Programación Matemática

Balbina Virginia Casas Méndez

Apuntes de la asignatura (primera parte: introducción a la programación no lineal convexa)

Máster en Técnicas Estadísticas

Curso 2010/11

Introducción

Introducción a la programación no lineal

Introducción

- Introducción a la programación no lineal

Convexidad

- Convexidad de conjuntos y funciones
- Generalizaciones de funciones convexas. Resultados básicos

Introducción

- Introducción a la programación no lineal

Convexidad

- Convexidad de conjuntos y funciones
- Generalizaciones de funciones convexas. Resultados básicos

Optimización en problemas sin restricciones

- Condiciones de primer y segundo orden
- Algoritmos

Introducción

- Introducción a la programación no lineal

Convexidad

- Convexidad de conjuntos y funciones
- Generalizaciones de funciones convexas. Resultados básicos

Optimización en problemas sin restricciones

- Condiciones de primer y segundo orden
- Algoritmos

Optimización con restricciones de igualdad y desigualdad

- Condiciones de optimalidad
- Interpretación económica. Sensibilidad
- Programación cuadrática: método símplex modificado

Introducción
Introducción a la programación no lineal
Convexidad
Convexidad de conjuntos y funciones
Generalizaciones de funciones convexas. Resultados básicos
Optimización en problemas sin restricciones
Condiciones de primer y segundo orden
Algoritmos
Optimización con restricciones de igualdad y desigualdad
Condiciones de optimalidad
Interpretación económica. Sensibilidad
Programación cuadrática: método símplex modificado
Ejercicios

Introducción
Introducción a la programación no lineal
Convexidad
Convexidad de conjuntos y funciones
Generalizaciones de funciones convexas. Resultados básicos
Optimización en problemas sin restricciones
Condiciones de primer y segundo orden
Algoritmos
Optimización con restricciones de igualdad y desigualdad
Condiciones de optimalidad
Interpretación económica. Sensibilidad
Programación cuadrática: método símplex modificado
Ejercicios
Bibliografía

Introducción
Introducción a la programación no lineal

Convexidad

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad

Ejercicios

Bibliografía

Planteamiento del problema

"Programación matemática": Robert Dorfman (Harvard, 1949).

- min $f(x)$

sujeto a:

- $g_i(x) \leq 0$ para $i = 1, \cdots, m$
- $h_i(x) = 0$ para $i = 1, \cdots, l$
- $x \in X$

donde $f, g_1, \cdots, g_m, h_1, \cdots, h_l$ son funciones definidas en $\mathbb{R}^n$, X es un subconjunto de $\mathbb{R}^n$ y x es un vector de n componentes, $x_1, \cdots, x_n$.

Los términos *función objetivo*, *restricción de desigualdad*, *restricción de igualdad*, *solución factible*, *región factible* o *solución óptima* se emplean en un sentido semejante al visto en programación lineal y entera.

Un ejemplo numérico

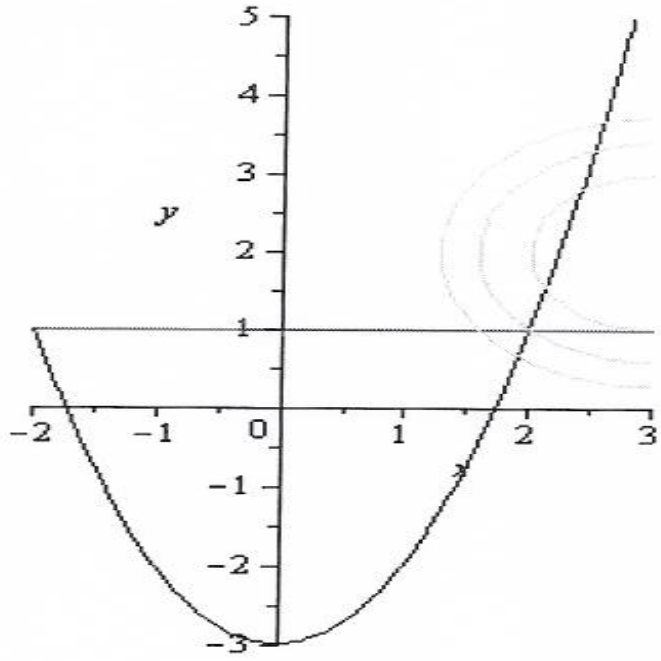
- $\min \ (x_1 - 3)^2 + (x_2 - 2)^2$

sujeto a:

- $x_1^2 - x_2 - 3 \leq 0$
- $x_2 - 1 \leq 0$
- $-x_1 \leq 0$

Gráficamente, dibujando la región factible y los contornos de la función objetivo, se obtiene que la solución óptima se obtiene en el punto (2,1), que proporciona un valor de la función objetivo igual a 2.

Dibujo del ejemplo numérico



Código en Maxima del ejemplo numérico

Maxima se distribuye bajo la licencia GNU-GPL y tanto el código fuente como los manuales son de libre acceso a través de la página web del proyecto: http://maxima.sourceforge.net

Tanto Maple como Maxima descienden del sistema Macsyma, desarrollado en el MIT (Massachusetts Institute of Technology) entre los años 1968 y 1982.

```
(%i1) load(draw)$
(%i2) draw2d(implicit((x-3)^2+(y-2)^2=2,x,-2,3,y,-
3,5),implicit((x-3)^2+(y-2)^2=3,x,-2,3,y,
-3,5),implicit(x^2-y-3=0,x,-2,3,y,-3,5),implicit(y=1,x,-2,3,y,-
3,5),implicit(x=0,x,
-2,3,y,-3,5));
```

Problemas de localización

Un *problema de localización* de centros de actividades puede referirse a la localización de máquinas o departamentos en una fábrica, la localización de almacenes desde los que se envían artículos a los consumidores, o la localización de servicios de emergencia (por ejemplo, puestos de bomberos o de policia) en una zona urbana.

Consideremos que hay n mercados con demandas y localizaciones conocidas. Esas demandas tienen que ser satisfechas por m almacenes de capacidades conocidas. El problema radica en determinar las localizaciones de los almacenes de forma que se minimice la distancia total ponderada por los envíos desde los almacenes a los mercados.

Variables y parámetros en el problema de localización

- (x_i, y_i): localización desconocida del almacén i para $i = 1, \cdots, m$
- c_i: capacidad del almacén i para $i = 1, \cdots, m$
- (a_j, b_j): localización conocida del mercado j para $j = 1, \cdots, n$
- r_j: demanda conocida del mercado j para $j = 1, \cdots, n$
- d_{ij}: distancia del almacén i al mercado j para $i = 1, \cdots, m; j = 1, \cdots, n$
- w_{ij}: unidades enviadas del almacén i al mercado j para $i = 1, \cdots, m; j = 1, \cdots, n$

Problema de localización de almacenes y planificación de envíos

- $\min \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} d_{ij}$

sujeto a:

- $\sum_{j=1}^{n} w_{ij} \leq c_i$ para $i = 1, \cdots, m$
- $\sum_{i=1}^{m} w_{ij} = r_j$ para $j = 1, \cdots, n$
- $w_{ij} \geq 0$ para $i = 1, \cdots, m; j = 1, \cdots, n$

El problema de localización es, en general, no lineal

w_{ij} y d_{ij} han de ser determinados y el problema anterior es, en general, no lineal. Diferentes distancias se pueden utilizar, por ejemplo:

$$d_{ij} = |x_i - a_j| + |y_i - b_j|,$$

$$d_{ij} = [(x_i - a_j)^2 + (y_i - b_j)^2].$$

Selección de cartera

Se desea invertir 10000 $ en dos tipos de acciones. Supongamos que W_i es la ganancia aleatoria anual por 1 $ invertido en la acción $i = 1, 2$, con ganancia esperada $E(W_1) = 0'10$, $E(W_2) = 0'05$, varianzas $V(W_1) = 0'15$, $V(W_2) = 0'06$ y covarianza $cov(W_1, W_2) = 0'07$. El objetivo es minimizar la varianza de la cartera garantizando una ganancia anual esperada de al menos el 9 %. Podemos definir las variables de decisión:

X_i: cantidad invertida en la acción $i = 1, 2$.

Teniendo en cuenta que la varianza de la cartera se calcula:

$V(W_1x_1 + W_2x_2) = V(W_1)x_1^2 + V(W_2)x_2^2 + 2cov(W_1, W_2)x_1x_2$, el problema de programación no lineal se puede formular como sigue:

El modelo de selección de cartera

- min $f(X_1, X_2) = 0'15X_1^2 + 0'06X_2^2 + 0'14X_1X_2$

sujeto a:

- $900 - 0'10X_1 - 0'05X_2 \leq 0$
- $10000 - X_1 - X_2 = 0$
- $X_1 \geq 0$, $X_2 \geq 0$

Solución óptima:

$X_1 = 8000\$$, $X_2 = 2000\$$ y $f = 1208 \times 10^4$.

La desviación típica de la cartera resulta, por tanto, $3475'63\$$.
Lo resolvemos con Solver de Excel:

$$[0.12080000\ 10^8, [x = 8000., y = 2000.0000000000046]]$$

Introducción

Convexidad

Convexidad de conjuntos y funciones

Generalizaciones de funciones convexas. Resultados básicos

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad

Ejercicios

Bibliografía

Convexidad de conjuntos y funciones

$\emptyset \neq S \subset \mathbb{R}^n$ se denomina un **conjunto convexo** si el segmento que une dos puntos cualesquiera de S también pertenece a S:

$$x_1, x_2 \in S,\ \lambda \in [0,1] \Rightarrow \lambda x_1 + (1-\lambda)x_2 \in S.$$

Sea $f : S \to \mathbb{R}$ con S un conjunto convexo. Se dice que es una **función convexa** (J. L. W. Jensen, 1905) sobre S si:

$$x_1, x_2 \in S,\ \lambda \in [0,1] \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2),$$

i. e., el valor de f en los puntos del segmento $\lambda x_1 + (1-\lambda)x_2$ es menor o igual que la altura de la cuerda que une los puntos $(x_1, f(x_1))$ y $(x_2, f(x_2))$.

El interés de la convexidad

- La convexidad tiene gran importancia en programación matemática porque permite obtener condiciones suficientes de optimalidad que en general son únicamente condiciones necesarias como las denominadas condiciones de Kuhn-Tucker.

- También asegura validez global a ciertas proposiciones que de otra manera sólo se verifican de manera local. **Ejercicio**: probamos que mínimos locales de funciones convexas son también mínimos globales.

Intuitivamente

Si x^* es un **mínimo local** de la función f definida en en un conjunto S, $f(x) \geq f(x^*)$ en todo punto x de S en un entorno E de x^*. Es un **mínimo global** si $f(x) \geq f(x^*)$ en todo punto x de S.

Sea x^* un mínimo local de f con f convexa. Si x^* no fuese mínimo global, entonces existiría algún $\overline{x} \in S$ con $f(\overline{x}) < f(x^*)$.

Tomamos el punto $x = \lambda x^* + (1 - \lambda)\overline{x}$ con λ tan próximo a 1 como sea necesario para que x próximo a x^*, esto es, x en el entorno E de x^*. Se verifica, aplicando la convexidad de f, que:

$$f(x) = f(\lambda x^* + (1 - \lambda)\overline{x}) \leq \lambda f(x^*) + (1 - \lambda) f(\overline{x}) < f(x^*),$$

lo cual contradice la optimalidad local de x^*. Por tanto, se ha de verificar que x^* es un óptimo global.

Funciones estrictamente convexas

Si $\lambda \in (0,1)$ y la desigualdad anterior se satisface estrictamente, entonces f es una **función estríctamente convexa**. Si f es (estríctamente) convexa, entonces $-f$ es **(estríctamente) cóncava** (B. de Finetti, 1954).

Comprobaremos con Maxima la convexidad y la convexidad estricta de funciones.

Maxima y la convexidad estricta

Con Maxima comprobamos que no es estrictamente convexa una función convexa:

```
(%i3) declare(lambda,real)$
(%i4) declare(a,real)$
(%i5) declare(b,real)$
(%i6) asumme(lambda>0,lambda<1);
(%o6) asumme(lambda>0,lambda<1)
(%i7) g:x+5;
(%o7) x+5
(%i8) l:%,x=lambda*a+(1-lambda)*b;
(%o8) a*lambda+b*(1-lambda)+5
(%i9) g:x+5;
(%o9) x+5
```

```
(%i10) ga:%,x=a;
(%o10) a+5
(%i11) g:x+5;
(%o11) x+5
(%i12) gb:%,x=b;
(%o12) b+5
(%i13) D:lambda*ga+(1-lambda)*gb;
(%o13) (a+5)*lambda+(b+5)*(1-lambda)
(%i14) is(I<D);
(%o14) false
(%i15) is(I<=D);
(%o14) true
```

TEOREMA Sean las funciones convexas sobre un conjunto convexo $S \subset \mathbb{R}^n$, f_i, $i = 1, \cdots, k$. Entonces, la función $f = \sum_{i=1}^{k} f_i$ es también convexa sobre S.

TEOREMA (*se puede ver en Hadley, 1964*) Si f es una función convexa sobre el ortante no negativo de $\mathbb{R}^n$, y el conjunto de puntos S que satisface $f(x) \leq b$ y $x \geq 0$ es no vacío, con $b \in \mathbb{R}$, entonces S es un conjunto convexo.

Sea $\emptyset \neq S \subset \mathbb{R}^n$ y sea $f : S \to \mathbb{R}$. Si la función f es diferenciable en algún punto $\overline{x} \in S$, entonces el vector **gradiente** en $\overline{x}$ es[1]:

$$\nabla f(\overline{x}) = \left(\frac{\partial f(\overline{x})}{\partial x_1}, \cdots, \frac{\partial f(\overline{x})}{\partial x_n}\right)'.$$

[1] $'$ denota, para matrices y vectores, "transpuesto".

Para una función dos veces diferenciable f se define la matriz **Hessiana** $H(x)$ de f en x como una matriz cuadrada donde la entrada en la fila i-ésima y la columna j-ésima es la segunda derivada parcial $\partial^2 f(x)/\partial x_i \partial x_j$:

$$H(x) = \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

Si $\partial f(x)/\partial x_i$, $i = 1, \cdots, n$ y $\partial^2 f(x)/\partial x_i \partial x_j$, $i, j = 1, \cdots, n$ son continuas, la matriz hessiana es simétrica.

Ejemplo: profundidad del terreno (*tomado de Schmidt and Devis, 1981*)

La siguiente función de dos variables describe la profundidad del terreno bajo el agua de un puerto. x e y miden en millas la distancia al este y al norte, respectivamente, desde el centro del puerto:

$$f(x,y) = -(5/10)x^2 + (1/10)y^3 - (2/10)y^2 + (4/100)xy$$

$$+(1/100)x + (3/100)y - (5/10).$$

Con Maxima determinamos el vector gradiente y calculamos la matriz Hessiana.

Cálculo de gradientes y hessianas con Maxima

```
(%i15) load (“vect")$
(%i16) grad(-(5/10)*x^2+(1/10)*y^3-
(2/10)*y^2+(4/100)*x*y+(1/100)*x+(3/100)*y-(5/10));
(%i17) express (%);
(%i18) ev (%, diff);
(%i19) hessian(-(5/10)*x^2+(1/10)*y^3-
(2/10)*y^2+(4/100)*x*y+(1/100)*x+(3/100)*y-(5/10),[x,y]);
```

Matrices definidas positivas y negativas

Sea Q una matriz simétrica $n \times n$. La matriz Q se denomina **definida positiva** si la **forma cuadrática** $x'Qx > 0$ para todo $x \in \mathbb{R}^n$ no nulo y **semi definida positiva** si $x'Qx \geq 0$ para todo $x \in \mathbb{R}^n$. Invirtiendo las desigualdades, se obtienen las definiciones de **definida negativa** y **semidefinida negativa**, respectivamente. La matriz Q se denomina **indefinida** si $x'Qx$ es positivo para algunos valores de x y negativo para otros.

TEOREMA (**Teorema de Sylvester**) (*ver en Peressini y otros, 1988*) Para la matriz simétrica $n \times n$ Q, definimos Δ_k como el determinante de la submatriz $k \times k$ de Q de la esquina superior izquierda, para $k = 1, \cdots, n$, y lo denominamos k-ésimo **menor principal** de Q. Entonces, Q es definida positiva si y sólo si $\Delta_k > 0$ para $k = 1, \cdots, n$ y definida negativa si y sólo si $(-1)^k \Delta_k > 0$ para $k = 1, \cdots, n$. En otro caso, Q es indefinida.

En el resultado anterior, si se reemplaza el signo $>$ por $\geq$ (o $<$ por $\leq$) no quiere decir que la matriz Q sea semidefinida positiva (o negativa). Lo vemos en el siguiente ejemplo tratado con Maxima[2].

```
(%i20) Q: matrix([2,2,2],[2,2,2],[2,2,1]);
(%i21) x:matrix([1/2,1/2,-1]);
(%i22) xt:matrix([1/2],[1/2],[-1]);
(%i23) xt.Q.x;
(%o23) -1
```

[2]Las matrices en xMaxima las introducimos por columnas; cada columna va entre corchetes y en los resultados se representan de la misma manera.

Determinación del carácter semidefinido positivo

TEOREMA Sea Q una matriz simétrica. Sea Q_r la matriz obtenida al eliminar las primeras r filas y r columnas de la matriz original Q, con $r = 0, \cdots, n-1$ y $Q_0 = Q$. Defínase Δ_{rk} el k-ésimo menor principal de Q_r. Entonces, Q es semidefinida positiva si y sólo si $\Delta_{rk} \geq 0$ para $r = 0, \cdots, n-1$, $k = 1, \cdots, n-r$; y semidefinida negativa si y sólo si $(-1)^k \Delta_{rk} \geq 0$ para $r = 0, \cdots, n-1$, $k = 1, \cdots, n-r$.

Por ejemplo, si $n = 3$ y

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix},$$

para determinar el carácter semidefinido positivo se necesita probar que:

$$\Delta_{01} = q_{11} \geq 0,\ \Delta_{02} = det \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \geq 0,\ \Delta_{03} = detQ \geq 0,$$

$$\Delta_{11} = q_{22} \geq 0,\ \Delta_{12} = det \begin{pmatrix} q_{22} & q_{23} \\ q_{32} & q_{33} \end{pmatrix} \geq 0,\ \text{y}\ \Delta_{21} = q_{33} \geq 0.$$

Este método requiere el cálculo de
$n + (n-1) + \cdots + 2 + 1 = n(n+1)/2$ determinantes.

Sea A una matriz cuadrada de orden m. Se dice que un escalar $\lambda \in \mathbb{R}$ es un **autovalor** de A si existe un vector $v \in \mathbb{R}^m$, $v \neq 0$ tal que $Av = \lambda v$, en cuyo caso se dice que v es un **autovector** de A asociado al autovalor λ.

TEOREMA (*ver en Peressini y otros, 1988*) La matriz simétrica Q es definida positiva (negativa) si y sólo si todos los autovalores de Q son positivos (negativos) y es semidefinida positiva (negativa) si y sólo si todos los autovalores de Q son no negativos (no positivos). La matriz Q es indefinida si y sólo si Q tiene al menos un autovalor positivo y al menos un autovalor negativo.

Convexidad de una función y definición de su Hessiana

TEOREMA (*demostración en Bazaraa y otros, 1993.*) Sea $f : S \to \mathbb{R}$ una función dos veces diferenciable sobre un conjunto no vacío, abierto y convexo $S \subset \mathbb{R}^n$. Entonces f es convexa si y sólo si la Hessiana es semidefinida positiva en cada punto de S. Si la Hessiana es definida positiva en cada punto de S, entonces la función f es estrictamente convexa. Pero si f es estrictamente convexa, entonces la Hessiana es semidefinida positiva en cada punto de S.

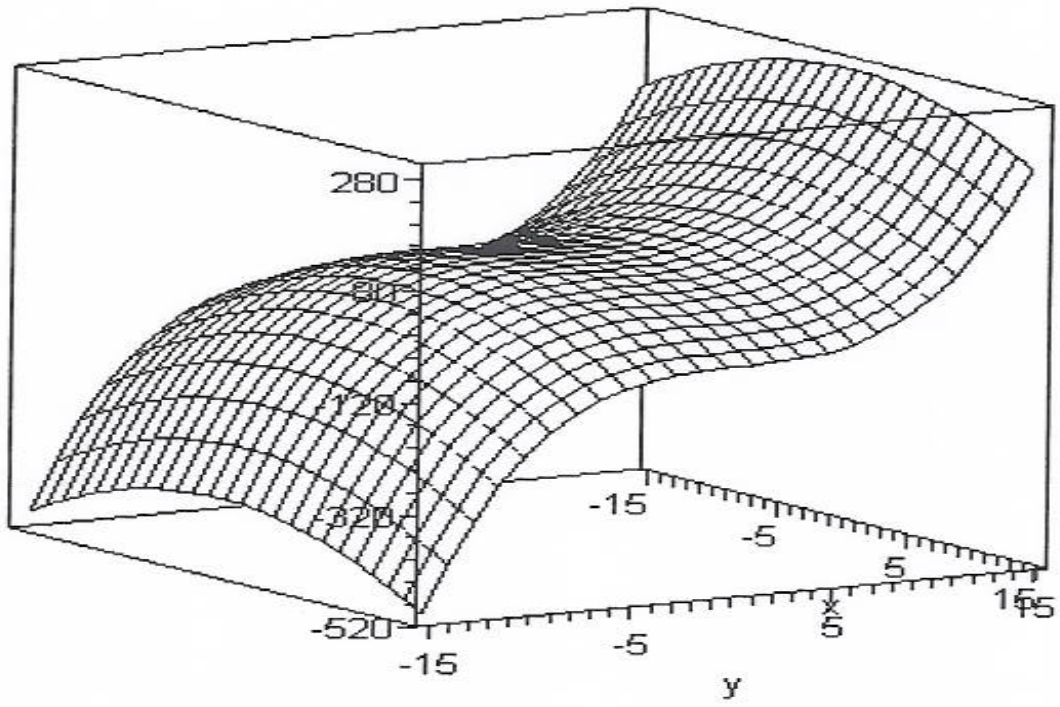
Por ejemplo, si consideramos la función estrictamente convexa $f(x) = \frac{1}{12}x^4$, la Hessiana es $H(x) = f''(x) = x^2$, que es semidefinida positiva en $x = 0$.

