Interpretación geométrica de las condiciones de optimalidad para funciones de dos variables

Geometric interpretation of the optimality conditions for functions of two variables

Oscar Piñón Jiménez,[1]* Aldo R. Sartorius Castellanos[1]

[1] Instituto Tecnológico de Minatitlán.
Bulevar Institutos Tecnológicos s/n, col. Buena Vista Norte. CP 96848. Minatitlán, Veracruz, México
* Correo-e: opinon@itmina.edu.mx

PALABRAS CLAVE:

valores críticos, valores extremos, valores característicos, transformación lineal

RESUMEN

La interpretación geométrica de las condiciones de optimalidad de primer y segundo orden para una función de una variable, se extiende y adapta para funciones de dos variables a través del uso de los conceptos de derivada direccional y formas cuadráticas. Se argumenta sobre la interpretación de la derivada direccional para justificar gráficamente las condiciones de primer y segundo orden. En la de segundo orden se usan los conceptos de diagonalización de matrices y formas cuadráticas para explicar la relación de la derivada direccional de segundo orden con la matriz Hessiana.

KEYWORDS:

critical values, extreme values, characteristic values, linear transformation

ABSTRACT

The geometric interpretation of the optimality conditions of first and second order for a function of one variable, extends and adapts to functions of two variables through the use of the concepts of directional derivative and quadratic forms. It is argued on the interpretation of the directional derivative to graphically justify conditions first and second order. In the second order, the concepts of diagonalization of matrices and quadratic forms are used to explain the relation of the second order directional derivative and the Hessian matrix.

1 INTRODUCCIÓN

La representación gráfica de una función de una variable independiente en un sistema de coordenadas rectangulares es una curva plana. De la interpretación geométrica de la derivada como la pendiente de la recta tangente a la curva, es claro que, en los valores extremos relativos de la función, máximo o mínimo, la recta tangente es horizontal y de pendiente igual a cero, como se muestra en la figura 1, en donde se observa un valor máximo en $x=-3$ y un valor mínimo en $x=3$.

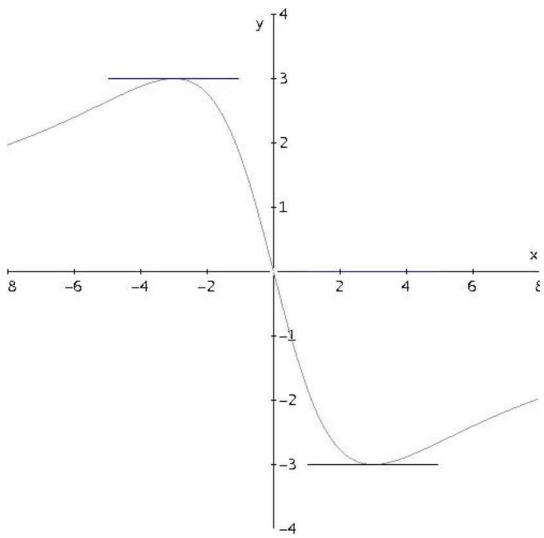

Figura 1. Gráfica de una función que tiene tangentes de pendiente cero en sus valores extremos

De la gráfica se justifica que analíticamente la ubicación de estos valores extremos se realice al igualar a cero la derivada y resolviendo la ecuación resultante para hallar los valores críticos; sin embargo, los valores críticos que se obtienen de esta forma no necesariamente ubican los valores máximos o mínimos. De hecho, la gráfica de la función también puede tener un comportamiento como el de la figura 2, donde existe un punto donde la pendiente de la recta tangente es igual a cero y no es un valor máximo o mínimo, sino un punto de inflexión.

La existencia de rectas de pendiente cero en los valores extremos y en un punto de inflexión hace que sea necesario examinar el comportamiento del signo de la derivada de la función a evaluar en la vecindad del punto crítico, para determinar si se trata de un valor máximo, mínimo o punto de inflexión.

