m1} & \cdots & g_{mn} + h_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} f_{11} + (g_{11} + h_{11}) & \cdots & f_{1n} + (g_{1n} + h_{1n}) \\ \vdots & \ddots & \vdots \\ f_{m1} + (g_{m1} + h_{m1}) & \cdots & f_{mn} + (g_{mn} + h_{mn}) \end{bmatrix}$$

$$= \begin{bmatrix} (f_{11} + g_{11}) + h_{11} & \cdots & (f_{1n} + g_{1n}) + h_{1n} \\ \vdots & \ddots & \vdots \\ (f_{m1} + g_{m1}) + h_{m1} & \cdots & (f_{mn} + g_{mn}) + h_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} f_{11} + g_{11} & \cdots & f_{1n} + g_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} + g_{m1} & \cdots & f_{mn} + g_{mn} \end{bmatrix} + \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mn} \end{bmatrix}$$

$$= (f + g) + h$$

3.) Sea la matriz $0 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$ tal que $\forall\, f \in \mathbb{R}^{mxn}$ se tiene:

$$f + 0 = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} f_{11} + 0 & \cdots & f_{1n} + 0 \\ \vdots & \ddots & \vdots \\ f_{m1} + 0 & \cdots & f_{mn} + 0 \end{bmatrix}$$

$$= \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix}$$

$$= f$$

$$\therefore\ \forall\, f \in \mathbb{R}^{mxn}, \exists\, 0 \in \mathbb{R}^{mxn}, tal\ que\ f + 0 = f$$

4.) $\forall\, f \in \mathbb{R}^{mxn}$sea $-f = \begin{bmatrix} -f_{11} & \cdots & -f_{1n} \\ \vdots & \ddots & \vdots \\ -f_{m1} & \cdots & -f_{mn} \end{bmatrix} \Rightarrow$

$$f + (-f) = \begin{bmatrix} f_{11} - f_{11} & \cdots & f_{1n} - f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} - f_{m1} & \cdots & f_{mn} - f_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Luego $\forall\, f \in \mathbb{R}^{mxn}$, $\exists -f \in \mathbb{R}^{mxn}$ tal que $f + (-f) = 0$

5.) *Sea* $f \in \mathbb{R}^{mxn}$, $\lambda, \beta \in \mathbb{R}$ *se tiene:*

$$(\lambda\beta)f = \begin{bmatrix} \lambda\beta f_{11} & \cdots & \lambda\beta f_{1n} \\ \vdots & \ddots & \vdots \\ \lambda\beta f_{m1} & \cdots & \lambda\beta f_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda(\beta f_{11}) & \cdots & \lambda(\beta f_{1n}) \\ \vdots & \ddots & \vdots \\ \lambda(\beta f_{m1}) & \cdots & \lambda(\beta f_{mn}) \end{bmatrix}$$

$$= \lambda \begin{bmatrix} \beta f_{11} & \cdots & \beta f_{1n} \\ \vdots & \ddots & \vdots \\ \beta f_{m1} & \cdots & \beta f_{mn} \end{bmatrix}$$

$$= \lambda(\beta f)$$

6.) Sea $f \in \mathbb{R}^{mxn}$, $\lambda, \beta \in \mathbb{R}$ se tiene:

$$(\lambda + \beta)f = \begin{bmatrix} (\lambda + \beta)f_{11} & \cdots & (\lambda + \beta)f_{1n} \\ \vdots & \ddots & \vdots \\ (\lambda + \beta)f_{m1} & \cdots & (\lambda + \beta)f_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda f_{11} + \beta f_{11} & \cdots & \lambda f_{1n} + \beta f_{1n} \\ \vdots & \ddots & \vdots \\ \lambda f_{m1} + \beta f_{m1} & \cdots & \lambda f_{mn} + \beta f_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda f_{11} & \cdots & \lambda f_{1n} \\ \vdots & \ddots & \vdots \\ \lambda f_{m1} & \cdots & \lambda f_{mn} \end{bmatrix} + \begin{bmatrix} \beta f_{11} & \cdots & \beta f_{1n} \\ \vdots & \ddots & \vdots \\ \beta f_{m1} & \cdots & \beta f_{mn} \end{bmatrix}$$

$$= \lambda f + \beta f$$

7.) Sea $f , g \in \mathbb{R}^{mxn}$, $\lambda \in \mathbb{R}$ se tiene:

$$\lambda(f + g) = \begin{bmatrix} \lambda(f_{11} + g_{11})(f_{11} + g_{11}) & \cdots & \lambda(f_{1n} + g_{1n}) \\ \vdots & \ddots & \vdots \\ \lambda(f_{m1} + g_{m1}) & \cdots & \lambda(f_{mn} + g_{mn}) \end{bmatrix}$$

$$= \begin{bmatrix} \lambda f_{11} + \lambda g_{11} & \cdots & \lambda f_{1n} + \lambda g_{1n} \\ \vdots & \ddots & \vdots \\ \lambda f_{m1} + \lambda g_{m1} & \cdots & \lambda f_{mn} + \lambda g_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda f_{11} & \cdots & \lambda f_{1n} \\ \vdots & \ddots & \vdots \\ \lambda f_{m1} & \cdots & \lambda f_{mn} \end{bmatrix} + \begin{bmatrix} \lambda g_{11} & \cdots & \lambda g_{1n} \\ \vdots & \ddots & \vdots \\ \lambda g_{m1} & \cdots & \lambda g_{mn} \end{bmatrix}$$

$$= \lambda \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix} + \lambda \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{mn} \end{bmatrix}$$

$$= \lambda f + \lambda g$$

8.) $\forall\ f\ \in \mathbb{R}^{mxn}, \exists I \in \mathbb{R}^{mxn}$ tal que $fI = f$

$$fI = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$$fI = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix}$$

$$fI = f$$

Por lo tanto se tiene que el conjunto de matrices $\mathbb{R}^{mxn}$ de orden mxn provisto de las dos operaciones de suma y producto es un espacio vectorial sobre$\mathbb{R}$.

1.2. ESPACIOS NORMADOS

Definición 2: Sea E un espacio vectorial sobre K ($\mathbb{R}$ ***o*** $\mathbb{C}$). Diremos que E es un espacio normado si existe una función:

$$\|.\|: E \to {R_0}^{+}$$

$$x \to \|x\|$$

Que cumple con las condiciones:

1.) $\|x\| \geq 0\ \forall x \epsilon\ E$

$\|x\| = 0 \leftrightarrow x = 0$

2.) $\|\lambda x\| = |\lambda| \|x\| \qquad \forall \lambda \epsilon\ k\ \text{ y } \forall x \epsilon\ E$

3.) $\|x + y\| \leq \|x\| + \|y\| \qquad \forall x, y \epsilon\ E$

Todo espacio normado ($E, \|.\|$) se puede convertir en un espacio métrico (E, d) donde se define la distancia.

$$d(x, y) = \|x - y\| \text{ .... (1)}$$

ESPACIO MÉTRICO:

Definición 3: Sea x un conjunto no vacío. Diremos que X es un espacio métrico si existe una función.

$$d: XxX \to \mathbb{R}$$
$$(x,y) \to d(x,y)$$

Que cumple con las condiciones:

1.) $d(x,y) \geq 0 \quad \forall x,y \in X$

$\quad d(x,y) = 0 \leftrightarrow x = y$

2.) $d(x,y) = d(y,x) \qquad \forall x,y \in X$

3.) $d(x,y) \leq d(x,z) + d(z,y) \qquad \forall x,y,z \in X$

Vemos que (1) cumple las condiciones de métrica:

$$d(x,y) = \|x-y\| \geq 0 \text{ Por ser norma}$$

1.) Si $\quad d(x,y) = 0$

$\|x-y\| = 0$

$\|x-y\| = 0$

$\to \quad x-y = 0$ Por 1.)

Si $\quad x = y \Rightarrow x - y = 0$

Luego $\quad \|x-y\| = 0$ Por 1.)

$d(x,y) = 0$

2.) $d(x,y) = \|x-y\|$

$= \|(-1)(y-x)\|$

$= |-1|\|(y-x)\|$ Por 2.)

$= \|(y-x)\|$

$= d(x,y)$

Por tanto $\quad d(x,y) = d(y,x)$

3.) Si $\quad d(x,y) = \|x-y\|$

$= \|x-z+z-y\|$

$\leq \|x-z\| + \|z-y\|$

$$= d(x,z) + d(z+y)$$

Por lo tanto $\quad d(x,y) \leq d(x,z) + d(z,y)$

Ejemplo 3:

Sea $E = \{f: [a,b] \to \mathbb{R}: f \text{ continua}\}$

Definiendo $\quad (f+g)(x) = f(x) + g(x)$

$$(\lambda f)(x) = \lambda f(x)$$

E con estas operaciones forman un espacio vectorial sobre$\mathbb{R}$.

Definimos: $\|f\| = \max\limits_{x\in[a,b]} |f(x)|$

Vemos que cumple las condiciones de norma

$\forall x \in [a,b] \qquad |f(x)| \geq 0$

1.) Entonces $\|f\| = 0 \leftrightarrow f = 0$

$\Rightarrow$ Supongamos que $\|f\| = 0$, por definición de máximo se sigue que:

$$|f(x)| \leq \max_{x\in[a,b]} |f(x)| = 0$$

Luego $|f(x)| = 0 \qquad \forall x \in [a,b]$

Y por tanto $f(x) = 0 \qquad \forall x \in [a,b]$

Luego f=0 elemento identidad de E.

$\Leftarrow$ Si $f = 0$

$|f(x)| = 0 \qquad \forall x \in [a,b]$

Y por tanto $\|f\| = \max\limits_{x\in[a,b]} |f(x)| = 0$

2.) $\|\lambda f\| = |\lambda|\|f\|$

$$\|\lambda f\| = \max_{x\in[a,b]}|\lambda f(x)|$$

$$= \max_{x\in[a,b]}|\lambda f(x)|$$

$$= \max_{x\in[a,b]}|\lambda||f(x)|$$

$$= |\lambda|\max_{x\in[a,b]}|f(x)|$$

$$= |\lambda|\|f\|$$

Por tanto: $\|\lambda f\| = |\lambda|\|f\|$

3.) $\|f + g\| \leq \|f\| + \|g\|$

$\forall\ x \in [a, b]$ **Se tiene**

$$|(f + g)(x)| = |f(x) + g(x)| \leq |f(x)| + |g(x)|$$

$$\leq \max_{x\in[a,b]}|f(x)| + \max_{x\in[a,b]}|g(x)|$$

$$\leq \|f\| + \|g\|$$

Luego $\forall x \in [a, b]$ **se tiene**

$$|(f + g)(x)| \leq \|f\| + \|g\|$$

Por tanto: $\max_{x\in[a,b]}|(f + g)(x)| \leq \|f\| + \|g\|$

$$\|f + g\| \leq \|f\| + \|g\|$$

1.3. FUNCIONES DIFERENCIABLES

Derivadas parciales de funciones de varias variables

Definición 4: Sea $f: U \to \mathbb{R}$ donde U es un subconjunto abierto de $\mathbb{R}^n$, y consideremos un punto $a \in U$. Para i = 1,....., n, la i-ésima derivada parcial de f en el punto a, que designamos indistintamente mediante

$$\partial_i f(\mathrm{a}) \text{ O también } \frac{\partial f}{\partial x_i} f(\mathrm{a}) \text{ o } f x_i$$

Es el valor del límite $\lim_{t \to 0} \frac{f(\mathrm{a}+\mathrm{t}e_i)-f(\mathrm{a})}{t}$ Cuando este límite existe.

En el caso particular en el que n = 2, se tiene para un punto $a = (x, y)$, la derivada parcial respecto a x viene dada por la siguiente expresión:

$$\partial_1 f(\mathrm{a}) = \frac{\partial f}{\partial x} f(\mathrm{a}) = \lim_{\mathrm{t} \to 0} \frac{\mathrm{f(x+t,y)-f(x,y)}}{\mathrm{t}}$$

La derivada parcial de f respecto de y se define de manera análoga como sigue:

$$\partial_2 f(\mathrm{a}) = \frac{\partial f}{\partial y} f(\mathrm{a}) = \lim_{t \to 0} \frac{f(\mathrm{x,y+t})-f(\mathrm{x,y})}{t}$$

Una situación análoga obtenemos cuando n = 3: en este caso $a =$ (x; y; z), y designamos la tercera derivada parcial, o derivada parcial respecto de z, mediante $\partial_3 f(\mathrm{a}) = f_z(\mathrm{a}) = f_z(x, y, z)$.

En lo que respecta al cálculo de las derivadas parciales, como la definición indica que la derivada parcial i-ésima es la derivada de una función real, con respecto de una variable real privilegiada x_i, cuando las otras permanecen constantes, se aplican las reglas de derivación de funciones reales de variable real.

Ejemplo 4. Calcular las dos derivadas parciales de la función $f(x, y) = 2x^2 - xy + y^2$ en el punto $p(1,0)$.

Solución

$$\frac{\partial f}{\partial x}(1,0) = \lim_{t\to 0}\frac{f(1+t,0)-f(1,0)}{t}$$
$$= \lim_{t\to 0}\frac{2(1+h)^2-(1+h).0+0^2-(2-0+0)}{t}$$
$$= \lim_{t\to 0}\frac{2(1+2t+t^2)-2}{t} = \lim_{t\to 0}\frac{2+4t+2t^2-2}{t} = \lim_{t\to 0}\frac{4t+2t^2}{t}$$
$$= \lim_{t\to 0} 4+2t = 4$$

$$\frac{\partial f}{\partial y}(1,0) = \lim_{t\to 0}\frac{f(1,0+t)-f(1,0)}{t} = \lim_{t\to 0}\frac{f(1,t)-f(1,0)}{t}$$
$$= \lim_{t\to 0}\frac{2.1^2-1.t+t^2-2}{t} = \lim_{t\to 0}\frac{2-t+t^2-2}{t} = \lim_{t\to 0}\frac{-t+t^2}{t} = \lim_{t\to 0} -1+t = -1$$

Continuidad y derivadas parciales:

La existencia de f_x depende del comportamiento de la función solo en la dirección del eje x, y la existencia de f_y del comportamiento de la función solo en la dirección del eje y, mientras que la continuidad depende del comportamiento de la función en todos los puntos del entorno. Esto significa que una función puede tener derivadas parciales en un punto aunque no sea continua en dicho punto. Es decir la existencia de las derivadas parciales no garantiza la continuidad de una función

Tenemos pues, que de la continuidad de las funciones de n variables en un punto dado, no se deriva la existencia de sus derivadas parciales en ese punto. Pero es más, cuando $n \geq 2$ incluso de la existencia de todas las derivadas parciales en cierto punto, no se deduce la continuidad en ese punto. (Recordemos que para n = 1, es decir, para las funciones de una variable, de la existencia de la derivada en un punto se deriva también que la función es continua en ese punto).

Ejemplo 5. Consideremos $f\colon \mathbb{R}^2 \to \mathbb{R}$ definida mediante $f(0,0) = 0$ y $f(x,y) = \frac{xy}{x^2+y^2}$, si $(x,y) \neq (0,0)$

En todo punto distinto de (0,0) la función tiene derivadas parciales, dadas por

$$\frac{\partial f}{\partial x}(x,y) = \frac{y^3-x^2y}{x^2+y^2}, \qquad \frac{\partial f}{\partial x}(x,y) = \frac{x^3-x^2y}{x^2+y^2}$$

En el origen (0,0) tenemos

$$\frac{\partial f}{\partial x}(0{,}0) = \lim_{t\to 0}\frac{f(0+t,0)-f(0,0)}{t} = \lim_{t\to 0}\frac{f(t,0)-f(0,0)}{t} = \lim_{t\to 0}\frac{\frac{t.0}{t^2+0^2}-0}{t} = 0$$

$$\frac{\partial f}{\partial y}(0{,}0) = \lim_{t\to 0}\frac{f(0,0+t)-f(0,0)}{t} = \lim_{t\to 0}\frac{f(0,t)-f(0,0)}{t} = \lim_{t\to 0}\frac{\frac{0.t^2}{0^2+t^2}-0}{t} = 0$$

Sin embargo, y aunque existen ambas derivadas parciales en todos los puntos, la función $f(\mathrm{x},\mathrm{y})$ no es continua (0,0), ya que no existe el limite en dicho punto. En efecto, si nos acercamos al punto mediante las rectas $y = mx$ resulta:

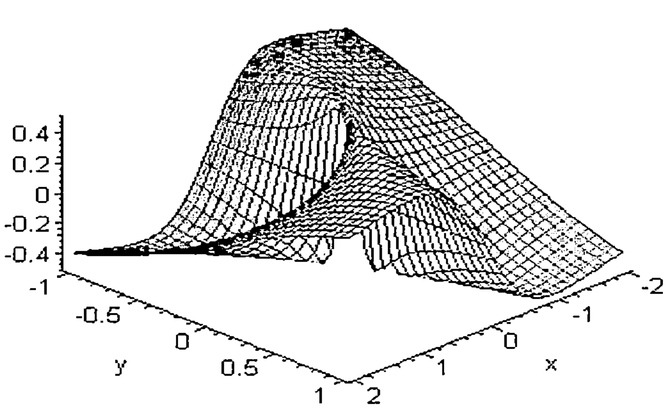
$$\lim_{(x,y)\to(0,0)}\frac{xy}{x^2+y^2}$$

$$= \lim_{\substack{(x,y)\to(0,0)\\ y=mx}}\frac{xy}{x^2+y^2} = \lim_{\substack{(x,y)\to(0,0)\\ y=mx}}\frac{xy}{x^2+y^2} = \lim_{\substack{(x,y)\to(0,0)\\ y=mx}}\frac{xmx}{x^2+(mx)^2}$$

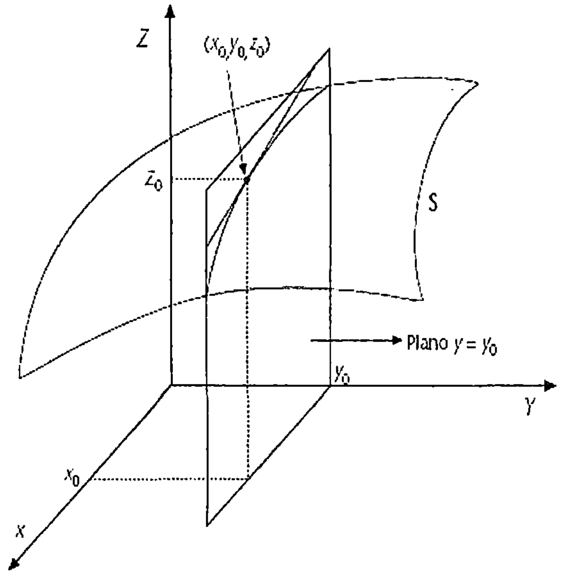
$$\lim_{\substack{(x,y)\to(0,0)\\ y=mx}}\frac{x^2m}{x^2+m^2x^2} = \lim_{\substack{(x,y)\to(0,0)\\ y=mx}}\frac{x^2m}{x^2(1+m^2)} = \lim_{x\to 0}\frac{m}{1+m^2} = \frac{m}{1+m^2}$$

Luego el límite no existe ya que depende del valor m.Es decir, según la recta por la que nos aproximemos al punto tendríamos un valor de límite u otro

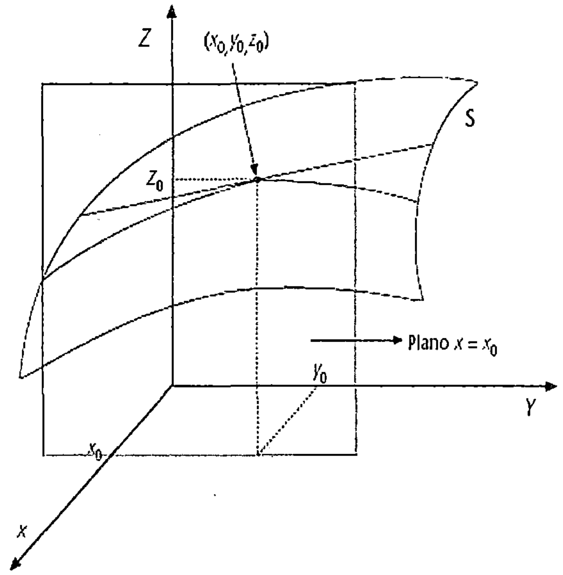
En lo que sigue revisamos la interpretación geométrica de las derivadas parciales



Interpretación geométrica de las derivadas parciales:



$\frac{\partial f}{\partial x}(x_0, y_0)$: pendiente en la dirección x



$\frac{\partial f}{\partial y}(x_0, y_0)$: pendiente en la dirección y

Desde el punto de vista geométrico, la función $g(x) = f(x, y_0)$ representa la curva que se obtiene de la intersección de la superficie $z = f(x, y)$ con el plano $y = y_0$.

$$\left.\begin{matrix} z = f(x,y) \\ y = y_0 \end{matrix}\right\} z = f(x, y_0) = g(x)$$

La derivada parcial de la función f, respecto de la variable x, en el punto p representa la pendiente de la tangente a la curva $g(x) = f(x, y_0)$ en el punto p correspondiente de la gráfica, es decir, la inclinación de la superficie en la dirección del eje X.

Análogamente, la función $g(y) = f(x_0, y)$ representa la curva que se obtiene de la intersección de la superficie $z = f(x, y)$ con el plano $x = x_0$.

$$\left.\begin{matrix} z = f(x,y) \\ x = x_0 \end{matrix}\right\} z = f(x_0, y) = g(y)$$

La derivada parcial de la función f , respecto de la variable y , en el punto p representa la pendiente de la tangente a la curva $g(y) = f(x_0, y)$ en el punto p correspondiente de la gráfica, es decir, la inclinación de la superficie en la dirección del eje Y.

Ejemplo 6. Hallar la pendiente a la superficie $f(x,y) = -\frac{x^2}{2} - y^2 + \frac{25}{8}$ en el punto $p\left(\frac{1}{2}, 1,2\right)$, en las direcciones de los ejes x e y.

<u>Solución</u>

$$\frac{\partial f}{\partial x} = -x \to \frac{\partial f}{\partial x}\left(\frac{1}{2}, 1\right) = -\frac{1}{2} \to tan\alpha = -\frac{1}{2}$$

$$\frac{\partial f}{\partial y} = -2y \to \frac{\partial f}{\partial y}\left(\frac{1}{2}, 1\right) = -2 \to tan\beta = -2$$

Definición (Derivadas parciales de orden superior) 5:

Consideremos $f : U \to \mathbb{R}$ diferenciable en cada punto del abierto $U \subset \mathbb{R}^n$. Para cada natural i = 1,....., n tenemos definida la función

$$\frac{\partial f}{\partial x_i} : U \to \mathbb{R}$$

Denotamos a las derivadas de segundo orden por

$$\frac{\partial^2 f}{\partial x_i \partial_{x_j}} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

Se ha derivado primero con respecto a x_j y luego con respecto a x_i.

$$\frac{\partial^2 f}{\partial {x_i}^2} = \frac{\partial^2 f}{\partial x_i \partial_{x_j}} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

Se ha derivado dos veces con respecto a x_i.

Ejemplo 7. Hallar las derivadas parciales de segundo orden de la función:

$$f(x,y) = x^2 y e^{x^2+y^2}$$

<u>Solución</u>

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(2xye^{x^2+y^2} + x^2y\left(2xe^{x^2+y^2}\right) = \frac{\partial}{\partial x}(2x^3y + 2xy)e^{x^2+y^2}\right)$$

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = (6x^2y + 2y)e^{x^2+y^2} + (2x^3y + 2xy)\left(2xe^{x^2+y^2}\right)$$
$$= (4x^4y + 10x^2y + 2y)e^{x^2+y^2}$$

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}\left(2xye^{x^2+y^2} + x^2y\left(2xe^{x^2+y^2}\right)\right) = (2x^3 + 2x)e^{x^2+y^2}$$
$$+ (2x^3y + 2xy)\left(2ye^{x^2+y^2}\right) = (4x^3y^2 + 2x^3 + 4xy^2 + 2x)e^{x^2+y^2}$$

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\left(x^2e^{x^2+y^2} + x^2y\left(2ye^{x^2+y^2}\right)\right) = \frac{\partial}{\partial x}\left((2x^2y^2 + x^2)e^{x^2+y^2}\right)$$

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = (4xy^2 + 2x)e^{x^2+y^2} + (2x^2y^2 + x^2)\left(2xe^{x^2+y^2}\right)$$
$$= (4x^3y^2 + 2x^3 + 4xy^2 + 2x)e^{x^2+y^2}$$

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\left(x^2e^{x^2+y^2} + x^2y\left(2ye^{x^2+y^2}\right)\right) = \frac{\partial}{\partial y}\left((2x^2y^2 + x^2)e^{x^2+y^2}\right)$$

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = 4x^2ye^{x^2+y^2} + (2x^2y^2 + x^2)\left(2ye^{x^2+y^2}\right) = (4x^2y^3 + 6x^2y)e^{x^2+y^2}$$

Diferenciabilidad de funciones de varias variables

Al generalizar un concepto de $\mathbb{R}$ a $\mathbb{R}^n$, tratamos de conservar las propiedades importantes que consideremos en el caso unidimensional. Por ejemplo, en $\mathbb{R}$ la existencia de la derivada en un punto x implica la continuidad en el mismo. Sin embargo, para funciones de dos variables, hemos visto que la existencia de las derivadas parciales en un punto no implica la continuidad en ese punto, ni siquiera la existencia de todas las derivadas direccionales implica la continuidad. Por esta razón las derivadas parciales, al igual que las derivadas direccionales, son una extensión en cierto modo poco satisfactoria del concepto de derivada unidimensional. Por tanto, parece natural el deseo de tener una noción de derivada para funciones de varias variables que implique la continuidad. Introducimos ahora una generalización más conveniente que implica la continuidad y, al propio tiempo, nos permite extender los principales teoremas de la teoría de la derivada unidimensional a las funciones de varias variables. El concepto que mejor sirve a tal propósito es la noción de diferencial.

Nos proponemos definir la diferenciabilidad de una función en un punto, generalizando la noción de derivabilidad de funciones reales, como sigue. Recordemos que dada $f : I \to \mathbb{R}$, donde I es un intervalo abierto de $\mathbb{R}$ y $a \in I$, definimos la derivada de f en el punto a, que designamos indistintamente.

$$f^{'}(a) \text{ o también } \frac{df}{dt}(a),$$

Como el valor del límite

$$\lim_{t \to 0} \frac{f(a+t)-f(a)}{t},$$

Cuando existe y es finito. Supongamos que f es derivable en el punto a, y definamos la función $p(t)$, en un entorno reducido $B(0, \in)$, suficientemente pequeño, mediante

$$p(t) = \frac{f(a+t)-f(a)}{t} - f^{'}(a) \qquad \text{............ (1)}$$

Es claro que $\lim_{t \to 0} p(t) = 0$ y podemos escribir, despejando,

$$f(a+t) - f(a) = f^{'}(a)t + tp(t), \qquad \lim_{t \to 0} p(t) = 0$$

Sea $f : U \to \mathbb{R}$, donde $U \subset \mathbb{R}^n$ es un conjunto abierto. La función f es diferenciable en un punto $a \in U$, cuando existen un vector $A = (A_1, A_2, \dots, A_n)$ y una función $p: \mathrm{B}(0, \epsilon) \to \mathbb{R}$ tales que, para todo $v = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{R}^n$ con $a + v \in \mathrm{B}(a, \epsilon) \subset U$, se verifica

$$f(a+v) - f(a) = \langle A, v \rangle + \|v\| p(v), \qquad \lim_{v \to 0} p(v) = 0 \qquad \text{............ (2)}$$

Decimos que f es diferenciable en U cuando es diferenciable en todo $a \in U$, introduciendo la función $r(v) = \|v\| p(v)$, que llamamos resto, la formula (2) se puede escribir como

$$f(a+v)-f(a)=\langle A,v\rangle+r(v) \qquad , \lim_{v\to 0}\frac{r(v)}{\|v\|}=0 \qquad \text{............} (3)$$

Estamos entonces definiendo que una función es diferenciable en un punto a cuando su incremento se puede aproximar por una transformación lineal de la forma $t(v)=\langle A,v\rangle$.