Ejemplos de definición y convexidad

EJEMPLO (*profundidad del terreno, continuación*) Debido a la irregularidad del terreno no cabe esperar que la Hessiana sea (semi)definida positiva. Sólo en algunas regiones el terreno es definido negativo (o semidefinido negativo) y por tanto f estrictamente cóncava (o cóncava).

Con Maxima y aplicando el teorema de Sylvester podemos comprobarlo. También haciendo la representación gráfica. Se tiene concavidad estricta para $y < \frac{83}{125}$.

Dibujo del terreno del puerto



Dibujo del terreno del puerto con Maxima

```
(%i1) wxplot3d(-(5/10)*x^2+(1/10)*y^3-
(2/10)*y^2+(4/100)*x*y+(1/100)*x+(3/100)*y-(5/10), [x,-15,15],
[y,-15,15])$
```

Convexidad de la función distancia euclídea

La función $f(a,b) = \sqrt{a^2 + b^2}$ juega un papel importante en los problemas de localización. Estudiamos con Maxima si es convexa.

```
(%i24) H:hessian((a^2+b^2)^(1/2),[a,b]);
(%i25) eigenvalues(%);
(%i26) determinant(H);
(%i27) expand(%);
```

Obtenemos autovalores iguales o mayores que cero, por tanto la Hessiana es semidefinida positiva y así la función distancia Euclídea es convexa.

Introducción

Convexidad

Convexidad de conjuntos y funciones

Generalizaciones de funciones convexas. Resultados básicos

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad

Ejercicios

Bibliografía

Generalizaciones de funciones convexas

DEFINICIÓN Sea S un subconjunto no vacío y convexo de $\mathbb{R}^n$ y sea $f : S \to \mathbb{R}$.

- Se dice que f es **cuasiconvexa** (B. de Finetti, 1954) si para todo $x, y \in S$ se verifica:

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\} \ \forall \lambda \in (0,1).$$

f es **cuasicóncava** si $-f$ es cuasiconvexa:

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\} \ \forall \lambda \in (0,1).$$

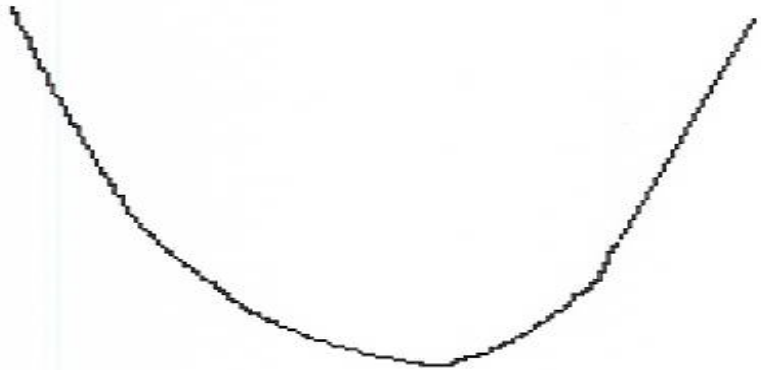
- Se dice que f es **estrictamente cuasiconvexa** si para todo $x, y \in S$ con $x \neq y$ se verifica:

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\} \ \forall \lambda \in (0,1).$$

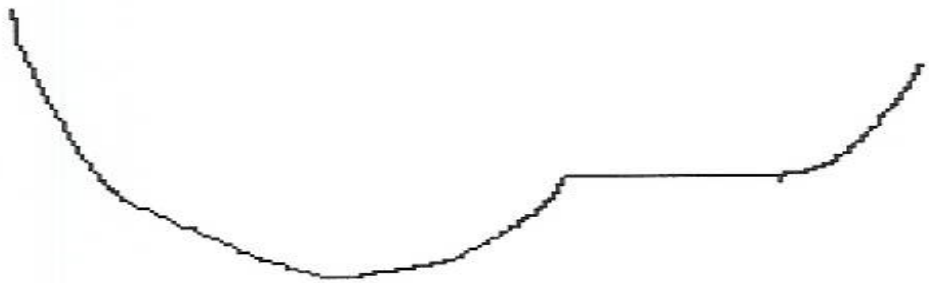
f es **estrictamente cuasicóncava** si $-f$ es estrictamente cuasiconvexa.

Función cuasiconvexa

Función ni cuasiconvexa ni cuasicóncava



Función convexa, no estrictamente convexa



Función cuasiconvexa, no estrictamente cuasiconvexa

Geométricamente, la cuasiconvexidad de f en cada punto $\overline{x}$ significa que para cada punto x de un conjunto convexo S con $f(x) \leq f(\overline{x})$, el valor de la función en todos los puntos del conjunto $(x, \overline{x})$ para los que f está definida es menor o igual que $f(\overline{x})$.

Es decir, la gráfica de la función relativa al intervalo $(x, \overline{x})$ está por debajo de la recta $y = f(\overline{x})$.

Ejemplo: La función $f(x) = \frac{1}{x}$ es convexa y cuasi-cóncava en $\mathbb{R}_+$.

La podemos dibujar con wxMaxima:

```
wxplot2d([1/x], [x,0,10], [y,0,10])$
```

Las funciones estrictamente cuasiconvexas y estrictamente cuasicóncavas son especialmente importantes en programación no lineal, pues aseguran que un mínimo local o un máximo local de una de tales funciones sobre un conjunto convexo son, respectivamente, un mínimo global y un máximo global[3].

TEOREMA Sea S un subconjunto no vacío y convexo de $\mathbb{R}^n$ y sea $f : S \to \mathbb{R}$. f es cuasiconvexa si y sólo si

$$S_\alpha = \{x \in S : f(x) \leq \alpha\} \text{ es convexo, } \forall \alpha \in \mathbb{R}.$$

Demostración:

[3] También aparecen en teoría de juegos no cooperativos representando funciones de utilidad cuando se prueban resultados de existencia de equilibrios de Nash en juegos en forma estratégica: *Teorema de Nash*.

“ $\Rightarrow$ ” Sea $\alpha \in \mathbb{R}$ y $x, y \in S_\alpha$. Entonces $f(x) \leq \alpha$ y $f(y) \leq \alpha$. Por la cuasiconvexidad, si $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\} \leq \alpha,$$

por lo que $\lambda x + (1 - \lambda)y \in S_\alpha$ y así S_α es convexo.

“ $\Leftarrow$ ” Sean $x, y \in S$. Supongamos sin pérdida de generalidad que $f(x) \leq f(y)$. Entonces $x \in S_{f(y)},\ y \in S_{f(y)}$. Como este conjunto $S_{f(y)}$ es convexo, $\lambda x + (1 - \lambda)y \in S_{f(y)}$ para todo $\lambda \in [0, 1]$. Por tanto,

$$f(\lambda x + (1 - \lambda)y) \leq f(y) = \max\{f(x), f(y)\}$$

y así f es cuasiconvexa. $\square$

TEOREMA Sea S un subconjunto no vacío y convexo de $\mathbb{R}^n$ y sea $f : \mathbb{R}^n \to \mathbb{R}$. Supongamos que f es estrictamente cuasiconvexa y consideremos el problema

- $\min \quad f(x)$

sujeto a:

- $x \in S$

Si $\overline{x}$ es una solución óptimo local, entonces $\overline{x}$ es una solución óptima global única.

Es inmediato comprobar que para cualquier función f definida en un subconjunto convexo S de $\mathbb{R}^n$ con valores en $\mathbb{R}$ se verifica:

- f convexa en $S \Rightarrow f$ es cuasiconvexa en S.
- f estrictamente convexa en $S \Rightarrow f$ es estrictamente cuasiconvexa en S.
- f estrictamente cuasiconvexa en $S \Rightarrow f$ es cuasiconvexa en S.

Análogas relaciones se cumplen entre cóncavas, cuasicóncavas, estrictamente cóncavas y estrictamente cuasicóncavas. En Bazaraa y otros (1993) se puede profundizar en el estudio de generalizaciones de funciones convexas.

Un ejemplo: estudio de la cuasiconvexidad y la cuasiconcavidad

Consideremos la función $y = f(x_1, x_2) = e^{x_1 x_2}$ con $(x_1, x_2) \in \mathbb{R}^2$. Vamos a tomar $\alpha = e$ y consideramos los conjuntos:

$$\{(x_1, x_2) \in \mathbb{R}^2 \,:\, e^{x_1 x_2} \leq e\} = \{(x_1, x_2) \in \mathbb{R}^2 \,:\, x_1 x_2 \leq 1\} \text{ y}$$

$$\{(x_1, x_2) \in \mathbb{R}^2 \,:\, e^{x_1 x_2} \geq e\} = \{(x_1, x_2) \in \mathbb{R}^2 \,:\, x_1 x_2 \geq 1\}.$$

Estos conjuntos no son convexos (lo podemos dibujar con Maxima) y así la función $f(x_1, x_2) = e^{x_1 x_2}$ no es cuasicóncava ni cuasiconvexa y por tanto no es cóncava ni convexa.

Condiciones para la cuasiconvexidad de funciones diferenciables

PROPOSICIÓN Sea S un subconjunto abierto, no vacío y convexo de $\mathbb{R}^n$ y f una función definida de S en $\mathbb{R}$ y diferenciable en S. Se verifica que:

- La función f es cuasiconvexa en S si y sólo si para cualesquiera $x, y \in S$ tales que $f(y) \leq f(x)$ se tiene que $\nabla f(x)(y - x) \leq 0$.
- Si para cualesquiera $x, y \in S$ con $x \neq y$ tales que $f(y) \leq f(x)$ se tiene que $\nabla f(x)(y - x) < 0$ (o equivalentemente, si $\nabla f(x)(y - x) \geq 0$, en particular cuando $\nabla f(x) = 0$, se tiene que $f(y) > f(x)$) entonces la función f es estrictamente cuasiconvexa en S.

EJEMPLO Consideremos la función $f(x_1, x_2) = ax_1^2 + bx_2^2$ con $(x_1, x_2) \in \mathbb{R}^2$, $a > 0$, $b < 0$. No es cuasiconvexa pues si tomamos, por ejemplo, $x = (1, -2)$ e $y = (1, 3)$, donde $f(x) \geq f(y)$, se obtiene que:

$$\nabla f(x)(y - x) = 2ax_1(y_1 - x_1) + 2bx_2(y_2 - x_2) = -20b > 0.$$

Podemos dibujar la función con wxMaxima para visualizar el resultado.

plot3d($3 * x^2 - 4 * y^2$, [x,-5,5], [y,-75,75], [plot_format,gnuplot])$

Un trabajo reciente en generalización de funciones convexas:

Y. X. Zhao, S. Y. Shou-Yang and L. Coladas Uría (2010). "Characterizations of r-Convex Functions". Journal of Optimization Theory and Applications Vol. 145, pp. 186-195.

Abstract:

This paper discusses some properties of r-convexity and its relations with some other types of convexity. A characterization of convex functions in terms of r-convexity is given without assuming differentiability. The concept of strict r-convexity is introduced. For a twice continuously differentiable function f, it is shown that the strict r-convexity of f is equivalent to a certain condition on $\nabla^2 f$ (the Hessian matrix). Further, it is shown that this condition is satisfied by quasiconvex functions satisfying a less stringent condition.

Introducción

Convexidad

Optimización en problemas sin restricciones
Condiciones de primer y segundo orden
Algoritmos

Optimización con restricciones de igualdad y desigualdad

Ejercicios

Bibliografía

Justificación del estudio de estos problemas:

- En algunos problemas de programación las variables de decisión no están restringidas, como es el caso de los problemas de localización más simples.
- Las condiciones de optimalidad para los problemas restringidos son extensiones naturales de las de los problemas no restringidos.

Optimalidad global y local

DEFINICIÓN Sea f una función definida en $S \subset \mathbb{R}^n$. Un punto $\overline{x} \in S$ es

- un **mínimo global** de f en S si $f(\overline{x}) \leq f(x)$ para todo $x \in S$,
- un **mínimo local** de f si existe un $\epsilon > 0$ tal que $f(\overline{x}) \leq f(x)$ para todo $x \in S$ para el cual $x \in N_\epsilon(\overline{x}) = \{y : ||y - \overline{x}|| \leq \epsilon\}$, esto es, x está en un entorno de $\overline{x}$ de radio ϵ.

Reemplazando el signo $\leq$ por el signo $<$ obtenemos las definiciones de **mínimo global estricto** y **mínimo local estricto**.

Condición necesaria de optimalidad local

TEOREMA (*La demostración en Bazaraa y otros, 1993*) Supongamos f diferenciable en $\overline{x}$. Si $\overline{x}$ es un mínimo local, entonces la condición de primer orden es que $\overline{x}$ es un punto crítico (o estacionario), esto es

$$\nabla f(\overline{x}) = (\frac{\partial f(\overline{x})}{\partial x_1}, \cdots, \frac{\partial f(\overline{x})}{\partial x_n})' = 0.$$

Además, si f es dos veces diferenciable en $\overline{x}$ y $\overline{x}$ es un mínimo local, una condición de segundo orden es que la Hessiana $H(\overline{x})$ es semidefinida positiva.

Esta condición no es suficiente, como podemos ver si consideramos la función $f(x) = x^3$.

wxplot2$d([x^3], [x, -5, 5], [y, -125, 125])$

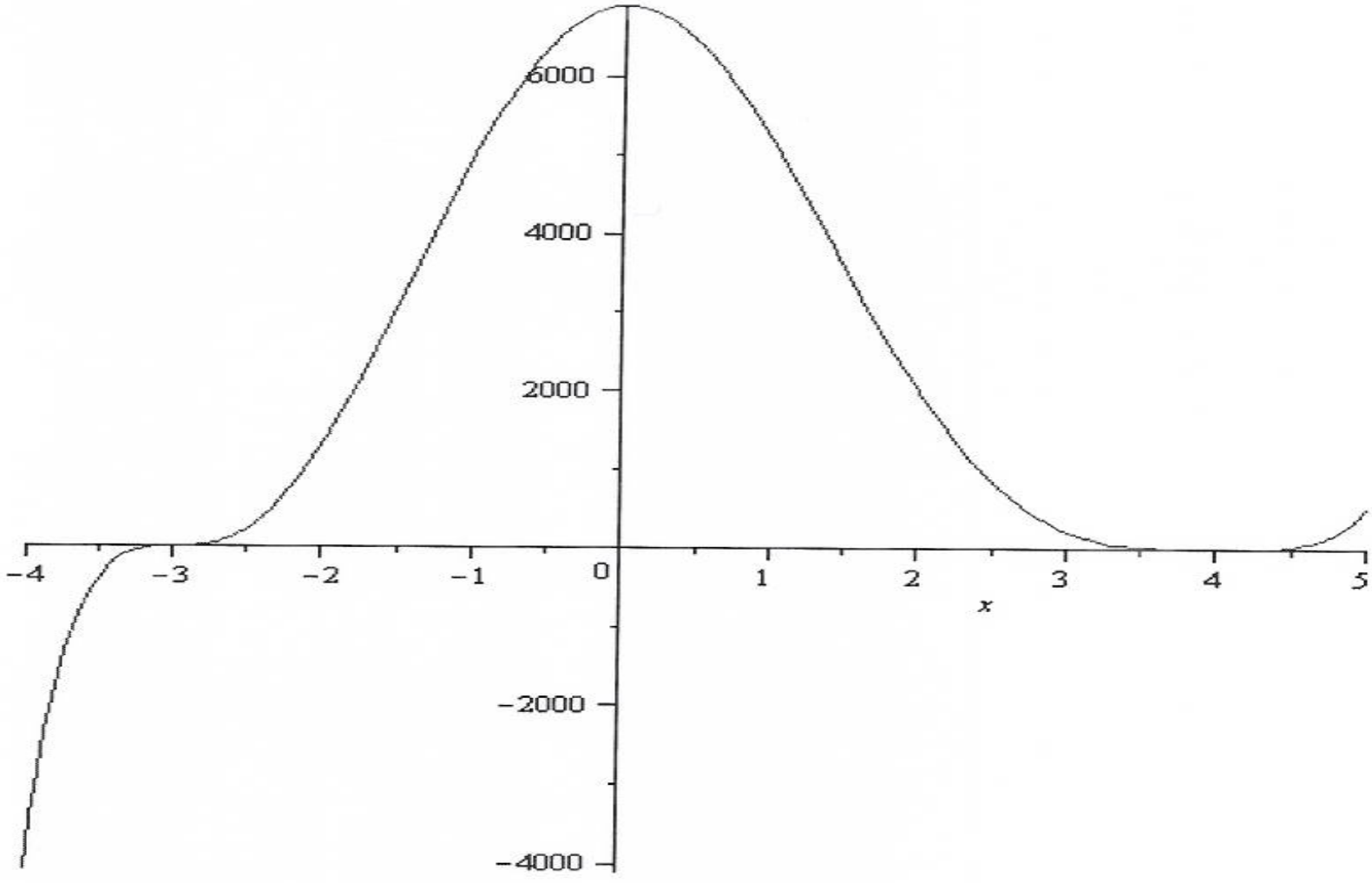
Condición suficiente de optimalidad local

TEOREMA (*La demostración en Bazaraa y otros, 1993*)
Supongamos f dos veces diferenciable en $\overline{x}$. Si $\nabla f(\overline{x}) = 0$ y la Hessiana $H(\overline{x})$ es definida positiva, entonces
$\overline{x}$ es un mínimo local.

EJEMPLO Análisis de la función $f(x) = (x-4)^4(x+3)^3$.

wxplot2d($[(x-4)^4 * (x+3)^3], [x, -4, 5], [y, -4000, 8000]$)$

Representación gráfica



El ejemplo con Maxima

```
(%i1) f1:diff((x-4)^4*(x+3)^3,x);
(%o1) 4*(x-4)^3*(x+3)^3+3*(x-4)^4*(x+3)^2
(%i2) solve(4*(x-4)^3*(x+3)^3+3*(x-4)^4*(x+3)^2=0);
(%o2) [x=-3,x=0,x=4]
(%i3) f2:diff((x-4)^4*(x+3)^3,x,2);
(%o3) 12*(x-4)^2*(x+3)^3+24*(x-4)^3*(x+3)^2+6*(x-4)^4*(x+3)
(%i4) g0:%,x=0;
(%o4) -4032
(%i5) f2:diff((x-4)^4*(x+3)^3,x,2)$
(%i6) g4:%,x=4;
(%o6) 0
(%i7) f2:diff((x-4)^4*(x+3)^3,x,2)$
(%i8) g3:%,x=-3;
(%o8) 0
```

```
(%i9) f:(x-4)^4*(x+3)^3$
(%i10) f4:%,x=4$
(%i11) epsilon:0.1$
(%i11) f:(x-4)^4*(x+3)^3$
(%i12) F4:%,x=4+epsilon$
(%i13) is(f4<F4);
(%o13) true
(%i14) f:(x-4)^4*(x+3)^3$
(%i15) FF4:%,x=4-epsilon$
(%i16) is(f4<FF4);
(%o16) true
(%i7) f:(x-4)^4*(x+3)^3$
(%i18) f3:%,x=-3$
```

```
(%i19) f:(x-4)^4*(x+3)^3$
(%i20) F3:%,x=-3+epsilon$
(%i21) is(f3<F3);
(%o21) true
(%i22) f:(x-4)^4*(x+3)^3$
(%i23) FF3:%,x=-3-epsilon$
(%i24) is(f3<FF3);
(%o24) false
```

De los cálculos anteriores y las condiciones, se deduce que 0 es un máximo local, 4 un mínimo local y -3 ni mínimo ni máximo local. La gráfica confirma estos resultados.

Existencia de soluciones óptimas

Consideremos el problema:

$$\min_{x \in X} f(x)$$

Dos posibilidades:

- El conjunto $\{f(x) \mid x \in X\}$ no está acotado inferiormente y no hay solución óptima.
- El conjunto $\{f(x) \mid x \in X\}$ está acotado inferiormente:
  - Existe un mínimo global si f es continua y X es compacto (**Teorema de Weierstrass**)

Introducción

Convexidad

Optimización en problemas sin restricciones
- Condiciones de primer y segundo orden
- Algoritmos

Optimización con restricciones de igualdad y desigualdad

Ejercicios

Bibliografía

Generalidades de estos algoritmos de optimización

Utilizar la condición necesaria $\nabla f(x) = 0$ para determinar puntos estacionarios tiene la desventaja de la eventual dificultad de resolver las ecuaciones resultantes.