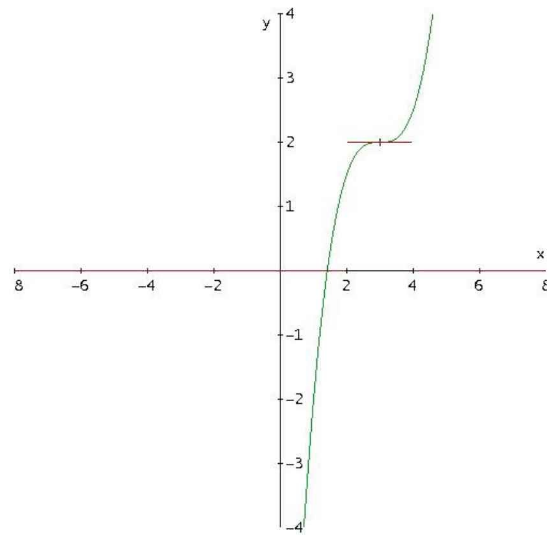

Figura 2. Función de una variable con un punto de inflexión en x=3

- Si el signo de la derivada pasa de negativo para valores menores al crítico a positivo para valores mayores, hay un valor mínimo en el punto crítico.
- Si el signo de la derivada pasa de positivo para valores menores al crítico a negativo para valores mayores, hay un valor máximo en el valor crítico.
- Si el signo de la derivada en la vecindad del punto crítico no cambia para valores menores y mayores al crítico, hay un punto de inflexión.

Las afirmaciones anteriores constituyen lo que se conoce como el criterio de la primera derivada para la determinación de valores extremos de una función de una variable independiente, y se pueden justificar si se observa la gráfica de la derivada de la función de la figura 1, mostrada en la figura 3, donde se nota que pasa de valores positivos a negativos en la vecindad del punto crítico que corresponde con un valor máximo y de valores negativos a positivos en el caso de un valor mínimo.

De manera similar, en la gráfica de la derivada de la función de la figura 2, mostrada en la figura 4, se nota que el signo en la vecindad del valor crítico no cambia si se trata de un punto de inflexión.

El criterio de la segunda derivada [1] utiliza la derivada de segundo orden de la función y examina el signo de la misma al evaluarla en el valor crítico.

- El signo de la segunda derivada es negativo en un extremo relativo máximo.

