Apuntes de Optimización

Denis Sauré
dsaure@dii.uchile.cl

Agosto, 2003

Capítulo 1

Complejidad

1.1. Complejidad de un Algoritmo

Para especificar el tamaño de un problema debemos especificar valores para sus parámetros. La asignación valores determinados para el conjunto de parámetros del problema se denomina una **instancia** del problema.

El tamaño de una instancia dada es una función $f(A)$ de los parámetros seleccionados, donde A denota el conjunto de parámetros. En particular diremos que una instancia tiene tamaño l si $f(A) = l$.

La complejidad de un algoritmo se formula en término del "tiempo de ejecución del mismo". Para realizar el análisis de complejidad se utiliza un modelo de computador que ejecuta los algoritmos a estudiar.

Considerando que comunmente es más difícil resolver instancias "grandes" de un problema, la complejidad de un algoritmo debe ser formulada en términos del tamaño de la instancia que resuelve. De la misma forma, existen instancias de igual tamaño tales que la resolución de una es más "fácil" que la otra.

Por estas razones la complejidad de un algoritmo se determina en función del tamaño de la instancias a resolver, considerando máximo tiempo de ejecución sobre todas las instancias de un mismo tamaño (peor caso).

Definición 1 *Diremos que un algoritmo tiene una complejidad de $O(f(l))$ si la relación entre el tamaño de la instancia y el tiempo de ejecución esta acotada por $Kf(l)$ donde K*

es una constante.

Definición 2 ***(clasificación de algoritmos)***

- *Diremos que un algoritmo es* ***polinomial*** *si su complejidad es $O(l^k)$ para algún k finito.*
- *Diremos que un algoritmo es* ***exponencial*** *si su complejidad no puede ser acotada por un polinomio en l.*

1.2. Complejidad de un Problema

La complejidad de un problema está dada por el mínimo de las complejidades de todos los algoritmos que lo resuelven.

La complejidad de un algoritmo existente solo entrega una cota superior a la complejidad de un problema.

1.3. Clasificación de Problemas

Para desarrollar una clasificación de problemas necesitamos 4 elementos:

- Una clase $\mathcal{C}$ de problemas legítimos.
- Una clase $\mathcal{C}_{\mathcal{A}} \subset \mathcal{C}$ de problemas ***fáciles***.
- Una clase $\mathcal{C}_{\mathcal{B}} \subset \mathcal{C}$ de problemas ***difícile***s.
- una relación ***no más difícil que*** entre pares de problemas legítimos.

Lemma 1 *Supongamos que P y Q son dos problemas legítimos.*

- *Si Q es* **fácil** *y P es* **no más difícl que** *Q, entonces P es* **fácil**.
- *Si P es* **difícil** *y P es* **no más difícl que** *Q, entonces Q es* **difícil**.

Para poder hacer una distinción entre problemas fáciles y difíciles se acostumbra enunciar problemas como **problemas de decisión**, es decir, uno para el cual la solución es la respuesta *si* o *no*.

- Problema de optimización:
$$\max\{cx :, x \in X\}$$
- Problema de decisión:
$$\text{Existe algún } x \in X \text{ tal que } cx \geq k?.$$

Definición 3 *$\mathcal{NP}$ es la clase de problemas de decisión con la propiedad que, para cualquier instancia para la cual la respuesta al problema de decisión es* si*, existe una prueba "rápida" (polinomial) de la respuesta.*

Esto implica que si un problema de decisión esta en $\mathcal{NP}$ el problema de optimización puede resolverse respondiendo un número polinomial de veces el problema de decisión.

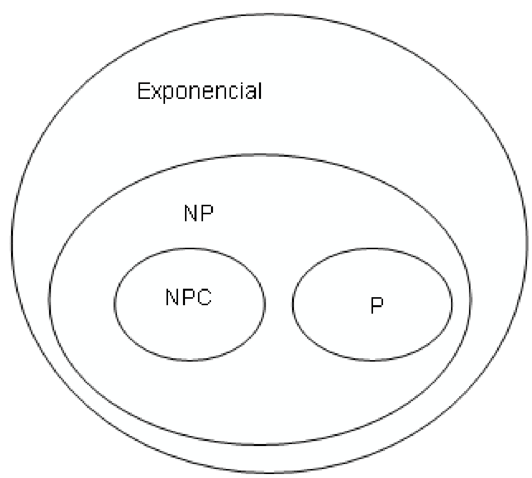
Definición 4 *$\mathcal{P}$ es la clase de problemas de decisión en $\mathcal{NP}$ para los cuales existe un algoritmo polinomial.*

Definición 5 *Si P y $Q \in \mathcal{NP}$, y si una instancia de P puede convertirse en tiempo polinomial en una instancia de Q, entonces P es* **reducible polinomialmente** *a Q.*

Definición 6 *$\mathcal{NPC}$, es el subconjunto de problemas $P \in \mathcal{NP}$ tales que $\forall Q \in \mathcal{NP}$, Q es* **reducible polinomialmente** *a P.*

Lemma 2 *Supongamos que los problemas $P, Q \in \mathcal{NP}$.*

- *Si $Q \in \mathcal{P}$ y P es* reducible polinomialmente *a $Q \rightarrow P \in \mathcal{P}$.*
- *Si $P \in \mathcal{NPC}$ y P es* reducible polinomialmente *a $Q \rightarrow Q \in \mathcal{NPC}$.*
- *Si $\mathcal{P} \cap \mathcal{NPC} \neq \emptyset \quad \Rightarrow \quad \mathcal{P} = \mathcal{NPC}$.*


Exponencial
NP
NPC
P

Capítulo 2

Optimización Irrestricta

2.1. Definiciones Básicas

Distinguiremos en un problema de optimización tres elementos:

- El **conjunto factible** o **conjunto de oportunidades**, que denotaremos con la letra X.
- La función objetivo, una función definida sobre X que toma valores numéricos, $f : X \to \mathbb{R}$. La función f es un índice que asociamos a cada elemento de X.
- Un criterio de decisión: buscaremos aquel elemento o elementos de X que tienen asociado un índice mayor (problemas de **maximización**) o bien un indice menor (problemas de **minimización**).

Representaremos un problema de óptimización de la siguiente forma:

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

No todos los problemas de optimización necesariamente tienen solución. Por ejemplo, el problema

$$\max_{x \in \mathbb{N}} f(x) = x$$

donde $\mathbb{N}$ son los números naturales, no tiene solución.

Definición 7 *Decimos que el problema*

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

tiene solución *si hay un cierto elemento* $x^* \in X$ *tal que se cumple que* $f(x^*) \leq f(x)$*, para todo* $x \in X$*. En este caso decimos que* x^* *es un* ***maximizador*** *del problema, y que* $f(x^*)$ *es el* ***máximo*** *del problema.*
Decimos que el problema

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

tiene solución *si hay un cierto elemento* $x^* \in X$ *tal que se cumple que* $f(x^*) \leq f(x)$*, para todo* $x \in X$*. En este caso decimos que* x^* *es un* ***minimizador*** *del problema, y que* $f(x^*)$ *es el* ***mínimo*** *del problema.*

Dada una función $f : X \to \mathbb{R}$ su recorrido esta definido como:

$$f(X) = \{c \in \mathbb{R} \,|\, \exists\, x \in X : f(x) = c\}$$

Lemma 3 *El problema* $\min_{x \in X} f(x)$ *tiene solución* $\Leftrightarrow$ *el conjunto* $f(X)$ *tiene un elemento mínimo.*
El problema $\max_{x \in X} f(x)$ *tiene solución* $\Leftrightarrow$ *el conjunto* $f(X)$ *tiene un elemento máximo.*

Sea $A \subset \mathbb{R}$, a^* es un elemento máximo de A si $a^* \in A$ y $a^* \geq a\, \forall a \in A$. Sea $A \subset \mathbb{R}$, a^* es un elemento mínimo de A si $a^* \in A$ y $a^* \leq a\, \forall a \in A$.

Una condición necesaria para la existencia de un elemento máximo (mínimo) es que el conjunto este **acotado** superiormente (inferiormente). Una condición suficiente para la existencia de un elemento máximo (mínimo) es que el conjunto sea **acotado** y **cerrado**[1] (**compacto**).

Teorema 1 *(**(Weierstrass)**) Si el conjunto factible* X *es compacto y la función objetivo es continua, entonces tanto el problema de minimizar* f *sobre* X *como el de maximizarla tienen solución.*

2.1.1. Convexidad y Concavidad de conjuntos

Definición 8 *Sea* $X \subset \mathbb{R}^n$*. Decimos que* X *es un conjunto convexo si el segmento que une 2 puntos cualesquiera del conjunto esta contenido en el mismo. Es decir, dados* x'*,* $x'' \in X$ *y cualquier* $\lambda \in (0,1)$ *el punto* $\lambda x' + (1-\lambda)x'' \in X$

[1] Un conjunto es cerrado si contiene a los puntos limites de todas las secuencias que se pueden formar con sus elementos.

Definición 9 *Sean $x_i \in X$ con $i = 1, \ldots, k$, una combinación convexa de los puntos es tal que:*

$$x = \sum_{i=1}^{k} \alpha_i x_i$$

con $\alpha_i \geq 0\, \forall i$ y $\sum_{i=1}^{k} \lambda_i = 1$.

Un conjunto será convexo si cualquier combinación convexa de puntos de él está contenida en el conjunto. Así, un semiespacio $\{x \in \mathbb{R} \,|\, a^T x \geq b\}$ es convexo. Por otro lado, la intesección de conjuntos convexos es un conjunto convexo.

Sea $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Una consecuencia de la afirmación anterior es que el **poliedro** $P = \{x \in \mathbb{R}^n \,|\, Ax \geq b\} = \{x \in \mathbb{R}^n \mid \bigcap_{i=1}^{m} a_i^T x \geq b_i\}$ es convexo.

2.1.2. Convexidad y Concavidad de funciones

Definición 10 *Sea $f : X \to \mathbb{R}$, con X un conjunto convexo no vacío.*

- *Decimos que f es una función convexa $\Leftrightarrow$ dados 2 puntos cualesquiera $x', x'' \in X$ y $\forall\, \lambda \in (0, 1)$ se cumple la siguiente desigualdad:*

  $$f(\lambda x' + (1 - \lambda)x'') \leq \lambda f(x') + (1 - \lambda) f(x'')$$

  Si la desigualdad es estricta para $x' \neq x''$ entonces f es estrictamente convexa.