- Los algoritmos para problemas sin restricciones requieren que se suministre un punto inicial, x_0.
- Comenzando en x_0, se genera una sucesión $\{x_k\}$ que termina cuando el algoritmo no puede progresar más o bien cuando se considera que el punto solución se ha aproximado con una precisión suficiente.
- Para decidir cómo pasar de un iterante x_k al siguiente, x_{k+1}, los algoritmos utilizan información sobre la función, f, y sus derivadas en el iterante x_k o incluso en todos los anteriores, $x_0, x_1, \cdots, x_k$.
- Una estrategia para pasar de un iterante a otro, es la denominada de búsqueda, por medio del movimiento una cierta distancia a través de una adecuada dirección.

Métodos del gradiente

Supongamos que se desea maximizar una función f. Este método sirve para optimizar funciones continuas que son dos veces diferenciables. La idea general es **obtener puntos sucesivos, comenzando en un punto inicial dado, y desplazándose en la dirección del aumento más rápido** (maximización) de la función. La técnica se conoce como método del gradiente porque el gradiente de la función en un punto es lo que indica la tasa más rápida de aumento. La técnica que se presenta ahora es la denominada método de la **cuesta de mayor pendiente**.

La terminación del método de gradiente se efectúa en el punto donde el vector gradiente se anula. Como sabemos, ésta es sólo una condición necesaria de optimalidad. Por tanto, hay que destacar que la optimalidad no puede verificarse a menos que a priori se sepa que la función es cóncava o convexa.

TEOREMA Si f es convexa en $\mathbb{R}^n$, entonces cualquier mínimo local es mínimo global. Si además es diferenciable, todos los puntos críticos o estacionarios de f ($\nabla f(x) = 0$), si existen, son mínimos globales.

Esto es así pues:

TEOREMA $f : S \subset \mathbb{R}^n \to \mathbb{R}$ es convexa si y sólo si para todo $x, y \in S$, $f(y) \geq f(x) + \nabla f(x)(y - x)$.

Por tanto, en este caso de convexidad:

$\nabla f(x) = 0 \Longrightarrow x$ es mínimo global.

Motivación del método de gradiente

DEFINICIÓN Un vector d se dice que es una **dirección de crecimiento** de f en $\overline{x}$ si

$$\exists \delta > 0 \ : \ f(\overline{x} + \lambda d) > f(\overline{x}) \ \forall \lambda \in (0, \delta),^4$$

$$i.\ e.\ lim_{\lambda \to 0^+} \frac{f(\overline{x} + \lambda d) - f(\overline{x})}{\lambda} > 0.$$

LEMA Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es diferenciable en x, y supongamos que $\nabla f(x) \neq 0$. Entonces, la solución al problema de maximizar $lim_{\lambda \to 0^+} \frac{f(x+\lambda d)-f(x)}{\lambda}$ sujeto a $||d|| \leq 1$ está dada por $d^* = \dfrac{\nabla f(x)}{||\nabla f(x)||}$, que es entonces la dirección de ascenso más pronunciado.

---

[4] $(0, \delta)$ es el intervalo de posibles "longitudes de paso".

De acuerdo con todo lo anterior, seleccionaremos puntos sucesivos x_k y x_{k+1}, siendo x_0 el punto inicial, tales que

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \mathbf{r^k} \nabla \mathbf{f}(\mathbf{x_k})$$

donde r^k es un parámetro llamado **tamaño de paso óptimo**. Este parámetro r^k se determina de modo que x_{k+1} resulta en la mejora más grande en f. En otras palabras, si se define la función $h(r)$ de manera que

$$\mathbf{h}(\mathbf{r}) = \mathbf{f}(\mathbf{x_k} + \mathbf{r} \nabla \mathbf{f}(\mathbf{x_k}))$$

r^k es el valor de r que maximiza $h(r)$. El procedimiento termina cuando dos puntos sucesivos de ensayo x_k y x_{k+1} son aproximadamente iguales, lo cual equivale a tener $r^k \nabla f(x_k) \cong 0$.

Si se quiere **minimizar** la función objetivo $f(x)$, en cada iteración nos movemos en la dirección **opuesta** al gradiente.

Un ejemplo

Queremos maximizar la función
$f(x_1, x_2) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$.

Sea el punto inicial $x_0 = (1, 1)$. Entonces

$$\nabla f(x) = (4 - 4x_1 - 2x_2, 6 - 2x_1 - 4x_2) \text{ y } \nabla f(x_0) = (-2, 0).$$

PRIMERA ITERACIÓN:
El punto siguiente x_1 se obtiene considerando

$$x = (1, 1) + r(-2, 0) = (1 - 2r, 1).$$

Por consiguiente,

$$h(r) = f(1 - 2r, 1) = -2(1 - 2r)^2 + 2(1 - 2r) + 4.$$

El tamaño del paso óptimo que proporciona el valor máximo de $h(r)$ es con $r = 1/4$, lo que nos proporciona $x_1 = (1/2, 1)$.

Un ejemplo

SEGUNDA ITERACIÓN:

$$\nabla f(x_1) = (0, 1).$$

$$x = (1/2, 1) + r(0, 1) = (1/2, 1 + r).$$

$$h(r) = -2(1 + r)^2 + 5(1 + r) + 3/2.$$

El tamaño del paso óptimo que proporciona el valor máximo de $h(r)$ es con $r = 1/4$, lo que nos proporciona $x_2 = (1/2, 5/4)$.

Procediendo análogamente, se obtiene:

$$x_3 = (3/8, 5/4),\ x_4 = (3/8, 21/16),$$

$$x_5 = (11/32, 21/169) = (\mathbf{0'3437}, \mathbf{1'3125}) \text{ con}$$

$$\nabla f(x_2) = (-1/2, 0),\ \nabla f(x_3) = (0, 1/4),\ \nabla f(x_4) = (-1/8, 0),$$

$$\nabla f(x_5) = (0, 1/16).$$

Ya que $\nabla f(x_5) \cong 0$, puede terminarse en este punto. El óptimo exacto es $\mathbf{x}^* = (\mathbf{1/3}, \mathbf{4/3}) = (\mathbf{0'333}, \mathbf{1'333})$.

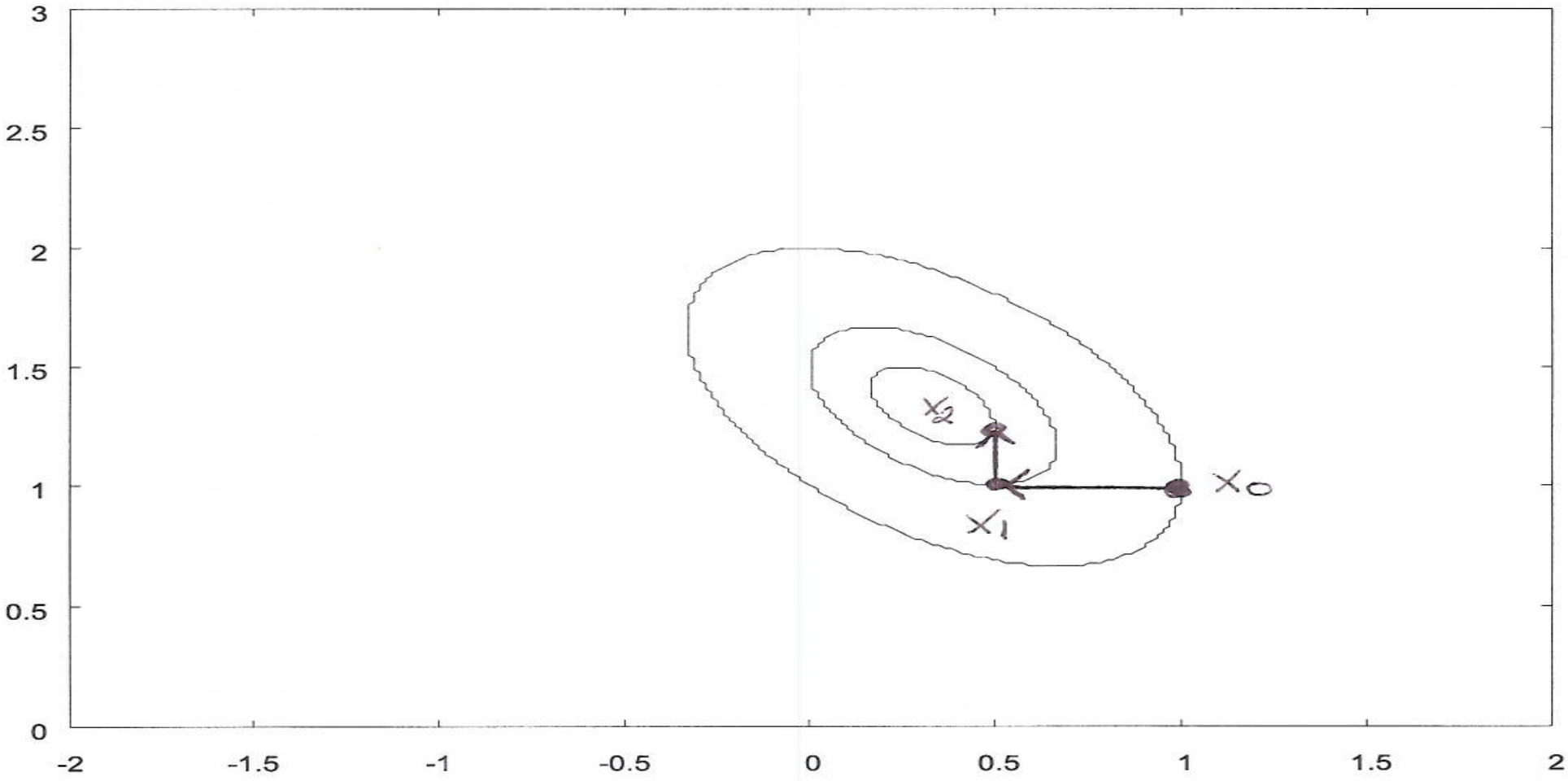
Con Maxima dibujamos las curvas de nivel de la función correspondientes a los distintos puntos encontrados.

```
(%i1) load(draw)$

(%i2) draw2d(implicit(4*x+6*y-2*x^2-2*x*y-2*y^2=4,x,-
2,2,y,0,3),implicit(4*x+6*y-2*x^2-2*x*y
-2*y^2=4.5,x,-2,2,y,0,3),implicit(4*x+6*y-2*x^2-2*x*y-
2*y^2=4.625,x,-2,2,y,0,3));
```

Curvas de nivel de la función

Vemos que los gradientes en dos puntos sucesivos son ortogonales.



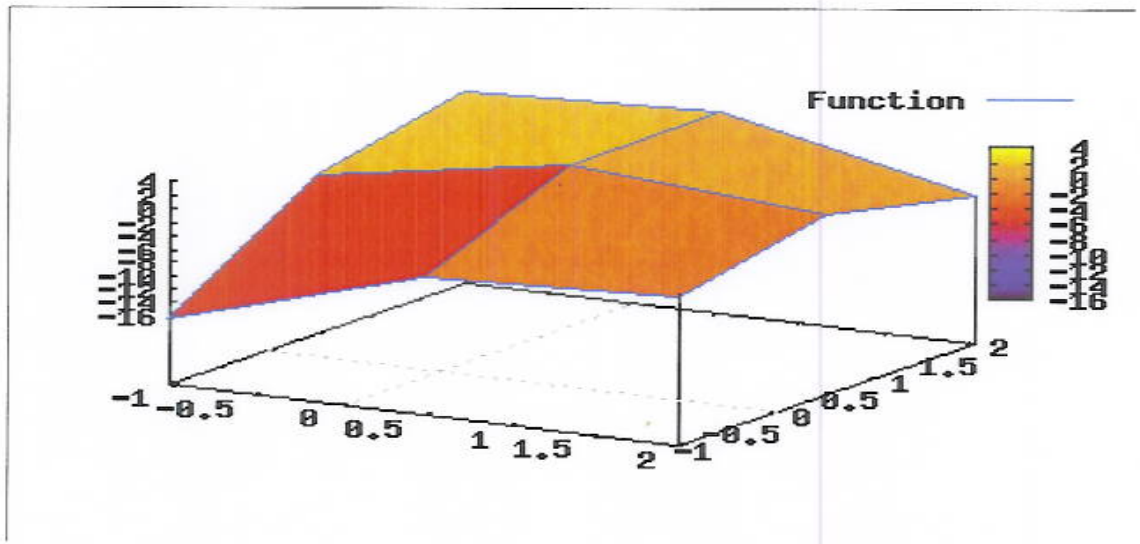
Gráfica 3 dimensional de la función

wxMaxima session

```
/*
wxMaxima 0.7.6 http://wxmaxima.sourceforge.net
Maxima 5.16.3 http://maxima.sourceforge.net
Using Lisp GNU Common Lisp (GCL) GCL 2.6.8 (aka GCL)
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.
```

```
(%i1) wxplot3d(4*x+6*y-2*x^2-2*x*y-2*y^2, [x,-1,2], [y,-1,2],
  [grid,2,2])$
```

(%t1)



(%i2)

Resumen del procedimiento de búsqueda del gradiente

Paso inicial: Elegir ϵ y cualquier solución de prueba inicial x'. Ir a la regla de detención.

Iteración:

- Expresar $f(x' + t\nabla f(x'))$ como función de t.
- Encontrar $t = t^*$ que maximiza $f(x' + t\nabla f(x'))$ para $t \geq 0$. Se puede utilizar un procedimiento de búsqueda en una dimensión.
- Establecer $x' = x' + t^*\nabla f(x')$. Pasar a la regla de detención.

Regla de detención: Evaluar $\nabla f(x')$. Verificar si $| \dfrac{\partial f}{\partial x_j} | \leq \epsilon$ para todo $j = 1, \cdots, n$. Si es así, el proceso se detiene con la x' actual como aproximación al óptimo x^* buscado. En otro caso, se realiza otra iteración.

Una aplicación a la estadística de la programación no lineal: construcción de modelos

- Sea un conjunto de m pares de datos input-output (y_i, z_i), $i = 1, \cdots, m$, del sistema físico.
- Se conjetura una relación input/output $z = h(x, y)$, donde x es un vector de parámetros desconocidos y h es conocida.
- Encontraremos los x que mejor emparejan los datos en el sentido de que minimizan la suma de errores al cuadrado:

$$\sum_{i=1}^{m} \|z_i - h(x, y_i)\|^2 .$$

- Ejemplo de modelo lineal: ajustar los pares de datos por una aproximación polinómica cúbica. Tomando

$$h(x, y) = x_3 y^3 + x_2 y^2 + x_1 y + x_0,$$

donde $x = (x_0, x_1, x_2, x_3)$ es el vector de coeficientes desconocidos del polinomio cúbico.

El método de Newton para una variable

Se desea maximizar una función f de una variable y se considera un punto inicial x_0. Si x_1 es la solución de prueba generada en la primera iteración, la serie de Taylor truncada para x_1 es:

$$f(x_1) \cong f(x_0) + f'(x_0)(x_1 - x_0) + \frac{f''(x_0)}{2}(x_1 - x_0)^2.$$

Al haber fijado x_0 al inicio de la primera iteración, se tiene que $f(x_0), f'(x_0)$ y $f''(x_0)$ también son constantes. La función se puede maximizar de la manera usual al igualar la primera derivada a cero y resolver para x_1:

$$f'(x_1) \cong f'(x_0) + f''(x_0)(x_1 - x_0) = 0$$

de donde

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Resumen del método de Newton

Paso inicial: Elegir ϵ y una solución de prueba inicial x_i. Establecer $i = 1$.

Iteración i:

- Calcular $f'(x_i)$ y $f''(x_i)$. (El cálculo de $f(x_i)$ es opcional.)
- Establecer $x_{i+1} = x_i - \dfrac{f'(x_i)}{f''(x_i)}$.

Regla de detención: Si $|x_{i+1} - x_i| \leq \epsilon$ debe detenerse. De otra manera, se redefine $i = i + 1$.

Ejemplo del método de Newton para una variable

Queremos maximizar la función cóncava

$$f(x) = 12x - 3x^4 - 2x^6.$$

Por inspección elegimos $x_1 = 1$ y $\epsilon = 0'00001$.

Iteración i	x_i	$f(x_i)$	$f'(x_i)$	$f''(x_i)$	x_{i+1}
1	1	7	-12	-96	0'875
2	0'875	7'8439	-2'1940	-62'733	0'84003
3	0'84003	7'8838	-0'1325	-55'279	0'83763
4	0'83763	7'8839	-0'0006	-54'790	0'83762

Tras cuatro iteraciones, el método converge a $x = 0'83762$.

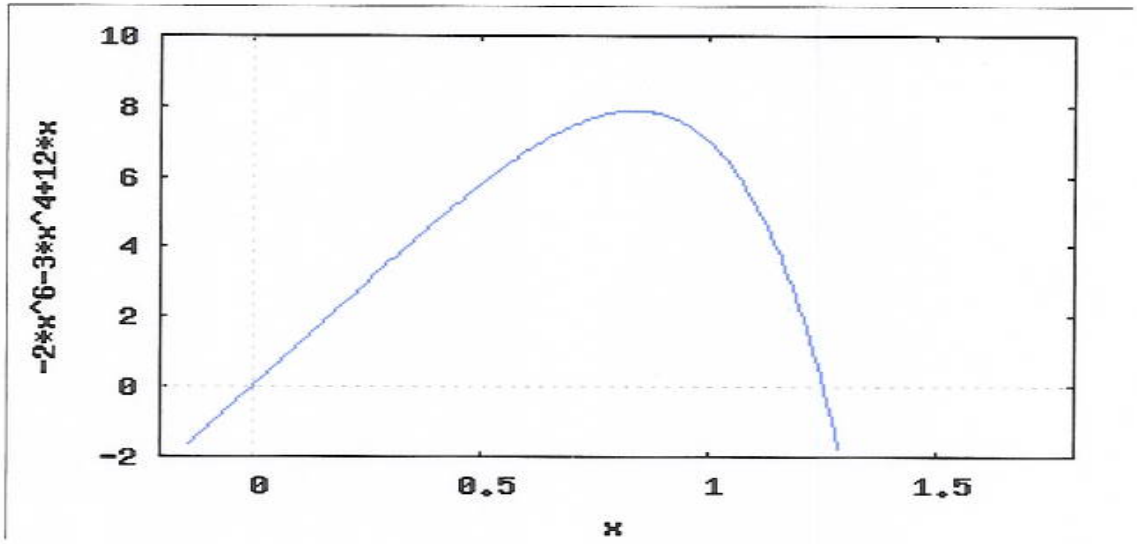
Gráfica de la función

wxMaxima session

```
/*
wxMaxima 0.7.6 http://wxmaxima.sourceforge.net
Maxima 5.16.3 http://maxima.sourceforge.net
Using Lisp GNU Common Lisp (GCL) GCL 2.6.8 (aka GCL)
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.
```

```
(%i1) wxplot2d([12*x-3*x^4-2* x^6], [x,-0.2,1.8], [y,-2,10],
  [nticks,8])$
```

(%t1)



(%i2)

Método de Newton para varias variables

Se desea maximizar una función f, ahora de varias variables y se considera un punto inicial $x = (x_1, \cdots, x_n)$. Si x' es la solución de prueba generada en la primera iteración, razonando como en el caso de una variable, la solución x' que maximiza la función de aproximación tiene la forma:

$$x' = x - [Hf(x)]^{-1} \nabla f(x),$$

donde $Hf(x)$ es la matriz hessiana, $n \times n$, de f evaluada en el punto inicial x y $[Hf(x)]^{-1}$ es la inversa de esta matriz y además estamos suponiendo que el vector gradiente $\nabla f(x)$ se escribe como vector columna.