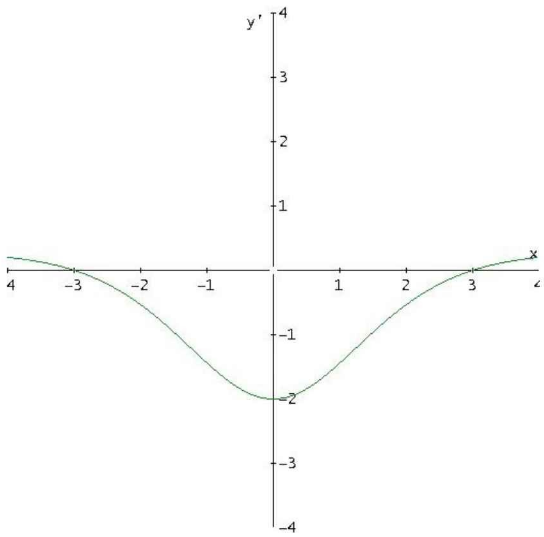

Figura 3. Gráfica de la derivada de la función de la figura 1

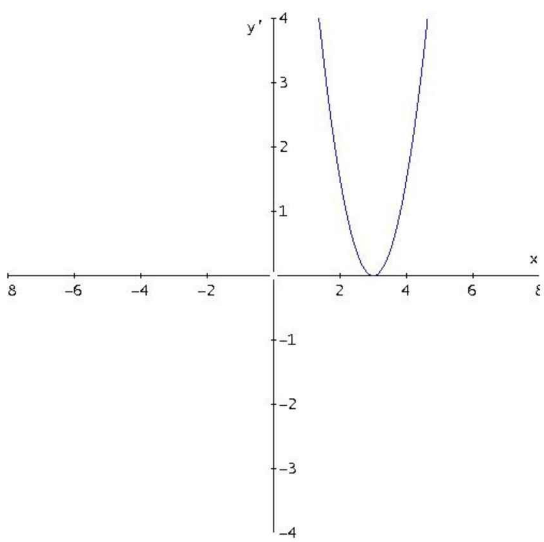

Figura 4. Gráfica de la derivada de la función de la figura 2

- El signo de la segunda derivada es positivo en un extremo relativo mínimo.
- Si se trata de un valor crítico, obtenido de igualar a cero la primera derivada, y la derivada de segundo orden es igual a cero, es necesario determinar la concavidad de la curva (signo de la segunda derivada) en la vecindad del valor crítico para determinar si es un punto de inflexión, o bien, un máximo o mínimo.

El análisis con la ayuda de las gráficas de las condiciones de primer y segundo orden es un recurso muy utilizado en la literatura relacionada con el caso de funciones de una variable independiente; sin embargo, en el caso de funciones de varias variables no es usual el enfoque geométrico para explicar estas condiciones de optimalidad.

Por lo anterior, en este trabajo se usa como referencia el análisis de una sola variable para ampliar su interpretación y aplicación al caso de funciones de dos variables. Este propósito requiere la aplicación de los conceptos: derivada direccional y formas cuadráticas, como se detalla en los siguientes apartados.

2 INTERPRETACIÓN GEOMÉTRICA DE LA CONDICIÓN DE PRIMER ORDEN $\nabla F(X*)=0$

La gráfica de una función de dos variables es una superficie en el espacio geométrico tridimensional y, como ya se ha señalado, la aplicación del concepto de derivada direccional es importante para la explicación de la condición de optimalidad de primer orden.

En la figura 5 se muestra la gráfica de una función de dos variables. Se puede notar que en sus puntos es posible trazar rectas tangentes a la superficie en diferentes direcciones, y si se trata de un valor extremo máximo, las rectas tangentes en todas las direcciones posibles deben ser horizontales o paralelas al plano *xy*. Es decir, la derivada direccional en todas las direcciones debe ser igual a cero.

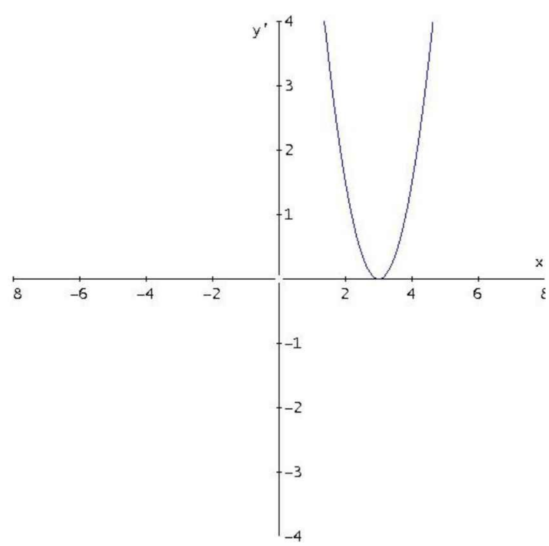

Figura 5. Función de dos variables con tangentes de pendiente cero en un valor extremo máximo

En la figura 6, la gráfica de la función de dos variables muestra que, también en un valor extremo