Definición 6: Una función $f(x,y)$ de dos variables, definida en un conjunto abierto $U\subset\mathbb{R}^2$, es llamada **diferenciable** en un punto $a=(x,y)$ cuando existen dos constantes A,B y una función $r(h,k):\ \mathrm{B}(0,\epsilon)\to\mathbb{R}$, tales que si, $v=(h,k), a+v\in\ \mathrm{B}(a,\epsilon)\subset U$, es decir

$$f(x+h,y+k)-f(x,y)=Ah+Bk+r(h,k), \qquad \lim_{v\to 0}\frac{r(h,k)}{\sqrt{h^2+r^2}}=0$$

Observemos finalmente, que en la definición utilizamos la norma euclidiana usual. Despejando $r(v)$ en (3) , obtenemos que f es diferenciable en $a\in U$ si y sólo si existe un vector A tal que

$$\lim_{v\to 0}\frac{1}{\|v\|}(f(a+v)-f(a)-\langle A,v\rangle)=0$$

Como este límite no depende de la norma en $\mathbb{R}^n$ (porque todas las normas son equivalentes), la definición de diferenciabilidad no depende de la norma elegida. A continuación, el resultado que estábamos buscando.

Teorema 1. Sea $f:\ U\to\mathbb{R}$, donde $U\subset\mathbb{R}^n$ es un conjunto abierto, y $a\in U$. Si f es diferenciable en el punto a, entonces es continua en a, y, dado un vector v cualquiera, existe la derivada direccional $(\partial f/\partial v)(a)$, que verifica

$$\frac{\partial f}{\partial v}(a)=\langle A,v\rangle \text{ .... } (4)$$

Donde $A=(A_1,A_2,\dots,A_n)$ es el vector de la definición de diferenciabilidad. En particular, existen las derivadas parciales, y se verifica

$$\frac{\partial f}{\partial x_1}(a)=A_1,\dots,\frac{\partial f}{\partial x_n}(a)=A_n$$

Demostración, La continuidad es inmediata, dado que

$$\lim_{v\to 0}\big(f(a+v)-f(a)\big)=\lim_{v\to 0}\big(\langle A,v\rangle+\|v\|p(v)\big)=0$$

Respecto de la derivada direccional, dado un vector v, si ponemos tv en (3), tenemos

$$f(a+tv)-f(a)=t\langle A,v\rangle+|t|\|v\|p(tv),$$

Dividiendo por t y tomando limite, obtenemos

$$\lim_{t\to 0}\frac{f(a+tv)-f(a)}{t}=t\langle A,v\rangle+\lim_{t\to 0}\frac{|t|}{t}\|v\|p(tv)=\langle A,v\rangle,$$

Porque $p(tv)\to 0\ (t\to 0)$, lo que demuestra (5). En particular, si $v=e_i$, obtenemos

$$\partial_i f(a)=\frac{\partial f}{\partial x_i}(a)=\langle A,e_i\rangle=A_i,\text{ concluyendo la demostración } \blacksquare$$

Observación 2. La fórmula (4) muestra que la derivada parcial $(\partial f/\partial v)(a)$ es una función lineal de v, cuando la función es diferenciable en a.

Estudiemos ejemplos de diferenciabilidad de algunas funciones sencillas a partir de la definición.

Ejemplo 8. Estudiar la continuidad y la diferenciabilidad en el origen de la siguiente función y, en su caso, hallar su diferencial en ese punto.

$$f(x,y)=\begin{cases}\dfrac{xy}{\sqrt{x^2+y^2}} & si\ (x,y)\neq(0,0)\\ 0\quad si\quad (x,y)=(0,0)\end{cases}$$

Solución

$$\lim_{(x,y)\to(0,0)}\frac{xy}{\sqrt{x^2+y^2}}=\lim_{(x,y)\to(0,0)}x.\frac{y}{\sqrt{x^2+y^2}}=0.acot=0=f(0,0)$$

Luego la función es continua (0,0).

Calculamos las derivadas parciales aplicando la definición:

$$\frac{\partial f}{\partial x}(0{,}0)=\lim_{h\to 0}\frac{f(0+h,0)-f(0{,}0)}{h}=\lim_{h\to 0}\frac{f(h,0)-0}{h}=\lim_{h\to 0}\frac{\frac{h.0}{\sqrt{h^2+0^2}}}{h}=0$$

$$\frac{\partial f}{\partial y}(0{,}0)=\lim_{h\to 0}\frac{f(0{,}0+k)-f(0{,}0)}{k}=\lim_{h\to 0}\frac{f(0,k)-0}{k}=\lim_{h\to 0}\frac{\frac{0.k}{\sqrt{0^2+k^2}}}{k}=0$$

Luego, $\lambda(h,k)=0.h+0.k=0$

Por otro lado, $\Delta f(0{,}0)=f(0+h,0+k)-f(0{,}0)=f(h,k)-0=\frac{hk}{\sqrt{h^2+k^2}}$

Luego, $\lim\limits_{(h,k)\to(0,0)}\frac{r(h,k)}{\|(h,k)\|}=\lim\limits_{(h,k)\to(0,0)}\frac{\Delta f-\lambda(h,k)}{\sqrt{h^2+k^2}}=\lim\limits_{(h,k)\to(0,0)}\frac{\frac{h.k}{\sqrt{h^2+k^2}}-0}{\sqrt{h^2+k^2}}$

$$=\lim_{(h,k)\to(0,0)}\frac{hk}{h^2+k^2}=\lim_{\substack{k=mh\\ h\to 0}}\frac{hmh}{h^2+m^2h^2}=\lim_{\substack{k=mh\\ h\to 0}}\frac{h^2m}{h^2(1+m^2)}=\frac{m}{1+m^2}$$

Luego el limite no existe, por depender de m, y en consecuencia la función no es diferenciable en (0, 0). Al no ser diferenciable resulta que $df(0{,}0)$ no existe, con lo cual $\lambda(h,k)=0h+0k=0$ no tiene ninguna significación.■

Ejemplo 9. Estudiar la continuidad y la diferenciabilidad en el origen de la siguiente función y, en su caso, hallar su diferencial en ese punto.

$$f(x,y)=\sqrt{x^2+y^2}$$

Solución

La función es continua en (0,0), en efecto:

$$\lim_{(x,y)\to(0,0)}\sqrt{x^2+y^2}=\sqrt{0^2+0^2}=\sqrt{0}=0$$

Y dicho límite no existe, puesto que $\lim\limits_{h\to 0^+}\frac{|h|}{h}=1$ y $\lim\limits_{h\to 0^-}\frac{|h|}{h}=-1$

Luego, la función no es diferenciable en el origen por no existir las derivadas parciales en dicho punto y ser la existencia de las derivadas parciales en un punto p una condición necesaria para la diferenciabilidad de la función en dicho punto. En consecuencia resulta que $df(0,0)$ no existe. ■

EL GRADIENTE DE UNA FUNCIÓN DE VARIAS VARIABLES

Definición 7: El vector gradiente de la función $f(x_1, x_2, x_3, \ldots., x_n)$ en el punto $(a_1, a_2, \ldots, a_n)$ se define como el vector que tiene por componentes las derivadas parciales de f en este punto. El vector gradiente se denota por $\nabla f(a_1, a_2, \ldots, a_n)$. Si se supone que se evalúa en el punto $(a_1, a_2, \ldots, a_n)$, tenemos:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \ldots.. \frac{\partial f}{\partial x_n}\right)$$

Es importante remarcar el hecho de que el gradiente es un vector.

Ejemplo 10; Dada la función $f(x, y) = x^2 + y^2 + 5xy + 2x - 3y - 20$, determinar el gradiente $\nabla f(x, y)$

Solución, de acuerdo a la definición del vector gradiente se tiene que:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x + 5y + 2{,}2y + 5x - 3)$$

Curvas de Nivel y el vector Gradiente

Proposición 1: El gradiente de una función f en un punto en particular es perpendicular a la curva de nivel de f en el mismo punto.

Prueba:

Consideremos que X_0 es el punto de interés. La curva de nivel que pasa a través de x_0 es una $x|f(X) = f(X_0)$. Consideremos la curva paramétrica $\gamma(t)$ es la curva de nivel que pasa a través de X, por lo tanto podemos asumir que $\gamma(0) = X = [x_1, x_2, x_3, \ldots, x_n]^T$. Entonces tenemos

$$f(\gamma(0)) = f(X_0) = C$$

Ahora vamos a calcular la derivada de la función con respecto al parámetro t utilizando la regla de la cadena

$$\frac{df\big(x_1(t), x_2(t), \dots, x_n(t)\big)}{dt} = \frac{\partial f(X)}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f(X)}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f(X)}{\partial x_n}\frac{dx_n}{dt} = 0$$

$$= \left[\frac{\partial f(X)}{\partial x_1}, \frac{\partial f(X)}{\partial x_2}, \dots, \frac{\partial f(X)}{\partial x_n}\right]\begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \cdot \\ \cdot \\ \cdot \\ \frac{dx_n}{dt} \end{bmatrix} = 0$$

$$= [\nabla f(X_0)]^T \gamma(X_0) = 0$$

De la formula anterior podemos ver que $f'\big(X(t)\big)$ es el producto escalar del gradiente y la derivada de la curva de nivel. El producto escalar de vectores, que también puede ser escrito como $\|\nabla f(x_0)\|\,\|y'(0)\|\cos\theta$ y la única posibilidad de que este sea cero es que el ángulo entre ambos vectores sea 90^o. Por lo tanto la el ángulo entre la tangente a la curva de nivel $\gamma'(X_0)$ forma un ángulo de 90^o con la dirección del vector gradiente. En la figura 2.1 podemos ver las curvas de nivel en un punto y la dirección del menos gradiente.■

Propiedades del gradiente:

- Si $\nabla f(x_1, x_2, \dots, x_n) = 0$ entonces $D_u f(x_1, x_2, \dots, x_n) = 0$ para todo u.
- La dirección de máximo crecimiento de f viene dada por $\nabla f(x_1, x_2, \dots, x_n)$, el valor máximo de $D_u f(x_1, x_2, \dots, x_n)$ es $\|\nabla f(x_1, x_2, \dots, x_n)\|$.
- Sea $u = f(x_1, x_2, \dots, x_n)$ una función diferenciable en el punto $((x_1)_0, (x_2)_0, \dots, (x_n)_0)$ y $\nabla f((x_1)_0, (x_2)_0, \dots, (x_n)_0) \neq 0$ entonces $\nabla f(x_1, x_2, \dots, x_n)$ es normal a la superficie que pasa por $((x_1)_0, (x_2)_0, \dots, (x_n)_0)$

El Hessiano de una función de varias variables

Definición 8: Si la función $z = f(x,y)$ tiene derivadas parciales de segundo orden en un conjunto $D \subset \mathbb{R}^2$ se denomina Hessiano de f en el punto $(x,y)\epsilon D$ al determinante:

$$H(x,y) = \begin{vmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{vmatrix}$$

CONDICIONES SUFICIENTES PARA LA EXISTENCIA DE EXTREMO.

Sea f una función con derivadas parciales primeras y segundas continuas en un conjunto abierto que contiene un punto (a,b) para el que $f_x(a,b) = 0$ y $f_y(a,b) = 0$. Si $H(a,b)$ representa el Hessiano de f en (a,b) entonces se verifica:

1. Si $H(a,b) > 0$ y $f_{xx}(a,b) > 0$ entonces f alcanza en el punto (a,b) un mínimo relativo.

2. Si $H(a,b) > 0$ y $f_{xx}(a,b) < 0$ entonces f alcanza en el punto (a,b) un máximo relativo.

3. Si $H(a,b) < 0$ entonces la función tiene un punto de silla en (a,b) .

4. Si $H(a,b) = 0$ entonces no se tiene información por esta vía.

Si $H(a,b) > 0$, entonces $f_{xx}(a,b)$ y $f_{yy}(a,b)$ deben tener el mismo signo.
Esto significa que se puede remplazar $f_{xx}(a,b)$ por $f_{yy}(a,b)$ en las dos primeras partes del criterio.

Ejemplo 11: Encontrar los extremos relativos de

$$f(x,y) = x^3 + 4xy - 2y^2 + 1$$

Solución, Comenzamos buscando los puntos críticos de f . Puesto que

$$f_x(x,y) = -3x^2 + 4y, f_y(x,y) = 4x - 4y$$

están definidas para todo x e y, los únicos puntos críticos son aquellos para los cuales ambas derivadas parciales primeras son nulas. Para localizar estos puntos, hacemos $f_x(x,y)$ y $f_y(x,y)$ cero y obtenemos el sistema de ecuaciones siguiente:

$$\left.\begin{aligned} -3x^2 + 4y &= 0 \\ 4x - 4y &= 0 \end{aligned}\right\}$$

De la segunda ecuación vemos que $x = y$, y sustituyendo en la primera obtenemos dos Soluciones:$y = x = 0$ e $y = x = 4/3$. Como

$$f_{xx}(x,y) = -6x, \qquad f_{yy}(x,y) = -4 \quad \text{y} \quad f_{xy}(x,y) = 4$$

se sigue que para el punto crıtico $(0,0)$, $H(0,0) = -16 < 0$ y, por el criterio de las derivadas parciales segundas, concluimos que $(0,0,1)$ es un punto de silla de f.

Para el punto crıtico $(4/3\,,4/3)$, se tiene $H\left(\frac{4}{3},\frac{4}{3}\right) = 16 > 0$ y como

$$f_{xx}\left(\frac{4}{3},\frac{4}{3}\right) = -8 < 0$$

Concluimos que $f(4/3\,,4/3)$ es un máximo relativo. ■

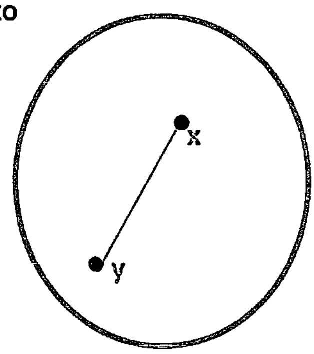
1.4. CONJUNTO CONVEXO

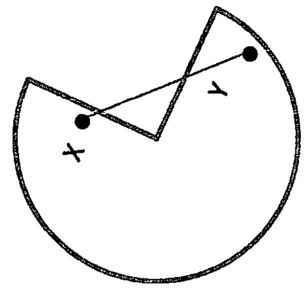
Definición 9: Dado S un subconjunto de $\mathbb{R}^n$, S es llamado Conjunto Convexo si para todo $x, y \in S$ y para todo $\lambda \in [0,1]$; se tiene:

$$z = \lambda x + (1-\lambda)y \ \in S$$

El conjunto de puntos z se llama segmento cerrado y la igualdad anterior se denomina combinación lineal convexa de x y y.

El gráfico de abajo en $\mathbb{R}^2$, se muestra un conjunto convexo y un conjunto no convexo





Convexo **no convexo**

Ejemplo 12: Demostrar que el conjunto $C = \{(x, y) \in \mathbb{R}^2 / \ y \geq x\}$ es convexo.

Demostración

$$C = \{(x, y) \in \mathbb{R}^2 / \ y \geq x\}$$

y sean $x, y \in C$ dos puntos cualesquiera. Tal que

$$x = (x_1, x_2) \in C \, , y = (y_1, y_2) \in C$$

y además sea un $\lambda \in [0,1]$

$$x\lambda + (1 - \lambda)y \ \in C$$

$$(x_1, x_2)\lambda + (1 - \lambda)(y_1, y_2)$$

En otras palabras demostraremos que: $\big(x_1\lambda + (1 - \lambda)y_1, x_2\lambda + y_2(1 - \lambda)\big) \in C$

Pero como

$$x = (x_1, x_2) \in C \ \rightarrow \ x_2 \geq x_1 \ \ \rightarrow x_2\lambda \geq x_1\lambda, \qquad \ldots\ldots. (1)$$

$$y = (y_1, y_2) \in C \ \rightarrow \ y_2 \geq y_1 \ \ \rightarrow (1 - \lambda)y_2 \geq (1 - \lambda)y_1 \quad \ldots\ldots. (2)$$

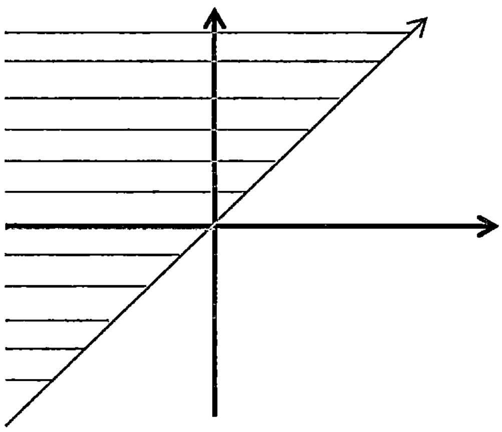
Sumamos (1) y (2):

$$x_2\lambda + (1 - \lambda)y_2 \geq x_1\lambda + (1 - \lambda)y_1$$

$$\rightarrow (x_1\lambda + (1-\lambda)y_1, x_2\lambda + (1-\lambda)y_2) \in C$$

Por lo tanto C es un conjunto convexo y su grafica será:

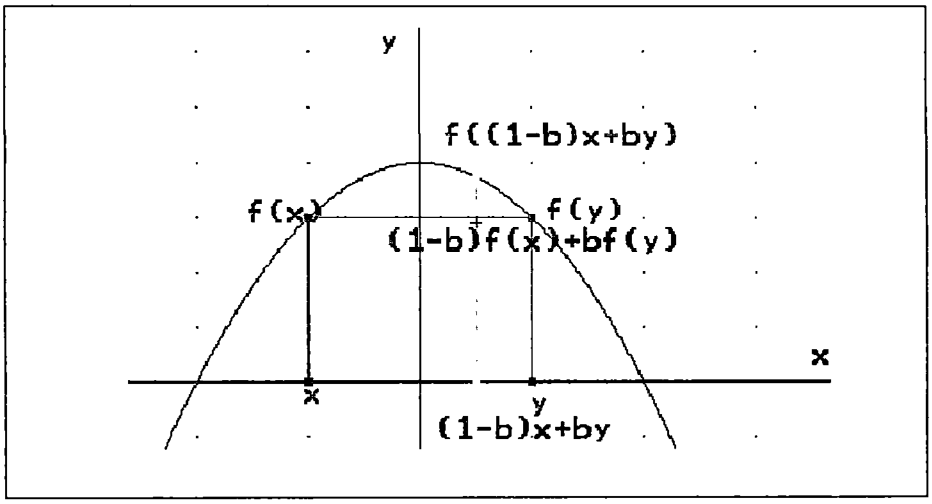
$y \geq x$



FUNCION CONCAVA

Una función $f: S \subset \mathbb{R}^n \rightarrow \mathbb{R}$, se dice cóncava sobre el conjunto cóncavo S si $f\big((1-b)x + by\big) \geq (1-b)f(x) + bf(y) \quad \forall\, x, y \in S$, para cualquier escalar $b \in [0,1]$.

En la figura siguiente se representa gráficamente una función cóncava:

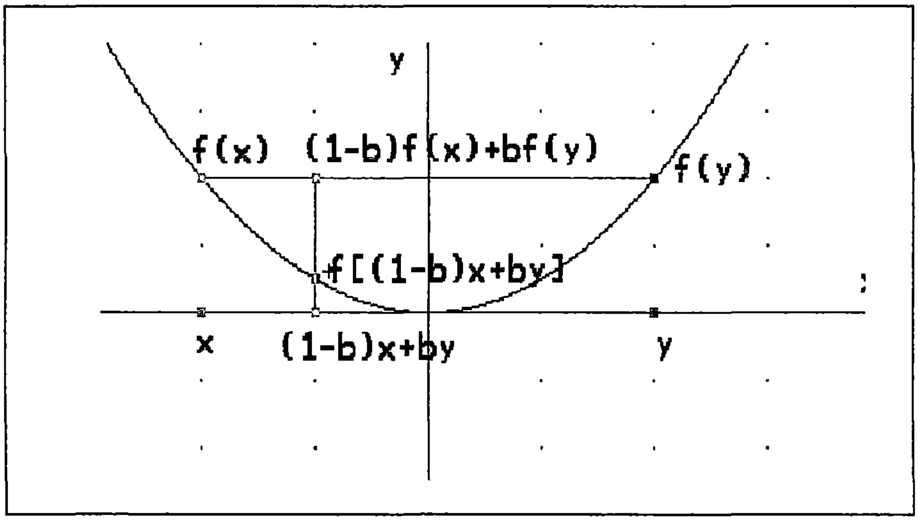


Propiedad 1: si f es concava en S, entonces f es convexa en S y si f es convexa en S, entonces f es cóncava en S.

1.5. FUNCIONES CONVEXAS

Una función $f: S \subset \mathbb{R}^n \to \mathbb{R}$, se dice convexa sobre el conjunto convexo S si $f\big((1-b)x+by\big) \leq (1-b)f(x)+bf(y) \quad \forall\, x,y \in S$, para cualquier escalar $b \in [0,1]$.

En la figura siguiente se representa gráficamente una función convexa:



Ejemplo 13: $f: \mathbb{R} \to \mathbb{R}$

$x \to m.x+b$ funciones afines

Solución

Sea $x,y \in \mathbb{R}$y $\quad \lambda \in [0,1]$

$$\begin{aligned} f(\lambda x+(1-\lambda)y) &= m[\lambda x+(1-\lambda)y]+b[\lambda+(1-\lambda)] \\ &= m(\lambda x)+(1-\lambda)my+b(\lambda)+(1-\lambda)b \\ &= \lambda(mx+b)+(1-\lambda)(my+b) \\ &= \lambda f(x)+(1-\lambda)f(y) \end{aligned}$$

Por lo tanto f es convexo. ■

Teorema 2: Sea S un conjunto convexo en $\mathbb{R}^n$, sea $f: S \to \mathbb{R}$ una función convexa. Luego el conjunto $T=\{x \in S / f(x) \leq \alpha\}$ es un conjunto convexo, donde α es un número real.

Demostración

Consideremos un número α real cualquiera. Como f es convexa en un conjunto S (convexo en $\mathbb{R}^n$), vamos a demostrar que el conjunto T es convexo.

Consideremos dos puntos cualesquiera $x, y \in T$ y sea un $b \in [0,1]$, como dichos puntos son del conjunto, entonces se verifica que: $f(x) \leq \alpha, f(y) \leq \alpha$ pues $x, y \in T$.

Como además f es convexa se verifica que:

$$f[(1-b)x + by] \leq (1-b)f(x) + bf(y) \leq (1-b)\alpha + b\alpha = \alpha$$

Luego se cumple que $\forall\, b \in [0,1]$ y $\forall\, x, y \in T$, $f[(1-b)x + by] \in T$, por lo tanto el conjunto T es convexo. ■

Teorema 3: Sea S un conjunto convexo en $\mathbb{R}^n$, sea $f: S \to \mathbb{R}$ una función cóncava. Luego el conjunto $T = \{x \in S / f(x) \geq \alpha\}$ es un conjunto convexo, donde α es un número real.

Demostración

Consideremos un número α real cualquiera. Como f es convexa en un conjunto S (convexo en $\mathbb{R}^n$), vamos a demostrar que el conjunto T es convexo.

Consideremos dos puntos cualesquiera $x, y \in T$ y sea un $b \in [0,1]$, como dichos puntos son del conjunto, entonces se verifica que: $f(x) \geq \alpha, f(y) \geq \alpha$ pues $x, y \in T$.

Como además f es convexa se verifica que:

$$f[(1-b)x + by] \geq (1-b)f(x) + bf(y) \geq (1-b)\alpha + b\alpha = \alpha$$

Luego se cumple que $\forall\, b \in [0,1]$ y $\forall\, x, y \in T$, $f[(1-b)x + by] \in T$, por lo tanto el conjunto T es convexo. ■

Teorema 4: Toda función lineal $h(x)$ definida en un conjunto convexo S de $\mathbb{R}^n$ es a la vez convexa y cóncava en su dominio.

Demostración

Si $h(x)$ es lineal entonces $h(\alpha x) = \alpha h(x)$; $h(x+y) = h(x) + h(y)$.

Por lo tanto $h[\alpha x+\beta y]=h[\alpha x]+h[\beta y]=\alpha h(x)+\beta h(y)$ y siendo $\alpha+\beta=1$, $\alpha\geq 0$ y $\beta\geq 0$, luego $\beta=1-\alpha$, se tiene:

$h[\alpha x]+h[(1-\alpha)y]=\alpha h(x)+(1-\alpha)h(y)$. ■

El siguiente resultado proporciona una condición necesaria y suficiente para que una función diferenciable sea convexa.

Teorema 5: Sea S un conjunto abierto convexo de R^n y $f: S\to\mathbb{R}$ una función diferenciable en S.Entonces f es convexa si y solo si $f(x)\geq f(y)+(\nabla f(y), x-y)$, para todo $x, y\in S$.

Demostración

Sea f convexa y $x,y\in S$, entonces en particular $\alpha\in[0,1[$, se tiene que

$$f\big(y+\alpha(x-y)\big)\leq f(y)+\alpha(f(x)-f(y)),\text{ luego}$$

$$\frac{f\big(y+\alpha(x-y)\big)-f(y)}{\alpha}\leq f(x)-f(y)$$

Tomando límite cuando $\alpha\to 0$ y por la diferenciabilidad de f tenemos $(\nabla f(y), x-y)\leq f(x)-f(y)$, esto es $f(x)\geq f(y)+(\nabla f(y), x-y)$.

Recíprocamente, supongamos que satisface la desigualdad

$$f(x)\geq f(y)+(\nabla f(y), x-y),\text{ para todo } x,y\in S. \qquad \dots\dots\dots(1)$$

Probaremos que es convexa. Sean $x,y\in S$ y $\alpha\in[0,1]$. Tomando los puntos x y $x+\alpha(y-x)$ en (1) y multiplicando por $(1-\alpha)$ se tiene

$$(1-\alpha)f(x)\geq(1-\alpha)f\big(x+\alpha(y-x)\big)-(1-\alpha)\alpha(\nabla f\big(x+\alpha(y-x)\big), y-x) \qquad \dots\dots\dots(2)$$

Análogamente, tomando ahora los puntos y y $x+\alpha(y-x)$ en (1) y multiplicado por α se tiene:

$$\alpha f(y)\geq\alpha\, f\big(x+\alpha(y-x)\big)+\alpha(1-\alpha)(\nabla f\big(x+\alpha(y-x)\big), y-x) \qquad \dots\dots\dots(3)$$

Sumando (2) y (3), obtenemos:

$$\alpha f(y) + (1-\alpha)f(x) \geq \alpha f\big(x + \alpha(y-x)\big) + (1-\alpha)f\big(x + \alpha(y-x)\big)$$
$$= f\big(x + \alpha(y-x)\big)$$

Por lo tanto $f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha)f(x)$. ■

Teorema 6: Sea S un conjunto abierto convexo de $\mathbb{R}^n$ y $f: S \to \mathbb{R}$ una función dos veces diferenciable en S.Entonces f es convexa si y solo si la matriz Hessiana es semidefinida positiva en cada punto de S.

Demostración

Suponiendo que f es convexa y sea $x_0 \in S$. Se quiere probar que $X^T H(X_0) X \geq 0$ para cada $X \in \mathbb{R}^n$. Puesto que S es abierto, para cualquier $X \in \mathbb{R}^n$, será $X_0 + \alpha X \in S$ para $|\alpha| \neq 0$ suficientemente pequeño. Entonces, por el teorema anterior:

$$f(X_0 + \alpha X) \geq f(X_o) + \alpha[\nabla f(X_o)]^T X \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Y por ser dos veces diferenciable, se puede escribir:

$$f(X_0 + \alpha X) = f(X_o) + \alpha[\nabla f(X_o)]^T X + \frac{1}{2}\alpha^2 X^T H(X_o) X + \alpha^2 \cdot X^2 \cdot w(X_o, \alpha X) \ldots\ldots\ldots (6)$$

Restando (5) y (6), se tiene:

$$\frac{1}{2}\alpha^2 X^T H(X_o) X + \alpha^2 \cdot X^2 \cdot w(X_o, \alpha X) \geq 0$$

Dividiendo por α^2 y haciendo $\alpha \to 0$, resulta: $X^T H(X_o) X \geq 0$

Recíprocamente, suponiendo ahora que la matriz Hessiana es semidefinida positiva en cada punto de S. Sean X y $X_0 \in S$, la formula de Taylor de segundo orden proporciona:

$$f(X) = f(X_0) + [\nabla f(X_0)]^T (X - X_0) + \frac{1}{2}(X - X_0)^T H(X^*)(X - X_0) \ldots\ldots\ldots (7)$$

Siendo $X^* = \alpha X_0 + (1-\alpha)X$ y $\alpha \in (0,1)$. En consecuencia, $X^* \in S$ y entonces por hipótesis $H(X^*)$ es semidefinida positiva.