- *Decimos que f es una función cóncava $\Leftrightarrow$ $-f$ es convexa.*

  Decimos que f es una función estrictamente cóncava $\Leftrightarrow$ $-f$ es estrictamente convexa.

Para funciones con derivada continua la definición de función convexa (cóncava) implica que toda tangente al grafo de la función queda enteramente por bajo (sobre) de la misma. Esto significa que $\forall \overline{x} \in X$ y $\forall x \in X$ una función convexa cumple con:

$$f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x}) \leq f(x)$$

Para una función cóncava tendremos que:

$$f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x}) \geq f(x)$$

Donde $\nabla f(x)$ es el vector gradiente de f evaluado en x, es decir:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

y el símbolo $\cdot$ denota el producto punto entre vectores.

Teorema 2 *Sea $X \subset \mathbb{R}$ un conjunto convexo, y $f : X \to \mathbb{R}$ una función en $\mathcal{C}^1$ (diremos que una función f es $\mathcal{C}^n$ si sus derivadas parciales de orden n existen y son continuas).*

- *f es una función (estrictamente) convexa $\Leftrightarrow$ dados 2 puntos distintos $x', x'' \in X$ se cumple que:*
$$f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x}) \quad (<) \leq f(x)$$
- *f es una función (estrictamente) cóncava $\Leftrightarrow$ dados 2 puntos distintos $x', x'' \in X$ se cumple que:*
$$f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x}) \quad (>) \geq f(x)$$

Definición 11 *Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica.*

- *A es definida (semidefinida) positiva en X si $x^T A x > (\geq)\, 0 \quad \forall x \in X \quad x \neq 0$.*
- *A es definida (semidefinida) negativa en X si $x^T A x < (\leq)\, 0 \quad \forall x \in X \quad x \neq 0$.*

Teorema 3 *Sea $X \subset \mathbb{R}$ un conjunto convexo, abierto y no vacío, y $f : X \to \mathbb{R}$ una función en $\mathcal{C}^2$.*

- *f es una función convexa $\Leftrightarrow \forall\, x \in X$ la matriz hessiana $\nabla^2 f(x)$ es semidefinida positiva.*
- *f es una función cóncava $\Leftrightarrow \forall\, x \in X$ la matriz hessiana $\nabla^2 f(x)$ es semidefinida negativa.*

Recordamos que la matriz hessiana de la función $f(x)$ es la siguiente:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

Además, si $f(x) \in \mathcal{C}^2$ entonces $\nabla^2 f(x)$ es simétrica.

Definición 12 *Sea $X \subset \mathbb{R}$ un conjunto convexo, abierto y no vacío, y $f : X \to \mathbb{R}$ una función en $\mathcal{C}^2$*

- *si la matriz hessiana $\nabla^2 f(x)$ es definida positiva en $X \Rightarrow f$ es estrictamente convexa.*
- *si la matriz hessiana $\nabla^2 f(x)$ es definida negativa en $X \Rightarrow f$ es estrictamente cóncava.*

Estas son algunas propiedades que cumplen las funciones convexas.

1. Sean $f_1, f_2, \ldots, f_k : X \to \mathbb{R}$ funciones convexas sobre un conjunto convexo, abierto, no vacío $X \subset \mathbb{R}$.
   - $f(x) = \sum_{i=1}^{k} \alpha_i f_i(x)$ con $\alpha_i \geq 0$ es convexa sobre X.
   - $f(x) = \text{máx}\{f_1(x), f_2(x), \ldots, f_k(x)\}$ es convexa sobre X.
2. $f(x) = a^t x + b \Leftrightarrow f$ es convexa y cóncava a la vez.
3. $f(x) : X \to \mathbb{R}$ convexa sobre X y $g(x) : \mathbb{R} \to \mathbb{R}$ convexa no decreciente. $\Rightarrow \quad h(x) = g(f(x))$ convexa sobre X.
4. Si $f(x) : X \to \mathbb{R}$ convexa sobre X, entonces f es continua en el interior de X.
5. El conjunto $\{x | f(x) \leq \alpha\}$ es convexo $\forall\, \alpha \in \mathbb{R}$

Definición 13 *Sea $X \subset \mathbb{R}$ un conjunto convexo, abierto y no vacío.*

- *Decimos que $f : X \to \mathbb{R}$ es cuasiconvexa si $\forall x \in X$ el contorno inferior $\{y \in X : f(y) \geq f(x)\}$ es un conjunto convexo.*
- *Decimos que $f : X \to \mathbb{R}$ es cuasicóncava si $\forall x \in X$ el contorno superior $\{y \in X : f(y) \leq f(x)\}$ es un conjunto convexo.*

Toda función convexa es cuasiconvexa y toda función cóncava es cuasicóncava.

2.2. Condiciones de Optimalidad

Dado un punto $x \in \mathbb{R}^n$ definimos la bola de radio $\gamma > 0$ alrededor de x como el conjunto:

$$\{x' \in \mathbb{R}^n : \|x' - x\| < \gamma\}$$

Definición 14 *Dado $x \in \mathbb{R}^n$, decimos que un conjunto E es un* ***entorno*** *de x si existe algún $\epsilon > 0$ tal que E contiene a la bola de radio γ alrededor de x. Si $X \subset \mathbb{R}$ y $x \in X$, entonces un* ***entorno de*** *x* ***relativo a*** *X es la intersección entre X y un entorno de x en $\mathbb{R}^n$.*

Definición 15 *Sea $X \subset \mathbb{R}^n$ y $f : X \to \mathbb{R}$.*

- *Decimos que f alcanza un* ***mínimo local*** *en un punto $x^* \in X$ si hay un cierto entorno de x^* relativo a X tal que $f(x^*) \leq f(x)$, para todo punto x de aquel entorno. Si la desigualdad es estricta para todo punto del entorno distinto de x^*, entonces decimos que el mínimo local es* ***estricto****.*
- *Decimos que f alcanza un* ***máximo local*** *en un punto $x^* \in X$ si hay un cierto entorno de x^* relativo a X tal que $f(x^*) \geq f(x)$, para todo punto x de aquel entorno. Si la desigualdad es estricta para todo punto del entorno distinto de x^*, entonces decimos que el máximo local es* ***estricto****.*

Dado $X \subset \mathbb{R}^n$ decimos que un punto $x \in X$ es **interior** a X si existe un entorno de x que está enteramente contenido en el conjunto. Un subconjunto es **abierto** si, y sólo si, todos sus puntos son interiores. Un subconjunto es cerrado si, y sólo si, su complemento es un conjunto abierto.

Definición 16 *Si la función $f : X \to \mathbb{R}$ es diferenciable en X, un punto $\overline{x}$ se dice estacionario para f si $\nabla f(\overline{x}) = 0$.*

Definición 17 *La derivada direccional de f en x, en la dirección $d \in \mathbb{R}^n$, se define como:*

$$D_d f(x) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon}$$

Definiendo $\phi(\epsilon) = f(x + \epsilon d)$ se puede demostrar que:

$$D_d f(x) = \phi'(0)$$

más aun:

$$D_d f(x) = \nabla f(x)^T d$$

Este número representa la variación de f en la dirección d, para el punto x. Si $\nabla f(x)^T d < 0$, entonces f decrece en la dirección d. Si $\nabla f(x)^T d < 0$, entonces f crece en la dirección d. Es más: $\nabla f(x)$ es la dirección de máximo crecimiento y $-\nabla f(x)$ es la dirección de más decrecimiento.

Teorema 4 *Sea $f : X \to \mathbb{R}$ una función diferenciable. Si x^* es un punto interior del conjunto X y f alcanza un máximo o un mínimo local en x^*, entonces x^* es un punto estacionario.*

De no cumplirse la condición anterior tendríamos que $\nabla f(x^*)(x-x^*) < 0$ para algún $x \in X$ por lo que la función decrece en la dirección $d = (x - x^*)$, contradiciendo la condición de mínimo de x^*.

Notamos que la condición impuesta por el teorema es sólo necesaria. Por ejemplo considere la función $f(x) = x^3$ en el punto $x = 0$. Para caracterizar puntos mínimos o máximos locales necesitamos de condiciones más "fuertes".

Teorema 5 *Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función (cóncava) convexa, y supongamos que el punto $x^* \in X$ es un (maximizador) minimizador local de f. Entonces x^* es también un (maximizador) minimizador global.*

Observaciones:

- En un problema de minimización nos interesará que la función objetivo sea convexa, ya que en este caso sabremos que los mínimos locales son también globales.
- En un problema de maximización nos interesará que la función objetivo sea cóncava, ya que en este caso sabremos que los máximos locales son también globales.
- Notemos que el maximizador o minimizador local del teorema anterior no tiene por qúe ser un punto interior. El único requerimiento es que se trate de un conjunto convexo. Esto permite derivar resultados de globalidad en todo tipo de problemas de optimización, tengan o no restricciones.
- Si la función es estrictamente convexa (cóncava), entonces el minimizador (maximizador) a que se refiere el teorema es único.

Para mostrar que x^* es un minimizador global, debemos probar que, para todo $x \in X$, $f(x^*) \leq f(x)$. Sea x un punto arbitrario en X. Sabemos que, para cualquier $\lambda \in (0,1)$, $\lambda x + (1-\lambda)x^* \in X$. Notamos que, cuando λ es muy pequeño, el punto $(1-\lambda)x^* + \lambda x$ es muy cercano a 0 (acá ocupamos que el dominio es convexo). Consideremos un λ_γ tal que (recordamos que x^* es mínimo local):

$$f(x^*) \leq f((1-\lambda)x^* + \lambda x)$$

por concavidad

$$f((1-\lambda)x^* + \lambda x) \leq (1-\lambda)f(x^*) + \lambda f(x)$$

Combinando ambas desigualdades, obtenemos:

$$f(x^*) \leq (1-\lambda)f(x^*) + \lambda f(x) \Rightarrow \quad \lambda f(x^*) \leq \lambda f(x)$$

Notamos que λ es no negativo. El resultado para problemas de máximización sobre funciones cóncavas es similar.

