Optimización

Formulación matemática

Docente: Cristian Guarnizo Lemus

Contenido

1. Definición formal de la optimización
2. Conocimientos matemáticos.
3. Condiciones de optimalidad en problemas suaves sin restricciones.

Problema de optimización simple

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$ Función objetivo

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, \forall i \in E$$ Condiciones de igualdad

$$c_i(\boldsymbol{x}) \leq 0, \forall i \in I$$ Condiciones de desigualdad

Ejemplo:

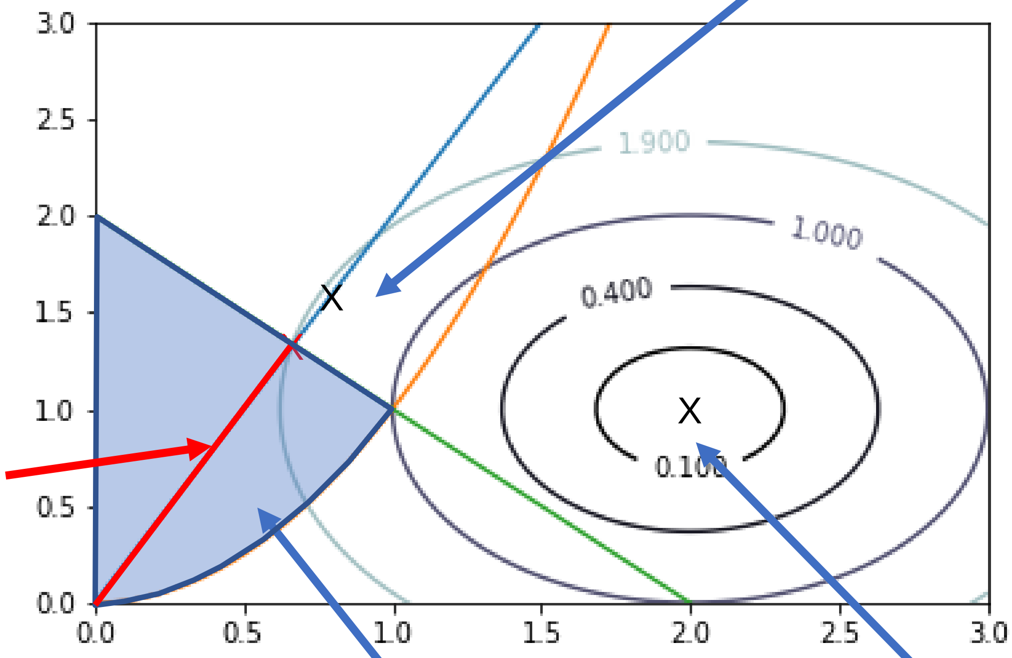
$$\min_{\boldsymbol{x}} (x_1 - 2)^2 + (x_2 - 1)^2$$

$$\text{s.t.} \quad x_2 - 2x_1 = 0$$

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

$$x_1 + x_2 \leq 2$$

Solución optima,
Condición de igualdad



Conjunto de soluciones
Factibles CI y CD

Conjunto de soluciones
de las restricciones de desigualdad

Solución optima
(sin restricciones)

Problema de optimización no-lineal

Formulación general:

$$\min_{\boldsymbol{x}\in D} f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, i \in E$$

$$c_i(\boldsymbol{x}) \leq 0, i \in I$$

$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^\intercal \in D \subseteq R^n$ un vector (punto)

D conjunto anfitrión

$f: D \rightarrow R$ función objetivo

$c_i: D \rightarrow R$ funciones de restricción $\forall i \in E \cup I$

E el conjunto índice de las restricciones de igualdad

I el conjunto índice de las restricciones de desigualdad

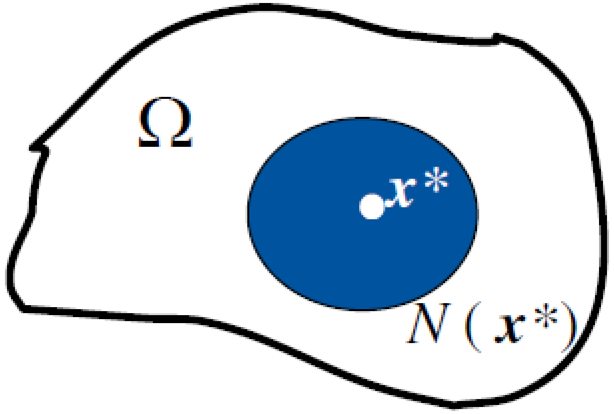
Las restricciones y el conjunto anfitrión define el conjunto factible, esto es, el conjunto de todas las soluciones factibles:

$$\Omega = \{x \in D | c_i(x) \leq 0 \ \forall i \in I, c_i(x) = 0 \ \forall i \in E\}$$