Por lo tanto: $(X - X_0)^T H(X^*)(X - X_0) \geq 0$

Concluyendo que de (7) resulta: $f(X) \geq f(X_0) + [\nabla f(X_0)]^T (X - X_0)$

Ya que la anterior desigualdad es válida para cada X y $X_0 \in S$, f es convexa por el teorema 5. ■

Este teorema es útil para probar la convexidad o concavidad de una función dos veces diferenciable. En especial, si la función es cuadrática, la matriz Hessiana es independiente del punto considerado.

1.6. HIPERPLANOS

En geometría, un **hiperplano** es una extensión del concepto de plano.

En un espacio unidimensional (como una recta), un hiperplano es un punto; divide una línea en dos líneas.

En un espacio bidimensional (como el plano *XY*), un hiperplano es una recta; divide el plano en dos mitades.

En un espacio tridimensional, un hiperplano es un plano corriente; divide el espacio en dos mitades.

Este concepto también puede ser aplicado a espacios de cuatro dimensiones y más, donde estos objetos divisores se llaman simplemente hiperplanos, ya que la finalidad de esta nomenclatura es la de relacionar la geometría con el plano.

Definición 10: Siendo $A = (a_1, a_2, \dots, a_n)$, $X = (x_1, x_2, \dots, x_n)^T$ y b un número real, se llama hiperplano en $\mathbb{R}^n$ al conjunto de puntos que satisfacen la ecuación lineal $AX = b$. Todo hiperplano divide los puntos $\mathbb{R}^n$ en dos semiespacios: $AX \leq b$ y $AX \geq b$.

En general, un hiperplano es un espacio afín de codimensión 1. Con respecto a este concepto se tienen los siguientes hechos.

a.) Siendo b un número real, se define como hiperplano en $\mathbb{R}^n$ al conjunto de puntos que satisfacen la ecuación lineal: $a_1x_1 + a_2x_{2+\dots+}a_nx_n = b$

Todo hiperplano divide los puntos del espacio R^n en dos semiespacios:

$$a_1x_1 + a_2x_{2+\dots+}a_nx_n \leq b$$

$$a_1x_1 + a_2x_{2+\dots+}a_nx_n \geq b$$

b.) **<u>Teorema</u> 7:** Todo semiespacio de $\mathbb{R}^n$ es un conjunto convexo.

Demostración

En efecto sean X_1 y X_2 dos puntos pertenecientes al semiespacio $AX \geq b$. Indiquemos con $X = \alpha X_1 + (1-\alpha)X_2$ un punto arbitrario del segmento definido por X_1 y X_2.Entonces:

$$AX = A\,[\alpha X_1 + (1-\alpha)X_2] = \alpha AX_1 + (1-\alpha)AX_2 \geq \alpha b + (1-\alpha)b = b$$

Por lo tanto, también X pertenece al semiespacio considerado, luego $AX \geq b$ es un conjunto convexo.

Análogamente se demuestra que el semiespacio $AX \leq b$ es un conjunto convexo.

En efecto sean X_1 y X_2 dos puntos pertenecientes al semiespacio $AX \leq b$. Indiquemos con $X = \alpha X_1 + (1-\alpha)X_2$ un punto arbitrario del segmento definido por X_1 y X_2.Entonces:

$$AX = A\,[\alpha X_1 + (1-\alpha)X_2] = \alpha AX_1 + (1-\alpha)AX_2 \leq \alpha b + (1-\alpha)b = b$$

Por lo tanto, también X pertenece al semiespacio considerado, luego $AX \leq b$ es un conjunto convexo. ■

Afirmación: La intersección de dos conjuntos convexos es convexa.

Demostración

Sean A y B dos conjuntos convexos,

Dados $x, y \in A \cap B$, y sea $\alpha \in \,]0,1[$, luego

$\alpha x + (1-\alpha)y \in A$, puesto que A es convexo, además también

$\alpha x + (1-\alpha)y \in B$, puesto que B es convexo

De donde se tiene que $\alpha x + (1-\alpha)y \in A \cap B$

Por lo tanto $A \cap B$ es convexo■

Corolario 1: El hiperplano *AX = b* es un conjunto convexo.

Demostración

Por la afirmación anterior se sabe que la intersección de dos conjuntos convexos es convexo, y como se tiene que:

$\{AX = b\} = \{AX \leq b\} \cap \{AX \geq b\}$

Usando teorema 1 se concluye la demostración. ■

1.7. FORMAS CUADRÁTICAS

Definición 11: Sea $A_{n\times n}$ una matriz simétrica. Definimos la forma cuadrática asociada a A, Q_A del siguiente modo:

$$Q_A: \mathbb{R}^n \to \mathbb{R}$$

$$X \to X.A.X^T$$

$$Q_A(x) = [x_1, x_2, x_3, \dots, x_n]\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Ejemplo 14: sea $A = \begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \\ 5 & 6 & 2 \end{pmatrix}$

$$Q_A(x) = [x_1, x_2, x_3]\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \\ 5 & 6 & 2 \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= [x_1, x_2, x_3]\begin{pmatrix} x_1 + 3x_2 + 5x_3 \\ 3x_1 + 4x_2 + 6x_3 \\ 5x_1 + 6x_2 + 2x_3 \end{pmatrix}$$

$$= x_1(x_1 + 3x_2 + 5x_3) + x_2(3x_1 + 4x_2 + 6x_3) + x_3(5x_1 + 6x_2 + 2x_3)$$

En forma general:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$Q_A(x) = a_{11}{x_1}^2 + a_{22}{x_2}^2 + a_{33}{x_3}^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3$$

Observación 3: Sea $f: \mathbb{R}^n \to \mathbb{R}$ con primeras y segundas derivadas parciales continuas, entonces el hessiano $Hf(z)$ es una matriz simétrica. Si x y x^* en $\mathbb{R}^n$, tendríamos que si $H = Hf(z)$ la forma cuadrática asociada a H evaluada en $(x - x^*)$ es:

$$Q_H(x - x^*) = (x - x^*)H(x - x^*)$$

$$= (x - x^*)Hf(z)(x - x^*)$$

Definición 12: supóngase que $A_{n\times n}$ es una matriz simétrica y que $Q_A(y) = y.A.y$ es la forma cuadrática asociada a A. entonces A y Q_A son llamadas:

a.) Definida positiva si: $Q_A(y) = y.A.y^T > 0, \forall\, y \in \mathbb{R}^n, y \neq 0$

b.) Semidefinida positiva si: $Q_A(y) = y.A.y^T \geq 0, \forall\, y \in \mathbb{R}^n$

c.) Definida negativa si: $Q_A(y) = y.A.y^T < 0, \forall\, y \in \mathbb{R}^n, y \neq 0$

d.) Semidefinida negativa si: $Q_A(y) = y.A.y^T \leq 0, \forall\, y \in \mathbb{R}^n$

e.) Indefinida si $\exists\, y \in \mathbb{R}^n$ tal que: $Q_A(y) = y.A.y^T > 0$ y si $\exists\, z \in \mathbb{R}^n$ tal que: $Q_A(z) = z.A.z^T < 0$.

1.8. MATRICES DEFINIDAS Y SEMIDEFINIDAS

Observación 4:

- Una matriz simétrica cuyas entradas son todas positivas no es necesariamente una matriz definida positiva.

Ejemplo 15: $A = \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}$

$Q_A(x) = {x_1}^2 + {x_2}^2 + 8x_1x_2$, si $x = (-1,1)$

$Q_A(x) = 1 + 1 - 8 < 0$

Por lo tanto A no es definida positiva

- Una matriz simétrica con algunas componentes negativas puede ser definida positiva.

Ejemplo 16: $A = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$

$$\begin{aligned} Q_A(x) &= {x_1}^2 + {x_2}^2 - 2x_1x_2 \\ &= {x_1}^2 - 2x_1x_2 + {x_2}^2 + 3{x_2}^2 \\ &= (x_1 - x_2)^2 + 3{x_2}^2 \end{aligned}$$

Sea $x = (x_1, x_2) \neq (0,0)$.Puede ocurrir que $x_1 = x_2$ o $x_1 \neq x_2$. En el caso en que $x_1 = x_2$ entonces $Q_A(x) = 3{x_2}^2 > 0$ porque $x \neq 0$.

En el caso en que $x_1 \neq x_2 \rightarrow Q_A(x) = (x_1 - x_2)^2 + \underbrace{3x_2}_{\geq 0}$

- La matriz diagonal

$$A = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

$$Q_A(x) = d_1{x_1}^2 + d_2{x_2}^2 + d_3{x_3}^2$$

a.) Si d_1, d_2, d_3 son positivas entonces A es definida positiva porque si $x = [x_1 \quad x_2 \quad x_3] \neq 0$ entonces por lo menos un $x_1 \neq 0$.

$$Q_A(x) = \underbrace{d_1{x_1}^2}_{\geq 0} + \underbrace{d_2{x_2}^2}_{\geq 0} + \underbrace{d_3{x_3}^2}_{\geq 0} \geq 0$$

b.) Si $d_1, d_2, d_3 \geq 0$, entonces A es semidefinida positiva. Supongamos que $d_1 > 0, d_2 = 0, d_3 > 0$.

$$Q_A(x) = d_1{x_1}^2 + d_2{x_2}^2 + d_3{x_3}^2 \geq 0, \forall x \in \mathbb{R}$$

Podemos encontrar un $x \neq 0$ tal que: $Q_A(x) = 0$. Sea $x = (0,1,0) \neq 0$.

c.) Si $d_i < 0$ para $i = 1,2,3 \rightarrow$ A es definido negativo.

$$A = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

Sea $x = (x_1, x_2, x_3) \neq 0$, entonces, por lo menos un $x_i = 0$.

$$Q_A(x) = \underbrace{\underbrace{d_1 x_1{}^2}_{+} + \underbrace{d_2 x_2{}^2}_{\leq 0} + \underbrace{d_3 x_3{}^2}_{\leq 0}}_{<0}$$

d.) Si $d_i \leq 0$ para $i = 1,2,3 \rightarrow$ A es semidefinida negativa. Sea $x = (x_1, x_2, x_3) \in R^3$ entonces:

$$Q_A(x) = \underbrace{\underbrace{d_1 x_1{}^2}_{\leq 0} + \underbrace{d_2 x_2{}^2}_{\leq 0} + \underbrace{d_3 x_3{}^2}_{\leq 0}}_{<0}$$

e.) Si al menos un d_i es positivo y al menos un d_i es negativo, entonces A es indefinida.

Supongamos que $d_1 > 0$ y $d_3 < 0$

$$Q_A(x) = \underbrace{d_1}_{+} x_1{}^2 + d_2 x_2{}^2 + \underbrace{d_3}_{-} x_3{}^2$$

Sea $\quad y = (6,0,0); \quad Q_A(y) = 36 d_1 > 0$

Sea $\quad z = (0,0,4); \quad Q_A(z) = 16 d_3 < 0$

- Si una matriz $A_{2\times2}$ simétrica es definida positiva, entonces los elementos de la diagonal son también positivos. Ósea $a_{11} > 0$ y $a_{22} > 0$.

$$A = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

$$Q_A(x) = a_{11} x_1{}^2 + a_{22} x_2{}^2 + 2a_{12} x_1 x_2 > 0, \forall x = (x_1, x_2) \neq (0,0)$$

Si $x = (1,0) \rightarrow$ para este $x, 0 < Q(1,0) = a_{11}$

Si $y = (0,1) \rightarrow$ para este $y, 0 < Q(0,1) = a_{22}$

- En forma general si $A_{n\times n}$ es una matriz simétrica entonces: $a_{ii} > 0, \forall\, i = 1, \dots, n$

Lo recíproco no necesariamente se cumple porque podemos encontrar matrices simétricas con todos los elementos de la diagonal positivos, que no son definidas positivas, como por ejemplo:

$$A = \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}; \quad a_{11} = a_{22} > 0$$

Y para cada $x = (-1,1)$ tenemos:

$$Q_A(x) = (-1)^2 + (1)^2 + 2(4)(-1)(1) < 0$$

Luego Q_A no es definida positiva a pesar de que los elementos de la diagonal de A son positivos.

Teorema 8: Una matriz $A_{2\times 2}$ simétrica:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{es}$$

(a) Definida positiva $\leftrightarrow a_{11} > 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$

(b) Definida negativa$\leftrightarrow a_{11} < 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$

<u>**Demostración**</u>

$\Leftarrow$ Supongamos que $a_{11} > 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$ demostrar que A es Definida positiva

Sea $x = (x_1, x_2) \neq (0,0)$ se tiene $(x_1, 0)\ y\ \ (x_1, x_2)$ de donde $x_1 \neq 0$ y $x_2 \neq 0$

Caso I: si $x = (x_1, 0)$ donde $x_1 \neq 0$

$Q_A(x) = a_{11}{x_1}^2 + a_{22}{x_2}^2 + 2a_{12}x_1x_2$, para el x escogido anteriormente se tiene

$$Q_A(x) = a_{11}{x_1}^2 > 0, \text{ para } x = (x_1, 0) \text{ donde } x_1 \neq 0$$

Caso II: si $x = (x_1, x_2)$ donde $x_2 \neq 0$, podemos expresar x como : $x = (tx_2, x_2)$ donde $t \in \mathbb{R}$.

$$Q_A(x) = a_{11}{x_1}^2 + a_{22}{x_2}^2 + 2a_{12}x_1x_2$$

⇒Supongamos que A es Definida positiva, demostrar que

$a_{11} > 0,$ $\qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$

Por hipótesis $Q_A(x) > 0, \ \forall x \in \mathbb{R}^2, x \neq 0$

En particular para $x = (x_1, x_2) \neq (0,0)$ con $x_2 \neq 0$, se tiene que

$Q_A(x) = {x_2}^2\sigma(t) > 0$

$$\therefore \sigma(t) > 0, \forall t \in \mathbb{R}$$

$a_{11}t^2 + 2a_{12}t + a_{22} > 0, \quad \forall t \in \mathbb{R}$, entonces

- Discriminante $< 0 \rightarrow 4{a_{12}}^2 - 4a_{11}a_{22} < 0 \rightarrow -4(a_{11}a_{22} - {a_{12}}^2) < 0$
- $a_{11} > 0$

$$Q_A(x) = a_{11}(tx_2)^2 + a_{22}{x_2}^2 + 2a_{12}(tx_2)x_2$$

$$= a_{11}t^2{x_2}^2 + a_{22}{x_2}^2 + 2a_{12}t{x_2}^2$$

$$= {x_2}^2[a_{11}t^2 + a_{22} + 2a_{12}(t{x_2}^2)]$$

$$= {x_2}^2[a_{11}t^2 + 2a_{12}t + a_{22}]$$

$$Q_A(x) = {x_2}^2\sigma(t) \ldots\ldots\ldots.(1)$$

Si demostramos que $\sigma(t) > 0 \Rightarrow$ por (1) se tendría que $Q_A(x) > 0 \ \forall x = (x_1, x_2)$ con $x_2 \neq 0$.

$$\sigma(t) = a_{11}t^2 + 2a_{12}t + a_{22}$$

$$\sigma(t)' = 2a_{12}t + 2a_{21}$$

Punto crítico: $2a_{11}t + 2a_{12} = 0 \rightarrow t = \frac{-a_{12}}{a_{11}}$, $\sigma(t)'' = 2a_{11} > 0$.......... Por hipótesis

$\therefore$ σ tiene un maximizador global estricto en $t^* = \frac{-a_{12}}{a_{11}}$, $\sigma(t) \geq \sigma(t^*)$, $\forall t \in \mathbb{R}$.

$$\sigma(t^*) = a_{11}\left(\frac{-a_{12}}{a_{11}}\right)^2 + 2a_{12}\left(\frac{-a_{12}}{a_{11}}\right) + a_{12}$$

$$= \frac{{a_{12}}^2}{a_{11}} + 2\frac{{a_{12}}^2}{a_{11}} + a_{22}$$

$$= \frac{-{a_{12}}^2}{a_{11}} + a_{22}$$

$$= \frac{-{a_{12}}^2 + a_{11}a_{22}}{a_{11}} = \frac{|A|}{a_{11}} > 0, \text{ por hipótesis}$$

De esto y se tiene $\sigma(t) \geq \sigma(t^*) > 0,\ \forall t \in \mathbb{R}$.

$\therefore$ De (1) se tiene que para $x = (x_1, x_2)$ con $x_2 \neq 0$ se cumple:

$Q_A(x) = {x_2}^2\,\sigma(t) > 0$ ■

Teorema 9: Si A es una matriz simétrica $n \times n$ y si A_k es el k-esimo menor principal de A para $1 \leq k \leq n$, entonces:

(a) A es definida positiva $\leftrightarrow$ $A_k > 0$ para $k = 1, \dots, n$

(b) A es definida negativa $\leftrightarrow$ $(-1)^k A_k > 0$ para $k = 1, \dots, n$

Esto es los menores principales alternan en signo, comenzando con $A_1 < 0$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix}$$

$$A_1 = a_{11},\ A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\ A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \text{......}, A_n = A$$

Observación 5:

- Si A_{3x2} es simétrica tal que $A_1 > 0,\ A_2 > 0, A_3 = 0$, entonces A es semidefinida positiva.
- En forma general si $A_{n\times n}$ es simétrica tal que $A_1 > 0,\ A_2 > 0, \dots \dots \dots \dots, A_{n-1} > 0$ y $A_n = 0$ entonces A es semidefinida positiva.
- En forma análoga. Si $A_{n\times n}$ es simétrica tal que $(-1)^k, A_k > 0$ para $k = 1,.., n-1$ y $A_n = 0$ entonces A es semidefinida negativa.
- No es verdadero que si $A_{n\times n}$ es simétrica entonces A es semidefinida positiva $\leftrightarrow$ $A_k \geq 0, \forall\ k = 1, \dots., n$ veremos esto con un ejemplo. Sea

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1/2 \end{pmatrix} \qquad A_1 = 1, A_2 = 0,\ \ A_3 = 0$$

Si tomamos $x = (1,1,-2)$ entonces

$$Q_A(x) = XAX^T = (1,1,-2)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

$$= (1,1,-2)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -2$$

Por tanto A no es semidefinida positiva.

- Si $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ es una matriz simétrica para la cual $A_2 = |A| < 0$ entonces A es indefinida $Q_A(x) = a_{11}{x_1}^2 + a_{22}{x_2}^2 + 2a_{12}x_1x_2$

$$|A| = a_{11}a_{22} - {a_{12}}^2 < 0$$

Caso I $\quad a_{11} = a_{22} = 0 + a_{12}$

Caso II Al menos uno de los términos $a_{11}\ o\ a_{22}$ es diferente de cero.

<u>Supongamos que se tiene el caso I</u>, entonces

$Q_A(x) = 2a_{12}x_1x_2$, si $x = (1,1)$ $\ Q_A(1,1) = 2a_{12}$

$$si\ y = (-1,1)\ \ Q_A(-1,1) = -2a_{12}$$

Por tanto para este caso se tiene que A es indefinida.

<u>Supongamos que se tiene el caso II, $y\ que\ a_{11} \neq 0$</u>

$$Q_A(x) = a_{11}{x_1}^2 + a_{22}{x_2}^2 + a_{12}x_1x_2$$

$$Q_A(x) = a_{11}\left[\left({x_1}^2 + \frac{2a_{12}}{a_{11}}x_1x_2 + \frac{{a_{12}}^2{x_2}^2}{{a_{11}}^2}\right) + \left(\frac{a_{22}{x_2}^2}{a_{11}} - \frac{{a_{12}}^2{x_2}^2}{a_{11}}\right)\right]$$

$$= a_{11}\left[\left({x_1}^2 + \frac{2a_{12}x_2}{a_{11}}\right)^2 + \left(\frac{a_{11}a_{22} - {a_{12}}^2}{{a_{11}}^2}\right){x_2}^2\right]$$

$$= \frac{1}{a_{11}}[(a_{11}x_1 + a_{12}x_2)^2 + A_2{x_2}^2]$$

Sea $x = (a_{12} - a_{11})$

$$Q_A(x) = \frac{1}{a_{11}}[(a_{11}a_{12} - a_{12}a_{11})^2 + A_2(-a_{11})^2]$$

$Q_A(x)$ tiene signo opuesto a a_{11}

Sea $y = (1,0)$

$$Q_A(y) = \frac{1}{a_{11}}[{a_{11}}^2 + A_2(0)] = a_{11}$$

Por tanto A es definida ■

Ejemplo 22. Minimice la función

$$f(x_1, x_2, x_3) = {x_1}^2 + {x_2}^2 + {x_3}^2 - x_1x_2 + x_2x_3 - x_1x_3$$

En efecto, tenemos que

$$\nabla f(x_1, x_2, x_3) = (2x_1 - x_2 - x_3, 2x_2 - x_1 + x_3, 2x_3 + x_2 - x_1) = (0,0,0)$$

de donde tenemos

$$\begin{aligned} 2x_1 - x_2 - x_3 &= 0 \\ 2x_2 - x_1 + x_3 &= 0 \\ 2x_3 + x_2 - x_1 &= 0 \end{aligned}$$

Donde $A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$ $Ax = 0, |A| = 4 \neq 0$, entonces el sistema anterior

tiene solución única en $x_1{}^+ = x_2{}^+ = x_3{}^+ = 0$, con el punto critico $(0,0,0)$.

$$Hf(x_1, x_2, x_3) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

Donde $\Delta_1 = 2$, $\Delta_2 = 3$ y $\Delta_3 = 4$. Por lo que $Hf(x_1, x_2, x_3)$ es definido positivo para todo x en R^3. Por tanto el punto critico $(0,0,0)$ es un minimizador global estricto de f.

Ejemplo 23. Encontrar el minimizador global de

$$f(x_1, x_2, x_3) = e^{x-3} + e^{y-x} + e^{x^2} + z^2$$

En efecto, hallamos los puntos críticos

$\nabla f(x, y, z) = \left(e^{x-y} - e^{y-x} + 2xe^{x^2}, -e^{x-y} + e^{y-x}, 2z\right) = (0,0,0)$, de donde tenemos.

$$e^{x-y} - e^{y-x} + 2xe^{x^2} = 0$$

$$-e^{x-y} + e^{y-x} = 0$$

$$2z = 0$$

Sumando las dos primeras igualdades tenemos, $2xe^{x^2}$, entonces $x^* = 0$ y de la ultima igualdad tenemos que $z^* = 0$, luego de la segunda igualdad tenemos que $y^* = 0$, por lo que el punto critico es (0; 0; 0), luego

$$Hf(x, y, z) = \begin{pmatrix} e^{x-y} + e^{y-x} + 2xe^{x^2} + 4x^2 & -e^{x-y} - e^{y-x} & 0 \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

De donde se tiene $\Delta_1 = e^{x-y} + e^{y-x} + 2xe^{x^2} + 4x^2e^{x^2} > 0$, $\Delta_2 = \left(2e^{x^2} + 4x^2e^{x^2}\right)\left(e^{x-y} + e^{y-x}\right) > 0$ y $\Delta_3 = 2\Delta_2 > 0$, entonces $Hf(x, y, z)$ es definido positivo para todo $(x, y, z) \in \mathbb{R}^3$. Por tanto el punto critico $(0,0,0)$ es un minimizador global de f. ■

CAPÍTULO 2

OPTIMIZACIÓN NO LINEAL CON RESTRICCIONES NO NEGATIVAS

El problema general de programación no lineal restringido, puede definirse como:

Optimizar $z = f(\boldsymbol{x})$	siendo: f la función objetivo
Sujeto a: $\mathrm{g}_i(\boldsymbol{x}) \leq 0 \quad i: 1,2, \ldots\ldots, m$	g_i las restricciones de desigualdad
$\mathrm{g}_i(\boldsymbol{x}) \geq 0 \quad i: m+1, \ldots\ldots, p$	
$\mathrm{h}_i(\boldsymbol{x}) = 0 \quad i: p+1, \ldots\ldots, r$	h_i Las restricciones de igualdad
$\boldsymbol{x} \geq 0 \leftrightarrow \boldsymbol{x_i} \geq \boldsymbol{0}$	$\boldsymbol{x}$ el vector de variables de elección

Se consideran que todas las funciones son dos veces diferenciables y al menos una de las mismas es no lineal.

No existe un algoritmo general para resolver modelos no lineales debido al comportamiento irregular de las funciones no lineales. Es por ello que en contraste con la programación lineal no se puede reducir el campo de elección al conjunto de puntos extremos de la región factible.

Sin embargo, se han determinado condiciones que bajo ciertos requisitos se convierten en condiciones de primer orden o necesarias, e inclusive en condiciones necesarias y suficientes. Estas son las condiciones de Karush – Kuhn – Tucker, que de aquí en adelante se indicarán como condiciones de KKT, y fueron desarrolladas independientemente por Karush y por Kuhn – Tucker.

2.1 Teorema 1: Supóngase que:

$$minf(x)$$
$$s.a.$$
$$g_1(x) \leq 0$$
$$\vdots$$
$$g_m(x) \leq 0$$
$$x \in C$$

Es un problema convexo para el cual existe un vector de sensitividad λ entonces:

$$MP = \{\{0\} = Inf\{f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)\}\}$$

Demostración

Si $Z \in D(h)$, entonces: $h(0) - \lambda(z) \leq h(z) \ldots \ldots (1)$

Sea $\bar{x} \in C$, tomemos $z = (g_1(\bar{x}), \ldots, g_m(\bar{x}))$

$$minf(x)$$
$$s.a.$$
$$g_1(x) \leq g_1(\bar{x})$$
$$\vdots$$
$$g_m(x) \leq g_m(\bar{x})$$
$$x \in C$$

Entonces $F(z) \neq 0$ porque $\bar{x} \in F(z)$

En (1) se tiene: $h(0) - \sum_{i=1}^{m} \lambda_i g_i(\bar{x}) \leq h(z) \ldots \ldots (2)$

Como $\bar{x} \in F(z)$ entonces: $h(z) \leq f(z) \ldots \ldots (3)$

De (2) y (3) tenemos que:

$$h(0) - \sum_{i=1}^{m} \lambda_i g_i(\bar{x}) \leq h(z) \leq f(\bar{x})$$

$$h(0) \leq f(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

Como $\bar{x}$ es arbitrario entonces:

$$h(0) \leq f(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i(\bar{x}), \ \forall \bar{x} \in C$$

De la igualdad anterior se tiene:

$$h(0) \leq \underset{\bar{x} \in C}{Inf} \left\{ f(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i(\bar{x}) \right\} \cong h(0) \leq \underset{x \in C}{Inf} \left\{ f(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i(x) \right\}$$

Demostraremos ahora es otra desigualdad:

$$\underset{x\in C}{Inf}\{f(x)+\textstyle\sum_{i=1}^{m}\lambda_i g_i(x)\} \leq \underset{x\in C}{Inf}\{f(x)+\textstyle\sum_{i=1}^{m}\lambda_i g_i(x) \;\; y \;\; g_i(x)\leq 0\}$$

$$\leq Inf\{f(x): x\in C \; y \;\; g_i(x)\leq 0\} = h(0)$$

Por lo tanto $Inf\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x): x\in C\} \leq h(0) \ldots (4)$

De (3) y (4) tenemos:

$$MP = h(0) = \underset{x\in C}{Inf}\left\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)\right\} \blacksquare$$

Definición 1: El lagrangiano $L(x,\lambda)$ de un problema convexo:

$$\begin{gathered} \min \; f(x) \\ s.a. \\ g_1(x)\leq 0 \\ \vdots \\ g_m(x)\leq 0 \\ x\in C \end{gathered}$$

Es la función definida por:

$$L(x,\lambda) = f(x)+\sum_{i=1}^{m}\lambda_i g_i(x), \quad x\in C \;, \lambda\geq 0$$

Nota: si (P) es un problema super consistente para el cual $h(0)$ es finito, entonces $h(z)$ es finito, $\forall\, z\in D(h)$ y exista un vector de sensitividad λ por el teorema 1 se tendría: $MP = h(0) = Inf\{f(x)+\sum_{i=1}^{m}\lambda_i g_i(x)\}$

Por lo tanto: $MP = h(0) = \underset{x\in C}{Inf} L(x,\lambda)$

2.2 Teorema 2: Karush – Kuhn – Tucker (forma punto silla)

Supóngase que (P) es un problema super consistente entonces:

$x^*\in C$ es solución de (P) $\leftrightarrow \exists\, \lambda^*\in R^m$ tal que

$(a)\lambda^*\geq 0 \quad , \quad \forall\, i = 1,2,\ldots,m$

$(b)\, L(x^*\;,\lambda)\leq L(x^*\;,\lambda^*)\leq L(x,\lambda^*)\,, \forall\in C$ y $\lambda\geq 0$

$(c)\, \lambda_i{}^* g_i(x^*) = 0 \qquad \forall\, i = 1,2,\ldots,m$ (Condición de holgura complementaria)

$\Longrightarrow$ Supongamos que x^* es solución de (P) por un teorema de karush-kuhn-tucker

$\exists\lambda^* \in R^m$, con $\lambda^* \geq 0$, tal que $f(x^*) = h(0) = \underset{x \in C}{Inf} L(x, \lambda^*)$........... (1)

De (1) tenemos que: $f(x^*) \leq L(x^*, \lambda^*)$. Además:

$$L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i{}^* g_i(x^*) \leq f(x^*)$$

Demostrar que x^* es factible de P.