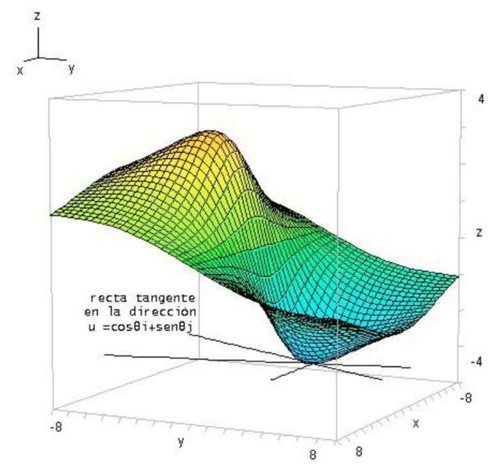

Figura 6. Función de dos variables con tangentes de pendiente cero en un valor extremo mínimo

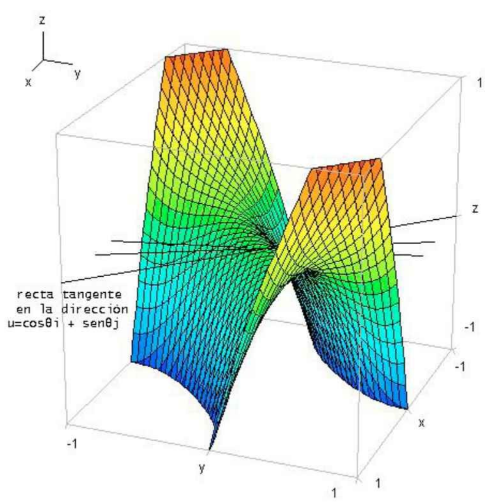

Figura 7. Función de dos variables con tangentes de pendiente cero en un punto silla

mínimo, las rectas tangentes a la superficie en todas las direcciones posibles deben ser horizontales o paralelas al plano *xy*.

En la figura 7 se observa un punto silla de una función de dos variables, donde el comportamiento es semejante al punto de inflexión de una función de una sola variable; por lo que se verifica nuevamente que la derivada direccional en todas las direcciones posibles es igual a cero.

Ahora bien, de la expresión de la derivada direccional en la dirección de un vector unitario *u* (1).

$$D_u f(x,y)=\partial f/\partial x\ cos\theta+\partial f/\partial y\ sen\theta \qquad (1)$$

Donde el vector unitario en el plano xy, que define la dirección en la que se deriva es:

$$u=cos\theta i+sen\theta j$$

Puesto que las componentes del vector unitario son funciones seno y coseno, que son continuas y tienen únicamente valores en el intervalo de -1 a 1; dependiendo del ángulo de la dirección en la que se deriva y considerando la expresión (1), concluimos que la derivada en todas las direcciones es cero, si $\partial f/\partial x$ y $\partial f/\partial y$ son iguales a cero. Es decir el gradiente de la función en los llamados valores críticos debe verificar $\nabla f(x*)=0$ (todas las derivadas parciales de primer orden de la función son iguales a cero en los puntos críticos).

Así, la interpretación geométrica de la derivada direccional justifica que, en el caso de las funciones de dos variables, la localización de los puntos críticos en donde pueden existir valores extremos o puntos silla, se realice igualando a cero todas las derivadas parciales y resolviendo el sistema de ecuaciones.

3 INTERPRETACIÓN GEOMÉTRICA DE LA CONDICIÓN DE SEGUNDO ORDEN ∇^2 F(X*)

Como se ha mencionado, en el caso de una sola variable, la exploración del signo de la segunda derivada evaluada en los valores críticos es un criterio que permite identificar si en éstos existe un valor máximo, mínimo o punto de inflexión. En el caso de una función de dos variables, se tiene que examinar el signo de la segunda derivada direccional en todas las direcciones; es decir, la derivada direccional en una dirección dada de la derivada direccional en esta misma dirección:

$$D_u(D_u f(x,y))=\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}cos\theta+\frac{\partial f}{\partial y}sen\theta\right)cos\theta+\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}cos\theta+\frac{\partial f}{\partial y}sen\theta\right)sen\theta$$