Ejemplo del método de Newton para varias variables

Consideremos el siguiente problema:

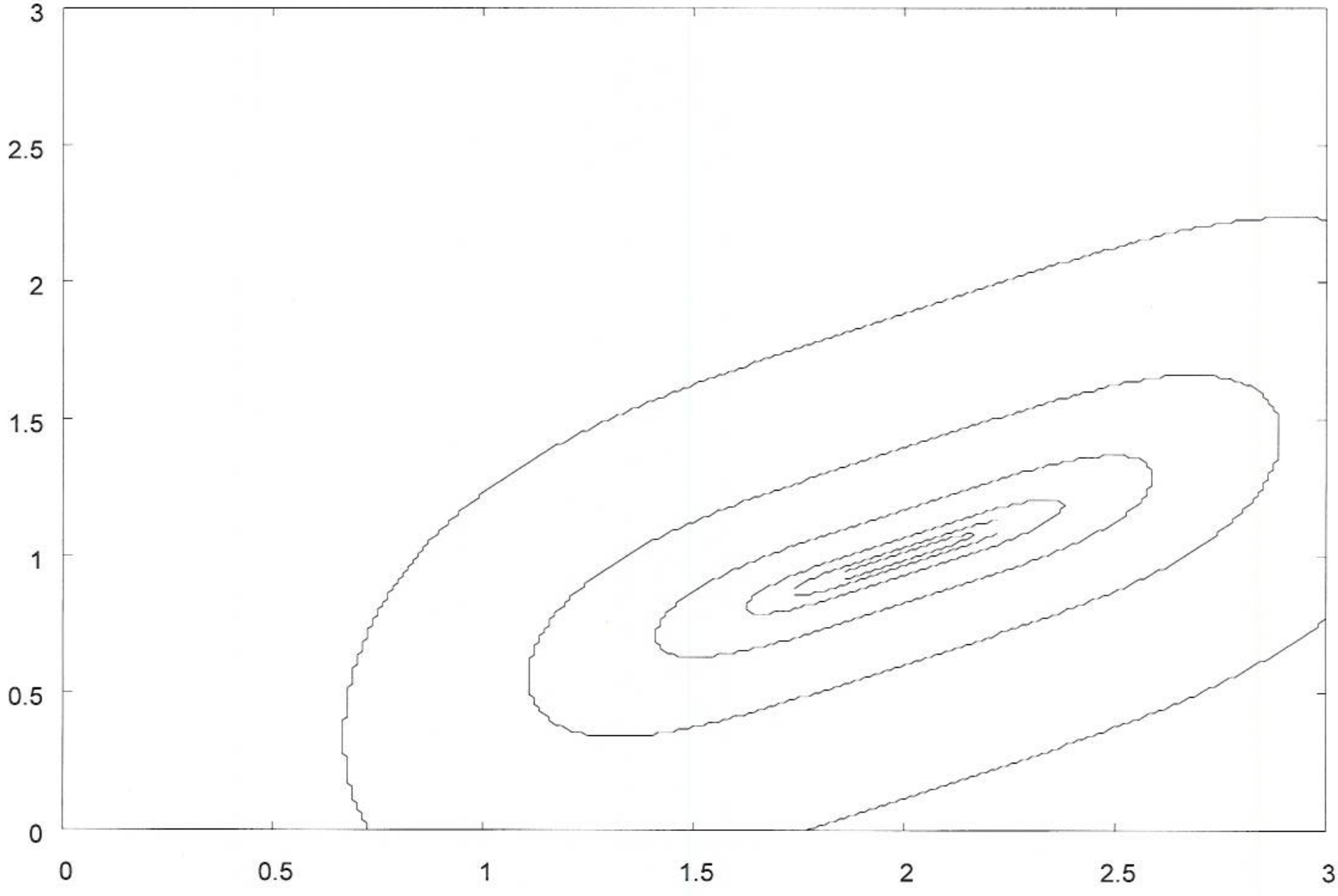
minimizar $(x_1 - 2)^4 + (x_1 - 2x_2)^2$.

El resumen de cálculos utilizados por el método de Newton está dado en la siguiente tabla. En cada iteración, x_{k+1} está dado por $x_{k+1} = x_k - Hf(x_k)^{-1}\nabla f(x_k)$. Tras 6 iteraciones, se obtiene el punto $x_7 = (1'83, 0'91)$. En este punto, $\mid \nabla f(x_7) \mid = 0'04$ y el procedimiento termina. Los puntos generados por el procedimiento pueden identificarse en el gráfico siguiente.

Resumen de los cálculos para el método de Newton

Iter. k	x_k $f(x_k)$	$\nabla f(x_k)$	$Hf(x_k)$	$\nabla f(x_k) Hf(x_k)^{-1}$
1	$(0,3)$ 52	(-44,24)	$\begin{pmatrix} 50 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'67,2'67)
2	$(0'67, 0'33)$ $3'13$	(-9'39,-0'04)	$\begin{pmatrix} 23'23 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'44,-0'23)
3	$(1'11, 0'56)$ $0'63$	(-2'84,-0'04)	$\begin{pmatrix} 11'50 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'30,-0'14)
4	$(1'41, 0'7)$ $0'12$	(-0'8,-0'04)	$\begin{pmatrix} 6'18 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'20,-0'1)
5	$(1'61, 0'8)$ $0'02$	(-0'22,-0'04)	$\begin{pmatrix} 3'83 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'13,-0'07)
6	$(1'74, 0'87)$ $0'005$	(-0'07,0)	$\begin{pmatrix} 2'81 & -4 \\ -4 & 8 \end{pmatrix}$	(-0'09,-0'04)
7	$(1'83, 0'91)$ $0'0009$	(0'0003,-0'04)		

Gráfica de las curvas de nivel de la función $(x_1 - 2)^4 + (x_1 - 2x_2)^2$



Algunos cálculos con Maxima para este mismo ejemplo

Vemos primero, que la matriz hessiana, H, de la función $f(x) = (x-2)^4 + (x-2y)^2$ es semidefinida positiva (su determinante, así como sus elementos H_{11} y H_{22} son mayores e iguales que cero), entonces f es convexa y como es diferenciable, eso implica que sus puntos críticos (aquéllos en los que se anula el gradiente) son mínimos globales.

```
(%i1) f:(x-2)^4+(x-2*y)^2;
(%i2) H:hessian(%,[x,y]);
(%i3) determinant(H);
```

A continuación obtenemos el mínimo (el punto (2,1)), que sale resolviendo el sistema resultante de igualar a cero la derivada parcial de f respecto de x e igualar a cero la derivada parcial de f respecto de y.

```
(%i4) e1:diff(f,x,1);
(%i5) e2:diff(f,y,1);
(%i6) algsys([e1,e2],[x,y]);
```

Con el comando plot3d hacemos el gráfico en tres dimensiones de la función.

```
(%i7) plot3d((x-2)^4+(x-2*y)^2,[x,-10,10],[y,-10,10]);
```

Y realizamos los cálculos correspondientes a la primera iteración del método de Newton, tomando como punto inicial (0,3): evaluamos en ese punto la función f, su gradiente, su matriz hessiana y multiplicamos el gradiente evaluado en el punto por la inversa de la hessiana evaluada en el punto. El resultado se lo restamos al punto inicial para evaluar el nuevo punto.

```
(%i8) H:hessian(f,[x,y]);
(%i9) h:%,x=0,y=3;
(%i10) I:invert(h);
(%i11) load("vect")$
(%i12) grad(f);
(%i13) express(%);
(%i14) ev(%,diff);
(%i15) %,x=0,y=3;
(%i16) g:matrix([-44,24]);
(%i17) g.I;
(%i18) 3-8/3;
```

Sobre la convergencia de los algoritmos

- Los métodos que usan el procedimiento de Newton son rápidos bajo condiciones de continuidad diferenciable de f y definición positiva de la Hessiana en un entorno del óptimo. En general, es necesario empezar en un punto inicial adecuado suficientemente próximo al punto estacionario. El método es costosos en el sentido de que deben calcular y almacenar la matriz Hessiana. Podemos decir que una alternativa son los llamados procedimientos de cuasi-Newton que, de diferentes maneras, aproximan la matriz Hessiana.

  Cuando f es suave cerca de la solución x^* y cuando $Hf(x^*)$ es no singular, el método converge cuadráticamente a x^*, esto es, el error ϵ_k definido como

  $$\epsilon_k := || x_k - x^* ||$$

  satisface $\epsilon_{k+1} \leq C\epsilon_k^2$ para alguna constante $C > 0$.

- El método del gradiente tiene la ventaja de su simplicidad. Si la función f no verifica ciertas condiciones de regularidad, la convergencia del método es lenta. Este algoritmo sirve de base para otros más eficientes

En Bazaraa y otros (1993) se puede leer más acerca de la convergencia y de otros algoritmos para problemas de optimización sin restricciones.

Introducción

Convexidad

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad
- Condiciones de optimalidad
- Interpretación económica. Sensibilidad
- Programación cuadrática: método símplex modificado

Ejercicios

Bibliografía

Problema general de programación con restricciones de igualdad y desigualdad

Problema P:

- min $f(x)$

sujeto a:

- $g_i(x) \leq 0$ para $i = 1, \cdots, m$
- $h_i(x) = 0$ para $i = 1, \cdots, p$
- $x \in S$

Teorema de Kuhn-Tucker: condiciones necesarias

Karush (1939), Kuhn y Tucker (1951). Para el problema P sean f, g_i, $i = 1, \cdots, m$ y $h_i, i = 1, ..., p$ continuamente diferenciables. Supongamos que $\overline{x}$ es un punto factible que resuelve P localmente y tal que los vectores gradiente $\nabla g_i(\overline{x})$ para $i \in B = \{i : g_i(\overline{x}) = 0\}$ y $\nabla h_i(\overline{x})$ son linealmente independientes. Entonces existen escalares λ_i, $i = 1, \cdots, m$ y μ_i, $i = 1, \cdots, p$ tales que

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\overline{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\overline{x}) = 0$$

$$\lambda_i g_i(\overline{x}) = 0, \ i = 1, \cdots, m$$

$$\lambda_i \geq 0, \ i = 1, \cdots, m,$$

junto con las restricciones iniciales

$$g_i(\overline{x}) \leq 0, \ i = 1, \cdots, m,$$

$$h_i(\overline{x}) = 0, \ i = 1, \cdots, p.$$

Los escalares λ_i, $i = 1, \cdots, m$ y μ_i, $i = 1, \cdots, p$ se denominan **multiplicadores de Lagrange**. Las condiciones $\lambda_i g_i(\overline{x}) = 0$ para $i = 1, \cdots, m$ se denominan **condiciones de holgura complementaria** y establecen que si $g_i(\overline{x}) < 0$, entonces $\lambda_i = 0$ y si $\lambda_i > 0$ entonces $g_i(\overline{x}) = 0$. No hay restricciones de signo en los multiplicadores μ_i que corresponden a las restricciones de igualdad.

Condiciones suficientes involucran propiedades de convexidad generalizada en la función objetivo y en las funciones de los dos tipos de restricción.

El problema de selección de cartera revisitado: las condiciones de Kuhn-Tucker con Maxima

```
(%i1) load("vect")$
(%i2) f:0.15*x^2+0.06*y^2+0.14*x*y$
(%i3) g:900-0.1*x-0.05*y$
(%i4) h:10000-x-y$
(%i5) H:hessian(0.15*x^2+0.06*y^2+0.14*x*y,[x,y]);
```

(%o5) $H = \begin{pmatrix} 0.3 & 0.14 \\ 0.14 & 0.12 \end{pmatrix}$

```
(%i6) determinant(%);
(%o6) 0.0164
```

```
(%i7) grad(f)$
(%i8) express(%)$
(%i9) ev(%,diff);
(%o9) [0.14*y+0.3*x,0.12*y+0.14*x,0]
(%i10) grad(g)$
(%i11) express(%)$
(%i12) ev(%,diff);
(%o12) [-0.1,-0.05,0]
(%i13) grad(h)$
(%i14) express(%)$
(%i15) ev(%,diff);
(%o15) [-1,-1,0]
```

No incluimos la condición de holgura complementaria, que trivialmente se va a satisfacer al saturar la solución la restricción de menor o igual.

```
(%i16) load(simplex)$

(%i17) minimize_lp(1, [

0.3*x+0.14*y-0.1*l-m=0,

0.14*x+0.12*y-0.05*l-m=0,

900-0.1*x-(1/20)*y<=0,

10000-x-y=0,

x>=0,y>=0,m>=0,l>=0

]);

(%o17) [1,[y=2000,x=8000,m=40,l=26400]]
```

Interpretación geométrica de las condiciones de Kuhn-Tucker

Si consideramos las condiciones de K-T

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\overline{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\overline{x}) = 0$$

y definimos $B = \{i \ : \ g_i(\overline{x}) = 0\}$, se tiene

$$-\nabla f(\overline{x}) = \sum_{i \in B} \lambda_i \nabla g_i(\overline{x}) + \sum_{i=1}^{p} \mu_i \nabla h_i(\overline{x}),$$

es decir, el opuesto del vector gradiente de la función objetivo es *combinación lineal* de los gradientes de las restricciones de desigualdad saturadas y de los vectores gradientes de las restricciones de igualdad. (Si alguna restricción de desigualdad es no saturada, entonces el correspondiente multiplicador debería ser cero debido a la condición de holguras complementarias.)

Para el problema de la selección de cartera, en la solución óptima $\overline{x}$=(8000,2000), se tiene

$$\nabla g(\overline{x}) = (\frac{-1}{10}, \frac{-1}{20})', \ \nabla h(\overline{x}) = (-1,-1)', \ \nabla f(\overline{x}) = (2680, 1360)'.$$

Los valores encontrados de l=26400 y m=40 satisfacen

$$l\nabla g(\overline{x}) + m\nabla h(\overline{x}) = -\nabla f(\overline{x}).$$

Geométricamente, el vector $-\nabla f(\overline{x})$ está en el cono generado por los vectores $\nabla g(\overline{x})$ y $\nabla h(\overline{x})$.

Ejemplo numérico revisitado

- min $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$

sujeto a:

- $g_1(x_1, x_2) = x_1^2 - x_2 - 3 \leq 0$
- $g_2(x_1, x_2) = x_2 - 1 \leq 0$
- $-x_1 \leq 0$

El punto (3,2) es la solución del problema no restringido: la función f es convexa pues su hessiana es definida positiva y (3,2) es un punto crítico.

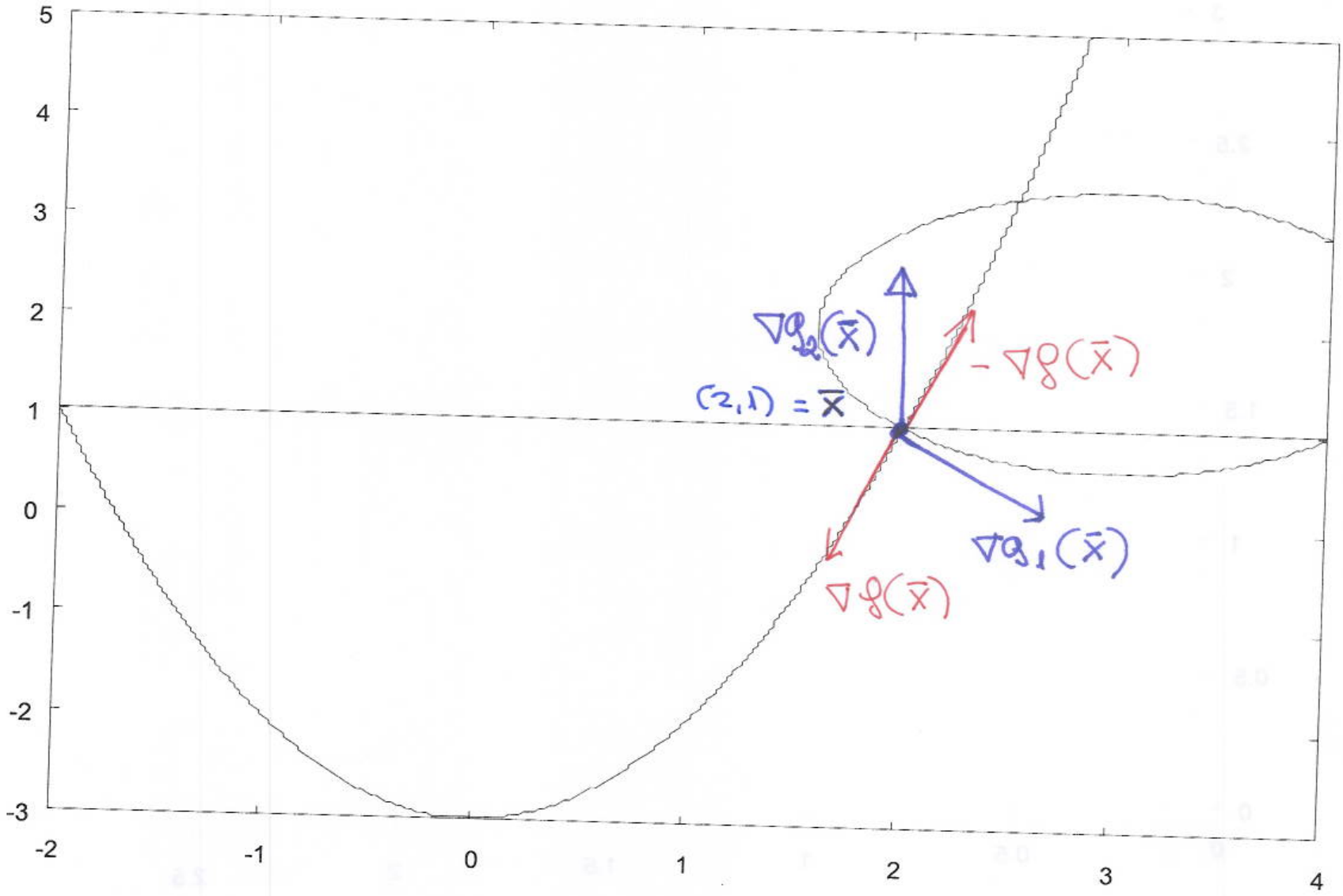
```
(%i1) f:(x-3)^2+(y-2)^2;
(%i2) hessian(%,[x,y]);
(%i3) e1:diff(f,x,1);
(%i4) e2:diff(f,y,1);
(%i5) algsys([e1,e2],[x,y]);
(%o5) [[x=3,y=2]]
```

El punto (2,1) es la solución del problema restringido.

Vemos que el vector $-\nabla f(2,1)$ está en el cono generado por los vectores $\nabla g_1(2,1)$ y $\nabla g_2(2,1)$.

```
(%i1) load(draw)$
(%i2)draw2d(implicit(y=1,x,-2,4,y,-3,5),implicit(x ^2-y-3=0,x,-
2,4,y,-3,5),implicit((x-3) ^2+(y-2) ^2=2,x,-2,4,y,-3,5));
```

Condiciones KT en el ejemplo numérico



Otro ejemplo: análisis gráfico y condiciones de optimalidad

- min $f(x_1, x_2) = x_1^2 - 10x_1 - x_2$

sujeto a:

- $g_1(x_1, x_2) = 6x_1 + 16x_2 - x_1^2 - 4x_2^2 \geq 9$
- $x_1 \geq 0, x_2 \geq 0$

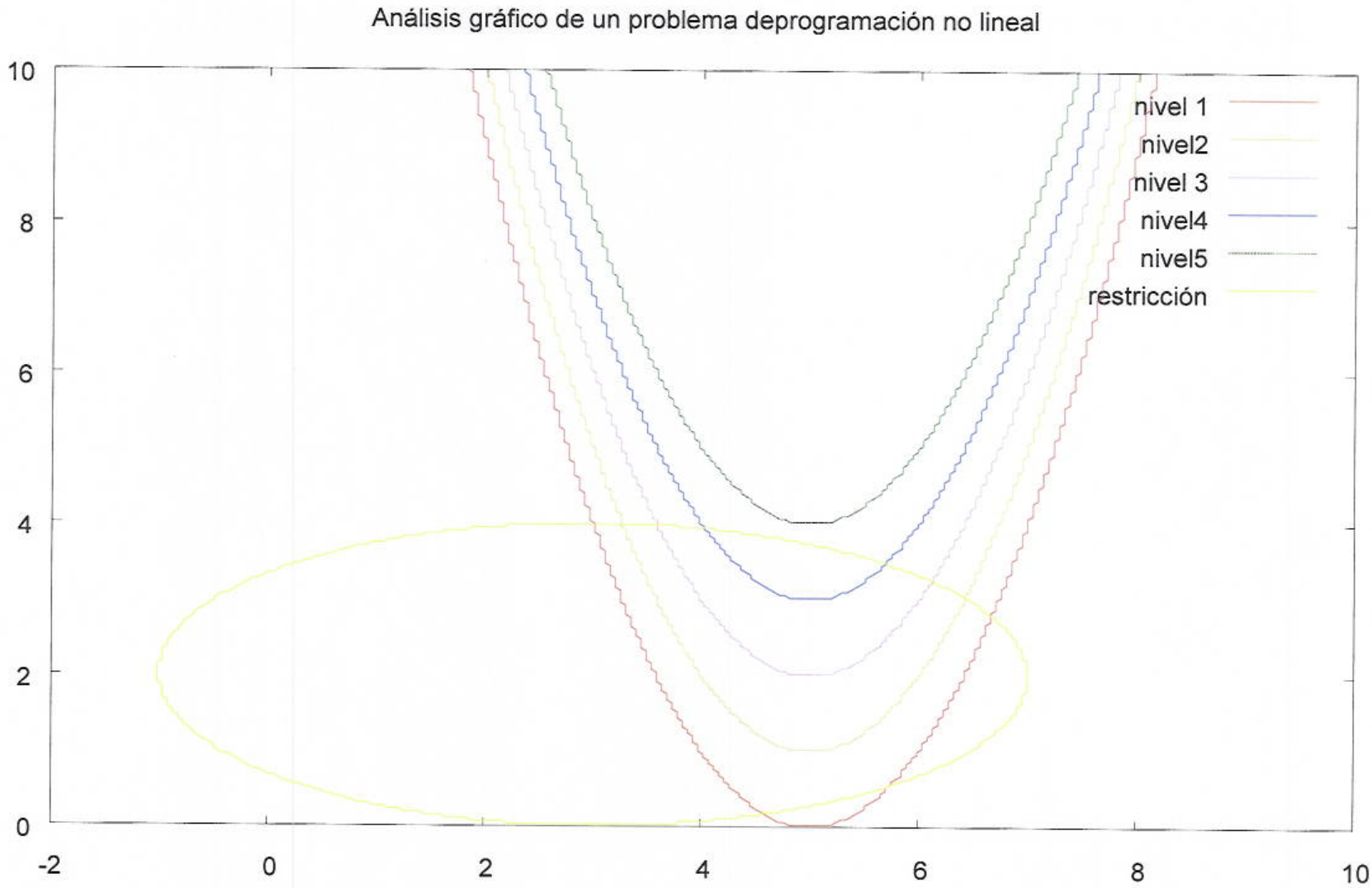
En este caso, la función objetivo es convexa pero no tiene puntos críticos:

```
(%i1) f:x^2-10*x-y;
(%i2) hessian(f,[x,y]);
(%i3) ec1:diff(f,x,1);
(%i4) ec2:diff(f,y,1);
(%i5) algsys([ec1,ec2],[x,y]);
(%o5) []
```