De las dos desigualdades anteriores se tiene:

$$f(x^*) = L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i{}^* g_i(x^*) \text{.......... (2)}$$

De esta desigualdad se tiene que:

$$\sum_{i=1}^{m} \underbrace{\lambda_i{}^*}_{\geq 0} \underbrace{g_i(x^*)}_{\leq 0} \text{.......... (3)}$$

Entonces $\lambda_i{}^* g_i(x^*) = 0 \qquad \forall\, i = 1,2, \dots, m$

De (1) se tiene: $f(x^*) \leq L(x, \lambda^*)$, $\forall\, x \in C$ y de (2) se tiene:

$$L(x^*, \lambda^*) = f(x^*) \leq L(x, \lambda^*),\ \forall\, x \in C \text{ .... (4)}$$

Demostrar que: $L(x^*, \lambda^*) - L(x^*, \lambda) \geq 0, \quad \forall\lambda \geq 0$

$L(x^*, \lambda^*) - L(x^*, \lambda) = f(x^*) + \sum_{i=1}^{m} \lambda_i{}^* g_i(x^*) - [f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x)] \geq 0$

Hemos demostrado que $\forall\lambda \geq 0$ se tiene que: $L(x^*, \lambda) \leq L(x^*, \lambda^*)$ .... (5)

Juntando (4) y (5) tenemos (b).

$\Longleftarrow$ Supongamos que $x^* \in R^m$ y que $\exists\lambda^* \in R^m$ tal que:

$(a)\lambda^* \geq 0 \quad , \quad \forall\, i = 1,2, \dots, m$

$(b)\ L(x^* \ ,\lambda) \leq L(x^* \ ,\lambda^*) \leq L(x, \lambda^*)$

Demostrar que: x^* es solución de P $\begin{cases} g_i(x^*) \leq 0 & \forall\, i = 1,2, \dots, m \\ f(x^*) \leq f(x)\ \forall x \in F \end{cases}$

Por (b) tenemos: $L(x^* \ ,\lambda) \leq L(x^* \ ,\lambda^*)\ \forall\, \lambda \geq 0$

Sea $\lambda^k = (\lambda^*{}_1, \lambda^*{}_2, \dots, \lambda^*{}_{k+1}, \dots, \lambda^*{}_m)$ el cual es ≥ 0.

Luego: $f(x^*) + \sum_{i=1}^{m} \lambda^k{}_i g_i(x^*) \leq f(x^*) + \sum_{i=1}^{m} \lambda_i{}^* g_i(x^*)$

$$\sum_{i=1}^{m} (\lambda^k - \lambda^*{}_i) g_i(x^*) \leq 0$$

$$g_k(x^*) \leq 0$$

Esto vale para $k = 1,2, \dots, m$. Por tanto:

x^* es factible para P. Demostraremos que $f(x^*) = L(x^*, \lambda^*)$, de (b) se tiene:

$f(x^*) = L(x^*, 0) \leq L(x^*, \lambda^*)$ .... (6), pero $L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i{}^* g_i(x^*)$

Por ser x^* factible entonces $g_i(x^*) \leq 0 \ \forall\, i = 1,2,\dots,m$, entonces en la igualdad anterior tendremos:

$$L(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{g_i(x^*)}_{\leq 0} \leq f(x^*) \dots\dots\dots\dots (7)$$

De (6) y (7), tenemos que $f(x^*) = L(x^*, \lambda^*) = \ f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \dots\dots\dots\dots$ (8)

Ya hemos demostrado que: $\lambda_i^* g_i(x^*) = 0, \ \ \forall\, i = 1,2,\dots,m$

Tenemos de (8) que: $f(x^*) = L(x^*, \lambda^*)$ y por (b) se tiene:

$$f(x^*) = L(x^*, \lambda^*) = inf\{L(x^*, \lambda^*) : x \in C\}$$
$$\leq inf\{L(x, \lambda^*) : x \in C\} \ \text{y} \ g_1(x) \leq 0, \dots, g_m(x) \leq 0$$
$$\leq inf\{f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) : x \in C\} \ \text{y} \ g_1(x) \leq 0, \dots, g_m(x) \leq 0$$
$$\leq inf\{f(x^*) : x \in C\} \ \text{y} \ g_1(x) \leq 0, \dots, g_m(x) \leq 0$$

Tenemos que:

$$f(x^*) \leq inf\{f(x^*) : x \in C\} \ \text{y} \ g_1(x) \leq 0, \dots, g_m(x) \leq 0$$

Por tanto: $f(x^*) = \min_{x \in F} f(x)$, esto indica que x^* es solución del problema P■

OBSERVACION: El teorema anterior nos asegura que si P es un problema superconsistente entonces $x^* \in C$ es solución de P si y solo si $\exists\, \lambda^* \geq 0$ en R^m tal que (x^*, λ^*) es un punto silla del lagrangiano y tal que la condición de holgura complementaria es satisfecha. Se demuestra que si (x^*, λ^*) es un punto silla del lagrangiano de cualquier problema convexo P entonces $h(0)$ es finito y las condiciones de holgura complementaria son satisfechas.

Definición 2: Un punto (x^*, λ^*) tal que $x^* \in C$ y $\lambda^* \geq 0$ es denominado punto silla del lagrangiano si: $L(x^*, \lambda) \leq \ L(x^*, \lambda^*) \leq L(x, \lambda^*), \forall\, x \in C$ y $\lambda \geq 0$

2.3 Teorema 3: Karush – Kuhn – Tucker (forma gradiente)

Supóngase que (P) es un problema convexo superconsistente tal que la función objetivo y las funciones g_i ; $i = 1,2,3,\dots,m$ tienen primeras derivadas parciales continuas. Si x^* es factible para (P) y además es un punto interior para C, entonces x^* es una solución (P) $\leftrightarrow$ $\exists\lambda^* \in R^m$ tal que:

(1) $\lambda_i^* \geq 0 \quad , \quad \forall\, i = 1,2,\dots,m$

(2) $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$

(3) $\lambda_i^* g_i(x^*) = 0 \qquad \forall\, i = 1,2,\dots,m$

Demostración

x^* es factible para para P y $x^* \in C^0$

$\Longrightarrow$ Supongamos que x^* es solución de (P) por el teorema de karush-kuhn-tucker $\exists\lambda^*$ en R^m tal que se cumple:

(1) $\lambda_i^* \geq 0 \quad , \quad \forall\, i = 1,2,\dots,m$

(2) $L(x^*\ ,\lambda) \leq L(x^*\ ,\lambda^*) \leq L(x,\lambda^*)\ , \forall \in C$ y $\lambda \geq 0$

(3) $\lambda_i^* g_i(x^*) = 0 \qquad \forall\, i = 1,2,\dots,m$

Sea $\psi : C \to R$

$x \to L(x,\lambda^*) = f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x)$

Esto indica que x^* es un mínimo global para la función ψ como $x^* \in C$ y ψ tiene primeras derivadas parciales continuas, entonces:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$$

$\Longleftarrow$ Supongamos que $\exists\ \lambda^* \in R^m$ tal que se cumple (1), (2), (3). Demostraremos que x^* es solución de (P).

Sea x factible para (P), entonces $x \in P$ y $g_1(x) \leq 0, \dots, g_m(x) \leq 0$

$f(x) \geq f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) + \sum_{i=1}^{m} \lambda_i^* [g_i(x^*) + \nabla g_i(x^*)(x - x^*)]$

$$= f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \left[\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*)\right](x - x^*)$$

Por (3) y (2) $\sum_{i=1}^{m}\lambda_i{}^*\nabla g_i(x^*)=0$ y $\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i{}^*\nabla g_i(x^*)=0$

Hemos demostrado que: $f(x)\geq f(x^*), \forall\ x$ factibles para (P).

Por lo tanto x^* es solución de (P) ■

2.4. Condiciones de Karush – Kuhn – Tucker

Dado que las condiciones KKT es el resultado analítico más importante en programación no lineal, se presentan algunas de las condiciones necesarias y suficientes que deben cumplir los candidatos a solución óptima del problema de optimización. Estas condiciones son las llamadas condiciones de Karush – Kuhn – Tucker.

Para mayor claridad, se dan a continuación las definiciones correspondientes a los problemas de minimización y maximización de forma separada.

Dado el problema

$$\begin{cases} \qquad\qquad Optimizar \quad f(x_1,x_2,\dots,x_n) \\ Sujeto\ a \quad h_i(x_1,x_2,\dots,x_n)=0 \quad i=1,\dots,m \\ \qquad\qquad g_j(x_1,x_2,\dots,x_n)\leq \quad j=1,\dots,m \end{cases} \qquad (\alpha)$$

Con $f,h_i,g_j: A\rightarrow\mathbb{R}$ funciones de clase $C^1(A)$ y $A\subseteq\mathbb{R}^n$ un conjunto abierto. Diremos que $x^*=(x^*{}_1,x^*{}_2,\dots,x^*{}_n)\in A$ es un punto de Karush – Kuhn – Tucker para el problema α si y solo si $\exists\ \lambda_1,\dots,\lambda_m,\mu_1,\dots,\mu_m\in\mathbb{R}$, de forma que se cumplen las siguientes condiciones:

a) Condición estacionaria:

$$\nabla f(x^*{}_1,x^*{}_2,\dots,x^*{}_n)+\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*{}_1,x^*{}_2,\dots,x^*{}_n)+\sum_{j=1}^{p}\mu_j\nabla g_j(x^*{}_1,x^*{}_2,\dots,x^*{}_n)=0$$

b) Condición de factibilidad:

$$h_i(x^*{}_1,x^*{}_2,\dots,x^*{}_n)=0, \qquad i=1,\dots,m$$

$$g_j(x^*{}_1,x^*{}_2,\dots,x^*{}_n)\leq 0, \qquad j=1,\dots,p$$

c) Condición de holgura:

$$\mu_j g_j(x^*{}_1,x^*{}_2,\dots,x^*{}_n)=0, \qquad j=1,\dots,p$$

d) Condición de signo: una vez que se cumplen las condiciones anteriores el punto es de $Minimo\ (CKKTMin) \Leftrightarrow \mu_j \geq 0 \quad j = 1, \dots, p$

O de $Maximo\ (CKKTMax) \Leftrightarrow \mu_j \geq 0 \quad j = 1, \dots, p$

En ambos casos los valores $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ son los llamados multiplicadores y existe uno por cada restricción del problema: el multiplicador λ_i esta asociado a la restricción de igualdad $h_i(x_1, x_2, \dots, x_n) = 0$ para $i = 1, \dots, m$ y el multiplicador u_j esta relacionado con la restricción de desigualdad $g_j(x_1, x_2, \dots, x_n) \leq 0$ para $j = 1, \dots, p$.

Los valores $(\lambda_1, \dots, \lambda_m)$ son los multiplicadores de Langrage y los valores $(\mu_1, \dots, \mu_p)$ son los multiplicadores de Karush – Kuhn – Tucker.

De la condición de holgura se deduce que si una restricción de desigualdad es no activa en el punto solución, entonces el multiplicador de Karush – Kuhn – Tucker asociado debe tomar el valor 0.

Los puntos $x^* \in A \cap \Omega$, siendo Ω el conjunto factible del problema, que cumplen la condición estacionaria se dice que son puntos críticos o estacionarios. Esta condición estacionaria se dice que son puntos críticos y estacionarios. Esta condición estacionaria suele expresarse en términos de la llamada función lagrangiana definida utilizando la función objetivo y las restricciones como

$$L(x_1, x_2, \dots, x_n, \lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p) = f(x_1, x_2, \dots, x_n) + \sum_{i=1}^{m} \lambda_i h_i\,(x_1, x_2, \dots, x_n)$$
$$+ \sum_{j=1}^{p} \mu_j g_j\,(x_1, x_2, \dots, x_n)$$

o forma vectorial

$$L(x, \lambda, u) = f(x) + \lambda^T h(x) + \mu^T g(x) \qquad (\beta)$$

Siendo

$$\lambda = (\lambda_1, \dots, \lambda_m)$$
$$u = (u_1, \dots, u_m)$$
$$h(x) = \big(h_1(x_1, x_2, \dots, x_n), \dots, h_m(x_1, x_2, \dots, x_n)\big)$$
$$g(x) = \big(g_1(x_1, x_2, \dots, x_n), \dots, g_p(x_1, x_2, \dots, x_n)\big)$$

En forma vectorial la condición estacionaria se puede expresar de forma más compacta como $\nabla_x L(x, \lambda, \mu) = 0$

Donde el subíndice indica que estamos derivando respecto a las componentes de x.

Para diferenciar los puntos estacionarios de problemas sin restricciones de los correspondientes a problemas con restricciones, a estos últimos se les suele añadir el adjetivo de condicionados.

¿Cómo encontrar los puntos de KKT?

Para la búsqueda práctica de puntos que cumplan las condiciones de CKKT o puntos de Karush – Kuhn – Tucker, ya sean de mínimo o máximo, primero hay que resolver el sistema de ecuaciones compuesto por: la condición estacionaria, la condición de factibilidad para las restricciones de igualdad y la condición de holgura.

$\frac{\partial f}{\partial x_k}(x^*_1, x^*_2, \dots, x^*_n) + \sum_{i=1}^{m} \lambda_i \frac{\partial h_i}{\partial x_k}(x^*_1, x^*_2, \dots, x^*_n) + \sum_{j=1}^{p} \mu_i \frac{\partial g_j}{\partial x_k}(x^*_1, x^*_2, \dots, x^*_n) = 0 \quad k = 1, \dots, n$

$$h_i(x^*_1, x^*_2, \dots, x^*_n) = 0 \qquad i = 1, \dots, m$$
$$u_j g_j(x^*_1, x^*_2, \dots, x^*_n) = 0 \qquad j = 1, \dots, p$$

Este sistema está compuesto por $(n + m + p)$ ecuaciones $(n + m + p)$ incógnitas (las n coordenadas de $x^* = (x^*_1, x^*_2, \dots, x^*_n)$, los m multiplicadores de Lagrange λ_i y los p multiplicadores de Karush – Kuhn – Tucker μ_j.)

La forma usual de resolver el sistema es comenzar por la condición de holgura complementaria, ya que dichas ecuaciones nos proporcionan dos opciones

$$\mu_j g_j(x^*_1, x^*_2, \dots, x^*_n) = 0 \Leftrightarrow \begin{cases} \mu_j = 0 \\ g_j(x^*_1, x^*_2, \dots, x^*_n) = 0 \end{cases}$$

Para p restricciones de desigualdad tendríamos 2^p casos.

Una vez resuelto el sistema, para ver cuál o cuáles de las soluciones obtenidos son puntos de KKT hay que, por una parte comprobar que son puntos factibles (notar que solo quedaría por comprobar si $g_j(x^*_1, x^*_2, \dots, x^*_n) \leq 0$) y por otra que sus multiplicadores de Karush – Kuhn – Tucker asociados tienen todos el mismo signo, bien todos ≥ 0 para puntos minimo o bien todos ≤ 0 para puntos de máximo.

Casos particulares:

Las anteriores condiciones de Karush – Kuhn – Tucker se aplican al problema general de optimización, sin embargo se obtienen expresiones más simplificadas de las mismas cuando se aplican a problemas sin restricciones o cuando el problema solo tiene restricciones de igualdad.

1.- Problemas sin restricciones ($m = p = 0$): si el problema no tienen restricciones de ningún tipo, los multiplicadores no son necesarios y tampoco las condiciones relacionadas con ellos. La única condición que queda es la estacionaria.

$$\nabla f(x^*{}_1, x^*{}_2, \dots, x^*{}_n) = 0 \Leftrightarrow \frac{\partial f}{\partial x_k}(x^*{}_1, x^*{}_2, \dots, x^*{}_n) = 0 \qquad k = 1,2,\dots,n$$

Que coinciden para ambos objetivos de maximizar y minimizar.

Si además $n = 1$, es decir, el problema es optimizar una función real de variable real, la condición estacionaria nos conduce a un resultado bien conocido del cálculo diferencial $\acute{f}(x^*) = 0$

2.- Problemas de lagrange ($P = 0$): Si el problema solo tienen restricciones de igualdad el problema considerado es un problema clásico de lagrange y las condiciones de Karush – Kuhn – Tucker se obtienen eliminando aquellas ecuaciones relacionadas con restricciones de desigualdad, quedando por tanto la condición estacionaria y la condición de factibilidad

$$\frac{\partial f}{\partial x_k}(x^*{}_1, x^*{}_2, \dots, x^*{}_n) + \sum_{i=1}^{m} \lambda_i \frac{\partial h_i}{\partial x_k}(x^*{}_1, x^*{}_2, \dots, x^*{}_n) = 0 \quad k = 1, \dots, n$$

$$h_i(x^*{}_1, x^*{}_2, \dots, x^*{}_n) = 0 \quad i = 1, \dots, m$$

Que es el resultado que proporciona el teorema clásico de los multiplicadores de lagrange. De nuevo las condiciones de KKT para ambos objetivos de maximizar y minimizar coinciden.

A continuación se presentan algunos ejemplos de búsqueda de puntos de Karush – Kuhn – Tucker.

Ejemplo 1: Encuentra los puntos de karush-kuhn-Tucker para el problema

$$\begin{cases} \quad optimizar \qquad x+y+z \\ sujeto\ a \qquad (y-1)^2\ +z^2 \leq 1 \\ \qquad x^2+(y-1)^2\ +z^2 \leq 3 \end{cases}$$

Solución

Se utilizaban los multiplicadores correspondientes para construir la función lagrangiana, expresando previamente las restricciones en la forma $g(x) \leq 0$

$L(x,y,z) = (x+y+z) + u_1((y-1)^2+z^2-1) + u_2(x^2+(y-1)^2+z^2+3)$ y se plantea cada una de las condiciones de KKT.

1. $(\nabla_x L = 0)$

$$\frac{\partial L}{\partial x} = 0 \Leftrightarrow 1 + 2u_2 x = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. [1]$$

$$\frac{\partial L}{\partial y} = 0 \Leftrightarrow 1 + 2u_1(y-1) + 2u_2(y-1) = 0 \ldots\ldots\ldots. [2]$$

$$\frac{\partial L}{\partial z} = 0 \Leftrightarrow 1 + 2u_1 z + 2u_2 z = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3]$$

2. $((y-1)^2+z^2-1) \leq 0$

   $(x^2+(y-1)^2+z^2-3) \leq 0$

3. $u_1 g_1(x) = 0 \Leftrightarrow u_1((y-1)^2+z^2-1) \ldots\ldots\ldots. [4].$

   $u_2 g_2(x) = 0 \Leftrightarrow u_2(x^2+(y-1)^2+z^2-3) \ldots. [5]$

4. $u_1, u_2 \geq 0 \Rightarrow \quad Para\ minimo$

   $u_1, u_2 \leq 0 \Rightarrow \quad Para\ maximo$

El sistema que permite localizar los puntos KKT estará formado por las tres ecuaciones que proporcionan la condición estacionaria y las dos condiciones de holgura.

A partir de las condiciones se obtienen cuatro casos:

$$\begin{cases} u_1 = 0 \Rightarrow \begin{cases} u_2 = 0 \\ x^2 + (y-1)^2 + z^2 - 3 = 0 \end{cases} \\ (y-1)^2 + z^2 - 1 = 0 \Rightarrow \begin{cases} u_2 = 0 \\ x^2 + (y-1)^2 + z^2 - 3 = 0 \end{cases} \end{cases}$$

De la ecuación [1] se deduce que $u_2 \neq 0$, ya que en caso contrario se llegaría a una contracción; por ello los casos $\boldsymbol{I}$ $\boldsymbol{y}$ $\boldsymbol{III}$ no pueden darse y quedan por comprobar los casos $\boldsymbol{II}$ $\boldsymbol{y}$ $\boldsymbol{IV}$.

1. ***Caso II*** $(u_1 = 0,\ x^2 + (y-1)^2 + z^2 - 3 = 0)$ sustituyendo el valor de $u_1 = 0$ en las ecuaciones del sistema se obtiene:

$$1 + 2u_2x = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [6]$$

$$1 + 2u_2(y-1) = 0 \ldots\ldots\ldots\ldots.. [7]$$

$$1 + u_2z = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [8]$$

$$x^2 + (y-1)^2 + z^2 - 3 = 0 \ldots [9]$$

Si restamos las ecuaciones [6] y [7] se obtiene

$$(1 + 2u_2x) + \big(1 + 2u_2(y-1)\big) = 0 \Leftrightarrow 2u_2(x - y + 1) = 0$$

Y como $u_2 \neq 0$, se llega a la conclusión de que

$$x - y + 1 = 0 \Rightarrow x = y - 1$$

Restando las ecuaciones [6] y [8] obtenemos

$$(1 + 2u_2x) - (1 + 2u_2z) = 0 \Leftrightarrow 2u_2(x - z) = 0$$

Y como $u_2 \neq 0$, se obtiene

$$(x - z) = 0 \Longrightarrow x = z$$

Con las relaciones que se han obtenido entre las variables x, y y z, utilizamos la ecuación [9]

$$x^2 + (y-1)^2 + z^2 - 3 = 0 \Longrightarrow 3x^2 - 3 = 0 \Leftrightarrow x = \pm 1$$

Y para las otras dos variables

$$z = x = \pm 1$$

$$y = x + 1 \Longrightarrow \begin{cases} y = 2 \\ y = 0 \end{cases}$$

Como $x \neq 0$, despejamos u_2 de la ecuación [6]

$$u_2 = -\frac{1}{2x} = \mp\frac{1}{2}$$

Finalmente y para este caso se han obtenido 2 puntos, que son junto con sus respectivos multiplicadores

$$P_1 = (1,2,1) \qquad u = \left(0, -\frac{1}{2}\right)$$

$$P_2 = (-1,0,-1) \qquad u = \left(0, \frac{1}{2}\right)$$

2. ***Caso IV*** $((y-1)^2 + z^2 - 1 = 0,\ \ x^2 + (y-1)^2 + z^2 - 3 = 0)$. En este ultimo caso, el sistema que hay que resolver es

$$1 + 2u_2 x = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [10]$$
$$1 + 2u_1(y-1) + 2u_2(y-1) = 0 \ldots\ldots [11]$$
$$1 + 2u_1 z + 2u_2 z = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [12]$$
$$(y-1)^2 + z^2 - 1 = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots [13]$$
$$x^2 + (y-1)^2 + z^2 - 3 = 0 \ldots\ldots\ldots\ldots\ldots [14]$$

Restando las ecuaciones [13] y [14] se obtiene el valor de x

$$((y-1)^2 + z^2 - 1) - (x^2 + (y-1)^2 + z^2 - 3) = 0 \Rightarrow x^2 - 2 = 0$$
$$\Rightarrow x^2 = 2 \Leftrightarrow x = \pm\sqrt{2}$$

Se sustituye ese valor en la ecuación [10] para calcular u_2.

$$u_2 = -\frac{1}{2x} = -\frac{1}{2(\pm\sqrt{2})} = \mp\frac{1}{2\sqrt{2}}$$

Si ahora se restan las ecuaciones [11] y [12] se obtiene

$$\big(1 + 2u_1(y-1) + 2u_2(y-1)\big) - (1 + 2u_1 z + 2u_2 z) = 0$$

Y entonces podemos agrupar y sacar factor común

$$2u_1(y-1-z) + 2u_2(y-1-z) = 0 \Leftrightarrow 2(u_1+u_2)(y-1-z) = 0$$

Esta ecuación nos proporciona dos opciones. La primera de ellas es

$$u_1 + u_2 = 0$$

Pero utilizando la ecuación [12]

$$1 + 2u_1 z + 2u_2 z = 0 \Leftrightarrow 1 + 2z(u_1+u_2) = 0 \Leftrightarrow 1 = 0$$

Que obviamente es imposible. Y la otra opción es

$$y - 1 - z = 0 \Leftrightarrow y - 1 = z$$

Y utilizando la ecuación [13] se obtiene el valor de z

$$(y-1)^2+z^2=1 \Rightarrow 2z^2=1 \Leftrightarrow z=\pm\frac{1}{\sqrt{2}}$$

Y por tanto el de y

$$y=1+z=1\pm\frac{1}{\sqrt{2}}$$

Queda por determinar el valor del multiplicador u_1 y para ello utilizamos la ecuación [12]

$$1+2u_1z+2u_2z=0 \Rightarrow u_1=-u_2-\frac{1}{2z}$$

Si se consideran los distintos valores que se han encontrado para u_2 (2 valores) y para z (otros dos valores) tendremos 4 casos posibles.

$$u_2=-\frac{1}{2\sqrt{2}} \qquad z=\frac{1}{\sqrt{2}} \Rightarrow u_1=-\frac{1}{2\sqrt{2}}$$

$$u_2=-\frac{1}{2\sqrt{2}} \qquad z=-\frac{1}{\sqrt{2}} \Rightarrow u_1=\frac{3}{2\sqrt{2}}$$

$$u_2=\frac{1}{2\sqrt{2}} \qquad z=\frac{1}{\sqrt{2}} \Rightarrow u_1=-\frac{3}{2\sqrt{2}}$$

$$u_2=\frac{1}{2\sqrt{2}} \qquad z=-\frac{1}{\sqrt{2}} \Rightarrow u_1=\frac{1}{2\sqrt{2}}$$

Se han obtenido cuatro puntos

$$P_3=\left(\sqrt{2},1+\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right) \qquad u=\left(-\frac{1}{2\sqrt{2}},-\frac{1}{2\sqrt{2}}\right)$$

$$P_4=\left(\sqrt{2},1-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right) \qquad u=\left(\frac{3}{2\sqrt{2}},-\frac{1}{2\sqrt{2}}\right)$$

$$P_5=\left(-\sqrt{2},1+\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right) \qquad u=\left(-\frac{3}{2\sqrt{2}},\frac{1}{2\sqrt{2}}\right)$$

$$P_6=\left(-\sqrt{2},1-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right) \qquad u=\left(\frac{1}{2\sqrt{2}},\frac{1}{2\sqrt{2}}\right)$$

Para determinar cuáles de estos puntos son de tipo KKT, hay que comprobar su factibilidad y el signo de sus multiplicadores. Se expone a continuación una tabla resumen con los resultados obtenidos

$P = (x, y, z)$	$u = (u_1, u_2)$	Factibilidad	Positividad /negatividad	CKKT
$P_1 = (1,2,1)$	$u = \left(1, -\frac{1}{2}\right)$	NO	–	–
$P_2 = (0, -1, -1)$	$u = \left(0, \frac{1}{2}\right)$	NO	–	–
$P_3 = \left(\sqrt{2}, 1 + \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$	$u = \left(-\frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$	SI	NEGATIVIDAD	Máximo
$P_4 = \left(\sqrt{2}, 1 - \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$	$u = \left(\frac{3}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$	SI	NO	–
$P_5 = \left(-\sqrt{2}, 1 + \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$	$u = \left(-\frac{3}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$	SI	NO	–
$P_6 = \left(-\sqrt{2}, 1 - \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$	$u = \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$	SI	POSITIVIDAD	Mínimo

Los puntos P_1 y P_2 no son factibles ya que no cumplen la primera restricción del problema

$$P_1 = (1,2,1) \Rightarrow g_1(P_1) = (y-1)^2 + z^2 = (2-1)^2 + (1)^2 = 2 \nleq 1$$

$$P_2 = (0, -1, -1) \Rightarrow g_1(P_2) = (y-1)^2 + z^2 = (0-1)^2 + (-1)^2 = 2 \nleq 1$$

Y por tanto no son puntos válidos para el problema.