Teorema 6 *Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función diferenciable.*

- *Si f es una función convexa, entonces cualquier punto crítico* ($\nabla f(x) = 0$) *es un minimizador global.*
- *Si f es una función cóncava, entonces cualquier punto crítico* ($\nabla f(x) = 0$) *es un maximizador global.*

2.3. Métodos de descenso

El objetivo de los métodos de descenso que se estudiarán a continuación es encontrar puntos estacionarios mediante aproximaciones sucesivas. La necesidad de estos procedimientos surge debido a la dificultad de caracterizar los puntos críticos de una función en forma analítica. Así estos métodos generan una sucesión de puntos $\{x^k\}$, $k = 1, 2, \ldots$ tal que la sucesión correspondiente $\{f(x^k)\}$, $k = 1, 2, \ldots$ sea decreciente.

Definición 18 *Sea $f(x) : \mathbb{R}^n \to \mathbb{R}$, diferenciable y $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ es una dirección de descenso* $\Rightarrow \nabla f(x) \cdot d < 0$. *Esto es, d es una dirección de descenso si forma un ángulo agudo con $-\nabla f(x)$.*

En cada iteración de un método de descenso, a partir de un x^k y de una dirección de descenso d^k para f en x^k construye x^{k+1} de la siguiente forma:

$$x^{k+1} = x^k + \lambda_k d^k$$

donde λ_k es el "paso" de la iteración, y se determina de forma que $f(x^k) < f(x^{k+1})$.

La siguiente es la estructura general de un método de descenso:

1. $k = 0$, elegir un punto de partida x^0.
2. Si $\nabla f(x^k) = 0$, (x^k es un punto crítico) o se alcanza algún otro criterio de detención, terminar. Si no, seguir.
3. Determinar una dirección de descenso d^k para f en x^k.
4. Determinar un paso $\lambda_k > 0$ tal que $f(x^k) < f(x^k + \lambda_k d^k)$.
5. $x^{k+1} = x^k + \lambda_k d^k$, incrementar k e ir al paso 2.

Se puede necesitar una sucesión infinita de puntos para encontrar un punto estacionario. En la práctica se utilizan distintos criterios de detención, como los siguientes:

- $\|\nabla f(x^k)\| \leq \epsilon$, con $\epsilon > 0$.
- $|\frac{\partial f(x^k)}{\partial x_i}| \leq \epsilon \quad \forall i \in \{1, \ldots, n\}$, con $\epsilon > 0$.
- $\|x^{k+1} - x^k\| \leq \epsilon$, con $\epsilon > 0$ durante las últimas T iteraciones.

Métodos particulares establecen reglas para la elección de d^k y λ_k. Estas reglas se establecen para que la sucesión de puntos converja a un punto estacionario, o en su defecto, para que todo punto de acumulación de la sucesión sea estacionario.

La elección de x^0 es muy importante, puesto que de ella puede depender que un procedimiento converja o no, para una instancia determinada.

Normalmente λ_k se calcula "minimizando" la función

$$h_k(\lambda) = f(x^k + \lambda d^k)$$

o sea, resolviendo un problema del tipo

$$\begin{array}{rl} & \min h_k(\lambda) \\ s.a. & \lambda \geq 0 \end{array}$$

Teorema 7 *Sea $f : \mathbb{R}^n \to \mathbb{R}$, diferenciable y sea $\{x^k\}$ la sucesión generada con λ^k tal que $\min_{\lambda \geq 0} h_k(\lambda) = h_k(\lambda_k)$. Si $\{x^k\}$ está contenida en un conjunto compacto, entonces la sucesión es finita y el último punto es estacionario para f, o la sucesión es infinita y todo punto de acumulación es punto estacionario de f.*

2.3.1. Método del gradiente

El método del gradiente, o método de descenso más pronunciado de Cauchy, elige $d^k = -\nabla f(x^k)$. En este método, el paso se escoge minimizando la función

$$h_k(\lambda) = f(x^k + \lambda d^k)$$

. Entonces

$$h_k(\lambda_k) = \min_{\lambda} h_k(\lambda)$$

Ejemplo: Resolver el problema de optimización $\{\min f(x) = (x-2)^2 + 2x + y^2 - y + 3\}$ utilizando el método del gradiente, partiendo desde el punto $(1,1)^T$).

Por el teorema de convergencia visto, sabemos que el procedimiento general puede o no converger. Si converge y f es convexa $\Rightarrow$ converge a un mínimo.

Notamos que las direcciones de descenso escogidas en iteraciones sucesivas de este método son ortogonales. Para esto basta ver la condición de optimalidad de primer orden en la selección del paso.

$$\nabla h'_k(\lambda) = 0 \Rightarrow \nabla f(x^k - \lambda \nabla f(x^k)) \nabla f(x^k) = 0 \Rightarrow \nabla f(x^{k+1}) \nabla f(x^k) = 0$$

2.3.2. Método de Newton

El método de Newton se basa en utilizar una aproximación de segundo orden de la función f y minimizar de manera exacta dicha función en cada iteración.

La aproximación utilizada es

$$g(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k)$$

La condición de primero orden en la optimización es

$$\begin{array}{rcl} & \nabla g(x) & = 0 \\ \Rightarrow & 0 & = \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) \\ \Rightarrow & -\nabla f(x^k) & = \nabla^2 f(x^k)(x - x^k) \\ \Rightarrow & x^{k+1} & = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k) \end{array}$$

La solución al sistema anterior se obtuvo suponiendo que $\nabla^2 f(x^k)$ no era singular. así podemos ver que la dirección que escoge el método de Newton en x^k es $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$ y que el paso es $\lambda = 1$ en cada iteración.

Ejemplo: Resolver el problema de optimización $\{\text{mín}\, f(x) = (x-2)^2 + 2x + y^2 - y + 3\}$ utilizando el método de Newton, partiendo desde el punto $(1,1)^T$).

Este método puede ser no convergente, por dos razones. En primer lugar, puede suceder que en algún punto x^k la matriz $\nabla^2 f(x^k)$ no sea invertible. Por otro lado, el utilizar un paso $\lambda = 1$ no asegura que $f(x^k) < f(x^{k+1})$.

Podemos decir que el método de Newton tiene tan solo convergencia local (Se comporta bien cerca del óptimo). Este método se puede globalizar permitiendo que el paso de aproximación sea escogido de forma de asegurar que $f(x^k) < f(x^{k+1})$.

2.4. Problemas

Problema 1

Sea el problema de minimización $Min\, f(x_1, x_2) = x_1^2 + x_2^2$. Si x_1 y x_2 pueden tomar cualquier valor real, resolver el problema aplicando el método del gradiente. Para dicho efecto, tome como punto de partida el par $(x_1 = 2, x_2 = 1)$.

Solución: Debemos resolver mín $f(x_1, x_2) = x_1^2 + x_2^2$ con $\vec{x}_0 = (2,1)$

- Método del gradiente: $\vec{x}_{k+1} = \vec{x}_k - \lambda_k \nabla f(\vec{x}_k)$

  Claramente, antes de comenzar a iterar se requiere calcular $\nabla f(\vec{x})$:

  $$\nabla f(\vec{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

  Ahora, comenzamos a iterar:

  - Iteración 1:
    - Vemos si estamos en el óptimo: $\nabla f(2,1) = (4,2) \neq \vec{0} \Rightarrow$ Seguimos.
    - Calculamos nuevo punto $\vec{x}_1$:

      $$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \lambda_0 \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

      Para calcular λ_0 minimizamos $h(\lambda)$:

      $$\begin{aligned} h(\lambda) &= f(\vec{x}_0 - \lambda \nabla f(\vec{x}_0)) \\ &= f\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix} - \lambda \begin{pmatrix} 4 \\ 2 \end{pmatrix}\right\} \\ &= (2-4\lambda)^2 + (1-2\lambda)^2 \end{aligned}$$

$$\frac{dh(\lambda)}{d\lambda} = 0 \Rightarrow \lambda_0 = \frac{1}{2}$$

(Si quieren pueden verificar que el punto crítico es óptimo comprobando que $d^2h/d\lambda^2 > 0$)

Entonces:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \vec{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Iteración 2:
  - Vemos si estamos en el óptimo: $\nabla f(0,0) = (0,0) = \vec{0} \Rightarrow$ Fin.

  $\therefore \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ es el mínimo de $x_1^2 + x_2^2$

- Método de Newton: $\vec{x}_{k+1} = \vec{x}_k - \{H_f(\vec{x}_k)\}^{-1}\nabla f(\vec{x}_k)$

  Claramente, antes de comenzar a iterar se requiere calcular $\nabla f(\vec{x})$ y $\{H_f(\vec{x})\}^{-1}$:

$$\nabla f(\vec{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

$$H_f(\vec{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow \{H_f(\vec{x})\}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

  Ahora, comenzamos a iterar:

- Iteración 1:
  - Vemos si estamos en el óptimo: $\nabla f(2,1) = (4,2) \neq \vec{0} \Rightarrow$ Seguimos.
  - Calculamos nuevo punto $\vec{x}_1$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \vec{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Iteración 2:
  - Vemos si estamos en el óptimo: $\nabla f(0,0) = (0,0) = \vec{0} \Rightarrow$ Fin.
  - Calculamos nuevo punto $\vec{x}_1$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \vec{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  $\therefore$ También se concluye que $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ es el mínimo de $x_1^2 + x_2^2$

Problema 2

Sea el problema de minimización $Min\, f(x_1,x_2) = \frac{(x_1-2)^2}{4} + (x_2-2)^2$. Comience a iterar con el método del gradiente e infiera que ocurre con la convergencia de la sucesión de descenso. Tome como punto de partida el par $(x_1 = 0, x_2 = 0)$.

Solución: Debemos resolver mín $f(x_1,x_2) = \frac{(x_1-2)^2}{4} + (x_2-2)^2$.