La cual simplifica a la siguiente expresión:

$$D_u(D_u f(x,y))=\frac{\partial^2 f}{\partial x^2}(cos\theta)^2+\frac{\partial^2 f}{\partial x\partial y}sen\theta cos\theta+\frac{\partial^2 f}{\partial y\partial x}cos\theta sen\theta+\frac{\partial^2 f}{\partial y^2}(sen\theta)^2$$

A su vez, la equivalencia de las derivadas mixtas de segundo orden reduce la expresión anterior a:

$$D_u(D_u f(x,y)) = \frac{\partial^2 f}{\partial x^2}(cos\theta)^2 + 2\,\frac{\partial^2 f}{\partial x \partial y} cos\theta sen\theta + \frac{\partial^2 f}{\partial y^2}(sen\theta)^2 \quad (2)$$

El análisis de esta ecuación (2) debe considerar el signo de la segunda derivada direccional en todas las direcciones posibles evaluada en los valores críticos y considerando que cada dirección es semejante al caso una variable. El criterio de la segunda derivada para las funciones de dos variables es:

- Si el signo de la segunda derivada direccional en todas las direcciones posibles en el valor crítico evaluado es negativo, la función tiene un valor extremo en este punto y es un máximo local estricto.
- Si el signo de la segunda derivada direccional en todas las direcciones posibles en el valor crítico evaluado es positivo, la función tiene un valor extremo en este punto y es un mínimo local estricto.
- Si la segunda derivada direccional evaluada en el valor crítico que se examina produce valores negativos en ciertas direcciones y valores positivos en otras, quiere decir que la función que se analiza en algunas direcciones es cóncava y en otras convexa; por lo tanto, se concluye que se trata de un punto silla que, por su forma geométrica, cumple con la unión de las dos posibilidades: cóncava en unas direcciones y convexa en otras.
- Si la segunda derivada direccional tiene direcciones donde produce resultados iguales a cero y valores positivos, o bien iguales a cero y valores negativos, se requiere examinar el signo de la segunda derivada direccional para puntos en la vecindad del valor crítico; o en su caso, explorar directamente los valores de la función para determinar si se trata de un punto silla, máximo o mínimo.

Hasta este punto, se ha establecido el criterio de la segunda derivada direccional con el concepto de la derivada direccional y con las gráficas de funciones de dos variables; sin embargo, en la literatura de optimización, el criterio que utiliza derivadas parciales de segundo orden se encuentra de la siguiente manera [2]:

Teorema: sea $x*$ un punto crítico de una función de dos variables o más con derivadas de segundo orden continuas, entonces:

- Si la hessiana de *f* representa una forma cuadrática definida negativa (positiva) en $x*$, entonces $x*$ es máximo (mínimo) local estricto.
- Si la hessiana de *f* es semidefinida negativa (positiva) en $x*$, y en un entorno del punto $x*$, entonces $x*$ es máximo (mínimo) local.
- Si la hessiana de *f* es indefinida en $x*$, entonces $x*$ es un punto silla.

Para mostrar la relación de este criterio en términos de la matriz hessiana, con el criterio enunciado en términos de la derivada direccional, es necesario reconocer que el lado derecho de la ecuación (2) se puede escribir en una forma cuadrática (3), donde las variables son *cosθ*, *senθ* :

$$D_u(D_u f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \begin{bmatrix} cos\theta \\ sen\theta \end{bmatrix} \cdot \begin{bmatrix} cos\theta \\ sen\theta \end{bmatrix} \quad (3)$$

Una vez que se identifica como forma cuadrática, se observa claramente que la matriz de las segundas derivadas parciales evaluadas en el valor crítico a examinar, es la hessiana evaluada en el punto crítico.