Los puntos P_4 y P_5 si son factibles, sin embargo no tiene multiplicadores de signo constante y por tanto tampoco son puntos de KKT.

Los puntos P_3 y P_6 son los únicos que cumplen con todas las condiciones para ser puntos de KKT; en el caso de P_3 seria un punto de máximo ya que $u \leq 0$, mientras que P_6 seria un punto de mínimo puesto que $u \geq 0$.

Ejemplo 2: Encuentra los puntos de karush-kuhn-Tucker del siguiente problema

$$\begin{aligned} &optimizar \quad && x^2 + y^2 \\ &sujeto\ a \quad && 2x + y - 2 = 0 \end{aligned}$$

Resolución

En este caso $m = 1$ y $p = 0$, es decir hay solamente una restricción de igualdad y el problema es de lagrange. La función lagrangiana del problema es

$$L(x, y, \lambda) = (x^2 + y^2) + \lambda(2x + y - 2)$$

Y las condiciones que debe cumplir un punto para ser de karush-kuhn-Tucker serán la condición estacionaria y la condición de factibilidad.

$$\text{(Condición estacionaria)} \Rightarrow \nabla_x L = 0 \quad \Rightarrow \begin{cases} \dfrac{\partial L}{\partial x} = 0 \Rightarrow 2x + 2\lambda_1 = 0 \\ \dfrac{\partial L}{\partial y} = 0 \Rightarrow 2y + \lambda_1 = 0 \end{cases}$$

$$\text{(Condición de factibilidad)} \Rightarrow \quad h(x, y) = 0 \quad \Rightarrow 2x + y - 2 = 0$$

El sistema es lineal y tiene como solución única y por tanto único punto de KKT

$$x = \frac{4}{5} \qquad y = \frac{2}{5} \qquad \lambda_1 = -\frac{4}{5}$$

Como el multiplicador está asociado a una restricción de igualdad y su signo no influye en el carácter del punto, el punto P es un punto de KKT que puede ser de máximo o de mínimo.

Ejemplo 3: Aplica las condiciones necesarias de primer orden al problema

$$\begin{array}{ll} \textit{minimizar} & xy + yz + zx \\ \textit{sujeto a} & x + y + z = 3 \end{array}$$

Resolución

Como solamente tiene una restricción de igualdad se trata de un problema de Lagrange; dicha restricción es lineal. De este modo, si el problema tuviera solución, es decir, si existiera el mínimo global de la función sobre el conjunto factible, sería el mínimo local, como debe ser factible, la consecuencia es que debe ser un punto que cumpla las condiciones karush-kuhn-Tucker de mínimo. Al ser un problema de lagrange estas condiciones se reducen a la condición estacionaria y a la condición de factibilidad.

La función lagrangiana para este problema es

$$L(x, y, z, \lambda) = (xy + yz + zx) + \lambda(x + y + z - 3)$$

Y las condiciones de karush-kuhn-Tucker proporcionan el siguiente sistema

$$\text{(Condición estacionaria)} \quad \Longrightarrow \quad \nabla_x L = 0 \quad \Rightarrow \begin{cases} \frac{\partial L}{\partial x} = 0 & \Rightarrow y + z + \lambda = 0 \\ \frac{\partial L}{\partial y} = 0 & \Rightarrow z + x + \lambda = 0 \\ \frac{\partial L}{\partial z} = 0 & \Rightarrow x + y + \lambda = 0 \end{cases}$$

$$\text{(Condición de factibilidad)} \quad \Longrightarrow \quad h(x,y,z) = 0 \quad \Rightarrow x + y + z - 3 = 0$$

Que es lineal y con solución única.

$$x = y = z = 1 \qquad \lambda = -2$$

El punto obtenido $P = (1,1,1)$ junto con el multiplicador asociado $\lambda = -2$ es el único que cumple las condiciones de KKT. Notar que aunque $\lambda_1 = -2 < 0$ y el objetivo sea de minimizar, el punto cumple las condiciones de KKT puesto que λ es un multiplicador asociado a una restricción de igualdad y no esta condicionado por su signo. No es posible determinar la naturaleza del punto, puesto que se cumplen las condiciones de KKT tanto para mínimo, como para máximo.

Ejemplo 4: Plantea y resuelve el problema de construir una caja de cartón rectangular de volumen máxima y área fija A.

Resolución

$$\begin{array}{ll} minimizar & xyz \\ sujeto\ a & xy + yz + zx = \frac{A}{2} \end{array}$$

Donde x, y, z son las dimensiones de la caja y $A > 0$ su área. Se han omitido las restricciones de positividad sobre las variables ya que por naturaleza del problema, los valores de estas variables deben > 0, es decir, serán inactivas en cualquier punto, en particular en el punto solución y por tanto los multiplicadores asociados a estas restricciones seria 0.

Comprobaremos que se cumple la hipótesis de regularidad en todos los puntos factibles, para ello calculamos el gradiente de la restricción

$$\nabla h(x,y,z) = \begin{pmatrix} y + z \\ x + z \\ x + y \end{pmatrix}$$

Y estudiamos su dependencia lineal en cada punto del conjunto factible. Como solo hay un vector, este será linealmente dependiente cuando sea el vector nulo, es decir

$$y + z = 0$$
$$x + z = 0$$
$$x + y = 0$$

Sistema que tiene por única solución el vector nulo

$$x = y = z = 0$$

Sin embargo este punto es infactible

$$0 + 0 + 0 = 0 \neq \frac{A}{2}$$

De donde se deduce que todos los puntos de Ω (conjunto factible) son regulares. Al cumplirse una de las hipótesis de cualificación de las restricciones, cualquier extremo local del problema que existiera debería cumplir las condiciones de karush-kuhn-Tucker. La función lagrangiana para este problema es

$$L(x, y, z, \lambda) = xyz + \lambda \left(xy + yz + zx - \frac{A}{2} \right)$$

Aplicando las condiciones de KKT se obtiene el sistema

(Condición estacionaria) $\Longrightarrow \quad \nabla_x L = 0 \quad \Rightarrow \begin{cases} \frac{\partial L}{\partial x} = 0 & \Rightarrow yz + \lambda(y + z) = 0 \\ \frac{\partial L}{\partial y} = 0 & \Rightarrow xz + \lambda(x + z) = 0 \\ \frac{\partial L}{\partial z} = 0 & \Rightarrow xy + \lambda(x + y) = 0 \end{cases}$

(Condición de factibilidad) $\Rightarrow \quad h(x, y, z) = 0 \quad \Rightarrow xy + yz + zx - \frac{A}{2} = 0$

El sistema anterior tiene como única solución, teniendo en cuenta que $x, y, z > 0$ a

$$x = y = z = \sqrt{\frac{A}{6}} \qquad \lambda = -\sqrt{\frac{A}{24}}$$

Y al ser un problema que solo tiene restricciones de igualdad, de momento con estos datos no es posible determinar si el punto es de mínimo o de máximo.

CONTRAEJEMPLOS:

El teorema de las condiciones necesarias de primero orden proporciona los requisitos que deben cumplir los extremos de un problema de optimización con restricciones bajo ciertas hipótesis de cualificación, sin embargo, es posible encontrar problemas en los que la solución óptima no cumple estas condiciones y también es posible encontrar puntos que cumplen las condiciones de Karush – Kuhn – Tucker pero que no son extremos de la función. Veamos algunos ejemplos ¨patológicos.

Ejemplo 1: Consideremos el problema

$$\begin{aligned} &maximizar \quad -x^2+y \\ &sujeto\ a \quad y^3=0 \end{aligned}$$

Su conjunto factible es

$$\Omega = \{(x,y) \in \mathbb{R}^2 : (x,0)\}$$

El siguiente es un problema equivalente

$$\begin{aligned} &maximizar \quad -x^2 \\ &sujeto\ a \quad x \in \mathbb{R} \end{aligned}$$

Y podemos obtener fácilmente que $x^* = 0$ es su solución óptima en y por equivalencia entre los dos problemas, será también la solución óptima del problema inicial.

Vamos a comprobar si $P = (0,0)$ es un punto de Karush – Kuhn – Tucker. Si esto fuera cierto, debería existir un multiplicador λ asociado a la restricción $y^3 = 0$, de forma que se cumplan las ecuaciones del sistema

(Condición estacionaria) $\Rightarrow \quad \nabla_x L = 0 \quad \Rightarrow \begin{cases} \frac{\partial L}{\partial x}(P) = 0 \ \Rightarrow -2x = 0 \\ \frac{\partial L}{\partial y}(P) = 0 \Rightarrow 1 + \lambda 3y^2 = 0 \end{cases}$

(Condición de factibilidad) $\Rightarrow \quad h(x,y,z) = 0 \quad \Rightarrow \quad y^3$

Sin embargo, este sistema no tiene solución, ya que de la última ecuación necesariamente $y = 0$ y al sustituir en la segunda obtenemos una contradicción.

El punto $P = (0,0)$ es un máximo local (de hecho es global puesto que la función objetivo siempre es ≤ 0 y solo se anula en $x = 0$) para el problema de optimización planteado, sin embargo no cumple las condiciones de Karush – Kuhn – Tucker.

¿Contradice este ejemplo el teorema de las condiciones de primer orden? La respuesta es no, ya que P no cumple ninguna de las hipótesis de cualificación de las restricciones necesarias en dicho teorema.

1. Hipótesis sin restricciones: Esta claro que esta hipótesis no se cumple por la presencia de la restricción: $y^3 = 0$.
2. Hipótesis de karlin: Esta condición tampoco se cumple, puesto que la única restricción del problema $h(x,y) = y^3$, es claramente no lineal.

3. Hipótesis de Slater: el conjunto factible es:

$$\Omega = \{(x, y) \in \mathbb{R}^2 : (x, 0)\}$$

Que es el eje OX, que en $\mathbb{R}^2$ es un conjunto convexo, pero su interior es vacío, puesto que cualquier bola de centro un punto de Ω tiene puntos fuera de Ω y tampoco se cumple esta hipótesis.

4. Hipótesis de Fiacco – McKormik: En este caso tenemos que comprobar si el punto es o no regular para las restricciones del problema, es decir, habrá que comprobar si el conjunto de vectores formado por los gradientes de las restricciones activas en P está formado por vectores linealmente independientes. Como solamente tenemos una restricción activa en P (por ser desigualdad); la familia de vectores estará formado por un único vector $\{\nabla h(x, y)\} = \{(0{,}3y^2)\}$

Y al evaluar en $P = (0{,}0)$ obtenemos

$$\{\nabla h(P)\} = \{(0{,}0)\}$$

Que por ser el vector nulo, es linealmente dependiente y como consecuencia el punto $P = (0{,}0)$ es no regular para las restricciones.

Si el planteamiento del problema cambiamos la restricción $y^3 = 0$ por la restricción equivalente $y = 0$.

La solución del problema es la misma, pero en este caso si se cumple la hipótesis de Karlin, ya que todas las restricciones son lineales. Ahora tendríamos que ser capaces de encontrar el valor del multiplicador λ y comprobar que el punto $P = (0{,}0)$ cumple las condiciones de KKT. El sistema con este cambio es

$$-2x = 0$$
$$1 + \lambda = 0$$
$$y = 0$$

Que tiene por solución

$$P = (0{,}0)$$
$$\lambda = -1$$

Y hemos encontrado el punto buscado y también su multiplicador correspondiente.

Ejemplo 2: Sea el problema de optimización

$$\begin{aligned} &Maximizar \quad f(x,y,z) = 0 \\ &Sujeto\ a \quad h_1(x,y,z) = (x-1)^2 + y^2 - 1 = 0 \\ &\qquad\qquad h_2(x,y,z) = (x+1)^2 + y^2 - 1 = 0 \end{aligned}$$

El conjunto factible para este problema es

$$\Omega = \{(x,y,z) \in \mathbb{R}^3 : (0,0,z)\}$$

Y la solución óptima del problema es cualquier punto deΩ, puesto que $f(x,y,z)$ es constante sobre él.

La condición estacionaria para este problema nos proporciona las siguientes ecuaciones

$$\text{(Condición estacionaria)} \;\Rightarrow\; \nabla_x L = 0 \;\Rightarrow\; \begin{cases} \frac{\partial L}{\partial x} = 0 \Rightarrow 2\lambda_1(x-1) + 2\lambda_2(x+1) = 0 \\ \frac{\partial L}{\partial y} = 0 \Rightarrow 1 + 2\lambda_1 y + 2\lambda_2 y = 0 \\ \frac{\partial L}{\partial x} = 0 \Rightarrow 0 = 0 \end{cases}$$

Siendo $L(x,y,z,\lambda_1,\lambda_2) = f + \lambda_1 f_1 + \lambda_2 f_2$ la función lagrangiana del problema.

Ninguno de los puntos de Ω, $(0,0,z)$, es solución del sistema anterior, puesto que al sustituir cualquiera de ellos en la segunda ecuación nos llevaría a una contradicción: ¡Ninguno de los extremos locales del problema cumple las condiciones KKT!

Podemos comprobar, como en el caso anterior, que ninguno de ellos cumple ninguna de las hipótesis de cualificación. Es un problema con restricciones, ambas no lineales y donde el conjunto factible Ω tiene interior vacío por ser una recta. Para comprobar si se cumple la hipótesis de regularidad observamos que el conjunto de los vectores que son gradiente de las restricciones activas (en este caso son todas puesto que es un problema con solo igualdades) está dado por

$$\{\nabla h_1(x,y,z), \nabla h_2(x,y,z)\} = \left\{ \begin{pmatrix} 2(x-1) \\ 2y \\ 0 \end{pmatrix}, \begin{pmatrix} 2(x+1) \\ 2y \\ 0 \end{pmatrix} \right\}$$

Y al evaluarlo en los puntos óptimos $(0,0,z)$ obtenemos

$$\left\{ \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \right\}$$

Que es una familia de vectores linealmente dependientes y por tanto ningún punto de la forma $(0,0,z)$ es regular.

También podría suceder que un punto donde las restricciones no cumplan ninguna de las hipótesis de cualificación, sea extremo local del problema y cumpla las condiciones de Karush – Kuhn – Tucker.

Consideremos, por ejemplo, las restricciones del ejemplo anterior, pero cambiando la función objetivo por $f(x,y,z) = x$.

En este caso la condición estacionaria para el problema es

$$\text{(Condición estacionaria)} \Rightarrow \quad \nabla_x L = 0 \quad \Rightarrow \begin{cases} \frac{\partial L}{\partial x} = 0 \Rightarrow 1 + 2\lambda_1(x-1) + 2\lambda_2(x+1) = 0 \\ \frac{\partial L}{\partial y} = 0 \Rightarrow 2\lambda_1 y + 2\lambda_2 y = 0 \\ \frac{\partial L}{\partial x} = 0 \Rightarrow 0 = 0 \end{cases}$$

Y teniendo en cuenta que el conjunto factible está formado por los puntos $(0,0,z)$, el sistema queda como

$$1 + 2\lambda_1 + 2\lambda_2 = 0$$
$$0 = 0$$
$$0 = 0$$

Que tiene por solución $\lambda_1 - \lambda_2 = \frac{1}{2}$

Tomando ahora cualquier solución de esta ecuación, por ejemplo $\lambda = (\lambda_1, \lambda_2) = \left(1, \frac{1}{2}\right)$

Vemos que todos los puntos extremos cumplen las condiciones de Karush – Kuhn – Tucker, sin embargo, como se ha comprobado, en ninguno de ellos las restricciones cumplen ninguna de las hipótesis de cualificación.

Todos estos ejemplos son atípicos y en general sucederá que los extremos locales del problema tendrán que cumplir las condiciones de Karush – Kuhn – Tucker, pero ilustran la necesidad de comprobar adecuadamente los resultados obtenidos.

Por ultimo hay que indicar que estas condiciones son necesarias, en el sentido de que bajo hipótesis del teorema, los extremos de un problema de optimización deben ser puntos de Karush – Kuhn – Tucker. Sin embargo, las condiciones no son suficientes, ya que podemos encontrar puntos que aun cumpliendo las condiciones de Karush – Kuhn – Tucker, no son extremos, por ejemplo la función $f(x) = x^3$ tiene como único punto estacionario $x = 0$, que no es extremo puesto que la función es siempre creciente.

Verificándose cierta hipótesis sobre las restricciones, las condiciones de KKT son condiciones necesarias para un mínimo o máximo local, respectivamente. Y dado que todo extremo global debe ser un extremo local, estas condiciones son asimismo condiciones necesarias para extremos globales, siempre que se cumpla con la cualificación de las restricciones sobre las cuales se trata a continuación.

2.5. Cualificación de Restricciones

Existen ciertos requisitos sobre las restricciones de un problema no lineal para que las condiciones de KKT sean condiciones necesarias para la existencia de un óptimo. Estos requisitos tienen la intención de salvar irregularidades que pueden existir en la frontera de la región factible cuando las restricciones son no lineales. Considerando un punto de frontera que es candidato a ser una solución $X^* \equiv (x_1^*, x_2^*, , \dots, x_n^*)$ el vector de diferenciales $dX \equiv (dx_1, dx_2, \dots, dx_n)$ que indica el movimiento en una dirección específica a partir del punto de frontera X^*.

La cualificación de las restricciones establece dos condiciones para los vectores dX:

a) Si la $j -$ ésima variable de elección tiene valor cero en el punto X^*, entonces $dx_j \geq 0$.

b) Si la $i -$ ésima restricción se satisface como igualdad en el punto X^*, entonces $dg_i(X^*) = \left(\frac{\partial g}{\partial x_1}\right) dx_1 + \left(\frac{\partial g}{\partial x_2}\right) dx_2 + \dots + \left(\frac{\partial g}{\partial x_n}\right) dx_n \leq 0$ para problemas de maximización y $dg_i(X^*) \geq 0$ para minimización; donde todas las derivadas parciales se calculan en X^*.

Todo vector que satisfaga(a) y (b) es considerado un vector prueba. Finalmente si existe un arco diferenciable que procede del punto X^*, está enteramente contenido en la región factible y es tangente al vector prueba dado, se le denomina arco de cualificación para dicho vector.

Con ello la cualificación de las restricciones se satisface si para cualquier punto X^*.sobre la frontera de la región factible existe un arco de cualificación para cada vector de prueba.

2.6. Programación Convexa y Programación Cuadrática

Un problema de optimización convexo, resuelve dos problemas similares:

- Minimizar una función convexa cuando el espacio de soluciones es un conjunto convexo.
- Maximizar una función cóncava siendo la región factible un conjunto convexo.

La importancia del análisis de este tipo de programación obedece a posibilidad de transformar las condiciones de KKT en condiciones suficientes para óptimos. Para ello es necesario enunciar los siguientes teoremas sobre óptimos locales y globales.

Teorema 4: Sean S un subconjunto convexo de R^n y $f(X)$ una función convexa en S. Dado el problema de minimizar $f(X)$ sujeto a X ∈ S, si X_0 es una solución óptima local, entonces X_0 es una solución óptima global. Además, si f(X) es estrictamente convexa, X_0 es la única solución óptima

Demostración: la demostración de la primera parte del teorema se hará por reducción al absurdo, sea X_0 una solución óptima local, entonces existe un entorno $N(X_0)$ tal que

$$f(X) \geq f(X_0) \qquad \forall X \in N(X_0) \cap S \qquad \text{(i)}$$

Suponiendo por el absurdo que X_0 no es una solución óptima global, entonces existe $X_1 \in S$ tal que $f(X_1) < f(X_0)$ y por la convexidad de f, para $\alpha \in (0, 1)$ se cumple que:

$$f(\propto X_1 + (1-\propto)X_0) \leq \propto f(X_1) + (1-\propto)f(X_0) < \propto f(X_1) + (1-\propto)f(X_0)$$

(ii)

Ahora, para $\alpha > 0$ suficiente pequeño

$$X = \propto X_1 + (1-\propto)X_0 \in N(X_0) \cap S$$

y en consecuencia la desigualdad (ii) contradice la (i) y la primera parte del teorema queda demostrada.

Finalmente, la demostración de la segunda parte del teorema también se hará por reducción al absurdo sea que X_0 una solución óptima local y f(X) estrictamente convexa, lo que implica que f(X) es convexa, se deduce por lo demostrado anteriormente que X_0 es una solución global.

Contrariamente, suponiendo que X_0 no es la única solución óptima global, entonces existe $X_2 \in S$ y $X_2 \neq X_0$ tal que $f(X_2) = f(X_0)$. Pero por la convexidad estricta de f se obtiene:

$$f\left(\frac{1}{2} X_2 + \frac{1}{2} X_0\right) < \frac{1}{2} f(X_2) + \frac{1}{2} f(X_0) = f(X_0)$$

Ahora, por la convexidad de *S*, se sigue que $\left(\frac{1}{2} X_2 + \frac{1}{2} X_0\right) \in S$ y la desigualdad anterior contradice la optimalidad global de X_0.

En consecuencia X_0 es el único punto en que f alcanza el mínimo global o absoluto.■

Nota: El caso de maximizar una función cóncava es similar al de minimizar una función convexa. En consecuencia, para el problema de maximizar una función cóncava *f(X)* sujeta a $X \in S$, siendo S un conjunto convexo de R^n, existe un teorema análogo al precedente.

Teorema 5: Dados $f: R^n \rightarrow R$ una función convexa diferenciable en R^n y S un conjunto convexo de R^n, se considera el problema de minimizar *f(X)* sujeto a $X \in S$. Entonces, $X_0 \in S$ es una solución óptima si y solo si $[\nabla f(X_0)]^T(X - X_0) \geq 0 \quad \forall X \in S$. Además, si S es un conjunto abierto, X_0 es una solución óptima si y solo si $\nabla f(X_0) = 0$.

Demostración: primeramente se demuestra la condición suficiente, supongamos que $[\nabla f(X_0)]^T(X - X_0) \geq 0 \;\; \forall X \in S$, por otra parte por convexidad de f se tiene

$$f(X) \geq f(X_0) + [\nabla f(X_0)]^T(X - X_0) \quad \forall X \in S,$$

resulta entonces $f(X) \geq f(X_0) \;\; \forall X \in S$ *(iii)*; por tanto X_0 es una solución óptima del problema.

Recíprocamente, sea $X_0 \in S$ una solución óptima. Por la convexidad de S, $\forall X \in S$ y $\alpha \in [0, 1]$ tenemos que $\propto X + (1-\propto)X_0 \in S$, o sea $X_0 + \propto (X - X_0) \in S$ y la desigualdad *(iii)* puede reescribirse

$$f(X_0 + \propto (X - X_0)) - f(X_0) \geq 0 \qquad \forall X \in S \text{ y } 0 \leq \alpha \leq 1,$$

y siendo f diferenciable se tiene:

$f(X_0 + \propto (X - X_0)) - f(X_0) = \propto [\nabla f(X_0)]^T (X - X_0) + \alpha \bullet X - X_0 \bullet \bullet (\alpha(X - X_0))$

luego: $\propto [\nabla f(X_0)]^T (X - X_0) + \alpha \bullet X - X_0 \bullet \bullet (\alpha(X - X_0)) \geq 0$

Dividiendo por α resulta:

$$[\nabla f(X_0)]^T (X - X_0) + \bullet X - X_0 \bullet \bullet ((X - X_0)) \geq 0$$

y tomando límite para $\alpha \to 0$

$$[\nabla f(X_0)]^T (X - X_0) \geq 0$$

Finalmente, si S es abierto todo punto es también un punto interior de S, luego si en X_0 la función alcanza un mínimo, por las conocidas condiciones necesarias de extremo, las derivadas parciales, que existen por la diferenciabilidad de f en S, deberán anularse en X_0, luego $\nabla f(X_0) = 0$ *(iv)*.

Recíprocamente, si se verifica *(iv)*, X_0 es una solución óptima. En efecto, de la desigualdad

$$f(X) \geq f(X_0)$$

completándose así la demostración■

Definición 1: Siendo $A = (a_1, a_2, \dots, a_n)$, $X = (x_1, x_2, \dots, x_n)^T$ y b un número real, se llama hiperplano en R^n al conjunto de puntos que satisfacen la ecuación lineal $AX = b$. Todo hiperplano divide los puntos R^n en dos semiespacios: $AX \leq b$ y $AX \geq b$

De esta forma la convexidad del espacio de soluciones se puede establecer verificando directamente la convexidad o la concavidad de las funciones que constituyen las restricciones.

Para el caso general de Maximización:

$$\begin{array}{ll} \text{Maximizar} & z = f(x_1, x_2, \dots, x_n) \\ \text{Sujeto a} & g_i(x_1, x_2, \dots, x_n) \leq r_1 \quad i: 1,2, \dots, m \\ & x_j \geq 0 \quad j: 1, 2, \dots, n \end{array}$$

El problema será de programación convexa si y solo si:

a) La función objetivo $f(x_1, x_2, \ldots, x_n)$ es cóncava en el octante no negativo.

b) Las funciones $g_i(x_1, x_2, \ldots, x_n)$ son convexas en el octante no negativo.

Para el caso general de Minimización:

$$\text{Minimizar} \quad z = f(x_1, x_2, \ldots, x_n)$$
$$\text{Sujeto a} \quad g_i(x_1, x_2, \ldots, x_n) \geq r_1 \quad i: 1,2, \ldots, m$$
$$x_j \geq 0 \quad j: 1, 2, \ldots, n$$

El problema será de programación convexa si y solo si:

a) La función objetivo $f(x_1, x_2, \ldots, x_n)$ es convexa en el octante no negativo.

b) Las funciones $g_i(x_1, x_2, \ldots, x_n)$ son cóncavas en el octante no negativo.

Las condiciones de KKT como condiciones necesarias y suficientes para extremos

De lo anterior se puede deducir que las condiciones de KKT son condiciones necesarias y suficientes si:

a) El problema es de programación convexa.

b) El punto óptimo cumple con la cualificación de las restricciones.

2.7. Programación Cuadrática

La programación cuadrática considera el problema de optimizar una función objetivo cuadrática sujeta a restricciones lineales y a condiciones de no negatividad. Este tipo de programación suele ser muy importante en el estudio de las Ciencias Económicas dado que las formulaciones de programas cuadráticos surgen de manera natural en muchas aplicaciones.

Un modelo de programación cuadrática se define de la siguiente manera:

$$\text{Optimizar} \quad f(X) = CX + X^T DX$$
$$\text{Sujeto a:} \quad AX \leq B$$
$$X \geq 0$$

Siendo:

$X = (X_1, X_2, \ldots, X_n)^T$

$C = (c_1, c_2, \ldots, c_n)$

$B = (b_1, b_2, \ldots, b_n)^T$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{r1} & a_{r2} & \dots & a_{rn} \end{pmatrix} \qquad D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{r1} & d_{r2} & \dots & d_{rn} \end{pmatrix}$$

Al cual se le pueden establecer ciertos requisitos para que las condiciones de KKT sean Condiciones necesarias y suficientes.