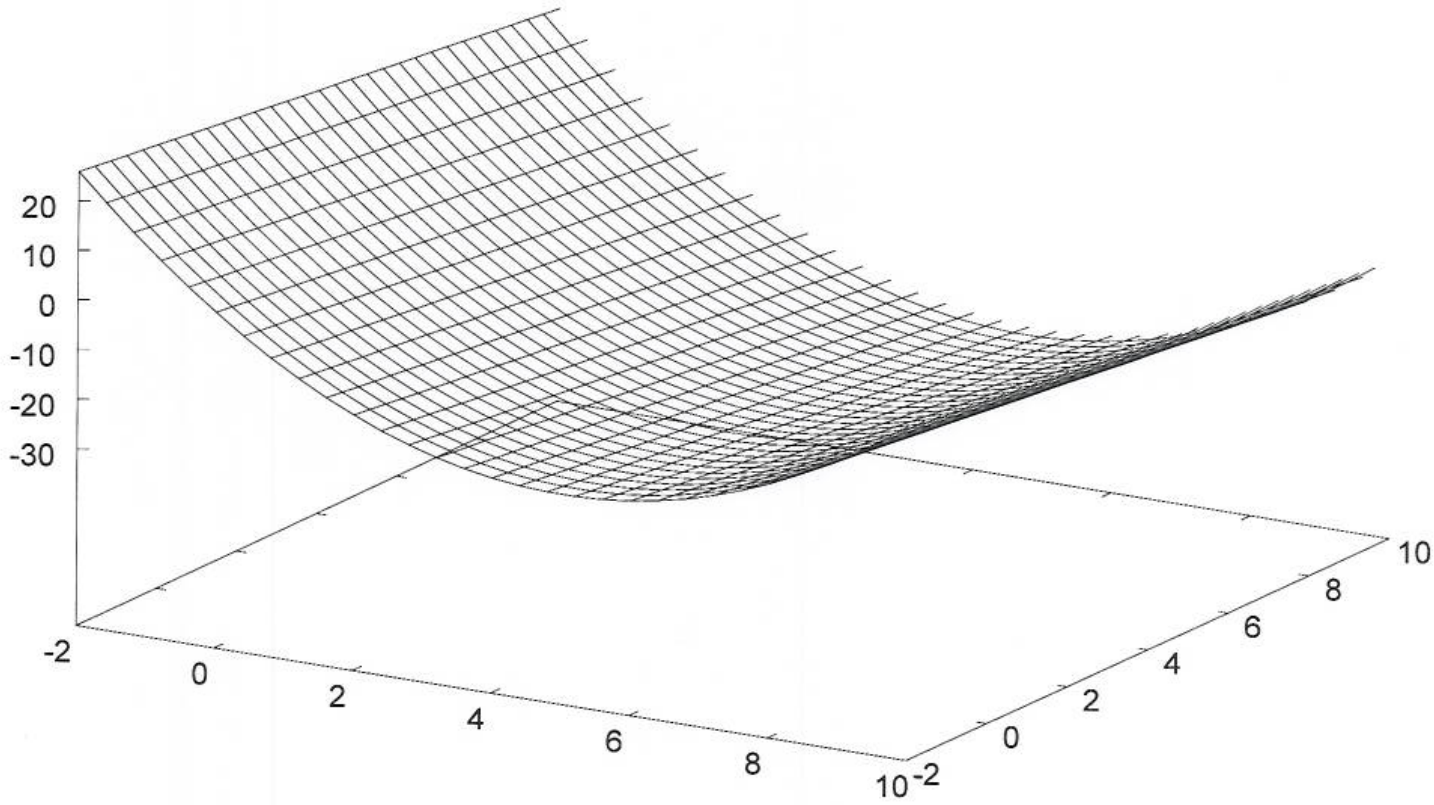
Dibujamos la región factible y las curvas de nivel de la función objetivo:

```
(%i6) load(draw)$
(%i7) draw2d(implicit(6*x+16*y-x^2-4*y^2=9,x,-2,10,y,-2,10));
(%i8) draw2d(key="nivel 1",line_width=1,color=red,implicit(x^2-
10*x-y=-25,x,-2,10,y,0,10),key= "nivel
2",color=orange,implicit(x^2-10*x-y=-26,x,-
2,10,y,0,10),key="nivel 3",color=
violet,implicit(x^2-10*x-y=-27,x,-
2,10,y,0,10),line_width=3,color=blue,key="nivel
4",implicit(x^2-10*x-y=-28,x,-2,10,y,0,10),key="nivel
5",color=green,implicit(x^2-10*x-y=-29,x,-2,10,y,0,10),key=
"restricción",color=yellow,implicit(6*x+16*y-x^2-4*y^2-9=0,x,-
2,10,y,0,10),title="Análisis gráfico de un problema de
programación no lineal");
```

Región factible y curvas de nivel del objetivo



Gráfica en tres dimensiones de la función objetivo



Dibujamos la gráfica en tres dimensiones del objetivo:

```
(%i9) draw3d(explicit(x^2-10*x-y,x,-2,10,y,-2,10));
```

Aplicamos las condiciones de K-T para encontrar el óptimo al problema restringido (no incluimos la condición de holgura complementaria, que trivialmente se va a satisfacer al saturar la solución la restricción de menor o igual):

```
(%i10) g:-6*x-16*y+x^2+4*y^2;
(%i11) diff(g,x,1);
(%i12) diff(g,y,1);
(%i13) e1:2*x-10+l*(-6+2*x);
(%i14) e2:-1+l*(-16+8*y);
(%i15) e3:6*x+16*y-x^2-4*y^2-9;
(%i16) algsys([e1,e2,e3],[x,y,l]);
```

(%o16) [[x=-0.99652830188679,y=1.916690856313498,
l=-1.500434404865334],[x=6.968114348543155,
y=1.747974797479748,l=-0.49598226655583],
[x=5.160442600276626,y=0.31681205392546,
l=-0.074263839811543],[x=4.8679706601467,
y=3.768522394307242,l=0.070680475103856]]

La solución óptima es x=4.8679706601467 e y=3.768522394307242. El valor de la función objetivo:

(%i17) f,x=4.8679706601467,y=3.768522394307242;
(%o17) -28.75109064772514

Cualificación de restricciones

Veamos un problema de programación no lineal que no satisface las condiciones de Kuhn-Tucker:
min $f(x, y) = -x$

sujeto a:

$g_1(x, y) = -(1 - x)^3 + y \leq 0$

$g_2(x, y) = -y \leq 0.$

Como muestra la representación gráfica de la región factible, la solución óptima para este problema es claramente $(x, y) = (1, 0)$.

```
(%i1)load(draw)$
(%i2) draw2d(implicit(-(1-x)^3+y=0,x,-1,2,y,-2,3),implicit(y=0,x,-
1,2,y,-2,3));
```

$$\nabla f(x,y) = (-1,0)',\ \nabla g_1(x,y) = (3(1-x)^2,1)',\ \nabla g_2(x,y) = (0,-1)',$$

de donde, aplicando la condición de Kuhn-Tucker,

$$-1 + \lambda_1 3(1-x)^2 = 0.$$

Tomando $x = 1$, se llegaría a $\lambda_1 = \infty$.

Lo que ocurre es la falta de independencia lineal de los gradientes $\nabla g_i(x,y)$, $i = 1,2$ en el punto $(x,y) = (1,0)$. Ya que

$$\nabla f(1,0) = (-1,0)',\ \nabla g_1(1,0) = (0,1)',\ \nabla g_2(1,0) = (0,-1)',$$

el sistema $\lambda_1 \nabla g_1(1,0) + \lambda_2 \nabla g_2(1,0) = -\nabla f(1,0)$ no tiene solución.

La función Lagrangiana asociada con el problema de optimización con restricciones: una formulación equivalente

Función Lagrangiana asociada con un problema de optimización:

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

donde $\lambda = (\lambda_1, \cdots, \lambda_m)$ y $\mu = (\mu_1, \cdots, \mu_p)$.

Las condiciones de Kuhn-Tucker se pueden expresar en términos de la Lagrangiana como sigue.

Esa formulación equivalente permite resolver también problemas de minimización.

- $\dfrac{\partial L(x,\lambda,\mu)}{\partial x_j} = \dfrac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \dfrac{\partial g_i(x)}{\partial x_j} + \sum_{i=1}^{p} \mu_i \dfrac{\partial h_i(x)}{\partial x_j} = 0,$
  $j = 1, \cdots, n$
- $\lambda_i \dfrac{\partial L(x,\lambda,\mu)}{\partial \lambda_i} = \lambda_i g_i(x) = 0,\ i = 1, \cdots, m$
- $\lambda_i \geq 0,\ i = 1, \cdots, m$
- $\dfrac{\partial L(x,\lambda,\mu)}{\partial \lambda_i} = g_i(x) \leq 0,\ i = 1, \cdots, m$
- $\dfrac{\partial L(x,\lambda,\mu)}{\partial \mu_i} = h_i(x) = 0,\ i = 1, \cdots, p.$

Un problema con función objetivo cuadrática y restricciones mixtas

Supongamos que se quiere minimizar $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$ sujeto a dos restricciones mixtas (esto es, lineal y no lineal) $g_1(x_1, x_2) = x_1^2 + x_2^2 - 5 \leq 0$, y $g_2(x_1, x_2) = x_1 + 2x_2 - 4 \leq 0$. Generaremos las condiciones de Kuhn-Tucker a través de la función Lagrangiana, nuevamente obviando la condición de holguras complementarias:

```
(%i1) L:(x-3)^2+(y-2)^2+l*(x^2+y^2-5)+r*(x+2*y-4);
(%i2) e1:diff(L,x,1);
(%i3) e2:diff(L,y,1);
(%i4) e3:diff(L,l,1);
(%i5) e4:diff(L,r,1);
(%i6) algsys([e1,e2,e3,e4],[x,y,l,r]);
(%o6) [[x=-2/5,y=11/5,l=-7/3,r=74/15],[x=2,y=1,l=1/3,r=2/3]]
```

Evaluamos la función objetivo en el óptimo y comprobamos que se verifica la cualificación de restricción:

```
(%i7) f:(x-3)^2+(y-2)^2$
(%i8) %,x=2,y=1;
(%i9) g1:(x^2+y^2-5)$
(%i10) diff(%,x)$
(%i11) %,x=2;
(%i12) diff(g1,y)$
(%i13) %,y=1;
(%i14) g2:(x+2*y-4)$
(%i15) diff(%,x)$
(%i16) %,x=2;
(%i17) diff(g2,y)$
(%i18) %,y=1;
```

```
(%i19) H:matrix([4,2],[1,2])$
(%i20) determinant(%);
(%o20) 6
```

Se ve fácilmente que menos el gradiente de la función f evaluado en el óptimo es combinación lineal de los gradientes de las restricciones evaluados en dicho óptimo.

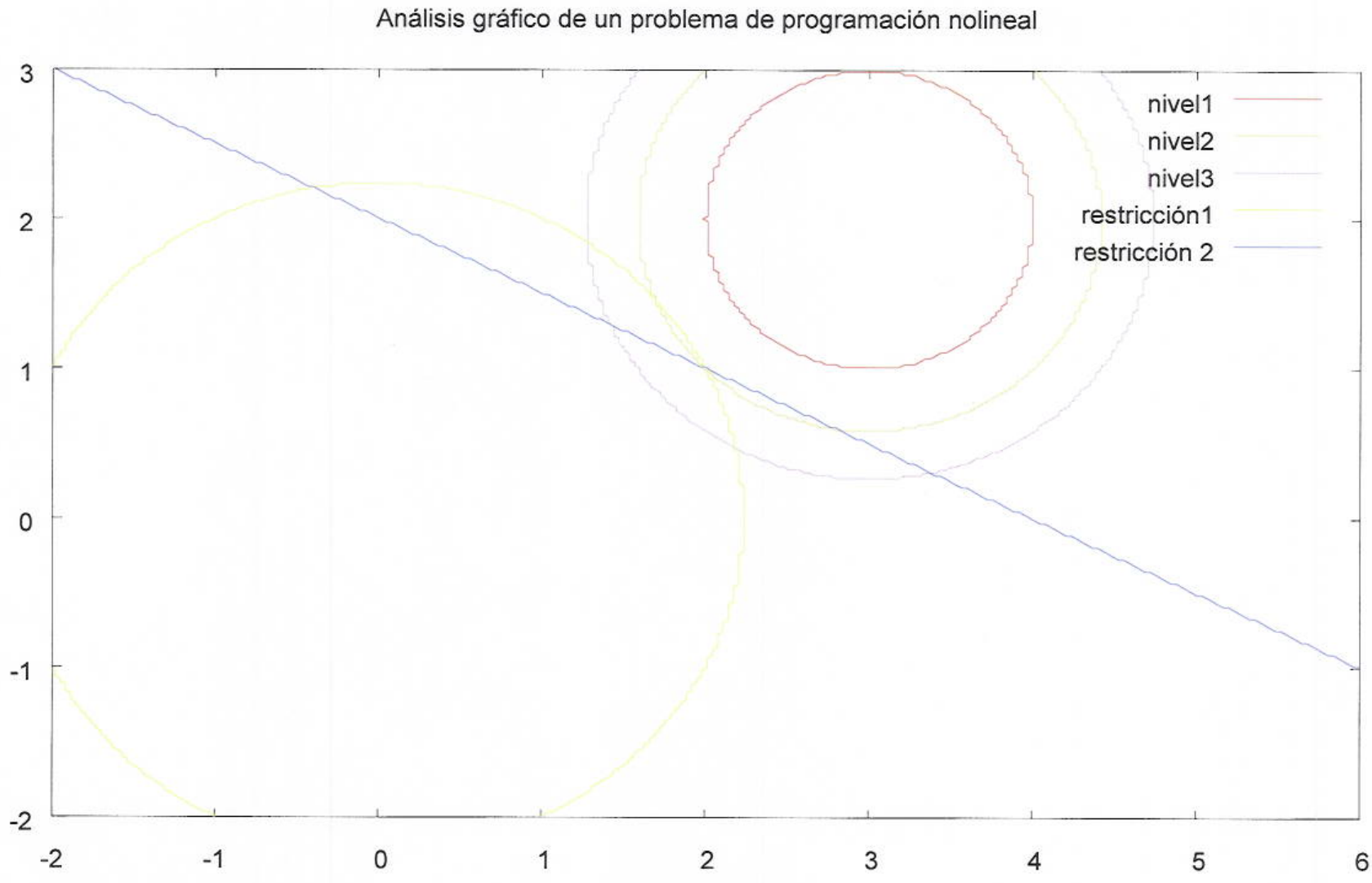
```
(%i21) diff(f,x,1)$
(%i22) %,x=2;
(%i23) diff(f,y,1)$
(%i24) %,y=1;
```

Por último, dibujamos la región factible y la función objetivo.

```
(%i25) load(draw)$
```

```
(%i26) draw2d(key="nivel
1",color=red,implicit((x-3)^2+(y-2)^2=1,x,-2,6,y,-2,3), key="nivel
2",color=orange,implicit((x-3)^2+(y-2)^2=2,x,-2,6,y,-2,3),
key="nivel
3",color=violet,implicit((x-3)^2+(y-2)^2=3,x,-2,6,y,-2,3),
key="restricción
1",color=yellow,implicit(x^2+y^2-5=0,x,-2,6,y,-2,3),
key="restricción 2",color=blue,implicit(x+2*y-4=0,x,-2,6,y,-2,3),
title="Análisis gráfico de un problema de programación no
lineal");
```

Conjunto factible, función objetivo y solución óptima del problema con objetivo cuadrático y restricciones mixtas



Una aplicación a la microeconomía: la función de producción de Cobb-Douglas

Consideremos una función de producción[5] simple que es definida por $q = f(x_1, x_2)$ donde $x_1 > 0$ y $x_2 > 0$ son las variables *inputs* (por ejemplo, trabajo y capital) y q es la cantidad de *output*. Una clase especial de estas funciones se conoce como función de producción de Cobb-Douglas (CD) y se define como $q = Ax_1^{a_1} x_2^{a_2}$ donde $a_1 > 0$ y $a_2 > 0$. La función CD tiene una propiedad interesante de homogeneidad: si los inputs se incrementan por un factor θ, se tiene que $f(\theta x_1, \theta x_2) = A(\theta x_1)^{a_1}(\theta x_2)^{a_2} = Ax_1^{a_1} x_2^{a_2}(\theta)^{a_1+a_2}$. Así, si $a_1 + a_2 = 1$, la función es lineal homogénea y se tiene una economía de escala constante. Si $a_1 + a_2 > 1$ (< 1) la economía de escala es creciente (decreciente).

[5] En Economía, es la relación que existe entre el producto obtenido y la combinación de factores que se utilizan en su obtención.

Consideramos un problema de optimización que involucra una función de producción CD. Supongamos que se desea un output q y que cada unidad de input x_1 cuesta c_1 y cada unidad de x_2 cuesta c_2. El problema es entonces minimizar el coste total $C = c_1 x_1 + c_2 x_2$ de producir q unidades sujeto a la restricción de que $A x_1^{a_1} x_2^{a_2} = q$.

La función Lagrangiana para el problema resulta:

$$L(x, m) = c_1 x_1 + c_2 x_2 + m(q - A x_1^{a_1} x_2^{a_2}).$$

Las condiciones de Kuhn Tucker en términos de la Lagrangiana resultan como sigue.

$$\frac{\partial L(x,m)}{\partial x_1} = c_1 - ma_1 x_1^{a_1-1} A x_2^{a_2} = 0$$

$$\frac{\partial L(x,m)}{\partial x_2} = c_2 - ma_2 x_2^{a_2-1} A x_1^{a_1} = 0$$

$$\frac{\partial L(x,m)}{\partial m} = q - A x_1^{a_1} x_2^{a_2} = 0.$$

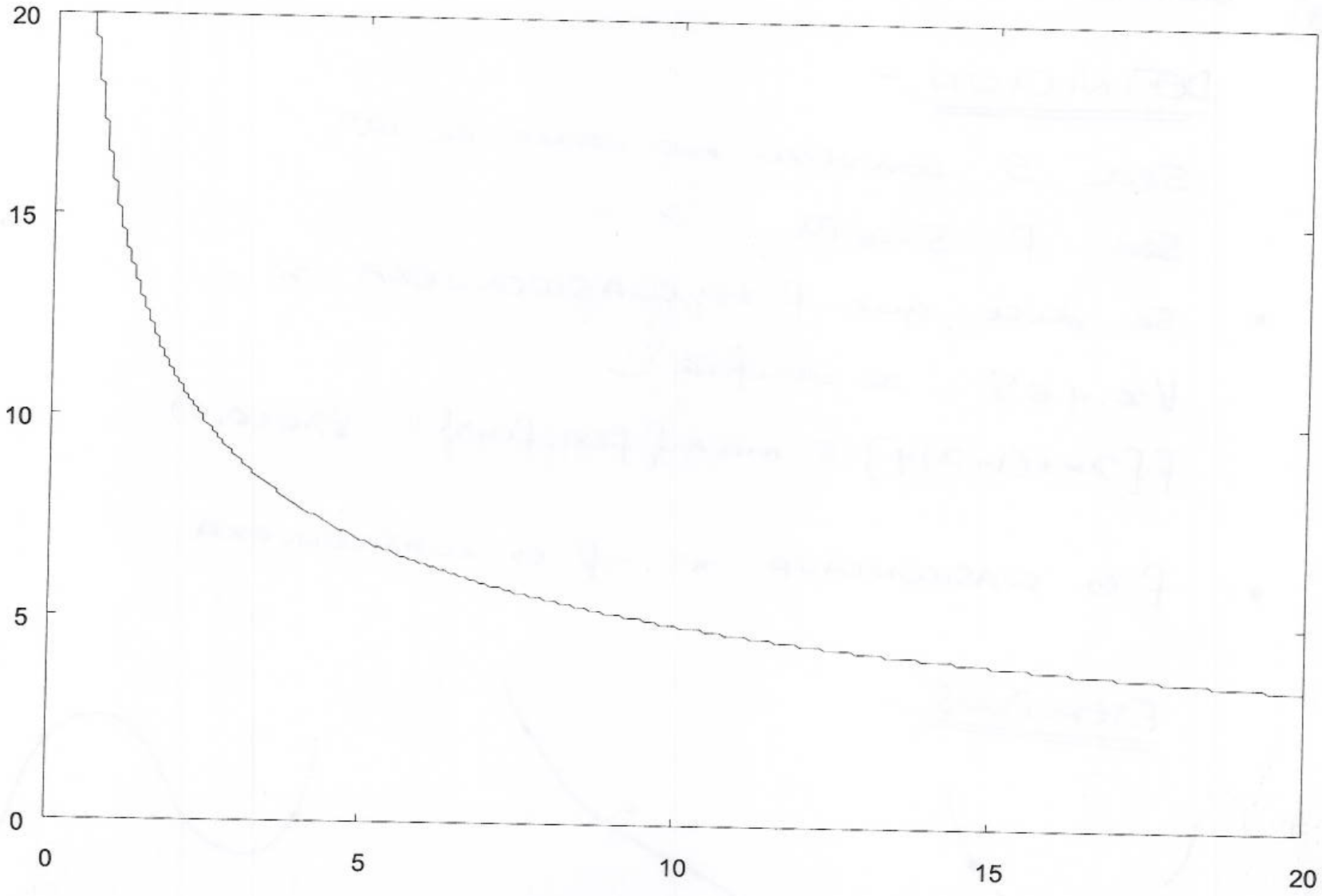
Esto nos proporciona un sistema de tres ecuaciones no lineales. Supongamos

$$a_1 = 1,\ a_2 = 2,\ c_1 = 3,\ c_2 = 4,\ A = 5,\ q = 1200.$$

La función de coste es lineal y la curva $q = A x_1^{a_1} x_2^{a_2}$ en el plano es decreciente y estrictamente convexa:

```
(%i1) load(draw)$
(%i12) draw2d(implicit(5*x^1*y^2=1200,x,0,20,y,0,20));
```