Claramente, como ya vimos, antes de comenzar a iterar con el método del gradiente, calculamos $\nabla f(\vec{x})$:

$$\nabla f(\vec{x}) = \begin{pmatrix} \frac{x_1}{2} - 1 \\ 2x_2 - 2 \end{pmatrix}$$

Ahora comenzamos a iterar

- Iteración 1:
  - Vemos si estamos en el óptimo: $\nabla f(0,0) = (-1,-2) \neq \vec{0} \Rightarrow$ Seguimos.
  - Calculamos nuevo punto $\vec{x}_1$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \lambda_0 \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

    Para calcular λ_0 minimizamos $h(\lambda)$:

$$\begin{aligned} h(\lambda) &= f(\vec{x}_0 - \lambda \nabla f(\vec{x}_0)) \\ &= f\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \lambda \begin{pmatrix} -1 \\ -2 \end{pmatrix} \right\} \\ &= \frac{(\lambda-2)^2}{4} + (2\lambda-2)^2 \end{aligned}$$

$$\frac{dh(\lambda)}{d\lambda} = 0 \Rightarrow \lambda_0 = \frac{10}{17}$$

  (Si quieren pueden verificar que el punto crítico es óptimo comprobando que $d^2h/d\lambda^2|_{10/17} > 0$)

  Entonces:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{10}{17} \begin{pmatrix} -1 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{10}{17} \\ \frac{20}{17} \end{pmatrix} \Rightarrow \vec{x}_1 = \begin{pmatrix} \frac{10}{17} \\ \frac{20}{17} \end{pmatrix}$$

- Iteración 2:

- Vemos si estamos en el óptimo: $\nabla f(10/17, 20/17) = (-12/17, 6/17) \neq \vec{0} \Rightarrow$ Seguimos.
- Calculamos nuevo punto $\vec{x}_2$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_2 = \begin{pmatrix} \frac{10}{17} \\ \frac{20}{17} \end{pmatrix} - \lambda_1 \begin{pmatrix} \frac{-12}{17} \\ \frac{6}{17} \end{pmatrix}$$

Para calcular λ_1 minimizamos $h(\lambda)$:

$$\begin{aligned} h(\lambda) &= f(\vec{x}_1 - \lambda \nabla f(\vec{x}_1)) \\ &= f\left\{ \begin{pmatrix} \frac{10}{17} \\ \frac{20}{17} \end{pmatrix} - \lambda \begin{pmatrix} \frac{-12}{17} \\ \frac{6}{17} \end{pmatrix} \right\} \\ &= \frac{((\frac{10}{17} - \frac{12}{17}\lambda) - 2)^2}{4} + ((\frac{20}{17} - \frac{6}{17}\lambda) - 2)^2 \end{aligned}$$

$$\frac{dh(\lambda)}{d\lambda} = 0 \Rightarrow \lambda_0 = \frac{5}{4}$$

(Si quieren pueden verificar que el punto crítico es óptimo comprobando que $d^2h/d\lambda^2|_{10/17} > 0$) Entonces:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_2 = \begin{pmatrix} \frac{10}{17} \\ \frac{20}{17} \end{pmatrix} - \frac{5}{4} \begin{pmatrix} \frac{-12}{17} \\ \frac{6}{17} \end{pmatrix} = \begin{pmatrix} \frac{50}{34} \\ \frac{25}{34} \end{pmatrix} \Rightarrow \vec{x}_2 = \begin{pmatrix} \frac{50}{34} \\ \frac{25}{34} \end{pmatrix}$$

- Iteración 3:
  - Vemos si estamos en el óptimo: $\nabla f(50/34, 25/34) = (-9/34, -18/34) \neq \vec{0} \Rightarrow$ Habría que continuar iterando.

Para ver que pasa con la convergencia, nos fijamos en las direcciones de descenso en las iteraciones determinados por el gradiente de la función:

- $d_0 = \nabla f(\vec{x}_0) = (1, 2)$
- $d_1 = \nabla f(\vec{x}_1) = (12/17, -6/17) = 6/17(2, -1)$
- $d_2 = \nabla f(\vec{x}_2) = (9/34, 18/34) = 9/34(1, 2)$

Y comprobamos el fenómeno de zig-zag (ver figura 2.1).

Problema 3

Considere el siguiente problema de optimización irrestricto:

$$\text{mín}\, f(x_1, x_2) = (x_1 - 2)^2 - 4 \cdot (x_2 - 5)^2$$

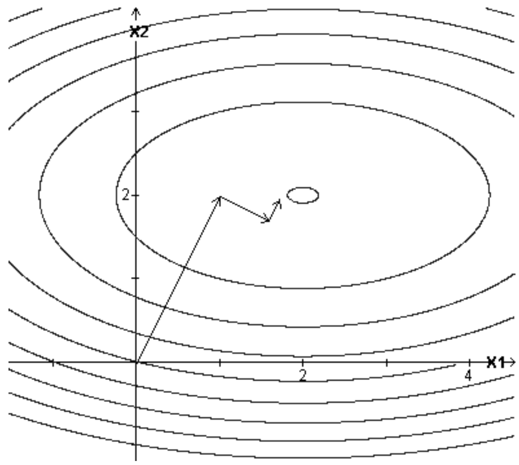


Figura 2.1: Fenómeno de Zig-Zag

- Resuelva mediante el método del gradiente usando como punto de partida el punto $(0, 5)$ y repita su procedimiento utilizando el punto $(0, 4)$. Comente los resultados obtenidos.
- Resuelva mediante el método de Newton partiendo de un punto (a, b) cualquiera. Comente y compare con respecto al método del gradiente.

Problema 4

Considere el siguiente problema de optimización irrestricto:

$$\text{mín}\, f(x_1, x_2) = x_2 \cdot e^{x_1 - 1} + \frac{x_2^2}{3} - x_1 - 2 \cdot x_2$$

- Analice su convexidad. Verifique si el punto $(1, 1)$ es estacionario y concluya sobre si es un mínimo global.

Capítulo 3

Optimización Restringida

Vamos a estudiar el siguiente problema (P):

$$\begin{array}{rcl} & \min f(x) & \\ s.a. & g_1(x) & \leq 0 \\ & \vdots & \\ & g_m(x) & \leq 0 \end{array}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y $g_i : \mathbb{R}^n \to \mathbb{R} \quad i = 1, \ldots, m$ continuamente diferenciables. En este caso vemos que $X = \{x \in \mathbb{R}^n : g_i \leq 0 \quad i = 1, \ldots, m\}$.

Queremos establecer condiciones necesarias y suficientes que nos permitan caracterizar mínimos locales y globales del problema (P). Para esto necesitamos unas definiciones previas:

Definición 19 *Sea $x \in X$, un vector $d \in \mathbb{R}^n$ es una dirección factible en X respecto a x, si $\exists\, \epsilon > 0$ tal que $x + \lambda d \in X \quad \forall \lambda \in (0, \epsilon]$.*

Entonces una dirección factible es tal que, si me muevo desde x es esa dirección, me mantengo en el conjunto X.

Definición 20 *Sea $x \in X$, el conjunto $D(x) = \{d \in \mathbb{R}^n : d \text{ es dirección factible}\}$ se conoce como el conjunto de direcciones factibles en X respecto a x.*

Con estas definiciones podemos establecer una primera condición de óptimalidad.

Teorema 8 *Sea x^* un mínimo local para el problema $(P) \Rightarrow \nabla f(x^*) \cdot d \geq 0 \quad \forall\, d \in D(x^*)$.*

Sea $d \in D(x^*)$ y $\epsilon > 0$ tal que $x^* + \lambda d \in X \quad \forall\, \lambda \in (0, \epsilon]$. Como x^* es un mínimo local, existe un $\gamma > 0$ tal que $f(x^*) \leq f(x)$, $\forall\, x \in X$ tal que $\|x - x^*\| < \gamma$.
Por otro lado, si $\nabla f(x^*) \cdot d < o$ entonces existe un $\epsilon^* < \text{mín}\{\epsilon, \gamma\}$ tal que $f(x^* + \lambda d) < f(x^*)$ $\forall\, \lambda \in (0, \epsilon^*]$. Pero para estos λ, $x^* + \lambda d \in X$ y esto contradice el hecho que x^* es mínimo local.

Para poder ocupar esta condición de óptimalidad en la práctica debemos caracterizar completamente el conjunto de direcciones factibles en un punto. Esto significa encontrar expresiones analíticas que definan este conjunto. Para esto necesitamos un par de definiciones adicionales:

Definición 21 *Una restricción de desigualdad $g_i(x)$ se dice activa en un punto factible x^* si $g_i(x^*) = 0$ y no activa, si $g_i(x^*) < 0$.*

Vemos que por definición las restricciones de igualdad son siempre activas. La factibilidad de una dirección en un punto x esta determinada por el comportamiento de las restricciones activas en ese punto. En particular una dirección $d \in \mathbb{R}^n$ será factible si al movernos en esa dirección el valor de las restricciones activas en x disminuye o se mantiene igual. Con esta interpretación es natural definir el siguiente conjunto.

Definición 22 *Sea $x \in X$, el conjunto de índices de restricciones activas o conjunto activo $I(x)$ es*

$$I(x) = \{i \in \{1, \ldots, m\} : g_i = 0\}$$

El cono tangente en x es

$$C(x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^T d \leq 0 \quad \forall\, i \in I(x)\}$$

Nuestra intuición nos dice que $C(x)$ describe al conjunto de direcciones factibles. La verdad es que todas las direcciones factibles están contenidas en $C(x)$, sin embargo existen elementos de $C(x)$ que no son direcciones factibles, es decir $D(x) \subseteq C(x)$.

Definición 23 *Sea $x \in X$, e $I(x) \neq 0$. Se dice que g_i diferenciables, con $i \in I(x)$, cumplen la condición de regularidad si $C(x) = Cl(D(x))$, la clausura de $D(x)$*

Para poder utilizar esta definición debemos centrar nuestro análisis en puntos tales que $C(x) = D(x)$. Una condición para que esto ocurra es que el punto en cuestión sea regular.

Definición 24 *Sea $x \in X$, se dice que x es regular para las restricciones si los vectores gradientes de las restricciones activas en x son linealmente independientes.*

Ejemplo:

$$\begin{array}{rl} \text{mín}\, f(x) & = (x-3)^2 + (y-2)^2 \\ s.a. \quad -x^2 - (y-1)^2 & \leq 0 \\ x - 2 & \leq 0 \\ -y & \leq 0 \end{array}$$

En el ejemplo anterior, el punto $(1, 2)$ no es un punto regular. Para puntos regulares podemos reescribir la condicion de óptimalidad de la siguiente forma:

Teorema 9 *Sea x^* un mínimo local para el problema $(P) \Rightarrow \nabla f(x^*) \cdot d \geq 0 \quad \forall\, d \in \mathbb{R}^n$ tal que $\nabla g_i(x^*) \cdot d \leq 0 \quad \forall\, i \in I(x^*)$*

El teorema anterior implica que, si x es un mínimo local, regular, entonces el siguiente sistema no tiene solución:

$$\begin{array}{rcl} \nabla f(x^*) \cdot d & < & 0 \\ \nabla g_i(x^*) \cdot d & \leq & 0 \quad \forall\, i \in I(x^*) \end{array} \tag{3.1}$$

esto es, si x es mínimo, entonces no puedo encontrar una dirección d de descenso, tal que la factibilidad se conserve.