Primero, se debe destacar que la linealidad de las restricciones garantiza que el espacio de soluciones sea un conjunto convexo.
Luego el problema queda reducido a determinar la concavidad o convexidad estricta de la función objetivo f de acuerdo si el problema es de maximización o minimización, respectivamente.

A partir de la definición de gradiente se tiene que:

$$\nabla f(X) = \nabla (CX + X^T DX) = C + 2DX$$

Siendo la matriz Hessiana *H* de f(X):

$H = \nabla^2 f(X) = \nabla (C + 2DX) = 2D$

Con lo cual si la matriz D es definida positiva, H también lo será, en cuyo caso f resultará estrictamente convexa.

De la anterior se deduce que en un problema de minimización las condiciones KKT son condiciones necesarias y suficientes si y solo si la matriz D es definida positiva, lo que equivale a que los auto valores de dicha matriz sean todos positivos. Asimismo, si el problema es de maximización las condiciones de KKT serán condiciones necesarias y suficientes si y solo si la matriz D es definida negativa. Lo que es equivalente a que la *f(X)* es estrictamente cóncava, siendo los autovalores de la matriz D todos negativos.

Por las condiciones expuestas, estos problemas son problemas de programación convexa aplicándose a ello todo lo expuesto en las secciones previas.

CAPÍTULO 3

UNA APLICACIÓN EN MICROECONOMIA SELECCIÓN DE CARTERAS DE ACTIVOS FINANCIEROS:

PRELIMINARES:

Los conceptos que se presentan a continuación constituyen a la base de la teoría moderna de las finanzas corporativas. Esencialmente, esta sección intenta deducir la relación entre las dos características básicas de un instrumento financiero: la rentabilidad y el riesgo.

Definición: La rentabilidad de un título se la puede definir como el beneficio económico que se obtiene a partir de la tenencia de dicho título durante un periodo determinado. Generalmente, se expresa en porcentajes.

A partir de la definición se deduce que la rentabilidad de un activo financiero es solo conocida luego que el periodo en el que se esté evaluando haya vencido, es decir, que no se puede fijar con exactitud cuál ha de ser la rentabilidad del mismo en el futuro.

Esta deducción puede así mismo considerarse en virtud de la relación inversa existente entre el precio de un activo y su rentabilidad. Con lo cual, dado que el precio exacto de los títulos, al igual que el de cualquier otra mercancía, es incierto hacia el futuro, entonces tampoco puede ser conocida su rentabilidad.

Es por ello que se suelen considerar dos términos distintos aunque muy relacionados: rentabilidad histórica y rentabilidad esperada.

La rentabilidad histórica de un activo financiero es la rentabilidad que el mismo ha reportado a lo largo de su existencia; generalmente se presenta las tasas de rentabilidad histórica por año y se construye un histograma con las mismas.

Esta información suele ser la base para estimar las tasas de rentabilidad esperada del mismo título para los periodos futuros, mediante diferentes tipos de modelos.

Mientras que, la rentabilidad esperada es el promedio ponderado de la rentabilidad esperada del activo con riesgo y del activo sin riesgo.

Asimismo, la incertidumbre da origen a la otra característica importante de cualquier activo financiero: su riesgo.

Definición: El riesgo es la medida de la posibilidad de perder o no poder ganar cierto valor económico. Esta característica obedece a que los activos financieros son títulos contingentes, es decir, compromisos de pagos futuros que no son 100% seguros.

A pesar de ser clara su definición, el riesgo es una medida de difícil interpretación en la práctica. Una manera de considerar el riesgo de la rentabilidad de los instrumentos financieros es en términos de la dispersión de la distribución de frecuencia de las tasas de rentabilidad históricas.

La dispersión de la distribución es una medida de cuanto se puede desviar una rentabilidad determinada de la rentabilidad media. Es por ello que se utilizan dos conceptos estadísticos como medidas de riesgo de un título: la varianza ($\sigma^2 \ o \, Var$) y su raíz cuadrada, **la desviación estándar** (SD o σ).

ELEMENTOS BASICOS DE ESTADISTICA:

Para entender a cabalidad las definiciones anteriores se hace necesario las siguientes definiciones básicas de estadística.

Definición: Variable es cada una de las características o cualidades que poseen las unidades de una población.

Considerando que dentro de las operaciones financieras de una empresa de inversión, no se pueden predecir con exactitud qué resultados (eventos) se van obtener, sino que, a lo más se puede describir cuales van a ser los resultados posibles y con qué probabilidad puede ocurrir cada uno de ellos, es decir en cuantificar mediante una función estos eventos o resultados, a estas funciones cuyos valores dependen de los posibles resultados se llaman **variables aleatorias,** formalmente se tiene la siguiente definición.

Definición: Una variable aleatoria X es una función definida sobre el espacio muestral Ω con valores reales. Es decir $X: \Omega \rightarrow \mathbb{R}$

Sea X una variable aleatoria con valores $x_1, x_2, \dots, x_n$. Entenderemos por $P(X = x_i)$ como la **probabilidad** del suceso.

Podemos distinguir dos tipos: discretas y continuas.

- **Discretas,** si toma un numero de valores finito o infinito numerable.

  Estas variables corresponden a experimento en los que se cuenta el número de veces que ha ocurrido un suceso. Para su descripción se especifica sus posibles valores con sus respectivas probabilidades.

  Por ejemplo en el experimento consiste en lanzar dos monedas, el espacio muestral es $\Omega = \{(c,c); (c,s); (s,c); (s,s)\}$, donde c representa cara y s representa sello. Sobre este espacio definimos la función:

  $$X: \Omega \rightarrow \mathbb{R}; \mathrm{X}(w) = numero\ de\ caras\ que\ aparecen$$

  Esta es una variable aleatoria discreta, puesto que toma los valores:

  $$X(s,s) = 0; X(c,s) = X(s,c) = 1; X(c,c) = 2$$

  Y las probabilidades que ocurran estos sucesos son:

  $$P(X = 0) = \frac{1}{4}; P(X = 1) = \frac{2}{4}; P(X = 2) = \frac{1}{4}$$

  Muchas veces es de interés conocer con que probabilidad una variable aleatoria toma valores que no sobrepasan un determinado número real x, es decir, la probabilidad acumulada de que la variable tome valores inferiores a ese x, a esta probabilidad se le llama **función de distribución**, en el caso de una variable discreta se define por:

  $$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

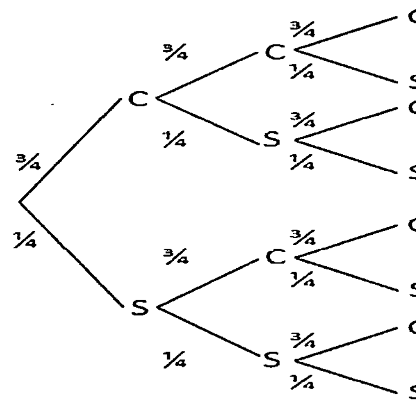
  Por ejemplo, la función de distribución para la variable aleatoria

  $$X = numero\ de\ caras\ que\ aparecen\ al\ lanzar\ dos\ veces\ una\ moneda$$

  Viene dada por:

$$F(x)=\begin{cases} 0, x<0 \\ \frac{1}{4}, 0\leq x<1 \\ \frac{3}{4}, 1\leq x<2 \\ 1, x\geq 2 \end{cases}$$

Ejemplo: Se sabe que una moneda sale cara 3 veces más a menudo que sello. Esta moneda se lanza a veces. Sea X el numero de caras que aparece, establecer la distribución de probabilidad de X y también la distribución acumulada.



$\#\ Caras$	X	$P(X)$
(c,c,c)	3	3/4
(c,c,s)	2	9/64
(c,s,c)	2	9/64
(c,s,s)	1	3/64
(s,c,c)	2	9/64
(s,c,s)	1	3/64
(s,s,c)	1	3/64
(s,s,s)	0	1/64

$\#\ Caras$	$P(X)$	$F(X)$
0	1/64	1/64
1	9/64	10/64
2	27/64	37/645
3	27/64	1

- **Continuas,** si toma cualquier valor de un intervalo real de la forma: $\langle a,b\rangle; \langle a,+\infty\rangle, \langle -\infty,b\rangle, \langle -\infty,+\infty\rangle$ o uniones de ellos. Por ejemplo, el peso de una persona, el tiempo de duración de un suceso, etc. Una variable aleatoria continua X, se describen por la llamada **función de densidad**, que es una función real no negativa $f(x)$ tal que:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

De forma que es posible calculara la probabilidad de que X tome valores en un cierto intervalo$[a,b]$, por integración

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Ejemplo: La densidad de la variable aleatoria continúa X esta dada por:

$$f(x)\begin{cases} \frac{1}{8}(x+1) & para\ 2 < x < 4 \\ 0 & en\ cualquier\ otro\ lugar \end{cases}$$

Encuentre $P(y < 3.2)$ y $P(2.9 < y < 3.2)$

Solución

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\int_{2}^{4} \frac{1}{8}\ (x+1)dx = \frac{1}{8}\left[\int_{2}^{4} xdx + \int_{2}^{4} dx\right] = \frac{1}{8}\left[\left(\frac{16}{2} - \frac{4}{2}\right) + (4-2)\right]$$

$$= \frac{1}{8}\ (6+2) = 1$$

Ahora $P(a < x < b) = \int_{a}^{b} f(x)dx$

$$P(y < 3.2) = \int_{2}^{3.2} f(x)dx = \int_{2}^{3.2} \frac{1}{8}(x+1)dx = \frac{1}{8}\left[\int_{2}^{3.2} xdx + \int_{2}^{3.2} dx\right]$$

$$= \frac{1}{8}\left[\left(\frac{(3.2)^2}{2} - \frac{4}{2}\right) + (3.2 - 2)\right] = \frac{1}{8}(3.12 + 1.2) = 0.54$$

$$P(2.9 < y < 3.2) = \int_{2.9}^{3.2} f(x)dx = \int_{2.9}^{3.2} \frac{1}{8}(x+1)dx = \frac{1}{8}\left[\int_{2.9}^{3.2} xdx + \int_{2.9}^{3.2} dx\right]$$

$$= \frac{1}{8}\left[\left(\frac{(3.2)^2}{2} - \frac{(2.9)^2}{2}\right) + (3.2 - 2.9)\right] = \frac{1}{8}(0.915 + 0.3) = 0.151875 \blacksquare$$

Ejemplo: El número de minutos de un vuelo de Piura a Trujillo se adelanta o atrasa con una variable aleatoria cuya densidad de probabilidad está dada por:

$$f(x)\begin{cases} \frac{1}{288}(36 - x^2) & para\ -6 < x < 6 \\ 0 & en\ cualquier\ otro\ lugar \end{cases}$$

Donde los valores negativos son indicativos de que un vuelo llegue adelantado y los valores positivos son del que el vuelo llegue retrasado. Encuentre la probabilidad de que cada uno de los vuelos llegara.

a.) Al menos 2 minutos adelantado

b.) Cualquier tiempo entre 1 y 3

Solución

a.) $\int_{-6}^{-2} \frac{1}{288}(36 - x^2)dx = \frac{1}{288}\left[\int_{-6}^{-2} 36dx - \int_{-6}^{-2} x^2dx\right]$

$$= \frac{1}{288}\left[\left(36(-2) - \frac{(-2)^3}{3}\right) - \left(36(-6) - \frac{(-6)^3}{3}\right)\right]$$

$$= \frac{1}{288}\left(\frac{-208}{3} + 144\right) = 0.2593$$

$$b.)\int_{-6}^{-1} \frac{1}{288}(36 - x^2)dx = \frac{1}{288}\left[\int_{-6}^{-1} 36dx - \int_{-6}^{-1} x^2dx\right]$$

$$= \frac{1}{288}\left[\left(36(-1) - \frac{(-1)^3}{3}\right) - \left(36(-3) - \frac{(-3)^3}{3}\right)\right]$$

$$= \frac{1}{288}\left(\frac{-107}{3} + 99\right) = 0.2199\blacksquare$$

La **función de distribución**, se define por

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Ejemplo: La duración de una vida total (en años) de perro de 5 años de una cierta raza es una variable aleatoria cuya función de distribución está dada por:

$$f(x) = \begin{cases} 0 & para\ \ x \leq 5 \\ 1 - \dfrac{25}{x^2} & para\ \ x > 5 \end{cases}$$

Encuentre la probabilidad de que un perro como esos de 5 años vivirá

a.) Más que 10 años

b.) Menos que 8 años

c.) Entre 12 y 15 años

Solución

x: Años de perro

a.) $P[x > 10] = 1 - P[x \leq 10] = 1 - \left(1 - \frac{25}{100}\right) = 1 - 0.75 = 0.25$

b.) $P[x < 8] = 1 - \frac{25}{x^2} = 1 - \frac{25}{64} = 0.6093$

c.) $P[12 \leq x \leq 15] = f(15) - f(12) = \left(1 - \frac{25}{15^2}\right) - \left(1 - \frac{25}{12^2}\right)$

$$= 0.8889 - 0.8263 = 0.0625\blacksquare$$

Como en este trabajo necesitamos describir variables aleatorias con medidas descriptivas similares a las que se tratan en Estadística Descriptiva, es decir mediante el valor esperado, desviación estándar o varianza las cuales pasamos a describir en la siguiente sección.

Valor esperado:

Sea X una variable aleatoria:

1. **Si X es discreta** y toma valores $x_1, x_2, \dots, x_n$ el valor esperado o esperanza matemática o promedio, se define por:

$$\mu = E(X) = \sum_{i \in \mathbb{N}} x_i P(X = x_i)$$

El valor esperado de un activo individual se define por:

$$E(R_j) = \sum_{i=1}^{n} p_i R_i$$

Ejemplo: El gerente de una bodega en una fábrica sabe, por haber estudiado sus registros, que la demanda diaria (número de veces que se usa) de cierta herramienta, tiene la siguiente distribución:

Demanda	0	1	2	3
Probabilidad	0.1	0.5	0.4	0.3

Determine el promedio y la varianza de la demanda en la bodega.

Solución

Se sabe que

$$E(X) = \sum_{i \in \mathbb{N}} x_i P(X = x_i)$$

$$E(X) = 0(0.1) + 1(0.5) + 2(0.4) + 3(0.3)$$

$$E(X) = 2.2$$

De donde se sigue la herramienta en un promedio de 2.2 veces diarias.

Por otra parte, usando la definición de la varianza

$$V(X) = E[(X - u)^2]P(X)$$

Se sigue que

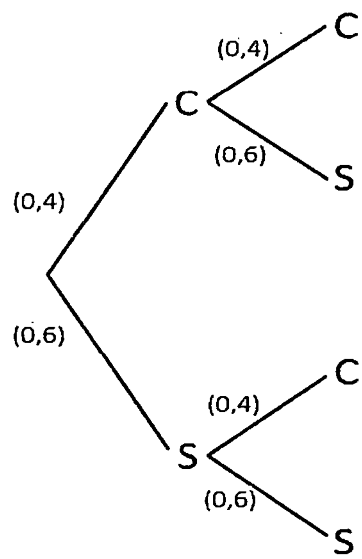
$$V[X] = (0 - 2.2)^2(0.1) + (1 - 2.2)^2(0.5) + (2 - 2.2)^2(0.4) + (3 - 2.2)^2(0.3)$$

$$V[X] = 0.48 + 0.72 + 0.016 + 0.192$$

$$V[X] = 1.408 \blacksquare$$

Ejemplo: La moneda de 25 centavos esta sesgada de manera que las probabilidades de caras y cruces son 0.40 y 0.60. Si se lanza 2 veces ¿Cuál es la covarianza de Z, el número de caras obtenidas en el primer tiro, W el número total de caras obtenidas?

Solución



$$0.4 \times 0.4 = 0.16$$
$$0.4 \times 0.6 = 0.24$$
$$0.6 \times 0.4 = 0.24$$
$$0.6 \times 0.6 = 0.36$$

W \ Z	0	1	
0	0.36		0.36
1	0.24	0.24	0.48
2		0.16	0.16
	0.6	0.4	

$$E(Z,W) = (0 \times 0 \times 0.36) + (1 \times 0 \times 0) + (1 \times 0 \times 0.24) + (1 \times 1 \times 0.24)$$

$$+(0 \times 2 \times 0) + (1 \times 2 \times 0.16) = 0.56$$

$$E(Z) = 0.4\ ; \quad E(W) = 0.8$$

$$COV(Z,W) = E(Z,W) - E(Z)E(W)$$

$$= 0.56 - (0.4)(0.8)$$

$$= 0.24\blacksquare$$

2. **Si X es continua,** el valor esperado o esperanza matemática o promedio, se define por:

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

Donde $f(x)$ es la función densidad de la variable aleatoria X.

Ejemplo: Obtenga el valor esperado de x cuya función de densidad de probabilidad esta dada por:

$$f(x)\begin{cases} \frac{1}{8}(x+1) & para\ \ 2 < x < 4 \\ 0 & en\ cualquier\ otro\ lugar \end{cases}$$

Solución

$$\int_2^4 x\left(\frac{1}{8}(x+1)\right)dx = \frac{1}{8}\left[\int_2^4 x^2dx\right] + \int_2^4 \frac{1}{8}xdx = \frac{1}{8}\left[\frac{4^3}{3} - \frac{2^3}{3}\right] + \frac{1}{8}\left[\frac{4^2}{2} - \frac{2^2}{2}\right]$$

$$= 18.6667 + 0.75 = 19.4167\blacksquare$$

La Desviación estándar:

Para muchos la palabra desviación estándar, no nos dice mucho. Sin embargo no se preocupe, es probable que si haya escuchado la palabra volatilidad del mercado, volatilidad del precio, lo cual significa que ya está familiarizado con el tema, ya que volatilidad lo podemos connotar como movimiento y significa lo mismo que desviación estándar sino que esta última palabra es usada en estricto sentido matemático.

¿Qué es desviación estándar?

Justamente la desviación estándar, en un conjunto de datos (precios en el caso del mercado de valores) es una medida de dispersión, que nos indica cuanto pueden alejarse los valores respecto al valor esperado (promedio), por lo tanto es útil para buscar probabilidades de que un evento ocurra, o en el caso del mercado bursátil, determinar entre que rango de precios puede moverse un determinado activo y determinar qué tipo de activos pueden ser más volátiles que otros.

Los operadores del mercado están interesados en la dirección del precio de un activo y en la velocidad de los movimientos del subyacente para determinar qué tan riesgoso o volátil puede llegar a ser un activo. Los mercados cuyos precios se mueven a alta velocidad son mercados de alta volatilidad.

Existen varias maneras de estimar la volatilidad o **desviación estándar**, por ejemplo:

Si X es **discreta** y toma valores $x_1, x_2, \dots, x_n, \dots$ la **desviación estándar o típica** se define por

$$\sqrt{\sigma_X^2} = \sqrt{\sum_{i \in \mathbb{N}} (x_i - \mu)^2 P(X = x_i)}$$

De otro lado:

$$\sum_{i} (x_i - \mu)^2 P(X = x_i) = \sum_{i \in \mathbb{N}} (x_i^2 - 2\, x_i \mu + \mu^2) P(X = x_i)$$

$$= \sum_{i \in \mathbb{N}} x_i^2 P(X = x_i) - 2\mu \sum_{i \in \mathbb{N}} x_i P(X = x_i) + \mu^2 \sum_{i \in \mathbb{N}} P(X = x_i)$$

Se sabe que $\mu = \sum_{i \in \mathbb{N}} x_i P(X = x_i)$; $\sum_{i \in \mathbb{N}} P(X = x_i) = 1$ de donde se tiene que:

$$\to \sum_{i \in \mathbb{N}} (x_i - \mu)^2 P(X = x_i) = \sum_{i \in \mathbb{N}} x_i^2 P(X = x_i) - 2\mu^2 + \mu^2$$

Por lo tanto:

$$\sqrt{\sigma_X^2} = \sqrt{\sum_{i} (x_i - \mu)^2 P(X = x_i)} = \sqrt{\sum_{i} x_i^2 P(X = x_i) - \mu^2}$$

La varianza, es simplemente la desviación estándar o típica elevada al cuadrado, es decir

$$\sigma_X^2 = \sum_{i} (x_i - \mu)^2 P(X = x_i) = \sum_{i} x_i^2 P(X = x_i) - \mu^2$$

Si X es **continua,** la **desviación estándar o típica** se define por

$$\sqrt{\sigma_X^2} = \sqrt{\int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx} = \sqrt{\int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2}$$

Y la **varianza** se define como

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{+\infty} x^2 f(x) - \mu^2$$

La **varianza** se interpreta de la misma manera que la desviación estándar: cuantifica la dispersión de una serie de datos. La interpretación de la desviación estándar y de la varianza es la misma, aunque obviamente las magnitudes serán distintas.

La varianza es mayor que la desviación estándar cuando la desviación estándar es mayor que uno.

La varianza es menor que la desviación estándar cuando la desviación estándar es menor que uno.

Como dato descriptivo es más frecuente el uso de la desviación estándar que el de la varianza (aunque se utilizan los dos)

Covarianza y coeficiente de correlación:

Cuando se trabajan con dos variables discretas, digamos por ejemplo, X e Y, se define como función de probabilidad conjunta $f(x_i, y_j)$ como la probabilidad de que ambas variables tomen los valores x_i, x_j simultáneamente, es decir

$$f(x_i, x_j) = [P(X = x_i) \cap P(Y = y_j)]$$

Los índices que reflejan la relación lineal entre las variables X e Y son los siguientes

La covarianza, cuando la variable aleatoria es discreta se define como:

$$\sigma_{XY} = \mu(XY) - \mu(X)\mu(Y); donde\ \mu(XY) = \sum_i \sum_j x_i\, x_j f(x_i, y_j)$$

Y su coeficiente de correlación se define como $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

El coeficiente de correlación ρ en la teoría de la cartera mide la relación que existe en las tasas de rendimiento de distintos activos a lo largo del tiempo.

ELEMENTOS BASICOS DE ECONOMIA:

Puesto que necesitamos elementos económicos en esta sección revisaremos algunas de sus definiciones básicas.

Inversionistas son las personas físicas o jurídicas que utilizan sus disponibilidades económicas para adquirir acciones o títulos negociables en el mercado financiero.

Bienes: objetos materiales que por sus características tienen la capacidad de satisfacer necesidades humanas.

Asociado con el consumo de un bien se tiene su utilidad, que es de dos tipos:

- **Utilidad total:** es la utilidad que proporciona toda la cantidad consumida del bien, es decir es la satisfacción total que tiene una persona por la posesión o consumo de un bien; esta aumenta a medida que se incrementa el número de unidades del bien.
- **Utilidad marginal:** Es el aumento de la utilidad total provocada o producida por un incremento de una unidad consumida, poseída o producida, es decir la utilidad marginal es decreciente en un periodo determinado.

Los bienes en los cuales estamos interesados son aquellos que se intercambian en los llamados **Mercados financieros**.

Un **Mercado financiero** es el conjunto de empresas que, dentro de un determinado marco, tratan de satisfacerse sus necesidades de pagos, movimiento de fondos, crédito y financiamiento de una parte y de reinversión de otra. En el mercado financiero las empresas vienen siendo la parte demandante y los bancos de la parte ofertante.

Existen los siguientes tipos de mercados financieros

- Mercado monetario
- Mercado de crédito
- Mercado que permite la movilización de los capitales de ahorro
- Mercado internacional de capitales

En este mercado financiero los bienes transados son las llamadas acciones, que son títulos o unidades de derecho de propiedad de una sociedad anónima de abierta o cerrada o de una sociedad encomiendita por acciones.

Pueden ser nominales o al portador, pudiendo diferenciarse series distintas por su valor nominal o por el contenido de sus derechos. Las acciones son títulos de renta variable y comprenden las acciones cotizadas en bolsa (acciones cotizadas), las acciones no cotizadas y otras formas de participación. Los valores de renta variable suelen generar ingresos en forma de dividendos.

En estas transacciones los accionistas corren riesgos, es decir la posibilidad que un evento salga fuera de los parámetros esperados. Desde el punto de vista de las finanzas un riesgo es la posibilidad de que el retorno de una operación tenga un valor diferente del retorno esperado. Sin embargo no todas las personas perciben el riesgo de la misma manera, de ahí que se han propuesto tres clasificaciones a las percepciones de las personas respecto del riesgo:

- **Adverso al riesgo:** Es la persona que no está dispuesta a aceptar más riesgo a menos que obtengan mayores retornos a cambio.
- **Indiferente al riesgo:** Es una persona que no tiene preferencias entre aceptar una inversión con un retorno y riesgo determinado a otra inversión con el mismo retorno pero un mayor nivel de riesgo.
- **Amante del riesgo:** es una persona que está dispuesta a aceptar mayores niveles de riesgo manteniendo igual el retorno de la inversión.

Estos riesgos son asumidos por los inversionistas en busca del mayor rendimiento de estas inversiones, el rendimiento se refiere al porcentaje de ganancia que se obtiene con respecto a la inversión durante cierto periodo de tiempo. Generalmente se genera un mayor rendimiento cuando se prolonga el plazo de la inversión. Algunos factores que reducen el nivel de rendimiento obtenido son las comisiones e impuestos requeridos para las transacciones y seguimiento de dicha inversión.

Si los inversionistas son personas jurídicas, es decir son empresas entonces los bienes transados reciben el nombre de activos, que son los bienes o derechos que la empresa posee y que pueden convertirse en dinero u otros medios líquidos equivalentes. Los activos que una empresa posee se clasifican dependiendo de su liquidez, es decir la facilidad, con la que ese activo puede convertirse en dinero. Por ello se dividen en:

- **Activo fijo:** son los activos utilizados en el negocio y no adquiridos con fines de venta, como maquinarias y bienes inmuebles.
- **Activo circulante:** son activos que se esperan que sean utilizados en un periodo inferior al año, como clientes o existencias.

Las cuentas de activos, tanto circulantes como fijos, se incluyen en el balance de situación.

La Cartera: Se denomina cartera a la combinación de activos o títulos financieros.

RIESGO Y RENDIMIENTO: TEORÍA DE LA ELECCIÓN

La teoría de la elección plantea que los inversionistas elegirán entre opciones que tienen diferentes combinaciones de riesgo y rendimiento.

Generalmente, para esta teoría el inversionista tiene aversión al riesgo.

Si suponemos que el riesgo puede medirse por medio de la desviación estándar del rendimiento (R), es decir $\sigma(R)$ y que el rendimiento se mide por el valor esperado $E(R)$.

El objetivo de la formación de carteras es reducir el riesgo mediante la diversificación; en otras palabras, podemos decir que la desviación estándar de los rendimientos sobre la cartera de activos $\sigma(R_p)$ puede ser menor que la suma de las desviaciones estándar provenientes de los activos individuales.

Pero, ¿Cómo se puede formar carteras que reduzcan el riesgo de un inversionista? La teoría de la cartera trata de la selección de carteras óptimas e eficientes, es decir, carteras que proporcionan el rendimiento más alto posible en cualquier grado específico de riesgo, o el riesgo más bajo posible en cualquier tasa de rendimiento. Entonces, para poder determinar las carteras óptimas debemos analizar los dos componentes elementales que las integran, a saber: rendimiento y riesgo.

La tasa de rendimiento de una cartera es el promedio ponderado de los rendimientos de los valores individuales de la cartera.

Rendimiento o retorno esperado de un activo cualquiera i

El Rendimiento o retorno de una inversión se mide como la ganancia o pérdida de valor experimentada en un periodo de tiempo determinado. El Rendimiento o retorno esperado del activo i , que se denota por $\bar{R}_i$, tiene que ver con las expectativas que se tiene hacia el futuro tomando en consideración los distintos escenarios de la economía. Se calcula usando la siguiente relación.

$$E(R_i) = \bar{R}_i = \sum R_{it} P_{it}$$

Dónde:

$$R_{it} \coloneqq \textit{Rendimiento del activo i cuando ocurre el evento t}$$

$$P_{it} \coloneqq \textit{Probabilidad de ocurrencia del rendimiento } R_{it}$$

Varianza de un activo cualquiera i

La varianza de un activo i, que se denota por σ_i^2 tiene que ver con la incertidumbre que tendrá el retorno de una inversión a lo largo del tiempo. Se calcula usando la siguiente relación.

$$\sigma^2(R_i) = \sigma_i^2 = \sum (R_{it} - \bar{R}_i)^2 P_{it}$$

Desviación estándar de un activo cualquiera i

La Desviación estándar de un activo cualquiera i, que se denota por σ_i, se define como la raíz cuadrada de la varianza del activo i. Es decir.

$$\sigma(R_i) = \sigma_i = \sqrt{\sigma_i^2}$$

Coeficiente de variación de un activo i

Se denota por V_i . Mide la dispersión de una variable aleatoria relativa a su valor esperado Se calcula usando la siguiente relación.

$$V(R_i) = V_i = \frac{\sigma_i}{R_i}$$

Ejemplo: El rendimiento de dos activos en una cierta economía se resume en el cuadro siguiente.