La curva de producción de Cobb-Douglas



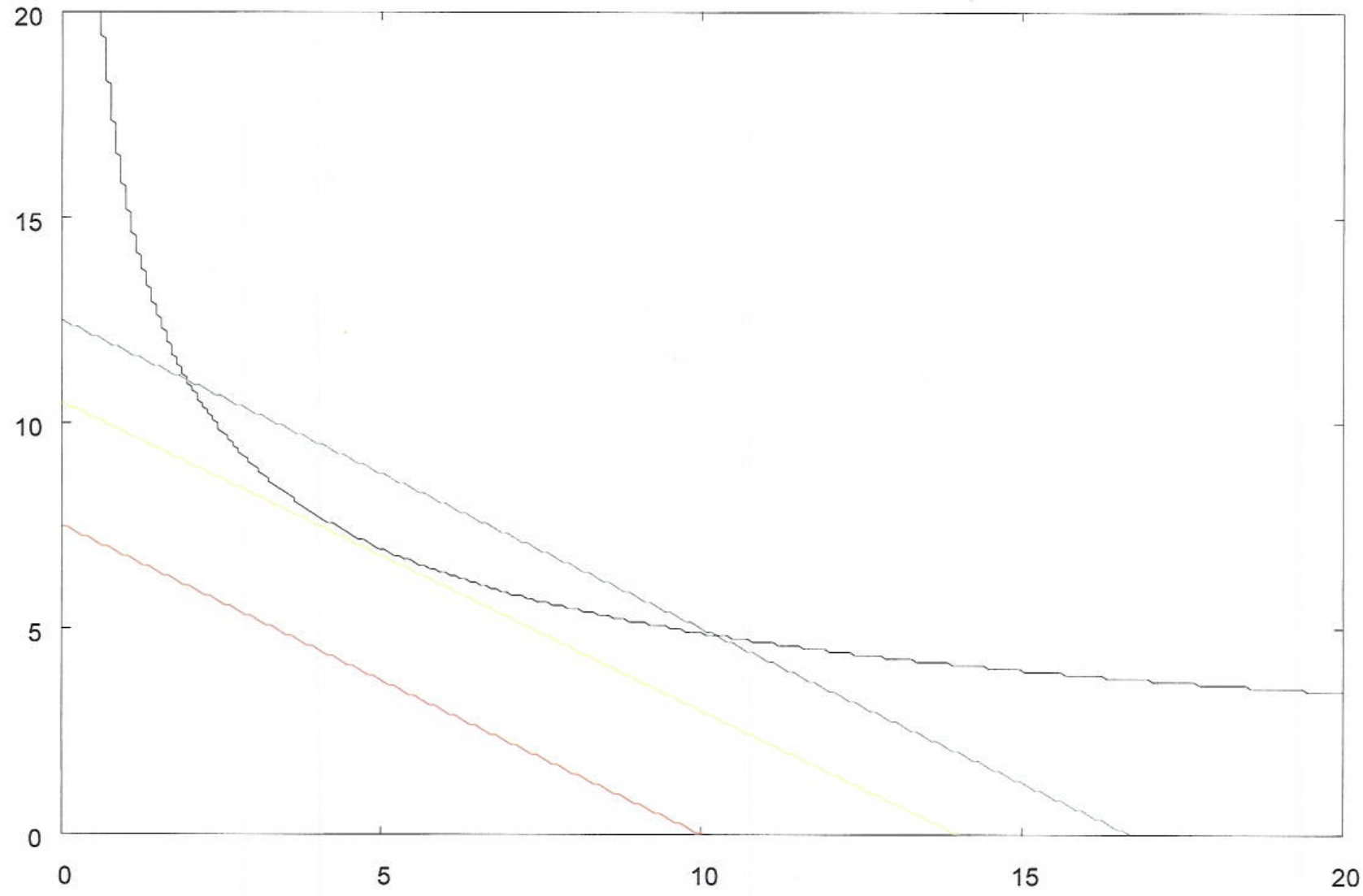
Por tanto, la solución encontrada al aplicar el teorema de Kuhn-Tucker nos va a proporcionar el único punto que minimiza la función de coste:

```
(%i3) e1:3-(m*5*x^1*y^2)/x;
(%i4) e2:4-(m*5*x*y^2*2)/y;
(%i5) e3:1200-5*x*y^2;
(%i6) algsys([e1,e2,e3],[x,y,m]);
(%o6 [[x=-4.107146613652231*%i-2.37126220299375,y=-
6.160719920478345*%i-3.556893304490063,m=-
0.010267866534131*%i-
0.0059281555074834],[x=4.107146613652231*%i-
2.37126220299375,y=6.160719920478345*%i-
3.556893304490063,m=0.010267866534131*%i-
0.0059281555074834],[x=4.742524394082468,
y=7.113786875376279,m=0.011856311118648]]
```

Dibujamos la curva de producción y la función objetivo:

```
(%i7)draw2d(implicit(5*x^1*y^2=1200,x,0,20,y,0,20),color=red,
implicit(3*x+4*y=30,x,0,20,y,0,20),color=yellow,
implicit(3*x+4*y=42,x,0,20,y,0,20),
color=green,implicit(3*x+4*y=50,x,0,20,y,0,20));
```

La función de producción CD y la función de coste



Introducción

Convexidad

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad
- Condiciones de optimalidad
- Interpretación económica. Sensibilidad
- Programación cuadrática: método símplex modificado

Ejercicios

Bibliografía

Interpretación económica de los multiplicadores de Lagrange. Análisis de sensibilidad

Hasta ahora, hemos tratado los multiplicadores de Lagrange como una herramienta algebraica que convierte un problema restringido en uno no restringido.

En el problema de la selección de cartera se obtuvo que $I = 26400$ y $m = 40$, mientras que en el ejemplo de la función objetivo cuadrática y restricciones mixtas se obtuvo que $I = 1/3$ y $r = 2/3$. Cabe preguntarse si estos valores tienen un significado económico con respecto a las restricciones con las que están relacionados. Como en programación lineal, los multiplicadores de Lagrange proporcionan información sobre la *mejora aproximada* en la función objetivo si un lado derecho de una restricción particular se relaja en una *pequeña cantidad*.

Si, por ejemplo, en el problema de selección de cartera, se incrementa el lado derecho de la primera restricción en una unidad, es decir, $900 - 0'10x_1 - 0'05x_2 \leq 1$ o, equivalentemente, $g_1(x_1, x_2) = 899 - 0'10x_1 - 0'05x_2 \leq 0$, y resolvemos de nuevo el problema, se obtiene la solución óptima $x_1 = 7980$, $x_2 = 2020$ y $f = 12053628$.

Se obtiene una reducción (mejora) en la función objetivo (es decir, en la varianza de la cartera) de 12080000-12053628=26372 unidades.

Esta diferencia, es muy próxima al multiplicador de Lagrange $\lambda = 26400$ obtenido como solución del problema original. Este resultado se formaliza en el siguiente resultado:

TEOREMA Considerar el problema P: min $f(x)$

sujeto a:

$g_i(x) \leq a_i,\ i = 1, \cdots, m$

$h_i(x) = b_i,\ i = 1, \cdots, p.$

Supongamos que para $a = (a_1, \cdots, a_m) = 0$, $b = (b_1, \cdots, b_p) = 0$ existe una solución $\overline{x}$ que satisface la cualificación de restricción con multiplicadores de Lagrange $\lambda_i \geq 0$, $i = 1, \cdots, m$, y μ_i, $i = 1, \cdots, p$. Entonces para cada $(a, b) \in \mathbb{R}^{m+p}$ en el entorno del $(0, 0)$ existe una solución $x(a, b)$, mínimo de P, que depende continuamente de (a, b) tal que $x(0, 0) = \overline{x}$. Además:

$$\frac{\partial f}{\partial a_i}(x(a, b))]_{0,0} = -\lambda_i,\ i = 1, \cdots, m$$

$$\frac{\partial f}{\partial b_i}(x(a, b))]_{0,0} = -\mu_i,\ i = 1, \cdots, p.$$

En el ejemplo de optimizar una función objetivo cuadrática con restricciones mixtas, si la primera restricción se escribe como $x_1^2 + x_2^2 - 5 \leq 0'1$ o equivalentemente $g_1(x_1, x_2) = x_1^2 + x_2^2 - 5'1 \leq 0$, el valor $\Delta a_1 \times \lambda_1 = 0'1 \times 1/3 = 1/30$ debería ser la *mejora aproximada* (es decir, la reducción) en el valor óptimo de la función objetivo $f(x_1, x_2)$.

Este incremento en el lado derecho de la primera restricción de 0'1 aumenta la región factible y mejora la función objetivo en aproximadamente 0'033 unidades.

Lo comprobamos viendo los resultados de Maxima. (La mejora *real* es igual a 2-1'968468797=0'031531203 unidades.)

```
(%i1) L:(x-3)^2+(y-2)^2+l*(x^2+y^2-5.1)+r*(x+2*y-4);
(%i2) e1:diff(L,x,1);
(%i3) e2:diff(L,y,1);
(%i4) e3:diff(L,l,1);
(%i5) e4:diff(L,r,1);
(%i6) algsys([e1,e2,e3,e4],[x,y,l,r]);
(%o6) [[x=(4*sqrt(2)*sqrt(19)-38)/(5*sqrt(2)*sqrt(19)),
y=(8*sqrt(2)*sqrt(19)+19)/(5*sqrt(2)*sqrt(19)),
l=-(4*sqrt(38)+19)/19,r=(32*sqrt(2)*sqrt(19)+266)/95],
[x=(4*sqrt(2)*sqrt(19)+38)/(5*sqrt(2)*sqrt(19)),
y=(8*sqrt(2)*sqrt(19)-19)/(5*sqrt(2)*sqrt(19)),
l=(4*sqrt(38)-19)/19,r=-(32*sqrt(2)*sqrt(19)-266)/95]]
```

```
(%i7) e:(4*(2)^(0.5)*(19)^(0.5)+38)/(5*(2)^(0.5)*(19)^(0.5));
(%o8) 2.032882800593795
(%i9) i:(8*(2)^(0.5)*(19)^(0.5)-19)/(5*(2)^(0.5)*(19)^(0.5));
(%o10) 0.9835585997031
(%i11) f:(x-3)^2+(y-2)^2;
(%i12) %,x=2.032882800593795,y=0.9835585997031;
(%o12) 1.968468797624824
```

Dualidad Lagrangiana

Considerar el problema P:

$\min \quad f(x)$

sujeto a:

$g_i(x) \leq a_i, \ i = 1, \cdots, m$

$h_i(x) = b_i, \ i = 1, \cdots, p$

$x \in S$.

Se define el **problema dual Lagrangiano D** como:

$$\max \ \mathbf{I}(\lambda, \mu) = \min[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \ : \ x \in S]$$

sujeto a:

$\lambda \geq 0$

Consideremos el problema de minimizar $f(x) = (x-2)^2 + 1$ sujeto a $g(x) = x - 1 \leq 0$ para el que la solución óptima es $\overline{x} = 1$. La Lagrangiana es $L(x, \lambda) = (x-2)^2 + 1 + \lambda(x-1)$, que es una función de dos variables (x, λ). Usando el comando de wxMaxima *wxplot3d*, dibujamos la superficie de la Lagrangiana para el rango $x = -2..3$ y $\lambda = 0..4$ y obtenemos el gráfico de la Lagrangiana que aparece en la siguiente figura.

```
(%i1) wxplot3d((y-2)^2+1+x*(y-1), [x,0,4], [y,-2,3], [grid,10,10])$
```

La gráfica sugiere que la Lagrangiana presenta un punto de silla en torno al punto $(x, \lambda) = (1, 2)$, es decir,

$$L(1, \lambda) \leq L(1, 2) \leq L(x, 2).$$

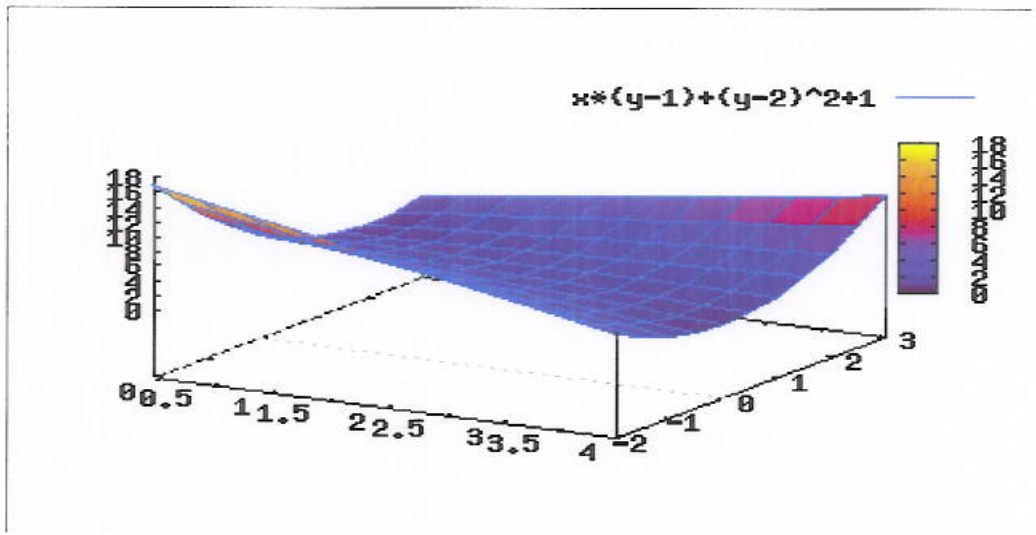
Gráfico de la Lagrangiana

wxMaxima session 1 / 1

```
/*
wxMaxima 0.7.6 http://wxmaxima.sourceforge.net
Maxima 5.16.3 http://maxima.sourceforge.net
Using Lisp GNU Common Lisp (GCL) GCL 2.6.8 (aka GCL)
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.
```

```
(%i1) wxplot3d((y-2)^2+1+x*(y-1), [x,0,4], [y,-2,3], [grid,10,10]);
```

(%t1)



(%o1)

(%i2)

Cuando aplicamos el test de los autovalores a la Hessiana de L, obtenemos uno positivo y uno negativo, lo que implica que la Lagrangiana tiene un punto de silla:

```
(%i2) L:(x-2)^2+1+lambda*(x-1)$
(%i3) H:hessian(%,[x,lambda])$
(%i4) eigenvalues(H);
(%o4) [[1 - sqrt(2), sqrt(2) + 1], [1, 1]]
```

Con los comandos adecuados de Maxima calculamos el punto de silla de la Lagrangiana, es decir, el valor de x que minimiza L y el valor de λ que maximiza L.

```
(%i5) f:(x-2)^2+1$
(%i6) g:x-1$
(%i7) L:f+lambda*g$
(%i8) diff(L,x,1);
(%o8) [lambda + 2 (x - 2)]
(%i9) xsol:solve(%,x);
(%o9) [x=2-lambda/2]
(%i10) L,x=xsol$
(%i11) diff(%,lambda);
(%o11) [1-lambda/2]
(%i12) solve(1-lambda/2,lambda);
(%o12) [lambda = 2]
(%i13) xsol:2-2/2;
(%o13) 1
```

Los cálculos prueban que $L(1,\lambda) \leq L(1,2) \leq L(x,2)$.

Si usamos las condiciones de Kuhn-Tucker haciendo uso de la formulación que hemos visto a través de la Lagrangiana, llegamos a los mismos resultados para x y λ:

```
(%i14) e1:diff(L,x,1)$
(%i15) e2:diff(L,lambda,1)$
(%i16) algsys([e1,e2],[x,lambda]);
(%o16) [[x = 1, lambda = 2]]
```

Para este problema, la solución obtenida usando el análisis de los puntos de silla de la Lagrangiana y las condiciones de Kuhn-Tucker coincide. No es una coincidencia. Cuando la función objetivo y las restricciones verifican ciertas condiciones de regularidad (por ejemplo, convexidad de las funciones involucradas) se verifica el siguiente teorema que muestra la relación entre el criterio de los puntos de silla y las condiciones K-T.

TEOREMA Un punto factible $\overline{x}$ del problema P y los multiplicadores $\overline{\lambda} \geq 0$ y $\overline{\mu}$ satisfacen las condiciones de Kuhn-Tucker si y sólo si satisfacen las condiciones de punto de silla

$$L(\overline{x}, \lambda, \mu) \leq L(\overline{x}, \overline{\lambda}, \overline{\mu}) \leq L(x, \overline{\lambda}, \overline{\mu}).$$

En el problema de minimizar $f(x) = (x-2)^2 + 1$ sujeto a $g(x) = x - 1 \leq 0$, usamos Maxima para diferenciar $L(x, \lambda)$ con respecto a x para un valor dado de λ. Se obtuvo $x(\lambda) = 2 - \lambda/2$ como el óptimo de x en términos de λ. Luego, se sustituyó $x(\lambda)$ en la Lagrangiana para obtener una nueva función $\mathbf{l}(\lambda) = \lambda^2/4 + 1 + \lambda(1 - \lambda/2)$. Maximizando $\mathbf{l}$ con respecto a $\lambda \geq 0$ se obtuvo $\lambda = 2$, y al sustituir en $x(\lambda)$ proporciona el valor óptimo de $x = 1$. De esta forma, en los óptimos obtenidos con este enfoque, $f(x = 1) = 2 = \mathbf{l}(\lambda = 2)$, es decir, min f=max$\mathbf{l}$.

Este enfoque de encontrar la solución óptima a través de una función relacionada como $\mathbf{l}(\lambda)$ tiene numerosas implicaciones en cuestiones teóricas y computacionales de la programación no lineal.

Teorema de dualidad

Bajo ciertas condiciones de regularidad (es decir, convexidad de las funciones f y g_i, linealidad de h_i en términos de x y cualificación de restricción), se tiene que:

$$min[f(x) \ : \ x \in S, \ g_i(x) \leq 0, \ i = 1, \cdots, m, \ h_i(x) = 0, \ i = 1, \cdots, p]$$

$$= max[\mathbf{I}(\lambda, \mu) \ : \ \lambda \geq 0].$$

Además, la función $\mathbf{I}(\lambda, \mu)$ es cóncava sobre cualquier subconjunto del dominio.

El teorema puede ser útil para resolver problemas de programación no lineal restringidos donde el vector de decisión x puede ser expresado en términos de (λ, μ) tras minimizar la función Lagrangiana $L(x, \lambda, \mu)$.

Si esto es posible, entonces $x = (\lambda, \mu)$ puede ser sustituido en L para obtener la función dual Lagrangiana $\mathbf{l}(\lambda, \mu)$, la cual es cóncava, y de esta forma su máximo global puede ser obtenido con los procedimientos habituales.

Finalmente, una vez que el óptimo $(\overline{\lambda}, \overline{\mu})$ son calculados, pueden ser sustituidos en x para general el valor óptimo del vector de decisión.

Una aplicación de la dualidad Lagrangiana: los problemas de control óptimo

El control óptimo se refiere al problema de encontrar una ley de control para un sistema dado y de tal forma que se satisfaga cierto criterio de optimalidad. Estos problemas incluyen una función de coste que es función del estado del sistema y de las variables de control. A menudo aparecen también restricciones en el problema.

Hay diversas formulaciones del problema de control óptimo, que varían dependiendo del criterio de optimalidad, del tipo de dominio de tiempo (continuo o discreto) y de la presencia de diversos tipos de restricciones.

Consideremos el problema del propietario de una mina que debe decidir qué cantidades quiere extraer de carbón. Posee los derechos de extracción desde la fecha 0 hasta la fecha T. A fecha 0 se estima que hay una cantidad x_0 de carbón en la tierra, y que la reserva instantánea de carbón x_t decrece en función de la cantidad u_t que el propietario decide extraer. El propietario extrae carbón a un coste de u_t^2/x_t y lo vende a un precio constante de p. Supongamos que el carbón remanente en el instante T no tiene valor. Escoge extraer las cantidades u_t que maximizan el beneficio. Una formulación del problema en tiempo discreto sería maximizar el objetivo

$$\sum_{t=0}^{T-1}(pu_t - \frac{u_t^2}{x_t})$$

sujeto a la condición dinámica en el estado del sistema:

$$x_{t+1} = x_t - u_t.$$

El control óptimo en origen estuvo estrictamente relacionado con la teoría del cálculo desarrollada por Johann Bernoulli (1667-1748), Isaac Newton (1642-1727), Leonhard Euler (1707-1793), Giuseppe Luigi Lagrange (1736-1813) y Andrien Legendre (1752-1833) entre otros. Avance importante en la teoría de control óptimo en el siglo XX fue por ejemplo el desarrollo de la programación dinámica por Richard Bellman (1920-1984).