Ahora bien, con la aplicación de transformaciones lineales para diagonalizar esta matriz Hessiana, se obtiene una expresión equivalente a la ecuación (2) de la derivada direccional de segundo orden sin el término mixto *cosθsenθ*, de la forma:

$$D_u\,(D_u f(x,y)) = \lambda_1\,(x_D)^2 + \lambda_2\,(y_D)^2 \quad (4)$$

Donde λ_1, λ_2 son los valores característicos de la matriz hessiana evaluada en el punto crítico; a su vez, las variables x_D, y_D se expresan en función de las variables *cosθ* , *senθ* con la transformación lineal correspondiente a la diagonalización.

$$\begin{bmatrix} x_D \\ y_D \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} cos\theta \\ sen\theta \end{bmatrix}$$

O bien

$$\begin{bmatrix} x_D \\ y_D \end{bmatrix} = \begin{bmatrix} acos\theta + bsen\theta \\ ccos\theta + dsen\theta \end{bmatrix}$$

El resultado de la diagonalización expresado en la ecuación (4) simplifica el análisis del signo de la segunda derivada direccional, ya que únicamente contiene términos propios al cuadrado, no hay términos mixtos y se tienen los siguientes casos:

- Si los valores λ_1, λ_2 de la matriz hessiana evaluada en un valor crítico son sólo positivos, conceptualmente la matriz es positiva definida; esto es equivalente a observar que en la ecuación (4), con estos valores característicos, la derivada direccional de segundo orden en todas las direcciones posibles es positiva, como en la figura 8 y se trata de un valor mínimo.

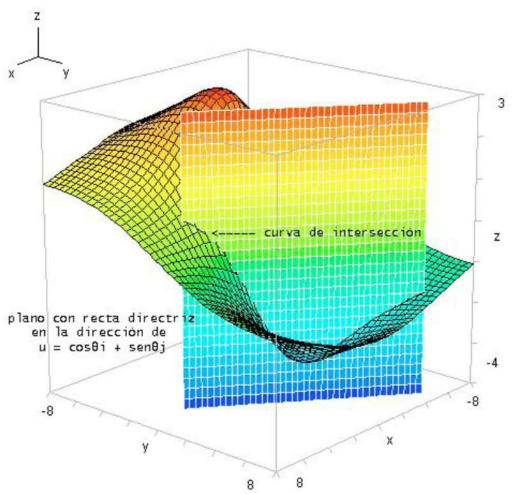

Figura 8. El signo de la segunda derivada direccional en todas las direcciones es positivo y es un mínimo

- Si la matriz hessiana evaluada en el valor crítico tiene todos sus valores característicos con signo negativo, la matriz es definida negativa y el signo de la segunda derivada direccional en la ecuación (4) es negativo en todas las direcciones, como en la figura 9, y se trata de un valor máximo.
- Si los valores característicos de la matriz hessiana son positivos y negativos, la matriz es indefinida y en el valor crítico hay un punto silla; por lo que el signo de la segunda derivada direccional de la ecuación (4) en ciertas direcciones será positivo y en otras negativo, lo que sólo se satisface con la forma geométrica del punto silla de la figura 10, cóncava en unas direcciones y convexa en otras.
- Si algún valor característico de la matriz hessiana es igual a cero y el otro positivo (o negativo), la matriz es semidefinida positiva (o negativa) y en la ecuación (4) un valor característico igual a cero permite que el resultado final de la derivada direccional pueda ser cero en algunas direcciones que anulen el término con el valor característico distinto de cero. En estos casos es necesario evaluar la matriz hessiana en la vecindad del valor crítico y caracterizar la matriz resultante nuevamente como definida negativa, positiva o indefinida, para determinar si se trata de un valor máximo, mínimo o punto silla.