ESTADO DE LA ECONOMÍA	PROBABILIDAD DE OCURRENCIA	RENDIMIENTO DEL ACTIVO 1 (R_{1t})	RENDIMIENTO DEL ACTIVO (R_{2t})
MALO	0.1	40	50
BUENO	0.6	75	60
EXCELENTE	0.3	90	80

Determine el rendimiento esperado, la varianza, la desviación estándar y el coeficiente de correlación para cada uno de los activos.

Solución

Para el activo 1

Rendimiento esperado

$$E(R_1) = (0{,}1)(40) + (0{,}6)(75) + (0{,}3)(90)$$

$$\Rightarrow E(R_1) = 76$$

Varianza

$$\sigma_1^2 = (40 - 76)^2(0{,}1) + (75 - 76)^2(0{,}6) + (90 - 76)^2(0{,}3)$$

$$\Rightarrow \sigma_1^2 = (36)^2(0{,}1) + 0{,}6 + (14)^2(0{,}3)$$

$$\Rightarrow \sigma_1^2 = 189$$

Desviación estándar

$$\sigma_1 = \sqrt{\sigma_1^2}$$

$$\Rightarrow \sigma_1 = \sqrt{189}$$

$$\Longrightarrow \sigma_1 = 13{,}7477$$

Coeficiente de correlación

$$V_1 = \frac{\sigma_1}{\bar{R}_1}$$

$$\Longrightarrow V_1 = \frac{13{,}7477}{76}$$

$$\Longrightarrow V_1 = 0{,}1$$

Para el activo 2

Rendimiento esperado

$$E(R_2) = (0{,}1)(50) + (0{,}6)(60) + (0{,}3)(80)$$

$$\Longrightarrow E(R_2) = 65$$

Varianza

$$\sigma_2^2 = (50 - 65)^2(0{,}1) + (60 - 65)^2(0{,}6) + (80 - 65)^2(0{,}3)$$

$$\Longrightarrow \sigma_2^2 = (-15)^2(0{,}1) + (-5)^2(0{,}6) + (15)^2(0{,}3)$$

$$\Longrightarrow \sigma_2^2 = 105$$

Desviación estándar

$$\sigma_2 = \sqrt{\sigma_2^2}$$

$$\Longrightarrow \sigma_2 = \sqrt{96{,}3}$$

$$\Longrightarrow \sigma_2 = 9{,}81$$

Coeficiente de correlación

$$V_2 = \frac{\sigma_2}{\bar{R}_2}$$

$$\Longrightarrow V_2 = \frac{9{,}81}{65}$$

$$\Rightarrow V_2 = 0{,}15$$

Covarianza entre dos activos i y j

La covarianza entre dos activos i y j, que se denota por σ_{ij} indica la, manera en que estos dos activos están correlacionados, es decir nos indica cómo será el comportamiento del activo i ante la variación del activo j. Se calcula usando la siguiente relación.

$$\sigma_{ij} = Cov(R_i; R_j) = \sum (R_{it} - \bar{R}_i)(R_{jt} - \bar{R}_j)P_t$$

La covarianza cumple lo siguiente:

i) $\sigma_{ij} = \sigma_{ji}$

ii) $\sigma_{ii} = \sigma_i^2$

Coeficiente de correlación lineal

Para su cálculo se usa la siguiente relación.

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Este coeficiente tiene las siguientes propiedades:

i) $\rho_{ij} \in [-1 - ,1]$

ii) Si $\rho_{ij} = -1$ se dice que los rendimientos de los dos activos i,j tienen una correlación perfecta negativa, es decir que cuando uno de ellos crece el otro decrece en la misma proporción.

iii) Si $\rho_{ij} = 1$ se dice que los rendimientos de los dos activos i,j tienen una correlación perfecta positiva, es decir que cuando uno de ellos crece el otro crece en la misma proporción.

iv) Si $\rho_{ij} = 0$ se dice que los rendimientos de los dos activos i,j no están correlacionados, es decir no existe ninguna relación entre estos dos activos.

Rendimiento de una cartera p

La tasa de rendimiento de una cartera p, que se denota por R_P para dos activos, tres activos y en general para k activos se definen como sigue.

Para dos activos: $R_P = wR_1 + (1-w)R_2$ donde R_1 es el activo riesgoso y R_2 es el activo libre de riesgo.

Para tres activos: $R_P = w_1R_1 + w_2R_2 + w_3R_3\,; \sum_{i=1}^{3} w_i = 1$

Para k activos: $R_P = w_1R_1 + w_2R_2 + \cdots + w_kR_k\,; \sum_{i=1}^{k} w_i = 1$

Dónde:

$R_P = rendimiento\ de\ \ la\ cartera\ p$; $w_i = porecentaje\ invertido\ en\ cada\ uno\ de\ los\ activos\ y\ 0 \leq w_i \leq 1$

De donde la tasa esperada de rendimiento sobre la cartera es:

Para dos activos: $E(R_P) = wE(R_1) + (1-w)E(R_2)$

Para tres activos: $E(R_P) = w_1E(R_1) + w_2E(R_2) + w_3E(R_3\,); \sum_{i=1}^{3} w_i = 1$

Para k activos: $E(R_P) = w_1E(R_1) + w_2E(R_2) + \cdots + w_kE(R_k\,); \sum_{i=1}^{k} w_i = 1$

Con $E(R_i) = \sum_{j=1}^{n} p_jR_j$ y además que generalmente **el riesgo de una acción individual** se mide por la **desviación estándar** o **raíz cuadrada de la varianza** de dicha acción.

De lo anterior se deduce que el riesgo de una cartera no solo depende del riesgo de los valores que forman la cartera, sino también de la relación que existe entre los mismos. Esta relación se puede medir mediante la covarianza de los posibles rendimientos de los valores implicados. Para una cartera de dos valores el término de covarianza será el siguiente:

$$\boldsymbol{\sigma}_{ik} = \sum_{i=1}^{n} p_i[(R_i - E(R_i))(R_k - E(R_k))]$$

Y recordando la definición del coeficiente de correlación se tiene que

$$\boldsymbol{\sigma}_{ik} = \rho_{ik}\boldsymbol{\sigma}_i\boldsymbol{\sigma}_k$$

Dónde:

$$\boldsymbol{\sigma}_{ik} = Covarianza\ del\ valor\ i\ y\ el\ valo\ k$$

$$\rho_{ik} = coeficiente\ de\ correlacion\ o\ la\ correlacion\ esperada\ entre\ los\ rendimientos\ posibles\ para\ los\ valores\ i\ y\ k$$

$$\boldsymbol{\sigma}_i = desviación\ estándar\ del\ valor\ i$$

$$\boldsymbol{\sigma}_k = desviación\ estándar\ del\ valor\ k$$

Entonces, para una cartera de dos activos podríamos derivar el riesgo de la siguiente manera:

La varianza se define como:

$$var(R_p) = \sum_{i=1}^{n} p_i[R_p - E(R_p)]^2$$

Para dos activos:

$$var(R_p) = \sum_{i=1}^{n} p_i[(wR_1 + (1-w)R_2) - (wE(R_1) + (1-w)E(R_2))]^2$$

Acomodando los términos tenemos que:

$$var(R_p) = \sum_{i=1}^{n} p_i\big[(wR_1 - wE(R_1)\big) + ((1-w)R_2 - (1-w)E(R_2))]^2$$

Llamemos $a = (wR_1 - wE(R_1); b = (1-w)R_2 - (1-w)E(R_2)$; entonces tenemos que:

$$var(R_p) = \sum_{i=1}^{n} p_i(a+b)^2$$

$$var(R_p) = \sum_{i=1}^{n} p_i(a^2 + 2ab + b^2)$$

La fórmula completa de la varianza sería:

$$var(R_p) = \sum_{i=1}^{n} p_i[\big((wR_1 - wE(R_1)\big)^2 + 2((wR_1 - wE(R_1))((1-w)R_2 - (1-w)E(R_2)\,) + ((1-w)R_2 - (1-w)E(R_2)\,)^2)]$$

Reduciendo términos:

$$var(i) = \sum_{i=1}^{n} p_i\big((R_1 - E(R_1)\big)^2$$

$$var(k) = \sum_{i=1}^{n} p_i\big((R_2 - E(R_2)\big)^2$$

$$Cov(i,k) = \sigma_{ik} = \sum_{i=1}^{n} p_i[(R_1 - E(R_1))(R_2 - E(R_2))]$$

De donde usando el hecho que:

$$\boldsymbol{\sigma}_{ik} = \rho_{ik}\boldsymbol{\sigma}_i\boldsymbol{\sigma}_k$$

Se deduce que la varianza de la cartera se puede rescribir como:

$$var(R_p) = w^2 var(i) + 2w(1-w)\rho_{ik}\boldsymbol{\sigma}_i\boldsymbol{\sigma}_k + (1-w)^2 var(k)$$

Y para una cartera de j activos la varianza es la siguiente:

$$varianza\ de\ la\ cartera = \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j\, \boldsymbol{\sigma}_{ij}$$

Cuando $i = j; \boldsymbol{\sigma}_{ij}$ es la covarianza consigo mismos, lo que es igual a la definición de la varianza de cada una de las acciones que forman el portafolio, y cuando $i \neq j$ se trata de la covarianza de la acción i y la acción j.

Puesto que:

$$\sum_{j=1}^{k} w_i w_k \boldsymbol{\sigma}_{ik} = w_i w_1 \boldsymbol{\sigma}_{i1} + w_i w_2 \boldsymbol{\sigma}_{i2} + \cdots + w_i w_k \boldsymbol{\sigma}_{ik}$$

Entonces:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j\, \boldsymbol{\sigma}_{ij} = \sum_{i=1}^{k}\{w_i w_1 \boldsymbol{\sigma}_{i1} + w_i w_2 \boldsymbol{\sigma}_{i2} + \cdots + w_i w_k \boldsymbol{\sigma}_{ik}\}$$

Realizando los cálculos respectivos:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j\, \boldsymbol{\sigma}_{ij} = (w_1{}^2\sigma_{11} + w_1 w_2 \sigma_{12} + \cdots + w_1 w_k \sigma_{1k}) +$$
$$+(w_1 w_2 \sigma_{12} + w_2{}^2\sigma_{22} + \cdots + w_2 w_k \sigma_{2k}) + \cdots + (w_1 w_k \sigma_{1k} + w_2 w_k \sigma_{2k}$$
$$+ \cdots + w_k{}^2\sigma_{kk})$$

Usando los cálculos anteriores, se tiene que en términos matriciales la varianza de una cartera es la sumatoria de la suma de los renglones o columnas de la matriz que contiene las combinaciones de los j activos.

$$\begin{pmatrix}
w_1{}^2\sigma_{11} & w_1w_2\sigma_{12} & w_1w_3\sigma_{13} & w_1w_4\sigma_{14} & \cdots & w_1w_k\sigma_{1k} \\
w_1w_2\sigma_{12} & w_2{}^2\sigma_{22} & w_2w_3\sigma_{23} & w_2w_4\sigma_{24} & \cdots & w_2w_k\sigma_{2k} \\
w_1w_3\sigma_{13} & w_2w_3\sigma_{23} & w_3{}^2\sigma_{33} & w_3w_4\sigma_{34} & \cdots & w_3w_k\sigma_{3k} \\
w_1w_4\sigma_{14} & w_2w_4\sigma_{24} & w_3w_4\sigma_{34} & w_4{}^2\sigma_{44} & \cdots & w_4w_k\sigma_{4k} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
w_1w_k\sigma_{1k} & w_2w_k\sigma_{2k} & w_3w_k\sigma_{3k} & w_4w_k\sigma_{4k} & \cdots & w_k{}^2\sigma_{kk}
\end{pmatrix}$$

Dónde:

w_i Es el porcentaje invertido en cada una de las acciones, y σ_{ij} Es la covarianza de cada pareja de acciones.

En la diagonal principal de esta matriz aparecen las varianzas de las posibles combinaciones por pareja de los valores en la cartera, y fuera de la diagonal principal se encuentran sus covarianzas.

Si sustituimos la covarianza de las distintas combinaciones de acciones por el producto entre el coeficiente de correlación y las desviaciones estándar de las acciones individuales

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

La matriz anterior queda como sigue:

$$\begin{pmatrix} w_1{}^2\sigma_{11} & w_1w_2\rho_{12}\sigma_1\sigma_2 & w_1w_3\rho_{13}\sigma_1\sigma_3 & w_1w_4\rho_{14}\sigma_1\sigma_4 & \cdots & w_1w_k\rho_{1k}\sigma_1\sigma_k \\ w_1w_2\rho_{12}\sigma_1\sigma_2 & w_2{}^2\sigma_{22} & w_2w_3\rho_{23}\sigma_2\sigma_3 & w_2w_4\rho_{23}\rho_{24}\sigma_2\sigma_4 & \cdots & w_2w_k\rho_{2k}\sigma_2\sigma_k \\ w_1w_3\rho_{13}\sigma_1\sigma_3 & w_2w_3\rho_{23}\sigma_2\sigma_3 & w_3{}^2\sigma_{33} & w_3w_4\rho_{34}\sigma_3\sigma_4 & \cdots & w_3w_k\rho_{3k}\sigma_3\sigma_k \\ w_1w_4\rho_{14}\sigma_1\sigma_4 & w_2w_4\rho_{24}\sigma_2\sigma_4 & w_3w_4\rho_{34}\sigma_3\sigma_4 & w_4{}^2\sigma_{44} & \cdots & w_4w_k\rho_{4k}\sigma_4\sigma_k \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_1w_k\rho_{1k}\sigma_1\sigma_k & w_2w_k\rho_{2k}\sigma_2\sigma_k & w_3w_k\rho_{3k}\sigma_3\sigma_k & w_4w_k\rho_{4k}\sigma_4\sigma_k & \cdots & w_k{}^2\sigma_{kk} \end{pmatrix}$$

Dónde:

$$\rho_{ij} = coeficiente\ de\ correlación\ de\ orden\ cero\ de\ los\ probables\ rendimientos\ para\ los\ valores\ i\ y\ j$$

$$\sigma_i = desviación\ estandar\ del\ activo\ i$$

$$\sigma_j = desviación\ estandar\ del\ activo\ j$$

$$w_i = porcentaje\ invertido\ en\ el\ activo\ i$$

$$w_j = porcentaje\ invertido\ en\ el\ activo$$

$$\sigma_{ii} = varianza\ del\ activo\ i$$

De las ecuaciones y matrices anteriores se desprenden dos observaciones importantes. **La primera** es que la varianza de una cartera de activos riesgosos no es meramente las sumas de las varianzas respectivas, sino también está presente la covarianza entre los rendimientos de los activos; **la segunda** es que la varianza de una cartera de activos depende de los coeficientes de correlación, y es el valor de este coeficiente el que determina el conjunto de oportunidades de cartera de un inversionista.

DIVERSIFICACION DE MERCADOS:

Las carteras de títulos o portafolios son combinaciones de activos financieros que se realizan con la intención de reducir el riesgo que estos poseen individualmente. A este fenómeno se le conoce como diversificación y su eficiencia para reducir el riesgo se evidencia en las grandes diferencias que existe entre la desviación estándar de un título individual y la desviación estándar de una cartera o un índice.

Con la opción de diversificar, todo inversor adverso al riesgo invertirá en carteras antes que en activos individuales, dada la posibilidad de obtener las mismas rentabilidades pero con menor riesgo. Es por ello que el análisis de los títulos se hace desde el punto de vista de la contribución de cada título a la rentabilidad esperada y al riesgo de la cartera.

Dado que la rentabilidad esperada de la cartera es el promedio aritmético de las rentabilidades esperadas de los activos que la componen, la rentabilidad esperada de un título es la medida adecuada de la contribución del título a la rentabilidad esperada de la cartera. Sin embargo, ni la varianza ni la desviación estándar son medidas adecuadas de la contribución de un título al riesgo de una cartera cuando la correlación entre cada par de activos que componen la cartera no es perfecta.

Para poder entenderlo con mayor claridad conviene analizar la fórmula de la varianza de una cartera compuesta por 2 títulos, la misma es:

$$var(cartera) = {x_1}^2{\sigma_1}^2 + 2x_1x_2\sigma_{1,2} + {x_2}^2{\sigma_2}^2 = {x_1}^2{\sigma_1}^2 + 2x_1x_2\rho_{1,2}\sigma_1\sigma_2 + {x_2}^2{\sigma_2}^2$$

Donde x_i es la participación del título i en el portafolio, ${\sigma_i}^2$ es la varianza del titulo i, y $\rho_{i,j}$ es igual a 1 (correlación perfecta positiva)

Puede asimismo hallarse la medida de riesgo para una cartera con n títulos (por simplificación se utiliza notación matricial):

$$\sigma^2{}_{cart} = X^TVX$$

Siendo: $X = (x_1, x_2, \ldots, x_n)^T$ el vector de participación relativa de cada titulo de cartera,

$$V = \begin{pmatrix} {\sigma_1}^2 & \ldots & \sigma_{1,n} \\ \sigma_{2,1} & \ldots & \sigma_{2,n} \\ \sigma_{3,1} & \ldots & \sigma_{3,n} \\ \sigma_{4,1} & \cdots & \sigma_{4,n} \\ \vdots & \cdots & \vdots \\ \sigma_{n,1} & \cdots & {\sigma_n}^2 \end{pmatrix}$$

$$la\ matriz\ de\ varianzas\ y\ covarianzas$$

Es evidente que mientras mayor sea la cantidad de activos en la cartera, mayor es la influencia sobre ${\sigma^2}_{cart}$ de las covarianzas entre activos y menor la de las varianzas de cada uno. Esto demuestra que en carteras bien diversificadas lo que adiciona cada título al riesgo de la cartera depende en mayor medida de la forma en que varía su rentabilidad respecto a la de los otros activos que conforman la cartera y en la menor medida de la variación de la rentabilidad de ese título con su propia rentabilidad promedio.

Asimismo, como $\sigma_{i,j} = \sigma_{j,i}$ entonces la matriz V es simétrica y ${\sigma^2}_{cart}$ es una forma cuadrática. Esto permite establecer problemas de programación matemática, específicamente de programación cuadrática para resolver el problema de agentes adversos al riesgo que consideran minimizar su riesgo (representado por la varianza de la cartera) sujeto a obtener un mínimo de rentabilidad.

Por lo tanto, si se considera que $E(R_i)$ es la rentabilidad esperada del titulo i y R_{cart} es la rentabilidad mínima que se desea obtener de la cartera, entonces continuando con la notación anterior, se puede establecer el problema de selección de cartera de mínima variancia de acuerdo al siguiente programa cuadrático:

$$Minimizar \qquad {\sigma^2}_{cart} = X^T V X$$

$$sujeto\ a: E(R_1) + E(R_2) + \cdots + E(R_n) \geq R_{cart}$$

$$x_1 + x_2 + \cdots + x_n = 1$$

$$x_1, \quad x_2 \quad , \ldots, \quad x_n \geq 0$$

Donde la primera restricción obedece a la rentabilidad mínima deseada y la siguiente al requisito de inversión total de fondos.

Para poder resolver este tipo de problemas, especialmente si la cartera contiene una gran cantidad de activos, conviene utilizar algún software como el Excel de Microsoft.

APLICACIÓN EN SELECCIÓN DE CARTERAS DE ACTIVOS FINANCIEROS:

Un inversor adverso al riesgo desea invertir en una cartera compuesto por las acciones de las empresas Alicorp S.A.A, Unión de cervecerías Backus & Johnston S.A.A y Gloria S.A, las cuales representaremos por A, B y C respectivamente.

Los rendimientos de sus acciones entre los años 2009 y 2012, fueron obtenidos de los boletines del CEMTRUM de la Pontificia Universidad Católica del Perú y se resumen en las siguientes tablas.

Tabla 1: RENDIMIENTO DE LAS ACCIONES DE ALICORP S.A.A

ALICORP S.A.A	2009	2010	2011	2012
UTILIDAD NETA(miles PEN)	222 759	285 647	322 510	315 613
ACCIONES EN CIRCULACION(miles)	854 580	854 580	854 580	854 580
UTILIDAD POR ACCION(UPA)	0,261	0,334	0,377	0,369
RIESGO PAIS	1,14%	1,14%	1,14%	1,14%

Veamos cómo se calcula la utilidad por acción (UPA).

$$UPA_{2009} = \frac{222\,759}{854\,580} = 0{,}26066$$

$$UPA_{2010} = \frac{285\,647}{854\,580} = 0{,}3342$$

$$UPA_{2011} = \frac{322\,510}{854\,580} = 0{,}377$$

$$UPA_{2012} = \frac{315\,613}{854\,580} = 0{,}3693$$

De donde el Rendimiento durante los cuatro años viene dado por

$$R_A = \frac{0{,}261 + 0.334 + 0{,}377 + 0{,}369}{4} = \frac{0{,}603}{4} \Rightarrow R_A = 0{,}15$$

Tabla 2: RENDIMIENTO DE LAS ACCIONES DE UNIÓN DE CERVECERÍAS BACKUS & JOHNSTON S.A.A

UNIÓN DE CERVECERÍAS BACKUS & JOHNSTON S.A.A	2009	2010	2011	2012
UTILIDAD NETA(miles PEN)	488´ 297	519´ 243	730´ 551	948´ 709
ACCIONES EN CIRCULACION(miles)	76´ 046	76´ 046	76´ 046	76´ 046
UTILIDAD POR ACCION(UPA)(56% UN)	3,60	3,82	5,38	6,99
RIESGO PAIS	1,17%	1,17%	1,17%	1,17%

Del mismo modo que en la tabla 01 se obtienen los valores UPA para la Tabla 2.

$$UPA_{2009} = \frac{488'\,297}{76'046} \times \frac{56}{100} = 3{,}5958$$

$$UPA_{2010} = \frac{519'243}{76'046} \times \frac{56}{100} = 3{,}824$$

$$UPA_{2011} = \frac{730'\,551}{76'046} \times \frac{56}{100} = 5{,}3795$$

$$UPA_{2012} = \frac{948'\,709}{76'046} \times \frac{56}{100} = 6{,}986$$

De donde el Rendimiento durante los cuatro años viene dado por

$$R_B = \frac{3,6 + 3,82 + 5,38 + 6,99}{4} = \frac{19,79}{4} \Rightarrow R_B = 4,95$$

Tabla 3: RENDIMIENTO DE LAS ACCIONES DE GLORIA S.A

GLORIA S.A	2009	2010	2011	2012
UTILIDAD NETA(miles PEN)	202 181	226 117	195 099	241 622
ACCIONES EN CIRCULACION(miles)	282 007	282 007	282 007	421 619
UTILIDAD POR ACCION(UPA)	0,717	0,802	0,692	0,573
RIESGO PAIS	1,17%	1,17%	1,17%	1,17%

Cuyos valores UPA se obtienen como sigue.

$$UPA_{2009} = \frac{202\ 181}{282\ 007} = 0,7169$$

$$UPA_{2010} = \frac{226\ 117}{282\ 007} = 0,8018$$

$$UPA_{2011} = \frac{195\ 099}{282\ 007} = 0,6918$$

$$UPA_{2012} = \frac{241\ 622}{421\ 619} = 0,573$$

De donde el Rendimiento durante los cuatro años viene dado por

$$R_C = \frac{0,717 + 0,802 + 0,692 + 0,573}{4} = \frac{2.784}{4}$$

$$\Rightarrow R_C = 0,696$$

Los rendimientos $R_A, R_B\ y\ R_C$ fueron obtenidos con un promedio de riesgo país de

$$promedio\ riego\ pais = \frac{1,14 + 1,17 + 1,17}{3} = \frac{3,48}{3} = 1,16$$

Es decir cuando la economía peruana tenía un escenario de excelente.

Construcción de la Matriz de Varianzas y Covarianzas V

Puesto que en la construcción de la matriz V se necesita calcular los rendimientos de cada uno de los activos en cada uno de los escenarios posibles. Si se considera que en un escenario bueno de la economía, estos rendimientos están constituidos por 80% del rendimiento obtenidos durante el escenario excelente de la economía, y que los rendimientos para un escenario malo, están constituidos por el 40% de los rendimientos obtenidos en un escenario excelente para la economía. Se tiene que:

i) **Rendimientos de los activos en un escenario bueno**

$$R_A = 0{,}34 \times 0{,}8 = 0{,}272$$
$$R_B = 4{,}95 \times 0{,}8 = 3{,}96$$
$$R_C = 0{,}696 \times 0{,}8 = 0{,}56$$

ii) **Rendimientos de los activos en un escenario malo**

$$R_A = 0{,}34 \times 0{,}4 = 0{,}14$$
$$R_B = 4{,}95 \times 0{,}4 = 1{,}98$$
$$R_C = 0{,}696 \times 0{,}4 = 0{,}28$$

Considerando que en el contexto actual de la economía mundial, los tres posibles escenarios: Malo, Bueno y Excelente, tienen probabilidades de ocurrencia de: 0,1; 0,6 y 0,3 respectivamente, mediante el siguiente cuadro, es posible obtener el rendimiento esperado para cada activo en este escenario mundial de la economía, como se verá a continuación.

Tabla 4: Rendimientos de los activos A, B y C en el contexto mundial

ESTADO DE LA ECONOMÍA	PROBABILIDAD DE OCURRENCIA	RENDIMIENTO DEL ACTIVO A (R_{At})	RENDIMIENTO DEL ACTIVO B (R_{Bt})	RENDIMIENTO DEL ACTIVO C(R_{Ct})
MALO	0,1	0,14	1,98	0,28
BUENO	0,6	0,27	3,96	0,56
EXCELENTE	0,3	0,34	4,95	0,696

Rendimiento esperado del activo A

$$E(R_A) = (0,1)(0,14) + (0,6)(0,27) + (0,3)(0,34) = 0,014 + 0,162 + 0,102$$

$$\Longrightarrow E(R_A) = 0,278$$

Rendimiento esperado del activo B

$$E(R_B) = (0,1)(1,98) + (0,6)(3,96) + (0,3)(4,95) = 0,198 + 2,376 + 1,485$$

$$\Longrightarrow E(R_B) = 4,059$$

Rendimiento esperado del activo C

$$E(R_C) = (0,1)(0,28) + (0,6)(0,56) + (0,3)(0,696) = 0,028 + 0,336 + 0,2088$$

$$\Longrightarrow E(R_C) = 0,573$$

Varianza del activo A

$$\sigma_A^2 = (0,14 - 0,278)^2(0,1) + (0,27 - 0,278)^2(0,6) + (0,34 - 0,278)^2(0,3)$$

$$\Longrightarrow \sigma_A^2 = 0,0019044 + 0,0000384 + 0,0011532$$

$$\Longrightarrow \sigma_A^2 = 0,0031 \ldots..(*_1)$$

Desviación estándar del activo A

$$\sigma_A = \sqrt{\sigma_A^2}$$

$$\Longrightarrow \sigma_A = \sqrt{0,0031}$$

$$\Longrightarrow \sigma_A = 0,0557$$

Varianza del activo B

$$\sigma_B^2 = (1,98 - 4,059)^2(0,1) + (3,96 - 4,059)^2(0,6) + (4,95 - 4,059)^2(0,3)$$

$$\Longrightarrow \sigma_B^2 = 0,43 + 0,00588 + 0,238$$

$$\Longrightarrow \sigma_B^2 = 0,6763 \ldots.(*_2)$$

Desviación estándar del activo B

$$\sigma_B = \sqrt{\sigma_1^2}$$

$$\Rightarrow \sigma_B = \sqrt{0{,}6763}$$

$$\Rightarrow \sigma_B = 0{,}8224$$

Varianza del activo C

$$\sigma_C^2 = (0{,}28 - 0{,}573)^2(0{,}1) + (0{,}56 - 0{,}573)^2(0{,}6) + (0{,}696 - 0{,}573)^2(0{,}3)$$

$$\Rightarrow \sigma_C^2 = 0{,}00858 + 0{,}0001 + 0{,}0045$$

$$\Rightarrow \sigma_C^2 = 0{,}013225 \ldots\ldots (*_3)$$

Desviación estándar del activo C

$$\sigma_C = \sqrt{\sigma_1^2}$$

$$\Rightarrow \sigma_C = \sqrt{0{,}0132}$$

$$\Rightarrow \sigma_C = 0{,}1149$$

De donde se sigue qué.

i) $\sigma_{AB} = \sigma_A\sigma_B = (0{,}0557)(0{,}8224) \Rightarrow \sigma_{AB} = 0{,}0458$

ii) $\sigma_{AC} = \sigma_A\sigma_C = (0{,}0557)(0{,}1149) \Rightarrow \sigma_{AC} = 0{,}0064$

iii) $\sigma_{BC} = \sigma_B\sigma_C = (0{,}8224)(0{,}1149) \Rightarrow \sigma_{BC} = 0{,}0945$

Luego la matriz de varianzas y covarianzas asociada a estos activos es la siguiente:

$$V = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix}$$

Sustituyendo $(*_1), (*_2), (*_3), (i), (ii)\ y\ (iii)$ se sigue:

$$V = \begin{pmatrix} 0{,}0031 & 0{,}0458 & 0{,}0064 \\ 0{,}0458 & 0{,}6763 & 0{,}0945 \\ 0{,}0064 & 0{,}0945 & 0{,}0132 \end{pmatrix}$$

Puesto que el inversionista desea invertir todos los fondos y espera un rendimiento mínimo de la cartera del 3%, conociendo que los rendimientos esperados delos activos A, B y C son 0,278; 4,059 y 0,573, respectivamente.