En ciertos métodos de resolución, los problemas de control óptimo se discretizan y se convierten en problemas de programación matemática no lineal.

El control óptimo se ha aplicado a diversos campos, como la aeronáutica (planificación del vuelo de aviones comerciales que minimice el combustible consumido por el avión en un vuelo entre dos aeropuertos asignados teniendo en cuenta las zonas de no sobrevuelo existentes y los datos meteorológicos como las tormentas y el viento previstos), el control de proceso, la robótica (planificación de trayectorias de robots que minimiza algún funcional de coste evitando posibles obstáculos presentes en el espacio de trabajo del robot, restricciones sobre la tarea a realizar, como por ejemplo operar en contacto con algunas piezas de trabajo, aplicar una fuerza asignada o moverse a una velocidad asignada durante parte de la tarea), la bioingeniería, la economía, las finanzas y las ciencias empresariales y continúa siendo un área de investigación activa dentro de la teoría de control.

Un ejemplo de problema de control discreto cuadrático lineal

Consideremos el problema de minimizar el objetivo

$$J = \sum_{i=0}^{N-1} (q_i x_i^2 + r_i u_i^2) + q_N x_N^2$$

sujeto a la condición dinámica del estado

$$x_{i+1} = a_i x_i + b_i u_i, \ i = 0, \cdots, N-1$$

donde $x_0, \{q_i\}_{i=0}^{N}, \{r_i\}_{i=0}^{N-1}, \{a_i\}_{i=0}^{N-1}, \{b_i\}_{i=0}^{N-1}$ son constantes dadas. Para resolver este problema de control como un problema de programación no lineal, se tratan todas las variables de estado x_i y todas las variables de control u_i como variables de decisión, es decir, se define el vector de decisión como $x = (x_1, \cdots, x_N, u_0, \cdots, u_{N-1})$.

Los comandos de Maxima para resolver el problema con $N = 2$ y los coeficientes $\{a_i\}_{i=0}^{N-1}, \{b_i\}_{i=0}^{N-1}$ todos iguales a 1, son como sigue. Los primeros, definen el objetivo y las condiciones dinámicas del estado

```
(%i1) c0:q0*x0^2+r0*u0^2$
(%i2) c1:q1*x1^2+r1*u1^2$
(%i3) f:c0+c1+q2*x2^2$
(%i4) h1:x1-x0-u0$
(%i5) h2:x2-x1-u1$
```

Se introduce la Lagrangiana:

```
(%i6) L:f+m1*h1+m2*h2$
```

y se diferencia con respecto a todas las variables de decisión del vector x:

(%i7) Lu0:diff(L,u0,1)$

(%i8) Lx1:diff(L,x1,1)$

(%i9) Lu1:diff(L,u1,1)$

(%i10) Lx2:diff(L,x2,1)$

Las variables de decisión son calculadas en términos de los dos multiplicadores de Lagrange m_i, $i = 1, 2$.

(%i11) solxu:algsys([Lx1,Lx2,Lu0,Lu1],[x1,x2,u0,u1])$

La dual Lagrangiana $\mathbf{l}(m_1, m_2)$ se calcula y maximiza tras sustituir $x(m_1, m_2)$ en L:

(%i12) el:L,x1=(m2-m1)/(2*q1),x2=-m2/(2*q2),u0=m1/(2*r0),u1=m2/(2*r1)$

```
(%i13) elm1:diff(el,m1,1)$
(%i14) elm2:diff(el,m2,1)$
```

Se encuentran los multiplicadores de Lagrange en términos de los parámetros del modelo como sigue:

```
(%i15) algsys([elm1,elm2],[m1,m2])$
```

Finalmente, la solución para el problema completo, o sea, los valores óptimos de u_0, u_1, x_1 y x_2 se calculan en términos de los parámetros del modelo. Una revisión cuidadosa de la estructura del vector de decisión óptimo indica que las variables de control óptimas u_i son función lineal de los estados x_i en cada instante de tiempo.

En Parlar y Vickson (1980) se resuelve como un problema de programación no lineal por medio de la dualidad Lagrangiana un problema de control en tiempo discreto cuadrático "vectorial" con $N = 25$ períodos de tiempo.

Algoritmos y optimización “convexa" con restricciones

En las secciones anteriores, se han estudiado algunos casos especiales de programación convexa y relativos a problemas no restringidos, a problemas con función objetivo cuadrática y restricciones lineales o las condiciones necesarias y suficientes de optimalidad como parte de la teoría del caso general.

La intención es ahora presentar una introducción al enfoque algorítmico para resolver problemas generales de programación convexa (donde la función objetivo f que se debe maximizar es cóncava y las funciones de restricción g_i son convexas).

No existe, como cabe esperar, un algoritmo único y los existentes se pueden ubicar, en su mayoría, dentro de una de las tres siguientes categorías.

Algoritmos de gradiente, en los que se modifica de alguna manera el procedimiento de búsqueda del gradiente estudiado para problemas sin restricciones para evitar que la trayectoria de búsqueda atraviese la frontera de restricción.

Algoritmos secuenciales no restringidos. Incluyen los denominados métodos de función de finalización y de función barrera. Estos algoritmos convierten el problema de optimización restringida original en una sucesión de problemas de optimización no restringida cuyas soluciones óptimas convergen a la solución óptima del problema original. Cada uno de estos problemas de optimización no restringida se pueden resolver por medio del procedimiento de búsqueda del gradiente ya visto. La conversión se logra al incorporar las restricciones a una función de penalización (o función barrera) que se resta de la función objetivo, con el fin de imponer una penalización al no cumplimiento de cualquier restricción.

Algoritmos de aproximación secuencial. Incluye métodos de aproximación lineal y aproximación cuadrática. Estos algoritmos sustituyen la función objetivo no lineal por una sucesión de aproximaciones lineales o cuadráticas. Para problemas de optimización linealmente restringidos, estas aproximaciones permiten la aplicación repetida de los algoritmos de programación lineal o cuadrática. Uno de estos algoritmos es el **algoritmo de Frank-Wolfe** (1956) para el caso de programación convexa linealmente restringida (de manera que las restricciones son $Ax \leq b$ y $x \geq 0$, en forma matricial).

Algoritmo de aproximación lineal secuencial (Frank-Wolfe)[6]

Se partirá de una solución inicial x^0. Haremos uso de la expansión de primer orden de Taylor como aproximación lineal de la función f, función objetivo a maximizar, en puntos x del entorno de x^0:

$$f(x) \approx f(x^0) + \nabla f(x^0)(x - x^0) = f(x^0) + \sum_{j=1}^{n} \frac{\partial f(x^0)}{\partial x_j}(x_j - x_j^0).$$

Como $f(x^0)$ y $\nabla f(x^0)x^0$ tienen valores fijos, se pueden eliminar de la expresión de la aproximación, con lo que finalmente consideramos como aproximación lineal de la función objetivo:

$$g(x) = \nabla f(x^0)x = \sum_{j=1}^{n} c_j x_j, \text{ donde } c_j = \frac{\partial f(x^0)}{\partial x_j}.$$

[6] M. Frank y P. Wolfe, "An algorithm for quadratic programming", en *Naval Research Logistics Quarterly* (1956), 3, 95-110.

A continuación se aplica el símplex al problema de programación lineal de maximizar $g(x)$ sujeto a las restricciones originales $Ax \leq b$ y $x \geq 0$, a fin de obtener su solución óptima, x_{LP}. El objetivo lineal necesariamente se incrementa cuando se hace un recorrido sobre el segmento de recta desde x_0 hasta x_{LP} (que se encuentra en la frontera de la región factible). La aproximación lineal puede no ser adecuada para un valor x lejano de x_0, y la función objetivo no lineal puede no continuar creciendo durante toda la trayectoria desde x_0 hasta x_{LP}. Por tanto, en lugar de aceptar x_{LP} como la siguiente solución de prueba, se elige el punto sobre este segmento de recta que maximiza la función objetivo no lineal. La sucesión de soluciones de prueba generadas por las iteraciones converge hacia una solución óptima para el problema original, y el algoritmo se detiene en cuanto las soluciones de prueba están lo suficientemente cerca entre sí como para poder concluir que se llegó a esta solución óptima.

Resumen del algoritmo de Frank-Wolfe

Paso inicial: se encuentra una solución de prueba factible inicial x^0, por ejemplo, aplicando los procedimientos de la programación lineal para encontrar una solución factible básica, como el de las dos fase. Se hace $k = 1$.

Iteración k:

- Para $j = 1, \cdots, n$, se evalúa

$$\frac{\partial f(x)}{\partial x_j} \text{ en } x = x^{k-1}$$

y se iguala c_j a este valor.

- Se encuentra una solución óptima x_{LP}^k para el problema de programación lineal de maximizar $g(x) = \sum_{j=1}^{n} c_j x_j$ sujeto a $Ax \leq b$ y $x \geq 0$.

- Para la variable t, $0 \leq t \leq 1$, se establece

$$h(t) = f(x) \text{ para } x = x^{k-1} + t(x_{LP}^{k} - x^{k-1}),$$

de manera que $h(t)$ proporciona el valor de $f(x)$ sobre el segmento de recta entre x^{k-1} (donde t=0) y x_{LP}^{k} (donde t=1). Se maximiza $h(t)$ en $0 \leq t \leq 1$ y se establece x^{k} igual al valor de x correspondiente.

Regla de detención: si x^{k-1} y x^{k} están lo suficientemente cerca, el proceso se detiene y se usa x^{k} como la estimación de la solución óptima. De otra manera se hace $k = k + 1$ y se realiza otra iteración.

Ejemplo de aplicación del algoritmo de Frank-Wolfe

Considerar el siguiente problema de programación convexa linealmente restringida:

Maximizar $f(x) = 5x_1 - x_1^2 + 8x_2 - 2x_2^2$

sujeto a $3x_1 + 2x_2 \leq 6$ y $x_1 \geq 0$ y $x_2 \geq 0$

Observamos que

$$\frac{\partial f}{\partial x_1} = 5 - 2x_1, \ \frac{\partial f}{\partial x_2} = 8 - 4x_2,$$

por lo que el máximo no restringido $x = (5/2, 2)$, no cumple las restricciones funcionales y necesitamos iterar para encontrar el máximo restringido.

Iteración 1: Como $x = (0, 0)$ es claro que es factible y corresponde a la solución inicial factible básica de las restricciones de programación lineal, se elige como la solución de prueba inicial x^0 para el algoritmo de Frank-Wolfe.

Al sustituir $x_1 = 0$ y $x_2 = 0$ en las expresiones de las derivadas parciales se obtiene $c_1 = 5$ y $c_2 = 8$, con lo que $g(x) = 5x_1 + 8x_2$ es la primera aproximación lineal de la función objetivo. Con la solución gráfica de este problema de programación lineal (ver la figura siguiente) se llega a $x_{LP}^1 = (0, 3)$. En la parte 3 del paso iterativo, los puntos en el segmento de recta que une a (0,0) y (0,3) se expresan por

$$(x_1, x_2) = (0, 0) + t[(0, 3) - (0, 0)] \text{ para } 0 \leq t \leq 1$$

$$= (0, 3t).$$

Esta expresión proporciona

$$h(t) = f(0, 3t) = 8 \times 3t - 2 \times (3t)^2 = 24t - 18t^2$$

de modo que el valor $t = t^*$ que maximiza $h(t)$ en el intervalo $0 \leq t \leq 1$ se puede obtener, en este caso, al establecer

$$\frac{dh(t)}{dt} = 24 - 36t = 0 \Rightarrow t^* = 2/3.$$

Este resultado conduce a la siguiente solución de prueba

$$x^1 = (0,0) + 2/3[(0,3) - (0,0)] = (0,2),$$

lo cual completa la primera iteración.

Iteración 2 Al sustituir $x_1 = 0$ y $x_2 = 2$ en las expresiones de las derivadas parciales se obtiene $c_1 = 5$ y $c_2 = 0$, con lo que $g(x) = 5x_1 + 8x_2$ es la segunda aproximación lineal de la función objetivo. Con la solución gráfica de este problema de programación lineal se llega a $x_{LP}^2 = (2,0)$.

En la parte 3 del paso iterativo, los puntos en el segmento de recta que une a (0,2) y (2,0) se expresan por

$$(x_1, x_2) = (0, 2) + t[(2, 0) - (0, 2)] \text{ para } 0 \leq t \leq 1$$

$$= (2t, 2 - 2t).$$

Esta expresión proporciona

$$h(t) = f(2t, 2 - 2t) = 8 + 10t - 12t^2,$$

de modo que el valor $t = t^*$ que maximiza $h(t)$ en el intervalo $0 \leq t \leq 1$ se puede obtener, en este caso, al establecer

$$\frac{dh(t)}{dt} = 10 - 24t = 0 \Rightarrow t^* = 5/12.$$

Este resultado conduce a la siguiente solución de prueba

$$x^2 = (0, 2) + 5/12[(2, 0) - (0, 2)] = (5/6, 7/6),$$

lo cual completa la segunda iteración.

En la figura, se puede observar cómo las soluciones de prueba adquieren valores de puntos alternados entre dos trayectorias que parecen cruzarse aproximadamente en el punto $x = (1, 3/2)$. Precisamente, este punto es la solución óptima, como se puede ver si se aplican las condiciones de Kuhn-Tucker.

```
(%i1) load(draw)$
(%i2) draw2d(key="Frontera
factible",color=black,implicit(3*x+2*y=6,x=0,3,y=0,4),
key="Segmento-segunda
iteración",color=red,implicit(x+y=2,x,0,3,y,0,4),
key="Aproximación lineal-primera
iteración",color=blue,implicit(5*x+8*y=24,x,0,3,y,0,4));
```

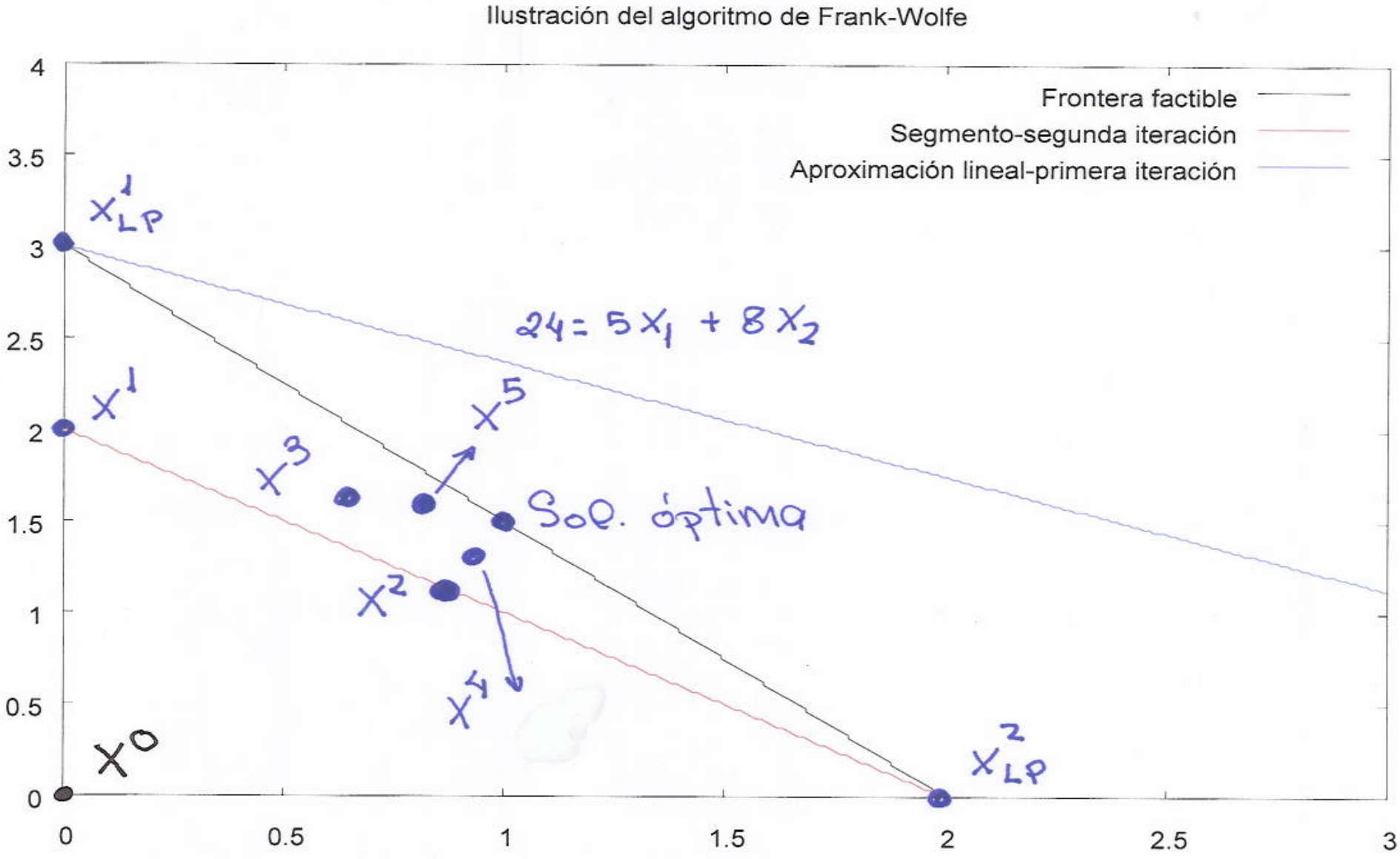


Ilustración del algoritmo de Frank-Wolfe
Frontera factible
Segmento-segunda iteración
Aproximación lineal-primera iteración
X^1_{LP}
$24 = 5X_1 + 8X_2$
X^1
X^5
X^3
Sol. óptima
X^2
X^4
X^0
X^2_{LP}
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3

Tabla de aplicación del algoritmo de Frank-Wolfe al ejemplo

k	x^{k-1}	c_1	c_2	x^k_{LP}	$h(t)$	t^*	x^k
1	(0,0)	5	8	(0,3)	$24t - 18t^2$	2/3	(0,2)
2	(0,2)	5	0	(2,0)	$8 + 10t - 12t^2$	5/12	(5/6,7/6)

El ejemplo anterior ilustra una característica común del algoritmo de Frank-Wolfe: que las soluciones de prueba alternan entre dos (o más) trayectorias. Cuando esto ocurre, se pueden extrapolar las trayectorías al punto aproximado en que se cruzan, para estimar una solución óptima. Esta estimación tiende a ser mejor que usar la última solución de prueba generada. Esto se debe a que la convergencia de las soluciones de prueba hacia la solución óptima tiende a ser lenta y por ello es posible que la última solución esté todavía alejada del óptimo.

El algoritmo de Frank-Wolfe es bastante sencillo y directo, pero es sólo un ejemplo de los algoritmos de aproximación secuencial. Muchos de ellos utilizan aproximaciones *cuadráticas* en lugar de *lineales* en cada iteración, porque la aproximación cuadrática proporciona un ajuste mejor al problema original y permite que la sucesión de soluciones converja más rápido hacia el óptimo.

Introducción

Convexidad

Optimización en problemas sin restricciones

Optimización con restricciones de igualdad y desigualdad
- Condiciones de optimalidad
- Interpretación económica. Sensibilidad
- Programación cuadrática: método símplex modificado

Ejercicios

Bibliografía

Los problemas de programación cuadrática convexa se definen como sigue:

- min $f(X) = X^t CX + P^t X$

sujeto a:

- $AX \leq b$
- $X \geq 0$,

donde $X = (x_1, \cdots, x_n)^t$ es el vector de variables, C es una matriz $n \times n$ simétrica definida positiva o semi definida positiva, $A = (a_{ij})$ es una matriz $m \times n$, $P = (p_1, \cdots, p_n)^t$ y $b = (b_1, \cdots, b_m)^t$.

El método de Wolfe

Las restricciones $AX \leq b$ se pueden escribir como $AX - b \leq 0$. Ya que f y todas las restricciones, que son lineales, son funciones convexas podemos aplicar las condiciones de Kuhn-Tucker. Wolfe propuso usar el método del símplex teniendo control sobre las variables que entran en la base, al cambiar de iteración, de forma que se cumplan las condiciones K-T.