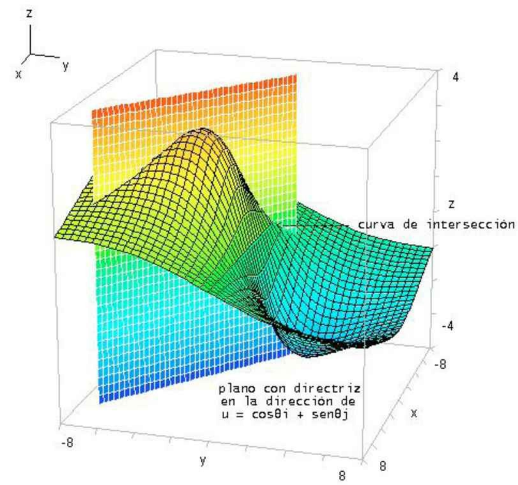

Figura 9. El signo de la segunda derivada direccional en todas las direcciones es negativo y es un máximo

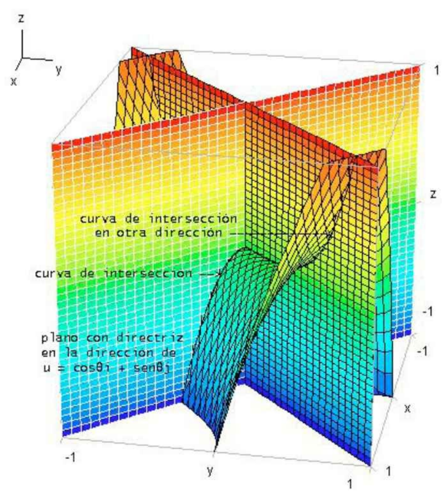

Figura 10. Punto silla de una función, el signo de la segunda derivada direccional en algunas direcciones es positivo y en otras negativo

4 CONCLUSIONES

La aplicación de la interpretación geométrica de la derivada direccional demuestra gráficamente las condiciones de optimalidad de primer y segundo orden para funciones de dos variables.

La aplicación de los conceptos de diagonalización de matrices y formas cuadráticas explica la relación existente entre la matriz hessiana y la derivada direccional de segundo orden en la localización de valores extremos de funciones de dos variables.

El uso de la representación gráfica para la explicación de conceptos o la solución de problemas es un recurso muy importante para mejorar la comprensión de temas, o en su caso, obtener soluciones. En este contexto, el trabajo aporta, con el detalle necesario, la interpretación geométrica de las condiciones de optimalidad de primer y segundo orden, la cual no aparece en la literatura relacionada con el enfoque geométrico que aquí se reporta con la integración y relación de diferentes conceptos matemáticos.

Resulta de interés extender esta interpretación geométrica al caso de más de dos variables, o bien, analizar el caso de la optimización restringida, haciendo uso de otros conceptos del cálculo vectorial y álgebra lineal adicionales a los usados en este trabajo.

REFERENCIAS

1. Larson, R. E., Hostetler, R. P., Edwards, B.H. *Cálculo*, volumen 1. México, DF: McGraw-Hill, 2000.
2. González Pareja, A. *Matemáticas con DERIVE en la economía y la empresa*. Madrid: RA-MA, 1995.

Acerca de los autores

Oscar Piñón Jiménez es Ingeniero Electromecánico en Producción por el Instituto Tecnológico de Minatitlán (1980). Obtuvo el grado de Maestro en Ingeniería en el Instituto Tecnológico y de Estudios Superiores de Monterrey en 1982; realizó estudios de Doctorado en Ingeniería Eléctrica en la Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional. Actualmente imparte cálculo vectorial, álgebra lineal y matemáticas avanzadas como profesor titular de tiempo completo del Instituto Tecnológico de Minatitlán, Veracruz, México. Su área de mayor interés se orienta hacia las matemáticas aplicadas a la planeación de la expansión de las redes de transmisión de energía eléctrica.

Aldo R. Sartorius Castellanos es Ingeniero Mecánico Electricista por la Universidad Veracruzana (2000). Obtuvo el grado de Maestría en Ciencias por la Universidad Central de las Villas, en Cuba, en 2002, así como el grado de Doctor en Ciencias en el área de Control Automático por la misma universidad en el año 2005. Actualmente trabaja como profesor titular de tiempo completo en el Instituto Tecnológico de Minatitlán, Veracruz, México. Sus áreas de interés se centran en la instrumentación industrial y el control avanzado de procesos.