Construcción de la función objetivo

Puesto que queremos minimizar el riesgo de la cartera entonces se debe minimizar la varianza de los activos A, B y C. Es decir debemos:

$$Minimizar \qquad \sigma^2{}_{cart} = X^T V X$$

Dónde:

$$X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Siendo $a, b \; y \; c$ las participaciones relativas de cada activo dentro de la cartera.

Luego

$$X^T V X \; = \begin{bmatrix} a \\ b \\ c \end{bmatrix}^T \begin{pmatrix} 0{,}0031 & 0{,}0458 & 0{,}0064 \\ 0{,}0458 & 0{,}6763 & 0{,}0945 \\ 0{,}0064 & 0{,}0945 & 0{,}0132 \end{pmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\Longrightarrow X^T V X$$

$$= \; (0{,}0031a + 0{,}0458b + 0{,}0064c \quad 0{,}0458a + 0{,}6763b + 0{,}0945c \quad 0{,}0064a + 0{,}0945b + 0{,}0132c) \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\Longrightarrow X^T V X = 0{,}0031a^2 + 0{,}0458ab + 0{,}0064ac + 0{,}0458ab + 0{,}6763b^2 + 0{,}0945bc$$
$$+ \, 0{,}0064ac + 0{,}0945bc + 0{,}0132c^2$$

$$\Longrightarrow X^T V X = 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$

De donde se obtiene que:

$$Minimizar \; 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$

Construcción de las restricciones

Como el inversionista desea invertir todos los fondos y espera un rendimiento mínimo de la cartera del 3%, conociendo que los rendimientos esperados delos activos A, B y C son 0,278, 4,059 y 0,573, respectivamente.

Se deducen las siguientes restricciones:

$$0{,}278a + 4{,}059b + 0{,}573c \geq 0{,}03$$

$$a + b + c = 1$$

$$a, b, c \geq 0$$

Planteamiento del problema

De las construcciones anteriores se deduce que debemos resolver el siguiente problema:

$$Minimizar\ 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$

Sujeto a:

$$0{,}278a + 4{,}059b + 0{,}573c \geq 0{,}03$$

$$a + b + c = 1$$

$$a, b, c \geq 0$$

Existencia de la Solución del problema

Para garantizar la solución del problema planteado se debe verificar que se cumplen las condiciones de KKT para lo cual primero se debe reescribir el problema de tal manera que las restricciones sean todas mayores o iguales dado que es un problema de minimización, a continuación se debe armar la función Lagrangiana. Es decir:

$$Minimizar\ 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$

Sujeto a:

$$0{,}278a + 4{,}059b + 0{,}573c \geq 0{,}03$$

$$a + b + c \geq 1$$

$$-a - b - c \geq -1$$

$$a, b, c \geq 0$$

De esta forma el Lagrangiano es:

$$L = 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$
$$+ \mu_1(1 - a - b - c) + \mu_2(-1 + a + b + c)$$
$$+ \mu_3(0{,}03 - 0{,}278a - 4{,}059b - 0{,}573c)$$

Derivando respecto a las variables de elección (a, b y c) y los multiplicadores de Lagrange y considerando el valor óptimo de las variables, se tiene:

$$\frac{\partial F}{\partial a} = 0{,}0062a + 0{,}0916b + 0{,}0128c - \mu_1 + \mu_2 - 0{,}278\mu_3$$

$$\frac{\partial F}{\partial b} = 0{,}0916a + 1{,}3526b + 0{,}189c - \mu_1 + \mu_2 - 4{,}059\mu_3$$

$$\frac{\partial F}{\partial c} = 0{,}0128a + 0{,}189b + 0{,}0264c - \mu_1 + \mu_2 - 0{,}573\mu_3$$

$$\frac{\partial F}{\partial \mu_1} = 1 - a - b - c$$

$$\frac{\partial F}{\partial \mu_2} = -1 + a + b + c$$

$$\frac{\partial F}{\partial \mu_3} = 0.03 - 0{,}278a - 4{,}059b - 0{,}573c$$

Como todas las variables de elección son distintas de cero en el óptimo, entonces las tres primeras derivadas deben ser iguales a cero, es decir:

$$\mu_1 - \mu_2 + 0{,}278\mu_3 = 0{,}0062a + 0{,}0916b + 0{,}0128c \ldots (*_1)$$

$$\mu_1 - \mu_2 + 4{,}059\mu_3 = 0{,}0916a + 1{,}3526b + 0{,}189c \ldots (*_2)$$

$$\mu_1 - \mu_2 + 0{,}573\mu_3 = 0{,}0128a + 0{,}189b + 0{,}0264c \ldots (*_3)$$

Luego, $-0{,}0854a - 1{,}261b - 0{,}1762c \leq 0$

$$\Rightarrow (0{,}0062 - 0{,}0916)a + (0{,}0916 + 1{,}3526)b + (0{,}0128 - 0{,}189)c \leq 0$$

$$\Rightarrow 0{,}0062a + 0{,}0916b + 0{,}0128c \leq 0{,}0916a + 1{,}3526b + 0{,}189c \ldots\ldots (1)$$

Si en $(*_1), (*_2)\ y\ (*_3)$ hacemos

$$k_1 = 0{,}0062a + 0{,}0916b + 0{,}0128c$$

$$k_2 = 0{,}0916a + 1{,}3526c + 0{,}189c$$

$$k_3 = 0{,}0128a + 0{,}189b + 0{,}0264c$$

De donde se obtiene:

$$\mu_1 - \mu_2 + 0{,}278\mu_3 = k_1 \Rightarrow \mu_1 - \mu_2 = k_1 - 0{,}278\mu_3 \ldots\ldots (2)$$

$$\mu_1 - \mu_2 + 4{,}059\mu_3 = k_2 \Rightarrow \mu_1 - \mu_2 = k_2 - 4{,}059\mu_3 \ldots\ldots (3)$$

De las ecuaciones $(2)\ y(3)$ se sigue de la ecuación (1) $k_1 \leq k_2 \Rightarrow k_2 - k_1 \geq 0 \ldots. (4)$

Por otro lado igualando las ecuaciones $(2)\ y(3)$ se obtiene:

$$k_1 - 0{,}278\mu_3 = k_2 - 4{,}059\mu_3 \Rightarrow \mu_3 = \frac{k_2 - k_1}{3{,}781}$$

$\Rightarrow$ Por la ecuación (4) se deduce que $\mu_3 \geq 0$

De la ecuación (2) se sigue que $\mu_1 = \mu_2 + k_1 - 0{,}278\mu_3$

$$\Rightarrow \mu_1 = \mu_2 + k_1 - 0{,}278\left(\frac{k_2 - k_1}{3{,}781}\right)$$

Puesto que podemos elegir $k_1 \leq 0{,}06848977581k_2 \Rightarrow \frac{0{,}278}{4{,}059}k_2 \geq k_1$

$$\Rightarrow 0{,}278k_2 \geq 4{,}059k_1 \Rightarrow \frac{0{,}278k_2 - 4{,}059k_1}{3{,}781} \geq 0$$

Luego podemos elegir $\mu_2 \geq \frac{0{,}278k_2 - 4{,}059k_1}{3{,}781} \geq 0 \Rightarrow \mu_2 + k_1 - 0{,}278\left(\frac{k_2 - k_1}{3{,}781}\right) \geq 0$

De donde $k_1 \geq 0$

En conclusión, se satisfacen las condiciones de Karush – Kuhn – Tucker, por lo tanto existe una solución óptima del problema planteado.

Solución del problema

La solución del problema fue hallada utilizando la herramienta de Solver de la planilla de cálculo de Excel de acuerdo a los siguientes pasos:

1.- Se utilizan nueve columnas: en las primeras seis columnas se establece el planteamiento del problema.

Las columnas siete y ocho (en nuestro caso las columnas I y J) es utilizada para establecer las formulas con las que trabaja Solver, por ello hay que utilizar formulas para el planteamiento de la función objetivo y las restricciones. Estas van a estar en función de las celdas que representan las variables; en nuestro ejemplo son las celdas C3, D3, E3, F3, G3 y H3.

Fórmula para la función objetivo:

$$= \mathrm{C4} * (\mathrm{C3} \wedge 2) + \mathrm{D4} * (\mathrm{D3} \wedge 2) + \mathrm{E4} * (\mathrm{E3} \wedge 2) + \mathrm{F4} * \mathrm{F3} + \mathrm{G4} * \mathrm{G3} + \mathrm{H4} * \mathrm{H3}$$

Fórmula para las restricciones:

Restricción 1: $= \mathrm{C6} * \mathrm{C3} + \mathrm{D6} * \mathrm{D3} + \mathrm{E6} * \mathrm{E3}$

Restricción 2: $= \mathrm{C7} * \mathrm{C3} + \mathrm{D7} * \mathrm{D3} + \mathrm{E7} * \mathrm{E3}$

2.- Paso seguido se selecciona Solver del menú de herramientas lo que despliega un cuadro de dialogo.

3.- En dicho cuadro de dialogo se deben seleccionar la celda objetivo (en nuestro caso es J3); si se desea maximizar, minimizar o dar un valor a la celda objetivo y las celdas que contienen las variables (C3, D3, E3, F3, G3 y H3) y finalmente se introducen las restricciones. Como Solver tiene una opción para las condiciones de no negatividad en el paso anterior estas no son introducidas.

4.- Posteriormente, se debe realiza un clic con el mouse sobre opciones lo cual despliega un nuevo cuadro de dialogo. Este contiene diversas opciones como tiempo máximo de cálculo, precisión, iteraciones, etc.

5.- En este cuadro lo relevante se ubica en la parte inferior izquierda. En primer lugar se debe desactivar (si es que esta con un tilde) la opción de Adoptar modelo lineal. Luego se debe activar con una tilde la opción Asumir no negativos que equivale a introducir las condiciones de no negatividad de las variables de elección. Finalmente en la sección titulada estimación se debe seleccionar GLN en lugar de SP lineal. Establecido todo lo anterior se da aceptar.

6.- Seguidamente, al volver al cuadro de Solver, se debe seleccionar Resolver.

7.- Finalmente Excel informa si se ha alcanzado una solución y da la opción a utilizar dicha solución hallada o volver a los valores iniciales de las celdas. Asimismo, da la posibilidad de obtener 3 informes: Respuestas, Sensibilidad y Límites.

En el ejercicio que se esta resolviendo, la solución dada por Excel es la siguiente:

A	B	C	AB	AC	BC		F.OBJETIVO	
0.80682315	0.0036982	0.18947965	0	0	0		0.00250115	
0.0031	0.6763	0.0132	0.0916	0.0128	0.189			
						L.IZQUIERDO	L. DERECHO	
0.278	4.059	0.573	0	0	0	0.34787967	0.03	R1
1	1	1	0	0	0	1.000001	1	R2

Con lo cual se tiene que la recomendación para el inversionista es que invierta 80,68% en acciones de la empresa ALICORP S.A.A, 0,37% en acciones de la empresa UNIÓN DE CERVECERÍAS BACKUS & JOHNSTON S.A.A y 18,95% en acciones de la empresa GLORIA S.A.

Usamos el algoritmo del gradiente, con condiciones de KKTT

$$Minimizar\ \ 0{,}0031a^2 + 0{,}0916ab + 0{,}0128ac + 0{,}189bc + 0{,}6763b^2 + 0{,}0132c^2$$

Sujeto a:

$$\mathrm{g}_1 = -0{,}278a - 4{,}059b - 0{,}573c + 0{,}03 \leq 0$$

$$\mathrm{g}_2 = \mathrm{a} + \mathrm{b} + \mathrm{c} - 1 = 0$$

$$\mathrm{g}_3 = -\mathrm{a} \leq 0$$

$$\mathrm{g}_4 = -\mathrm{b} \leq 0$$

$$\mathrm{g}_5 = -\mathrm{c} \leq 0$$

Si $x^* = \begin{bmatrix} a^* \\ b^* \\ c^* \end{bmatrix}$ es un mínimo local $\Longrightarrow\ \exists\ \{\mu_i,\ i = 1, \dots ,5\}\ talque$

1) $\mu_i \geq 0 \quad i = 1, \dots ,5$
2) $\mu_i \mathrm{g}_i = 0\ , i = 1, \dots ,5$
3) $\nabla f(a^*; b^*; c^*) + \sum \mu_i \nabla \mathrm{g}_i = 0$

$$\nabla f = (0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^*; 0{,}0916a^* + 1{,}3526b^* + 0{,}189c^*; 0{,}0128a^* + 0{,}189b^* + 0{,}0264c^*)$$

$$\nabla \mathrm{g}_1 = (-0{,}278; -4{,}059; -0{,}573)$$

$$\nabla \mathrm{g}_2 = (1; 1; 1)$$

$$\nabla \mathrm{g}_3 = (-1; 0; 0)$$

$$\nabla \mathrm{g}_4 = (0; -1; 0)$$

$$\nabla \mathrm{g}_5 = (0; 0; -1)$$

Las ecuaciones principales serían las siguientes:

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* - 0{,}278\mu_1 + \mu_2 - \mu_3 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* - 4{,}059\mu_1 + \mu_2 - \mu_4 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* - 0{,}573\mu_1 + \mu_2 - \mu_5 = 0$$

Ahora, empleamos las condición 2), debe satisfacer:

$\mu_1 g_1 = \mu_1(0{,}03 - 0{,}278a - 4{,}059b - 0{,}573c) = 0$

$\mu_2 g_2 = \mu_2(a + b + c - 1) = 0$

$\mu_3 g_{3=}\mu_3(-a) = 0$

$\mu_4 g_{4=}\mu_4(-b) = 0$

$\mu_5 g_{5=}\mu_5(-c) = 0$

Observe que en las tres últimas ecuaciones, a, b y c son números diferentes de cero, por lo cual, necesariamente, $\mu_3 = \mu_4 = \mu_5 = 0$

Finalmente, las ecuaciones principales serían

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* - 0{,}278\mu_1 + \mu_2 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* - 4{,}059\mu_1 + \mu_2 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* - 0{,}573\mu_1 + \mu_2 = 0$$

Teniendo en cuenta estas observaciones, se tienen los siguientes casos:

CASO I: si

$\mu_1 = \mu_2 = 0$

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* = 0$$

Resolviendo el sistema de ecuaciones:

a^*	b^*	c^*	R
0.0062	0.0916	0.0128	0
0.0916	1.3526	0.189	0
0.0128	0.189	0.0264	0

1	14.7741935	2.064516129	0
0.0916	1.3526	0.189	0
0.0128	0.189	0.0264	0

1	14.7741935	2.064516129	0
0	-0.00071613	-0.000109677	0
0	-0.00010968	-2.58065E-05	0

1	14.7741935	2.064516129	0
0	1	0.153153153	0
0	-0.00010968	-2.58065E-05	0

1	0	-0.198198198	0
0	1	0.153153153	0
0	0	-9.00901E-06	0

Luego los valores de $a^* = b^* = c^* = 0$

Reemplazamos en las condiciones:

$$0 \ngeq 0{,}03$$

$$0 \neq 1$$

No satisfacen las restricciones, por lo cual se descartan estos valores.

CASO II:

si $g_2 = 0$ y $\mu_1 = 0$ pero $g_2 = a^* + b^* + c^* - 1 \rightarrow a^* + b^* + c^* = 1$

Las ecuaciones serán:

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* + \mu_2 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* + \mu_2 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* + \mu_2 = 0$$

$$a^* + b^* + c^* = 1$$

Resolviendo el sistema de ecuaciones:

a^*	b^*	c^*	μ_2	**R**
0.0062	0.0916	0.0128	1	0
0.0916	1.3526	0.189	1	0
0.0128	0.189	0.0264	1	0
1	1	1	0	1

1	14.7741935	2.06451613	161.2903226	0
0.0916	1.3526	0.189	1	0
0.0128	0.189	0.0264	1	0
1	1	1	0	1

1	14.7741935	2.06451613	161.2903226	0
0	-0.00071613	-0.00010968	-13.77419355	0
0	-0.00010968	-2.5806E-05	-1.064516129	0
0	-13.7741935	-1.06451613	-161.2903226	1

1	0	-0.1981982	-284009.009	0
0	1	0.15315315	19234.23423	0
0	0	-9.009E-06	1.045045045	0

0	0	1.04504505	264774.7748	1

1	0	0	-307000	0
0	1	0	37000	0
0	0	1	-116000	0
0	0	0	386000	1

1	0	0	0	0.79533679
0	1	0	0	-0.09585492
0	0	1	0	0.30051813
0	0	0	1	2.5907E-06

Luego los valores de $a^* = 0.79533679$

$$b^* = -0.09585492$$

$$c^* = 0.30051813$$

$$\mu_2 = 2.5907\mathrm{E} - 06$$

Al reemplazamos en las condiciones, los valores satisfacen las desigualdades,

$$0.004225389 \geq 0{,}03$$

$$1 = 1$$

Pero el valor de b^* es negativo, por lo cual se descartan estos valores.

CASO III:

si $g_1 = 0$ y $\mu_2 = 0$ pero $g_1 = -0{,}278a - 4{,}059b - 0{,}573c + 0{,}03$

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* - 0{,}278\mu_1 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* - 4{,}059\mu_1 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* - 0{,}573\mu_1 = 0$$

$$0{,}278a^* + 4{,}059b^* + 0{,}573c^* = 0{,}03$$

Resolviendo el sistema de ecuaciones:

a^*	b^*	c^*	μ_1	**R**
0.0062	0.0916	0.0128	-0.278	0
0.0916	1.3526	0.189	-4.059	0
0.0128	0.189	0.0264	-0.573	0
0.278	4.059	0.573	0	0.03

1	14.7741935	2.06451613	-44.83870968	0
0.0916	1.3526	0.189	-4.059	0
0.0128	0.189	0.0264	-0.573	0
0.278	4.059	0.573	0	0.03

1	14.7741935	2.06451613	-44.83870968	0
0	-0.00071613	-0.00010968	0.048225806	0
0	-0.00010968	-2.5806E-05	0.000935484	0
0	-0.04822581	-0.00093548	12.46516129	0.03

1	0	-0.1981982	950.0900901	0

0	1	0.15315315	-67.34234234	0
0	0	-9.009E-06	-0.00645045	0
0	0	0.00645045	9.217522523	0.03

1	0	0	1092	0
0	1	0	-177	0
0	0	1	716	0
0	0	0	4.599	0.03

1	0	0	0	-7.12328767
0	1	0	0	1.15459883
0	0	1	0	-4.67058056
0	0	0	1	0.00652316

Luego los valores de $a^* = -7.12328767$

$$b^* = 1.15459883$$

$$c^* = -4.67058056$$

$$\mu_1 = 0.0065231612$$

Los valores de a^* y c^* son negativos, por lo cual se descarta este resultado.

CASO IV:

si $g_1 = g_2 = 0$

$$g_1 = -0{,}278a^* - 4{,}059b^* - 0{,}573c^* + 0{,}03 = 0$$

$$g_2 = a^* + b^* + c^* - 1 = 0$$

Luego

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* - 0{,}278\mu_1 + \mu_2 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* - 4{,}059\mu_1 + \mu_2 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* - 0{,}573\mu_1 + \mu_2 = 0$$

$$0{,}278a^* + 4{,}059b^* + 0{,}573c^* = 0{,}03$$

$$a^* + b^* + c^* = 1$$

Resolviendo el sistemas de ecuaciones

a^*	b^*	c^*	μ_1	μ_2	
0.0062	0.0916	0.0128	-0.278	1	0
0.0916	1.3526	0.189	-4.059	1	0
0.0128	0.189	0.0264	-0.573	1	0
0.278	4.059	0.573	0	0	0.03
1	1	1	0	0	1

1	14.7741935	2.06451613	-44.8387097	161.290323	0
0	-0.00071613	-0.00010968	0.04822581	-13.7741935	0
0	-0.00010968	-2.5806E-05	0.00093548	-1.06451613	0
0	-0.04822581	-0.00093548	12.4651613	-44.8387097	0.03
0	-13.7741935	-1.06451613	44.8387097	-161.290323	1

1	0	-0.1981982	950.09009	-284009.009	0
0	1	0.15315315	-67.3423423	19234.2342	0
0	0	-9.009E-06	-0.00645045	1.04504505	0
0	0	0.00645045	9.21752252	882.747748	0.03
0	0	1.04504505	-882.747748	264774.775	1

1	0	0	1092	-307000	0
0	1	0	-177	37000	0
0	0	1	716	-116000	0
0	0	0	4.599	1631	0.03
0	0	0	-1631	386000	1

1	0	0	0	-694269.406	-7.123287671
0	1	0	0	99771.6895	1.154598826
0	0	1	0	-369923.896	-4.670580561
0	0	0	1	354.642314	0.006523157
0	0	0	0	964421.613	11.63926941

1	0	0	0	0	1.255608601
0	1	0	0	0	-0.049511033
0	0	1	0	0	-0.206097568
0	0	0	1	0	0.006523078
0	0	0	0	1	1.20687E-05

Luego los valores de $a^* = 1.255608601$

$$b^* = -0.049511033$$

$$c^* = -0.206097568$$

$$\mu_1 = 0.006523078$$

$$\mu_2 = 1.20687\mathrm{E} - 05$$

Los valores de b^* y c^* son negativos, por lo cual se descarta este resultado.

CASO V:

$g_1 = g_2 = 0$ y $\mu_2 = 0$

$$0{,}0062a^* + 0{,}0916b^* + 0{,}0128c^* - 0{,}278\mu_1 = 0$$

$$0{,}0916a^* + 1{,}3526b^* + 0{,}189c^* - 4{,}059\mu_1 = 0$$

$$0{,}0128a^* + 0{,}189b^* + 0{,}0264c^* - 0{,}573\mu_1 = 0$$

Entonces $g_1 = g_2 = 0$

$$0{,}278a^* + 4{,}059b^* + 0{,}573c^* = 0{,}03 + h \ldots\ldots\ldots\ldots (1)$$

$$a^* + b^* + c^* = 1 \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Despejando "α" en la ecuación (2) se tiene:

$$a^* = 1 - b^* - c^*$$

Reemplazamos el valor de "α" en la ecuación (1)

$$0{,}278(1 - b^* - c^*) + 4{,}059b^* + 0{,}573c^* = 0{,}03 + h$$

$$0{,}278 + 3{,}781b^* + 0{,}295c^* = 0{,}03 + h$$

Luego despejamos "b"

$$b^* = -0{,}06559111 - 0{,}07802169C + 0{,}26448h \ \ldots\ldots (3)$$

Despejando el valor de b en las ecuaciones principales se tiene:

$$0.0062a^* + 0.00565321c^* - 0.278\mu_1 = 0.005031759$$

$$0.0916a^* + 0.083467866c^* - 4.059\mu_1 = 0.074449098$$

$$0.0128a^* + 0.0116539c^* - 0.573\mu_1 = 0.010397249$$

Resolviendo el sistema de ecuaciones:

a^*	c^*	μ_1	R
0.0062	0.00565321	-0.278	0.00503176
0.0916	0.083467866	-4.059	0.0744491
0.0128	0.0116539	-0.573	0.01039725

1	0.911808065	-44.83870968	0.81157406
0.0916	0.083467866	-4.059	0.0744491
0.0128	0.0116539	-0.573	0.01039725

1	0.911808065	-44.83870968	0.81157406
0	-5.3753E-05	0.048225806	0.00010891
0	-1.7243E-05	0.000935484	9.1008E-06

1	0.911808065	-44.83870968	0.81157406
0	1	-897.1731352	-2.02619309
0	-1.7243E-05	0.000935484	0

1	0	773.2109903	2.65907327
0	1	-897.1731352	-2.02619309
0	0	-0.014534675	-3.4938E-05

1	0	0	0.80044715
0	1	0	0.13041027
0	0	1	0.00240378

Entonces los valores de $a^* = 0.80044715$

$c^* = 0.13041027$

$\mu_1 = 0.130410274$

Despejando los valores de a y c en la ecuación (3) se tiene:

$$b^* = 0.069142581$$

Reemplazamos en las condiciones:

$$0.57789913 \geq 0,03$$

$$1 = 1$$

$a^*, b^*\ y\ c^* \geq 0$

Cumple las condiciones

Por lo tanto el punto $x^* = (0.80044715; 0.069142581; 0.13041027)$ es solución óptima

CONCLUSIONES

1. Las condiciones de Karush – Kuhn – Tucker proporcionan las condiciones necesarias y suficientes para la existencia problema de optimización no lineal con restricciones de no negatividad
2. Usando las varianzas y covarianzas de una cartera es posible establecer el problema de selección de cartera de mínima variancia
3. El problema de selección de cartera de mínima variancia para una cartera de inversión es un problema de programación cuadrática.
4. La solución del problema de selección de cartera de mínima variancia puede ser hallada utilizando la herramienta de Solver de la planilla de cálculo de Excel.
5. Empleando el algoritmo del gradiente, con las condiciones KKT, los valores los a, b y c varían menos del 0.1% en cada una de las variables, por lo cual es aceptable los valores obtenidos. Precisamente, el valor de "a " disminuye en 0.0064%, el valor de "b " aumenta en 0.065% y el valor de "c " disminuye en 0.059%

BIBLIOGRAFÍA

[1]Bazaraa M. y Sherali H.: Nonlinear programming: Theory and algorithms. Wiley. Third Edition. 2006.

[2]Bellman, R. : "Introduction to Matrix Analysis", McGraw-Hill Book Company, New York,1970

[3] Chiang, A.: Métodos fundamentales de economía matemática, 3.ª edición. McGraw-Hill, México, 1998.

[4]Mas-Colell, A., M.D. Winston, J.R. Green Microeconomic theory, Oxford University Press. 1995.

[5]Munkres, J.R: Topología, Segunda Edición, Prentice Hall, Madrid, 2002.

[6]Royden, H.L., Real Analysis, the MacMillan Company, New York 1971.

[7]Strang, G.: Algebra lineal y sus aplicaciones. Addison-Wesley Iberoamericana. 1986.

[8] Simmons, D.M., Nonlinear Programming for Operations Research, Prentice-Hall Inc., New Jersey (1975).

[9] Ochoa, S.I.: El Modelo de Markowitz en la Teoría de Portafolios de Inversión, tesis para optar grado de maestría, Instituto Politécnico, México, 2008