Finalmente, para caracterizar las condiciones necesarias que deben cumplir los mínimos locales regulares debemos echar mano al lemma de Farkas.

Lemma 4 ***(Farkas)*** *Sean $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Uno y sólo uno de los siguientes sistemas tiene solución.*

$$\begin{array}{rcl} A\mu = b & \qquad\qquad & d^T A \geq 0 \\ \mu \geq 0 & & d^T b < 0 \end{array}$$

Reconociendo términos, vemos que, dado que en un punto mínimo regular el sistema descrito en (3.1) no tiene solución, entonces $\exists\, \mu \geq 0 \in \mathbb{R}^m$ tal que:

$$\nabla f(x) = -\sum_{i}^{|I(x)|} \mu_i \nabla g_i(x)$$

Como la sumatoria solamente incluye los indices de las restricciones activas en un punto agregamos las restricciones de complementaridad.

$$\mu_i g_i(x) = 0$$

de forma de escribir las relacción encontrada de la siguiente forma:

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

Ahora estamos en condiciones de enunciar las condiciones necesarias para la óptimalidad local de puntos regulares.

Teorema 10 ***(Condición de Karush-Kuhn-Tucker)***

Sea x^* factible, mínimo local del problema (P) que cumple con la condición de regularidad. Entonces $\exists \mu \geq 0 \in \mathbb{R}^m$ tal que

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

$$\mu_i g_i(x) = 0 \quad \forall\, i \in \{1, \ldots m\}$$

La primera de las condiciones indicadas en el teorema anterior es conocida como la condición de tangencia. Las condiciones de Karush-Kuhn-Tucker (KKT de ahora en adelante) nos dicen que $-\nabla f(x)$ esta contenido en el cono generado por los ∇g_i de las restricciones activas en el punto. Es importante notar que solo consideramos las restricciones activas en el punto.

Debemos recalcar que las condiciones de KKT nos aseguran que de haber un punto mínimo regular este cumplirá con las condiciones. Perfectamente podría resultar que el sistema de condiciones de KKT no tiene solución, sin embargo el problema de mínimización (P) si tiene mínimos locales, los que en este caso serían puntos que no son regulares. De la misma forma cualquier punto que satisfaga las condiciones de KKT no es necesariamente un mínimo local.

Ejemplo:

$$\begin{array}{rrl} & \min f(x) & = -y \\ s.a. & x - 2(y-1)^3 & \leq 0 \\ & -x - 2(y-1)^3 & \leq 0 \end{array}$$

En el ejemplo anterior, el punto $(0,1)$ es un mínimo local para el problema, sin embargo no cumple las condiciones de KKT.

Ejemplo:

$$\begin{array}{rll} & \min f(x) & = -y \\ s.a. & -x^2+y & \leq 0 \\ & x^2+y^2-2 & \leq 0 \end{array}$$

En el ejemplo anterior, el punto $(0,0)$ cumple las condiciones de KKT, sin embargo no es un mínimo local del problema. Aun así podemos encontrar casos, para los cuales un punto que cumple las condiciones de KKT es un mínimo global.

Teorema 11 *Sea el problema (P) tal que $f(x)$ y $g_i(x)$ son diferenciables y convexas $\forall\, i \in \{1,\ldots,m\}$. Sea x^* una solución factible para (P). Si x^* satisface las condiciones de KKT, $\Rightarrow x$ es mínimo global para (P).*

Ejemplo:

$$\begin{array}{rll} & \min f(x) & = (x-3)^2+(y-2)^2 \\ s.a. & x^2-y-3 & \leq 0 \\ & y-1 & \leq 0 \\ & -x & \leq 0 \end{array}$$

En el ejemplo anterior, el punto $(2,1)$ cumple con las condiciones de KKT. Dado que tanto la función objetivo como las restricciones son convexas, concluimos que este punto es un mínimo global para el problema.

A continuación veremos como extender los resultados logrados hasta el momento para el caso de restricciones con formas de igualdad.

3.1. Condición de Lagrange

Consideremos el problema $(P1)$:

$$
\begin{array}{rll}
 & \text{mín}\, f(x) & \\
s.a. & h_1(x) & = 0 \\
 & \vdots & \\
 & h_m(x) & = 0
\end{array}
$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y $h_j : \mathbb{R}^n \to \mathbb{R} \quad j = 1, \dots, m$ continuamente diferenciables. En este caso vemos que $X = \{x \in \mathbb{R}^n : h_j \leq 0 \quad j = 1, \dots, m\}$.

Teorema 12 *Sea x^* mínimo local de $(P1)$ tal que x^* es regular. Entonces, existen $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ tales que*

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla h_j(x^*)$$

A diferencia de las condiciones de KKT no exigimos que los multiplicadores λ sean positivos. Esta condición puede inferirse fácilmente de las condiciones de KKT. Escribamos las condiciones $h(x) = 0$ de la siguiente forma:

$$h_j(x) = 0 \Leftrightarrow \begin{cases} h_j(x) \leq 0 \\ h_j(x) \geq 0 \end{cases}$$

Dado que x^* es mínimo local y regular, debe cumplir con las condiciones de KKT $\Rightarrow$ $\exists \alpha_1, \dots, \alpha_m \geq 0$ y $\beta_1, \dots, \beta_m \geq 0$ tales que:

$$
\begin{aligned}
0 &= \nabla f(x^*) + \sum_{j=1}^{m} \alpha_j \nabla h_j(x^*) - \sum_{j=1}^{m} \beta_j \nabla h_j(x^*) \\
0 &= \nabla f(x^*) + \sum_{j=1}^{m} \lambda_j \nabla h_j(x^*)
\end{aligned}
$$

Como λ_j es la resta de dos números positivos, vemos que puede ser positivo como puede ser negativo.

Si x es un punto regular del problema $(P1)$, la condición de Lagrange establece que en este punto el vector gradiente de $f(x)$ puede expresarse como una combinación lineal de los gradientes de las restricciones $h_j(x)$. Esto implica que la proyección de $f(x)$ sobre el subespacio tangente a la superficie factible en x es nula.

Ejemplo:

$$\begin{array}{rll} & \min f(x) & = (x-2)^2 + (y-1)^2 \\ s.a. & x + y & = 1 \end{array}$$

En el ejemplo anterior el punto $(1,0)$ cumple las condiciones de Lagrange. Este punto corresponde al mínimo global del problema.

Teorema 13 *En el problema $(P1)$ consideremos $f(x)$ convexa y $h_j(x)$ afines, es decir, $h_j(x) = a_j^T x + b_j$, $a_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$, y sea x^* factible para $(P1)$. Si x^* satisface la condición de Lagrange, entonces x^* es un mínimo global de $(P1)$.*

El teorema anterior considera funciones $h_j(x)$ afines para asegurar la convexidad del espacio de soluciones factibles (las funciones afines son convexas y cóncavas a la vez).

3.2. Técnica de los Multiplicadores de Lagrange

Sea $(P2)$ es siguiente problema de optimización:

$$\begin{array}{rlll} & \min f(x) & & \\ s.a. & g_i(x) & \leq 0 & i = 1, \ldots, m \\ & \vdots & & \\ & h_j(x) & = 0 & j = 1, \ldots, p \end{array}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R} \quad i = 1, \ldots, m$ y $h_j : \mathbb{R}^n \to \mathbb{R} \quad j = 1, \ldots, p$ continuamente diferenciables.

Definición 25 *Se define la función lagrangeana (o Lagrangeano) $L : h_j : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$*

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

Esta función nos entrega una manera conveniente de obtener y verificar las condiciones de KKT y/o Lagrange para el problema $(P2)$. Para ver esto veamos las formas de las derivadas parciales de la función respecto a cada uno de sus argumentos, en particular igualamos a 0 (desigualdad para las derivadas respecto a μ) cada una de estas componentes.

$$\begin{aligned} D_x L(x,\mu,\lambda) &= \nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) + \sum_{j=1}^{p} \lambda_j \nabla h_j(x) = 0 \\ D_{\mu_i} L(x,\mu,\lambda) &= g_i \leq 0 \quad \forall i = 1,\dots,m \\ D_{\lambda_j} L(x,\mu,\lambda) &= h_j \leq 0 \quad \forall j = 1,\dots,p \end{aligned}$$

Además agregamos condiciones de exclusión y factibilidad:

$$\begin{aligned} \mu_i &\geq 0 \quad \forall\, i = 1,\dots,m \\ \mu_i g_i(x) &= 0 \quad \forall\, i = 1,\dots,m \end{aligned}$$

Resolver este sistema, denominado sistema de Karush-Kuhn-Tucker, significa determinar x, μ y λ tales que, por una parte verifiquen las condiciónes de tangencia y exclusión, y por otra parte, entreguen un x factible para $(P2)$.

Las soluciones a este sistema nos entregan candidatos a óptimos locales.

Puede darse que el multiplicador asociada a una restricción activa en un mínimo local regular sea nulo.

Ejemplo:

$$\begin{aligned} & \min f(x) && = (x-2)^2 + y^2 \\ s.a. \quad & x + y && = 1 \\ & x^2 + y^2 && \leq 1 \end{aligned}$$

En el ejemplo anterior, el punto $(1,0)$ cumple con las condiciones de KKT. Allí el multiplicador asociado a las restricción de igualdad tiene valor nulo.

Notamos nuevamente que el sistema puede no tener solución y el problema tener mínimos locales. Esto ocurre cuando los mínimos son no regulares.

Ejemplo:

$$\begin{aligned} & \min f(x) && = -y \\ s.a. \quad & x - 2(y-1)^3 && \leq 0 \\ & -x - 2(y-1)^3 && \leq 0 \end{aligned}$$

En el ejemplo anterior, el punto $(0, 1)$ corresponde al mínimo global del problema, y sin embargo no satisface las condiciones de KKT.