Un ejemplo de aplicación del método de Wolfe

- max $f(X) = -2x_1^2 - 3x_2^2 - 4x_1x_2 + 6x_1 + 5x_2$

sujeto a:

- $x_1 + x_2 \leq 1$
- $2x_1 + 3x_2 \leq 4$
- $x_1 \leq 0'8$
- $x_1, x_2 \geq 0,$

Este problema equivale a:

- min $f(X) = 2x_1^2 + 3x_2^2 + 4x_1x_2 - 6x_1 - 5x_2$

sujeto a:

- $x_1 + x_2 - 1 \leq 0$
- $2x_1 + 3x_2 - 4 \leq 0$
- $x_1 - 0'8 \leq 0$
- $x_1, x_2 \geq 0.$

La función objetivo es de la forma $X^t CX + P^t X$, donde $X^t CX$ es definida positiva. La función objetivo (al ser suma de funciones convexas) y las restricciones (que son lineales) son convexas y éste es un problema de programación cuadrática convexo. Usamos el método de Wolfe. La función de Lagrange es:

$$L = 2x_1^2 + 3x_2^2 + 4x_1x_2 - 6x_1 - 5x_2 + \lambda_1(x_1 + x_2 - 1)$$
$$+\lambda_2(2x_1 + 3x_2 - 4) + \lambda_3(x_1 - 0'8).$$

Las condiciones de Kuhn-Tucker son:

$$4x_1 + 4x_2 - 6 + \lambda_1 + 2\lambda_2 + \lambda_3 = 0$$

$$4x_1 + 6x_2 - 5 + \lambda_1 + 3\lambda_2 = 0$$

$$x_1 + x_2 - 1 \leq 0, \;\; \lambda_1(x_1 + x_2 - 1) = 0$$

$$2x_1 + 3x_2 - 4 \leq 0, \;\; \lambda_2(2x_1 + 3x_2 - 4) = 0$$

$$x_1 - 0'8 \leq 0, \;\; \lambda_3(x_1 - 0'8) = 0$$

$$x_1, x_2, \lambda_1, \lambda_2, \lambda_3 \geq 0.$$

Las soluciones del sistema anterior van a ser la solución óptima del problema. Las determinamos usando la fase 1 del método de las dos fases. Consideramos el siguiente problema de programación lineal auxiliar:

$\min R_1$

s. a $4x_1 + 4x_2 + \lambda_1 + 2\lambda_2 + \lambda_3 = 6$

$4x_1 + 6x_2 + \lambda_1 + 3\lambda_2 + R_1 = 5$

$x_1 + x_2 + s_1 = 1, \ \lambda_1 s_1 = 0$

$2x_1 + 3x_2 + s_2 = 4, \ \lambda_2 s_2 = 0$

$x_1 + s_3 = 0'8, \ \lambda_3 s_3 = 0$

$x_1, x_2, \lambda_1, \lambda_2, \lambda_3, R_1, R_2, s_1, s_2, s_3 \geq 0.$

Ahora, se aplica el algoritmo del símplex con la salvedad de que en cada iteración se verifiquen las condiciones $\lambda_1 s_1 = 0$, $\lambda_2 s_2 = 0$ y $\lambda_3 s_3 = 0$.

		c_j	0	0	0	0	0	1	0	0	0
Base	x_B	c_B	x_1	x_2	λ_1	λ_2	λ_3	R_1	s_1	s_2	s_3
λ_3	6	0	4	4	1	2	1	0	0	0	0
R_1	5	1	4	6	1	3	0	1	0	0	0
s_1	1	0	1	1	0	0	0	0	1	0	0
s_2	4	0	2	3	0	0	0	0	1	1	0
s_3	0'8	0	**1**	0	0	0	0	0	0	0	1
z, z_j	5		4	6	1	3	0	1	0	0	0
$z_j - c_j$			4	6	1	3	0	1	0	0	0

La solución de esta iteración no satisface $\lambda_3 s_3 = 0$. Si aplicamos los criterios del símplex, debería entrar en la base x_2 y salir R_1, pero entonces sigue sin satisfacerse la condición $\lambda_3 s_3 = 0$. Por eso, en la siguiente iteración, entra en la base x_1 y sacamos s_3, de forma que la solución asociada es factible para el problema de programación lineal que estamos considerando y verifica las condiciones de holgura complementaria para el problema inicial.

Base	x_B	c_j c_B	0 x_1	0 x_2	0 λ_1	0 λ_2	0 λ_3	1 R_1	0 s_1	0 s_2	0 s_3
λ_3	2'8	0	0	4	1	2	1	0	0	0	-4
R_1	1'8	1	0	6	1	3	0	1	0	0	-4
s_1	0'2	0	0	**1**	0	0	0	0	0	0	-1
s_2	2'4	0	0	3	0	0	0	0	0	1	-2
x_1	0'8	0	1	0	0	0	0	0	0	0	1
z, z_j	1'8		0	6	1	3	0	1	0	0	-4
$z_j - c_j$			0	6	1	3	0	0	0	0	-4

Aplicando los criterios de entrada y salida del símplex, entra x_2 y sale s_1 de la base.

Las condiciones de holgura complementaria se siguen verificando.

Base	x_B	c_j c_B	0 x_1	0 x_2	0 λ_1	0 λ_2	0 λ_3	1 R_1	0 s_1	0 s_2	0 s_3
λ_3	2	0	0	0	1	2	1	0	-4	0	4
R_1	0'6	1	0	0	**1**	3	0	1	-6	0	2
x_2	0'2	0	0	1	0	0	0	0	1	0	-1
s_2	1'8	0	0	0	0	0	0	0	-3	1	1
x_1	0'8	0	1	0	0	0	0	0	0	0	1
z, z_j	0'6		0	0	1	3	0	1	-6	0	2
$z_j - c_j$			0	0	1	3	0	0	-6	0	2

No cambiamos R_1 por λ_2 ni R_1 por s_3 para seguir cumpliendo las condiciones de holgura complementaria se siguen verificando.

Cambiamos R_1 por λ_1.

Base	x_B	c_j c_B	0 x_1	0 x_2	0 λ_1	0 λ_2	0 λ_3	1 R_1	0 s_1	0 s_2	0 s_3
λ_3	1’4	0	0	0	0	-	1	-	-	0	-
λ_1	0’6	0	0	0	1	3	0	1	-6	0	2
x_2	0’2	0	0	1	0	-	0	-	-	0	-
s_2	1’8	0	0	0	0	-	0	-	-	1	-
x_1	0’8	0	1	0	0	-	0	-	-	0	-
z, z_j	0		0	0	0	0	0	0	0	0	0
$z_j - c_j$			0	0	0	0	0	-1	0	0	0

La solución buscada es $x_1 = 0'8$, $x_2 = 0'2$, $\lambda_1 = 0'6$, $\lambda_2 = 0$, $\lambda_3 = 1'4$, $s_1 = 0$, $s_2 = 1'8$, $s_3 = 0$ y $R_1 = 0$.

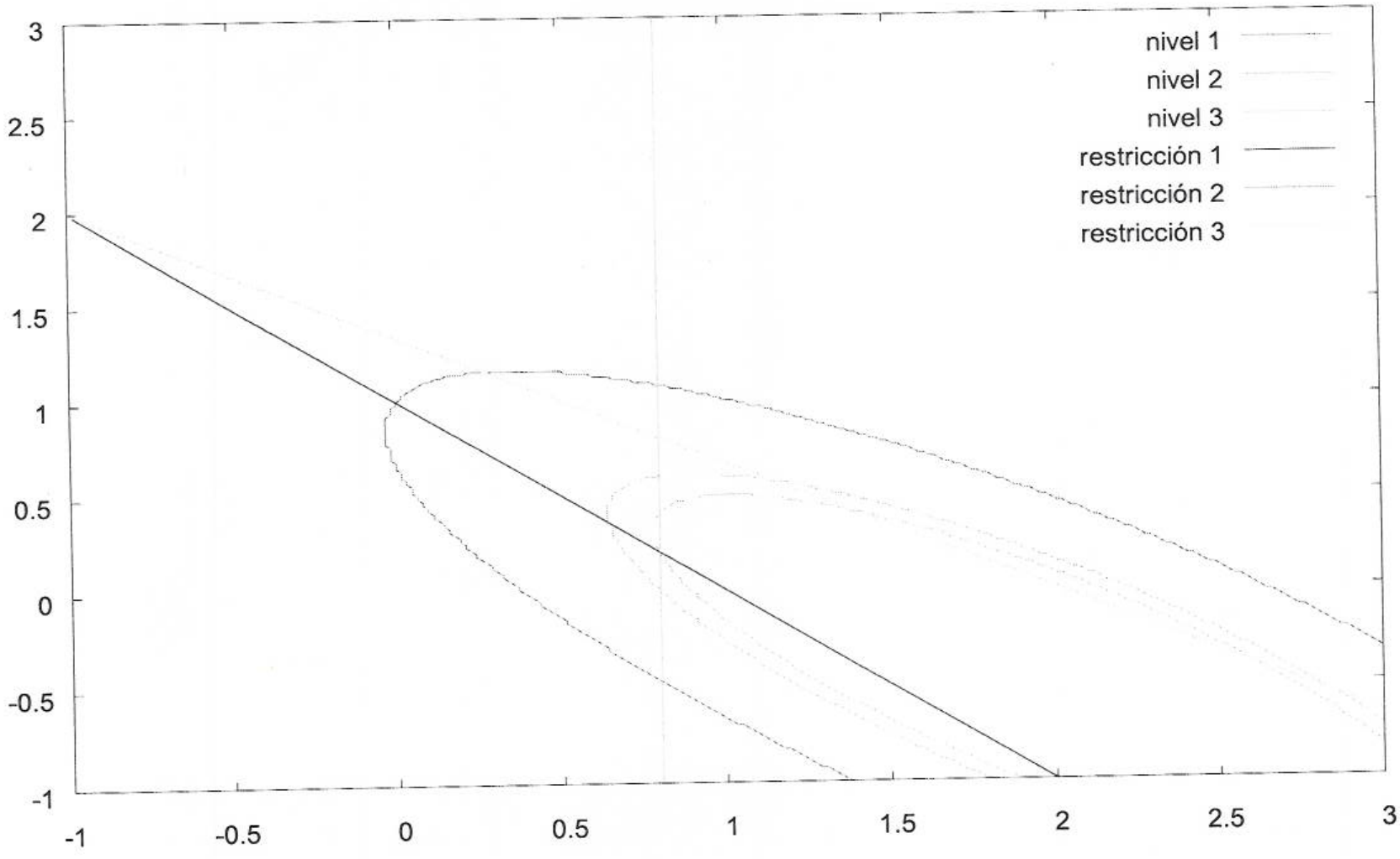
Notemos por la fila $z_j - c_j$ de la última tabla que el problema de programación auxiliar tiene múltiples óptimos.

Con Maxima comprobamos los resultados anteriores:

```
(%i1) f:-2*x^2-3*y^2-4*x*y+6*x+5*y;
(%i2) f,x=0.8,y=0.2;
(%o2) 3.76
(%i3) f,x=0.8,y=0;
(%o3) 3.520000000000001
(%i4) f,x=0,y=1;
(%o4) 2
(%i5) load(draw)$
```

```
(%i6) draw2d(
color= red, implicit(-2*x^2-3*y^2-4*x*y+6*x+5*y=2,x,-1,3,y,-1,3),
color = orange,
implicit(-2*x^2-3*y^2-4*x*y+6*x+5*y=3.52,x,-1,3,y,-1,3),
color = violet,
implicit(-2*x^2-3*y^2-4*x*y+6*x+5*y=3.76,x,-1,3,y,-1,3),
color = blue, implicit(x+y=1,x,-1,3,y,-1,3),
color = green, implicit(x=0.8,x,-1,3,y,-1,3),
color = yellow, implicit(2*x+3*y=4,x,-1,3,y,-1,3),
title=“análisis gráfica de un problema de programación
cuadrática");
```

Análisis gráfica de un problema de programación cuadrática



Programación no convexa

Los supuestos de programación convexa (*el objetivo f a maximizar es cóncavo y todas las restricciones g_i son convexas*) son muy convenientes pues aseguran que *cualquier máximo local lo es también global*. (Si el objetivo es minimizar una función convexa, un mínimo local lo es también global). Pero los problemas de programación no lineal que aparecen en la práctica muchas veces no cumplen estos supuestos y la cuestión es cuál es en este caso el mejor enfoque. Algunos de estos problemas son bastante más difíciles de resolver que otros. Problemas donde la región factible no es convexa por no serlo algunas de las funciones g_i, o si se trata con funciones no diferenciables, presentan una complicación importante. La meta de una parte de la investigación que se realiza en la actualidad es desarrollar procedimientos eficientes de optimización global para encontrar una solución óptima global para diferentes tipos de problemas de programación no convexa.

Ejercicios

- 1. Estudiar la convexidad, concavidad, convexidad o concavidad estricta de las siguientes funciones:
  - $f(x_1, x_2) = e^{x_1+x_2}$, $(x_1, x_2) \in \mathbb{R}^2$.
  - $f(x_1, x_2) = ln(x_1 + x_2)$, $(x_1, x_2) \in \mathbb{R}^2_{++}$.
  - $f(x_1, x_2) = ln(x_1 x_2)$, $(x_1, x_2) \in \mathbb{R}^2_{++}$.
  - $f(x_1, x_2, x_3) = 2x_1 + e^{x_2} + 3x_3$, $(x_1, x_2, x_3) \in \mathbb{R}^3$.
- 2. Estudia la cuasiconvexidad y cuasiconcavidad de las funciones:
  - $f(x_1) = ln(x_1)$, $x_1 \in \mathbb{R}_{++}$.
  - $f(x_1) = \dfrac{1}{x_1}$, $x_1 \in \mathbb{R}_{++}$.
  - $f(x_1, x_2) = e^{x_1+x_2}$, $(x_1, x_2) \in \mathbb{R}^2$.

- 3. Aplicando las condiciones para funciones diferenciables, probar que:
  - La función $f(x) = x^3$ con $x \in \mathbb{R}$ es cuasiconvexa.
  - La función $f(x) = e^x$ con $x \in \mathbb{R}$ es estrictamente cuasiconvexa.
- 4. Se desea minimizar la función $f(x_1, x_2) = x_1^2 + x_2^2 + \alpha x_1 x_2 + x_1 + 2x_2$.
  - Probar, resolviendo el sistema de ecuaciones $\nabla f = 0$ y estudiando la definición de la hessiana, que para $\alpha = 1$ existe mínimo y es único. Con wxMaxima, realiza el Gráfico 3D de la función (tomar *x* e *y* entre -5 y 5 y una cuadrícula de 10).
  - Probar, análogamente, que para $\alpha = 2$ no existe mínimo. Con wxMaxima, realiza el Gráfico 3D de la función (tomar *x* e *y* entre -15 y 15 y una cuadrícula de 2).
  - Probar que para $\alpha = 6$ no existe mínimo. Con wxMaxima, realiza el Gráfico 3D de la función (tomar *x* e *y* entre -15 y 15 y una cuadrícula de 2).

- 5. Dada la función $f(x) = x_1 + 3x_2^2 - 4x_1x_2 + 4x_1^2$, realícense tres iteraciones para minimizar la función usando el método del gradiente. Tómese como punto inicial $x_0 = (7, 9)$.
- 6. Dada la función $f(x) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$, realícense dos iteraciones para minimizar la función usando el método del gradiente. Tómese como punto inicial $x_0 = (0, 0)$.

- 7. Considerar el siguiente problema:
  - min $(x-1)^2+(y-1)^2$

  sujeto a:

  - $-x+y=\frac{1}{2}$
  - $x+y\leq 1$
  - $x,y\geq 0$

  1. Resolverlo gráficamente.
  2. Comprobar la respuesta utilizando la teoría de Kuhn-Tucker.
  3. Encontrar la función dual $\mathbf{l}(\lambda,\mu)$ y probar que es cóncava.
- 8. Considerar la función objetivo $f(x,y)=x^2+4y^3$. Encontrar todos los puntos que son máximos locales y mínimos locales para f sujeto a la restricción de igualdad $x^2+2y^2=1$.

Solución de algunos ejercicios

Ejercicio 1, primer apartado. Estudiar la convexidad, concavidad, convexidad o concavidad estricta de la siguiente función:

$$f(x_1, x_2) = e^{x_1+x_2},\ (x_1, x_2) \in \mathbb{R}^2.$$

Calculamos la matriz hessiana de la función:

$$Hf(x) = \begin{pmatrix} e^{x_1+x_2} & e^{x_1+x_2} \\ e^{x_1+x_2} & e^{x_1+x_2} \end{pmatrix} = e^{x_1+x_2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Como la matriz hessiana es semidefinida positiva, la función es convexa. Como la hessiana no es definida positiva, no podemos afirmar que f sea estrictamente convexa.

plot3d($exp(x + y), [x, -5, 5], [y, -5, 5], [plot_format, gnuplot]$)\$

Consideremos ahora dos conjuntos, dado $\alpha \in \mathbb{R}$.

$$\{(x_1, x_2) \in \mathbb{R}^2 \ : \ e^{x_1+x_2} \leq \alpha\}$$

$$= \begin{cases} \emptyset & \alpha \leq 0 \\ \{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 + x_2 \leq ln(\alpha)\} & \alpha > 0 \end{cases}$$

$$\{(x_1, x_2) \in \mathbb{R}^2 \ : \ e^{x_1+x_2} \geq \alpha\}$$

$$= \begin{cases} \mathbb{R}^2 & \alpha \leq 0 \\ \{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 + x_2 \geq ln(\alpha)\} & \alpha > 0 \end{cases}$$

Con (%i1) load(draw)$

(%i2)draw2d(implicit(x+y=5,x,-100,100,y,-100,100));

Vemos que esos dos conjuntos son convexos y la función es entonces cuasiconvexa y cuasicóncava.

Estudiamos la convexidad estricta haciendo uso de un resultado:

TEOREMA $f : S \subset \mathbb{R}^n \to \mathbb{R}$ es estrictamente convexa si y sólo si para todo $x, y \in S$, $x \neq y$, $f(y) > f(x) + \nabla f(x)(y - x)$, o bien $(\nabla f(y) - \nabla f(x))(y - x) > 0$.

Sean entonces $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$, $x \neq y$,

$$f(y_1, y_2) > f(x_1, x_2) + \nabla f(x_1, x_2)[(y_1, y_2) - (x_1, x_2)]$$

$$\Leftrightarrow e^{y_1+y_2} > e^{x_1+x_2} + (y_1 - x_1)e^{x_1+x_2} + (y_2 - x_2)e^{x_1+x_2}$$

$$\Leftrightarrow \frac{e^{y_1+y_2}}{e^{x_1+x_2}} = e^{(y_1-x_1)+(y_2-x_2)} > 1 + (y_1 - x_1) + (y_2 - x_2).$$

Tomando $x = (2, 1), y = (1, 2)$, se llegaría a $1 > 1$: absurdo. Entonces f no es estrictamente convexa.

Ejercicio 3, primer apartado Veamos que $f(x) = x^3$ donde $x \in \mathbb{R}$ es cuasiconvexa.

Por la caracterización de cuasiconvexidad para funciones diferenciables, habrá que probar que para todo $x, y \in \mathbb{R}$ con $f(x) \geq f(y)$ se tiene que $\nabla f(x)(y - x) \leq 0$.

Sean entonces $x, y \in \mathbb{R}$ con $f(x) = x^3 \geq f(y) = y^3$. Por la monotonía de la función f, se tendrá que $x \geq y$ e $y - x \leq 0$.

Por tanto, $\nabla f(x)(y - x) = 3x^2(y - x) \leq 0$, c.q.d.

Ejercicio 4, tercer apartado Sea la función $f(x_1, x_2) = x_1^2 + x_2^2 + 6x_1x_2 + x_1 + 2x_2$. Con Maxima, comprobamos, anulando el gradiente de la función (condición necesaria de primer orden) cuáles son los candidatos a mínimo local. La hessiana evaluada en los puntos obtenidos, NO es semidefinida positiva (condición necesaria de segundo orden).

```
(%i1) f:x^2+y^2+6*x*y+x+2*y;
(%i2) e1:diff(f,x,1);
(%i3) e2:diff(f,y,1);
(%i4) algsys([e1,e2],[x,y]);
(%o4) [[x=-5/16,y=-1/16]]
(%i5) hessian(f,[x,y]);
(%i6) determinant(%);
(%o6) -32
```

Referencias

- Barrodale, I.; Roberts, F. D. K. (1973): “An improved algorithm for discrete l_1 linear approximation". SIAM Journal on Numerical Analysis, 10: 839-848.

  Abstract. *By modifying the simplex method of linear programming, we are able to present an algoritm for l_1-approximation which appears to be superior computationally to any other known algorithm for this problem.*
- Bazaraa, M.S.; Sherali, H.; Shetty, C. (1993): “Nonlinear programming. Theory and algorithms". Wiley.
- Gill, P.E.; Murray, W.; Wright, M. H. (1981): “Practical optimization". Academic Press, New York.

- Hadley, G. (1964): “Nonlinear and dynamic programming". Addison-Wesley, Reading, Mass.
- Hillier, F.S.; Lieberman, G.J. (2002): “Investigación de operaciones". McGraw-Hill.
- Kasana, H.S.; Krishna, D.K. (2004): “Introductory Operations Research. Theory and Applications". Springer.
- Koenker, R. (2005): “Quantyle Regression". Cambridge University Press.
- Martín, Q.; Santos, M. T.; Santana, Y. (2005): “Investigación operativa: problemas y ejercicios resueltos". Pearson.
- Parlar, M. (2000): "Interactive operations research with Maple. Methods and models". Birkhäuser.

- Parlar, M.; Vickson, R. G. (1980): "An optimal control problem with piecewise-quadratic cost functional containing a “dead-zone"". Optimal Control Applications and Methods, 1: 361-372.
- Peressini, A. L.; Sullivan, F. E.; Uhl, J. J. (1988): “The mathematics of nonlinear programming". Springer-Verlag, New York.
- Rodríguez Riotorto, M. (2008): “Primeros pasos en Maxima". www.telefonica.net/web2/biomates
- Schmidt, J. W.; Davis, R. P. (1981): "Foundations of analysis in operations research". Academic Press, New York.
- Taha, H. (2004): “Investigación de operaciones". Ed. Pearson.