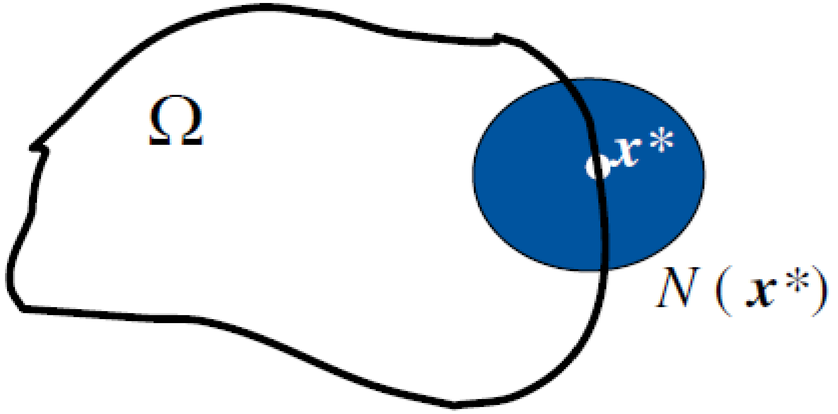
Formulación equivalente: $\boxed{\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})}$

Cual es una solución optima?

$$\boxed{\min_{x\in\Omega} f(x)}$$



Solución en el interior
de un conjunto factible



Solución en el perímetro
del conjunto factible

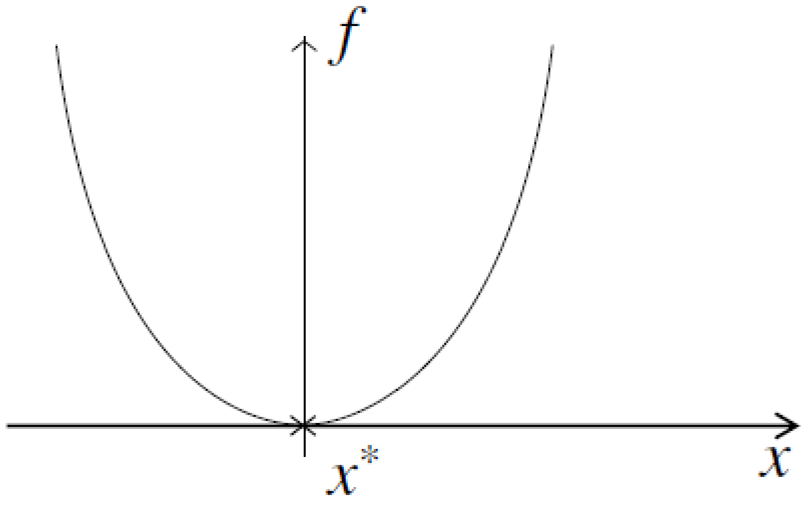
a) x^* es una solución local sí $x^* \in \Omega$ y una vecindad $N(x^*)$ de x^* existe: $f(x^*) \leq f(x) \forall x \in N(x^*) \cap \Omega$

b) x^* es una solución local estricta sí $x^* \in \Omega$ y una vecindad $N(x^*)$ de x^* existe: $f(x^*) < f(x) \forall x \in N(x^*) \cap \Omega$, $x \neq x^*$