3.3. Problemas

Problema 1

Sea el problema:

$$\text{máx } f(x_1, x_2) = x_1^2 + x_2^2$$

$$\begin{aligned} s.a \quad & x_1 \leq 1 \\ & x_1 \geq -1 \\ & x_2 \leq 1 \\ & x_2 \geq 0 \end{aligned}$$

- Resuelva gráficamente el problema. Verifique el cumplimiento de las condiciones de Khun-Tucker en el óptimo.
- Verifique si para los puntos $(x_1 = 0, x_2 = 1)$ y $(x_1 = 0, x_2 = 0)$ se cumplen las condiciones de Khun-Tucker. ¿Existen direcciones factibles para estos 2 puntos en los cuales se mejore el valor de la función objetivo?. Explique brevemente.

Solución:

- Al resolver gráficamente, dibujando las curvas de nivel, podemos ver que los óptimos del problema vienen dados por a=(1,1) y b=(-1,1)(ver figura 3.1). De las condiciones

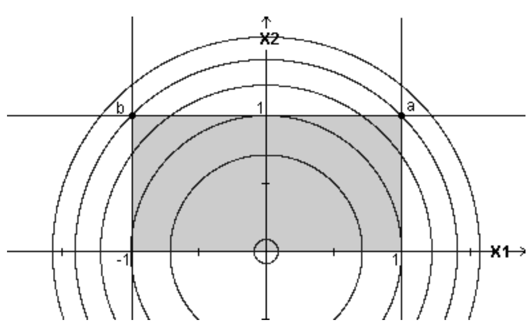


Figura 3.1: curvas de nivel

de KKT notamos 2 cosas:

1. Las condiciones están escritas para la forma estándar del problema, es decir, el problema es de minimización y lasrestricciones están escritas de la forma $g_i(\vec{x}) \leq 0$.
2. Claramente, antes de plantear las ecuaciones necesitamos conocer los gradientes de la función objetivo (∇f) y de las restricciones (∇g)

Entonces:

- Escribimos el problema en forma estándar

$$\text{mín} \quad \tilde{f} = -x_1^2 - x_2^2$$

$$\begin{array}{lcrrrcc} s.a & g_1: & x_1 & & -1 & \leq & 0 \\ & g_2: & -x_1 & & -1 & \leq & 0 \\ & g_3: & & x_2 & -1 & \leq & 0 \\ & g_4: & & -x_2 & & \leq & 0 \end{array}$$

- Calculamos los gradientes:

$$\nabla f(\vec{x}) = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$$

$$\nabla g_1(\vec{x}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \nabla g_2(\vec{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\nabla g_3(\vec{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \nabla g_4(\vec{x}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

- Ahora podemos imponer las ecuaciones sobre los 2 puntos óptimos encontrados: a y b.
  - a=(1,1)

    - $\mu_1 \cdot (x_1^a - 1) = 0$. Como $x_1^a = 1 \Rightarrow (x_1^a) = 0$ y por tanto la condición se cumple $\forall \mu_1 \in R$ y no podemos decir nada más.
    - $\mu_2 \cdot (-x_1^a - 1) = 0$. Como $x_1^a = 1 \Rightarrow (x_1^a) = 2 \neq 0$ y por tanto la condición solo se cumple si $\mu_2 = 0$.
    - $\mu_3 \cdot (x_2^a - 1) = 0$. Análogamente al primer caso, $\mu_3 \in R$.
    - $\mu_4 \cdot (-x_2^a) = 0$ Análogamente al segundo caso, $\mu_4 = 0$.

    La otra condición:

$$\begin{pmatrix} -2x_1^a \\ -2x_2^a \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Reemplazando los valores:

$$\begin{pmatrix} -2 \\ -2 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Resolviendo el sistema de ecuaciones: $\begin{matrix} \mu_1 = 2 & \mu_2 = 0 \\ \mu_3 = 2 & \mu_4 = 0 \end{matrix}$

Como $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0 \Rightarrow$ se cumple la condición de KKT para el punto a=(1,1).

- b=(-1,1)

  - $\mu_1 \cdot (x_1^b - 1) = 0 \Rightarrow \mu_1 = 0$
  - $\mu_2 \cdot (-x_1^b - 1) = 0 \Rightarrow \mu_2 \in R$
  - $\mu_3 \cdot (x_2^b - 1) = 0 \Rightarrow \mu_3 \in R$
  - $\mu_4 \cdot (-x_2^b) = 0 \Rightarrow \mu_4 = 0$

  Y la otra condición:

$$\begin{pmatrix} -2x_1^b \\ -2x_2^b \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  Reemplazando los valores:

$$\begin{pmatrix} +2 \\ -2 \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  Resolviendo el sistema de ecuaciones: $\begin{matrix} \mu_1 = 0 & \mu_2 = 2 \\ \mu_3 = 2 & \mu_4 = 0 \end{matrix}$

  Como $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0 \Rightarrow$ se cumple la condición de KKT para el punto b=(-1,1).

- Para verificar las condiciones de KKT, el sistema a resolver es el mismo anterior, pero evaluando en otros puntos

  - (0,1)

    - $\mu_1 \cdot (0 - 1) = 0 \Rightarrow \mu_1 = 0$
    - $\mu_2 \cdot (-0 - 1) = 0 \Rightarrow \mu_2 = 0$
    - $\mu_3 \cdot (1 - 1) = 0 \Rightarrow \mu_3 \in R$
    - $\mu_4 \cdot (-1) = 0 \Rightarrow \mu_4 = 0$

La otra condición (evaluada inmediatamente):

$$\begin{pmatrix} 0 \\ -2 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Resolviendo el sistema de ecuaciones: $\mu_3 = 2$. Como $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0 \Rightarrow$ se cumple la condición de KKT para el punto (0,1).

- (0,0)

  - $\mu_1 \cdot (0 - 1) = 0 \Rightarrow \mu_1 = 0$
  - $\mu_2 \cdot (-0 - 1) = 0 \Rightarrow \mu_2 = 0$
  - $\mu_3 \cdot (0 - 1) = 0 \Rightarrow \mu_3 = 0$
  - $\mu_4 \cdot (-0) = 0 \Rightarrow \mu_4 \in R$

  La otra condición (evaluada inmediatamente):

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Resolviendo el sistema de ecuaciones: $\mu_4 = 0$. Como $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0 \Rightarrow$ se cumple la condición de KKT para el punto (0,1).

Con respecto a las direcciones factibles, debemos decir que una dirección factible es aquella dirección (vector) que al movernos *un poco* (diferencial) sobre ella no nos *saldremos* del conjunto factible. Así, las direcciones factibles que mejoren el valor de la función objetivo las podemos hallar gráficamente:

- (0,1):Este punto no es óptimo local porque existen direcciones factibles que mejoran la función objetivo, siendo estas las direcciones (1,0) y (-1,0) como lo indica la figura 3.2.

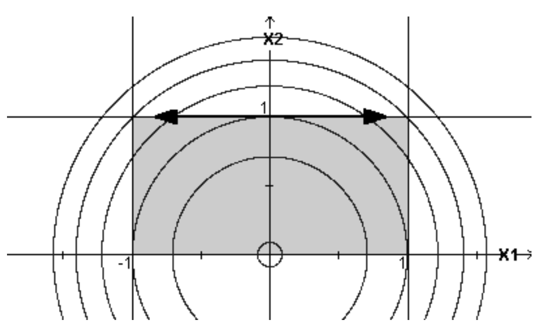


Figura 3.2: Análisis punto (0,1)

- (0,0):Este punto tampoco es óptimo local porque también existen direcciones factibles que mejoran la función objetivo, siendo estas todas las direcciones factibles como lo indica la figura 3.3 (el punto es un mínimo).

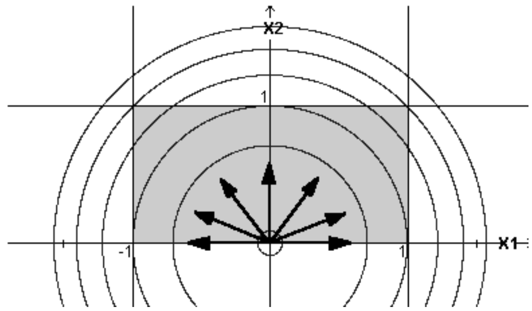


Figura 3.3: Análisis punto (0.0)

Problema 2

Sea el problema

$$\text{mín } f(x,y) = e^{-x} + e^{-2y}$$

$$\begin{array}{rrcl} s.a & x+y & = & 1 \\ & x & \geq & 0 \\ & y & \geq & 0 \end{array}$$

Encontrar el óptimo del problema utilizando la técnica de los multiplicadores de Lagrange.

Solución: La forma estándar equivalente del problema viene dada por:

$$\text{mín } f(x,y) = e^{-x} + e^{-2y}$$

$$\begin{array}{rrcll} s.a & x+y-1 & = & 0 & (h_1) \\ & -x & \leq & 0 & (g_1) \\ & -y & \leq & 0 & (g_2) \end{array}$$

Construimos la función Lagrangeana:

$$L = e^{-x} + e^{-2y} + \lambda(x+y-1) + \mu_1(-x) + \mu_2(-y)$$

Luego las condiciones de lagrange, nos llevan al siguiente sistema de ecuaciones:

$$\begin{array}{lr} \partial L/\partial x = 0: & -e^{-x} + \lambda - \mu_1 = 0 \\ \partial L/\partial y = 0: & -2e^{-2y} + \lambda - \mu_2 = 0 \\ \partial L/\partial \lambda = 0: & x+y-1 = 0 \\ \partial L/\partial \mu_1 = 0: & -x \leq 0 \\ \partial L/\partial \mu_2 = 0: & -y \leq 0 \end{array}$$

y las condiciones adicionales:

$$\mu_1, \mu_2 \geq 0$$

$$\mu_1(-x) = 0$$

$$\mu_2(-y) = 0$$

Estas condiciones adicionales nos dicen que hay 4 casos posibles:

- $\underline{x = 0, y = 0}$(Restricciones g_1 y g_2 activas:) En este caso, la restricción $x + y - 1 = 0$ no se satisface y por tanto este caso no es posible.
- $\underline{x = 0, \mu_2 = 0}$(Restricción g_1 activa:) En este caso, al resolver el sistema se obtiene que $\mu_1 = 2e^{-2} - 1 = -0,729 < 0$ y nuevamente no puede ser.
- $\underline{y = 0, \mu_1 = 0}$(Restricción g_1 activa:) En este caso, al resolver el sistema se obtiene que $\mu_2 = 2e^{-2} - 1 = -1,632 < 0$ y nuevamente no puede ser.
- $\underline{\mu_1 = 0, \mu_2 = 0}$ (Ni g_1 ni g_2 activas:) En este caso, al resolver el sistema de ecuaciones $\mu_1, \mu_2 \geq 0$ y el resultado óptimo viene dado por $(x = 0{,}436; y = 0{,}564)$ como lo indica la figura 3.4 .

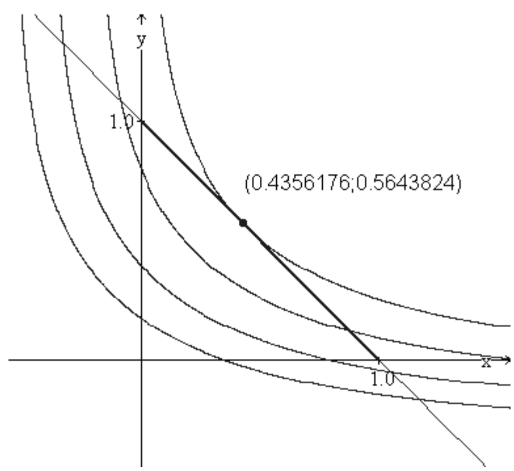


Figura 3.4: óptimo del problema: $(x = 0{,}436; y = 0{,}564)$

Problema 3

Considere el siguiente problema de optimización:

$$\text{mín}\, f(x, y) = x \cdot y$$

S.a.

$$x^2 + y^2 \leq 4$$

$$(x+2)^2 + (y-2)^2 \geq 4$$

$$x - y \leq 2$$

- Determine gráficamente el conjunto de soluciones factibles.
- Verifique si se cumplen las condiciones de KKT en los siguientes puntos: $(2^{\frac{1}{2}} - 2, 2 - 2^{\frac{1}{2}}); (1, -1); (0, 2); (0, 0)$.
- Puede usted afirmar que los puntos que cumplen KKT son óptimos locales?.
- A partir de lo anterior, puede afirmar usted que alguno de los puntos es un óptimo global?.

Problema 4

Considere el siguiente problema de optimización:

$$\text{mín}\, f(x_1, x_2) = x_1^2 + x_2^2 - 6x_1 - 8x_2$$

S.a.

$$x_1^2 + x_2^2 - 2x_1 - 4x_2 + 4 \geq 0$$

$$x_1 \leq 2$$

$$x_2 \leq 2$$

$$x_1, x_2 \geq 0$$

- Grafique el conjunto factible. Determine (gráficamente) la solución óptima, dibujando curvas de nivel para la función objetivo.
- Para el punto óptimo encontrado, se pueden ver las condiciones de KKT?. escríbalas y explique.

Problema 5

Considere el siguiente problema de optimización:

$$\text{mín}\, f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$$

S.a.

$$x_1^2 - 2x_1 + 2 \geq 0$$

$$x_1 - 2 \leq 0$$

$$x_1, x_2 \geq 0$$

- Grafique el conjunto factible. Es convexo?. Demuestre gráficamente y mediante un ejemplo numérico.
- Demuestre que las condiciones de KKT no son suficientes para determinar una solucíon optima.
- Si el signo de la primera restricción se cambia por $\leq$, son suficientes las condiciones de KKT en el nuevo problema?. Grafique el nuevo espacio factible. Es convexo?. Demuestre gráficamente.
- Determine la solución óptima del problema modificado analíticamente, con el apoyo del análisis gráfico.

Problema 6

Considere el siguiente problema de optimización:

$$\text{mín}\, f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$$

S.a.

$$x_1 \leq 2$$

$$x_1 - 2 \leq 0$$

$$x_1, x_2 \geq 0$$

- Grafique el conjunto factible y determine si $x_1 = 2yx_2 = 2$ es solución óptima. Determine si este punto cumple las condiciones de KKT
- Verifique si los puntos $(2, 1)$ y $(1, 1)$ cumplen las condiciones de KKT y determine si son o no óptimos. Explique la situación.

Problema 7

Considere el siguiente problema de optimización:

$$\text{máx} \, f(x_1, x_2) = x_1$$

S.a.

$$(x_1 - 2)^2 + (x_2 - 1)^2 = 5$$

$$(x_1 - 3)^2 + x_2^2 \leq 9$$

$$x_1, x_2 \geq 0$$

- Como modificaría el modelo para poder aplicar las condiciones de KKT
- Plantee las condiciones para el modelo modificado y encuentre los posibles puntos óptimos para este problema. explique la situación en cada punto y el por qué del cumplimiento o no de las condiciones de KKT (apóyese en el análisis gráfico).
- Ahora considere el problema de mín $f(x_1, x_2)$. Plantee las condiciones para el modelo modificado y encuentre los posibles puntos óptimos para este problema. explique la situación en cada punto y el por qué del cumplimiento o no de las condiciones de KKT (apóyese en el análisis gráfico).

Problema 8

Considere el siguiente problema de optimización:

$$\text{mín} \, f(x_1, x_2) = ln(x_1 + 1) + x_2$$

s.a.

$$2x_1 + x_2 \leq 3$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

- Plantee las condiciones de KKT para este problema.
- Son suficientes?.Por qué?
- El punto $x_1 = 0$,$x_2 = 1$ es óptimo?.
- Determine la solución óptima.

Problema 9

Considere el siguiente problema de optimización con restricciones:

$$\text{mín}\, f(x_1, x_2) = x_1 - 3x_2$$

s.a.

$$(x_1-1)^2 + (x_2-1)^2 \leq 1$$
$$x_1^2 + x_2^2 \leq 1$$
$$x_1 + x_2 \leq \frac{4}{3}$$
$$x_1, x_2 \geq 0$$

- Grafique el conjunto factible. Determine (gráficamente) los candidatos a óptimo.
- Verifique las condiciones de KKT para los siguientes puntos: $(1, 0)$; $(0, 1)$; $(0, 0)$;$(\frac{1}{2}, \frac{3}{4})$. Las condiciones de KKT, ¿son suficientes para asegurar la optimalidad?. De lo contrario, ¿qué se necesita?. Especifique en el problema.

Problema 10

Sea el siguiente problema no lineal restringido:

$$\text{máx}\, f(x_1, x_2) = \ln(x_1 + x_2)$$

s.a.

$$x_1 + 2x_2 \leq 5$$
$$x_1, x_2 \geq 0$$

- Vea analíticamente si en el problema planteado se cumple la condición de suficiencia. Diga si ésta condición se cumple o no para todo punto factible.
- Grafique la condición necesaria de KKT en el punto $(5, 0)$ y vea si cumple.
- Escriba la condición necesaria de KKT para este problema. (Deje planteado el sistema de ecuaciones).
- Diga cuales restricciones son activas y cuál es el valor de los multiplicadores Lagrange en el punto $(5, 0)$. Para hacer esto resuelve analíticamente la condición necesaria planteada en el punto anterior.
- Si ahora se cambia la restricción: $x_1 + 2x_2 \leq 5$ por $x_1 + x_2 \leq 5$. Vuelva a ver el punto $(5, 0)$ si se cumple la condición necesaria.

Problema 11

- Sea el siguiente problema de optimización:

$$\begin{array}{ll} \max & f(x) = x_1 \\ s.a: & \\ & \begin{array}{rcl} (x_1-1)^2 + x_2^2 & \geq & 1 \\ (x_1-2)^2 + x_2^2 & \leq & 4 \\ (x_1-3)^2 + x_2^2 & \geq & 1 \end{array} \end{array}$$

  - Encuentre la solución óptima gráficamente. Verifique si esta solución óptima cumple las condiciones de Karush-Kuhn-Tucker.
  - Encuentre 2 puntos más que cumplan las condiciones de Karush-Kuhn-Tucker. Explique que características tienen.

- Sea el siguiente problema de optimización:

$$\begin{array}{ll} \max & f(x) = x_2 \\ s.a: & \\ & \begin{array}{rcl} (x_1-1)^2 + (x_2-1)^2 & \leq & 1 \\ x_2 & \leq & 1 \end{array} \end{array}$$

  - Encuentre gráficamente la(s) solución(es) óptima(s). Verifique si cumple(n) las condiciones de Karush-Kuhn-Tucker. Comente la situación.
  - Encuentre otro punto que cumpla las condiciones de Karush-Kuhn-Tucker. Comente de que punto se trata.
  - Suponga que la función objetivo se cambió por:

$$\min f(x) = x_2$$

    Encuentre gráficamente la solución óptima y analice si cumple o no las condiciones de Karush-Kuhn-Tucker.

Solución:

- Primero debe transformarse el problema original en un problema en forma estándar, quedando:

$$\begin{array}{ll} \min & \tilde{f}(x) = -x_1 \\ s.a: & \\ & \begin{array}{rcll} 1-(x_1-1)^2 - x_2^2 & \leq & 0 & (g_1) \\ (x_1-2)^2 + x_2^2 - 4 & \leq & 0 & (g_2) \\ 1-(x_1-3)^2 - x_2^2 & \leq & 0 & (g_3) \end{array} \end{array}$$

Ahora escribamos $\nabla f(x)$ y $\nabla g_i(x)$:

$$\nabla f(x) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} -2 \cdot (x_1 - 1) \\ -2 \cdot x_2 \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 2 \cdot (x_1 - 2) \\ 2 \cdot x_2 \end{pmatrix}$$

$$\nabla g_3(x) = \begin{pmatrix} -2 \cdot (x_1 - 3) \\ -2 \cdot x_2 \end{pmatrix}$$

- La solución óptima corresponde al punto $(x_1, x_2) = (4, 0)$, el cual se puede ver en la figura 3.5. Para verificar las condiciones de KKT en ese punto se tiene:

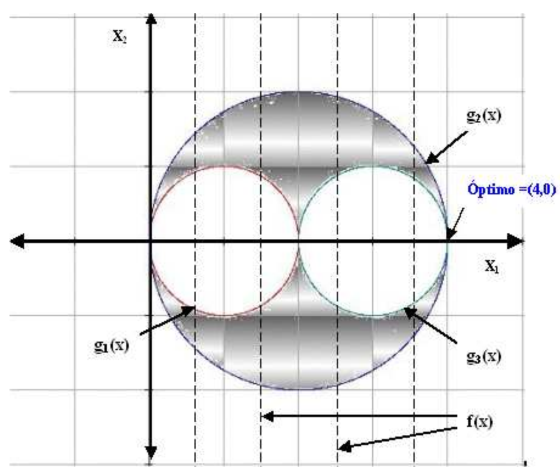


Figura 3.5: Gráfico asociado al problema inicial

  - Las restricciones g_2 y g_3 son activas en el punto $(4, 0)$, por lo tanto u_2 y $u_3 \in \mathbb{R}$.
  - La restricción g_1 es inactiva en el punto $(4, 0)$, por lo tanto $u_1 = 0$.

Utilizando una de las condiciones de KKT ($\nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$) se tiene que:

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 2 \cdot (4-2) \\ 2 \cdot 0 \end{pmatrix} + u_3 \begin{pmatrix} 2 \cdot (4-3) \\ 2 \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

entonces se tiene la siguiente ecuación:

$$-1 + 4u_2 - 2u_3 = 0$$

Luego, es posible encontrar un par de valores para $u_2 \geq 0$ y $u_3 \geq 0$ que cumplan la restricción, por ejemplo $u_2 = 1$ y $u_3 = \frac{3}{2}$. Luego, existen u_1, u_2 y $u_3 \geq 0$ entonces se cumple la condición de KKT para el punto $(4, 0)$.

- Otros dos puntos que cumplen con las condiciones de KKT pueden ser:
  - Punto $(0, 0)$: En este punto las restricciones g_1 y g_2 son activas, por lo tanto u_1 y $u_2 \in \mathbb{R}$. La restricción g_3 es inactiva, por lo tanto $u_3 = 0$. Ahora, utilizando $\nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

    $$\begin{pmatrix} \text{-1} \\ 0 \end{pmatrix} + u_1 \begin{pmatrix} 2 \cdot (0 - 1) \\ 2 \cdot 0 \end{pmatrix} + u_2 \begin{pmatrix} 2 \cdot (0 - 2) \\ 2 \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

    entonces se obtiene la siguiente ecuación:

    $$-1 + 2u_1 - 4u_2 = 0$$

    Luego, es posible encontrar un par de valores para $u_1 \geq 0$ y $u_2 \geq 0$ que cumplan la restricción, por ejemplo $u_1 = 1$ y $u_2 = \frac{1}{4}$. Luego, existen u_1, u_2 y $u_3 \geq 0$ entonces se cumple la condición de KKT para el punto $(0, 0)$.
  - Punto $(2, 0)$: En este punto las restricciones g_1 y g_3 son activas, por lo tanto u_1 y $u_3 \in \mathbb{R}$. La restricción g_2 es inactiva, por lo tanto $u_2 = 0$. Ahora, utilizando $\nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

    $$\begin{pmatrix} \text{-1} \\ 0 \end{pmatrix} + u_1 \begin{pmatrix} 2 \cdot (2 - 1) \\ 2 \cdot 0 \end{pmatrix} + u_3 \begin{pmatrix} 2 \cdot (2 - 3) \\ 2 \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

    entonces se obtiene la siguiente ecuación:

    $$-1 - 2u_1 + 2u_3 = 0$$

    Luego, es posible encontrar un par de valores para $u_1 \geq 0$ y $u_3 \geq 0$ que cumplan la restricción, por ejemplo $u_1 = 1$ y $u_2 = \frac{3}{2}$. Luego, existen u_1, u_2 y $u_3 \geq 0$ entonces se cumple la condición de KKT para el punto $(2, 0)$.

- Sea el siguiente problema de optimización:

$$\begin{array}{lrcl} \text{máx} & f(x) = x_2 & & \\ s.a: & & & \\ & (x_1 - 1)^2 + (x_2 - 1)^2 & \leq & 1 \\ & x_2 & \leq & 1 \end{array}$$

- Primero, transformemos el problema planteado a la forma estándar y calculemos $\nabla f(x)$ y $\nabla g_i(x)$. La forma estándar es:

$$\begin{array}{lll} \text{mín} & \tilde{f}(x) = -x_2 & \\ s.a: & & \\ & (x_1-1)^2 + (x_2-1)^2 - 1 & \leq 0 \\ & x_2 - 1 & \leq 0 \end{array}$$

Además:

$$\nabla \tilde{f}(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 2 \cdot (x_1-1) \\ 2 \cdot (x_2-1) \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Gráficamente se puede apreciar (ver Figura 3.6) que la solución óptima al problema corresponde a todo el segmento comprendido entre los puntos $(0,1)$ y $(2,1)$. Luego, para analizar las condiciones de KKT, revisemos 3 casos:

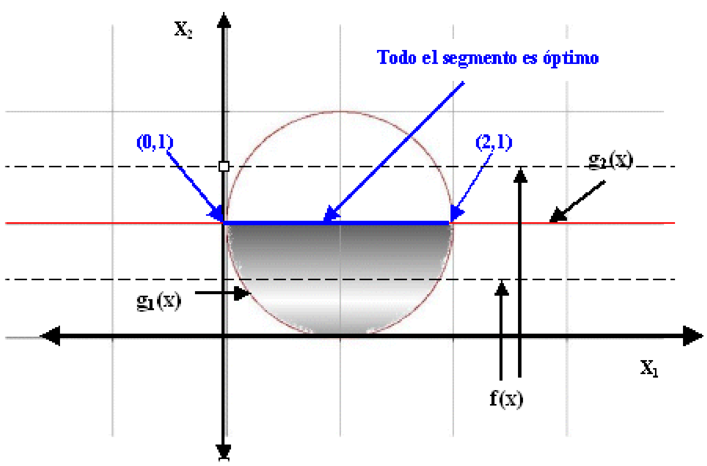


Figura 3.6: Gráfico asociado al problema inicial

  - Segmento completo exceptuando los extremos: en este caso, las restricción g_1 es inactiva, por lo tanto $u_1 = 0$, y la restricción g_2 es activa, luego $u_2 \in \mathbb{R}$. Ahora, utilizando $\nabla \tilde{f}(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

entonces se obtiene la siguiente ecuación:

$$-1 + u_2 = 0$$

Luego, $u_2 = 1$, y como existen u_1 y $u_2 \geq 0$ entonces se cumple la condición de KKT para cualquier punto del segmento.

- Punto $(0, 1)$: en este punto ambas restricciones son activas, por lo tanto u_1 y $u_2 \in \mathbb{R}$. Ahora, utilizando $\nabla \tilde{f}(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

$$\begin{pmatrix} 0 \\ \text{-1} \end{pmatrix} + u_1 \begin{pmatrix} 2 \cdot (0-1) \\ 2 \cdot (1-1) \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

entonces se obtienen las siguientes ecuaciones:

$$-2u_1 = 0$$

$$-1 + u_2 = 0$$

Luego, la solución a este sistema de ecuaciones es $u_1 = 0$ y $u_2 = 1$. Como existen u_1 y $u_2 \geq 0$ entonces se cumple la condición de KKT para el punto $(0, 1)$.

- Punto $(2, 1)$: en este punto ambas restricciones son activas, por lo tanto u_1 y $u_2 \in \mathbb{R}$. Ahora, utilizando $\nabla \tilde{f}(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

$$\begin{pmatrix} 0 \\ \text{-1} \end{pmatrix} + u_1 \begin{pmatrix} 2 \cdot (2-1) \\ 2 \cdot (1-1) \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

entonces se obtienen las siguientes ecuaciones:

$$2u_1 = 0$$

$$-1 + u_2 = 0$$

Luego, la solución a este sistema de ecuaciones es $u_1 = 0$ y $u_2 = 1$. Como existen u_1 y $u_2 \geq 0$ entonces se cumple la condición de KKT para el punto $(0, 1)$.

- No es posible encontrar otro punto que cumplas las condiciones de KKT, esto porque la función $(f(x))$ y las restricciones $(g_i(x))$ son convexas, luego, cualquier otro punto que cumpla las condiciones de KKT necesariamente será óptimo global del problema, pero estos (los óptimos globales) ya han sido determinados y analizada la condición de KKT en ellos.

- Suponga que la función objetivo se cambió por:

$$\text{mín } f(x) = x_2$$

Sobre el problema original solamente cambia el gradiente de la función objetivo, quedando:

$$\nabla \tilde{f}(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

La nueva solución se puede observar en la figura 3.7. En este punto las restricción

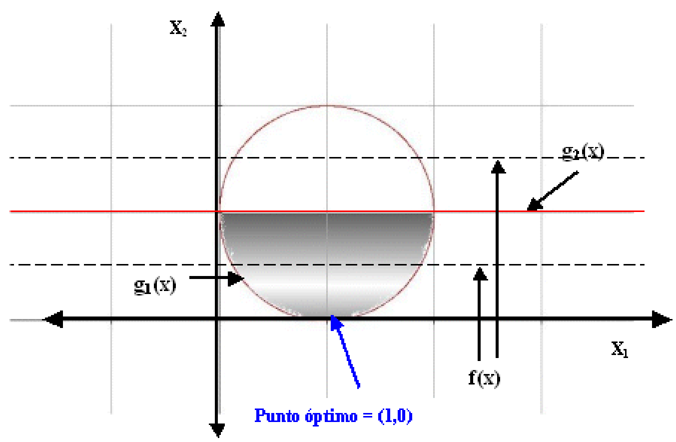


Figura 3.7: Gráfico asociado al problema modificado

g_1 es activa, por lo tanto $u_1 \in \mathbb{R}$. La restricción g_2 es inactiva en el punto, luego $u_2 = 0$. Ahora, utilizando $\nabla \tilde{f}(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0$ se tiene que:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_1 \begin{pmatrix} 2 \cdot (1-1) \\ 2 \cdot (0-1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

entonces se obtienen las siguientes ecuaciones:

$$1 - 2u_1 = 0$$

Luego, la solución a esta ecuación es $u_1 = \frac{1}{2}$. Como existen u_1 y $u_2 \geq 0$ entonces se cumple la condición de KKT para el punto $(0, 1)$.