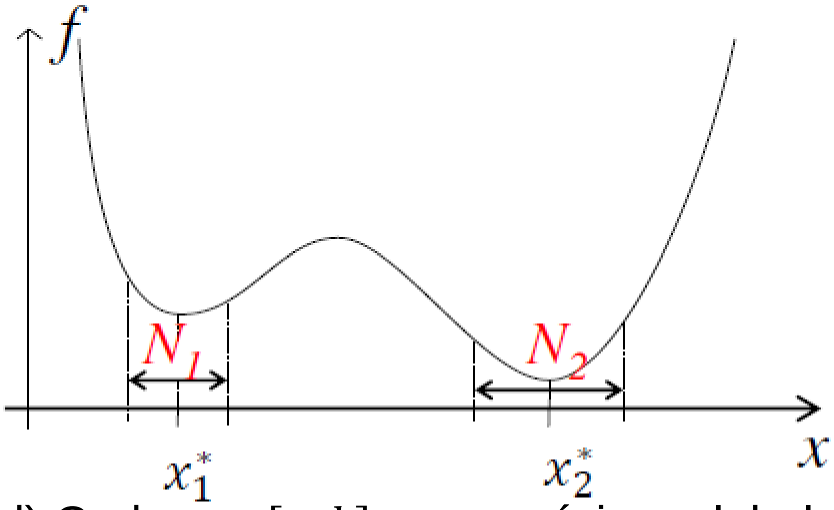
c) x^* es una solución global sí $x^* \in \Omega$ y $f(x^*) \leq f(x)$ $\forall x \in \Omega$

Solución optima: ejemplos

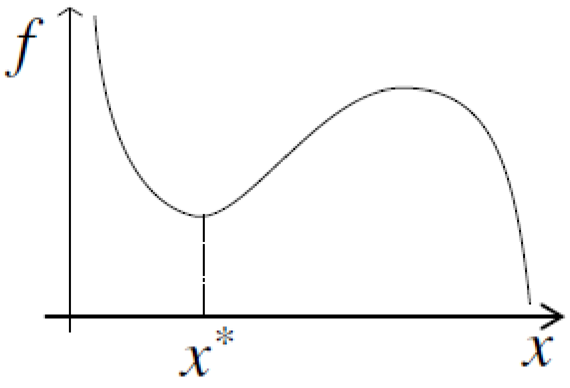
a) Mínimo global estricto



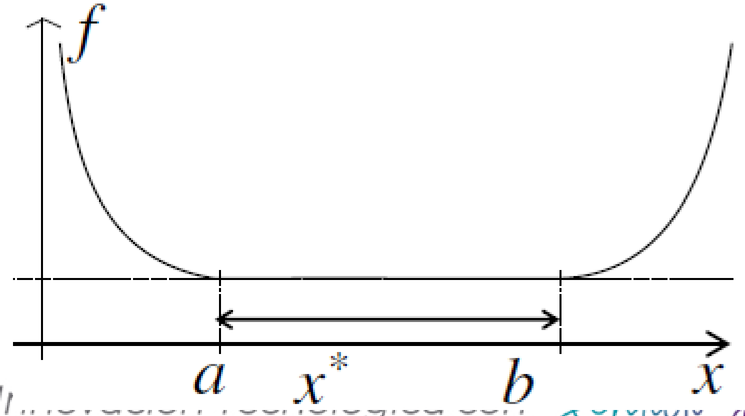
b) Dos mínimos locales estrictos, donde uno es un mínimo global estricto.



c) Un mínimo local estricto, no hay mínimo global.



d) Cada $x \in [a, b]$ es un mínimo global y local sin mínimo estricto.



Chequeo

- Escribir la definición general de un problema de optimización.
- Definición de una solución local y global de un problema de optimización?
- Son todas las soluciones locales también una solución global? Son todas las soluciones globales también una solución local?
- Que es el conjunto factible de un problema de optimización?
- Puede una solución ser el punto interior de un conjunto factible? Sobre su perímetro? Por fuera del conjunto factible? – Dibujar una figura explicativa.

Contenido

1. Definición formal de la optimización
2. Conocimientos matemáticos.
3. Condiciones de optimalidad en problemas suaves sin restricciones.

Problema de optimización no-lineal

Formulación general:

$$\min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, i \in E$$

$$c_i(\boldsymbol{x}) \leq 0, i \in I$$

$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^{\mathsf{T}} \in D \subseteq R^n$ un vector (punto)

D conjunto anfitrión

$f: D \rightarrow R$ función objetivo

$c_i: D \rightarrow R$ funciones de restricción $\forall i \in E \cup I$

E el conjunto índice de las restricciones de igualdad

I el conjunto índice de las restricciones de desigualdad

Las restricciones y el conjunto anfitrión define el conjunto factible, esto es, el conjunto de todas las soluciones factibles:

$$\Omega = \{x \in D | c_i(x) \leq 0 \ \forall i \in I, c_i(x) = 0 \ \forall i \in E\}$$

Formulación equivalente: $\boxed{\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})}$

Derivada direccional

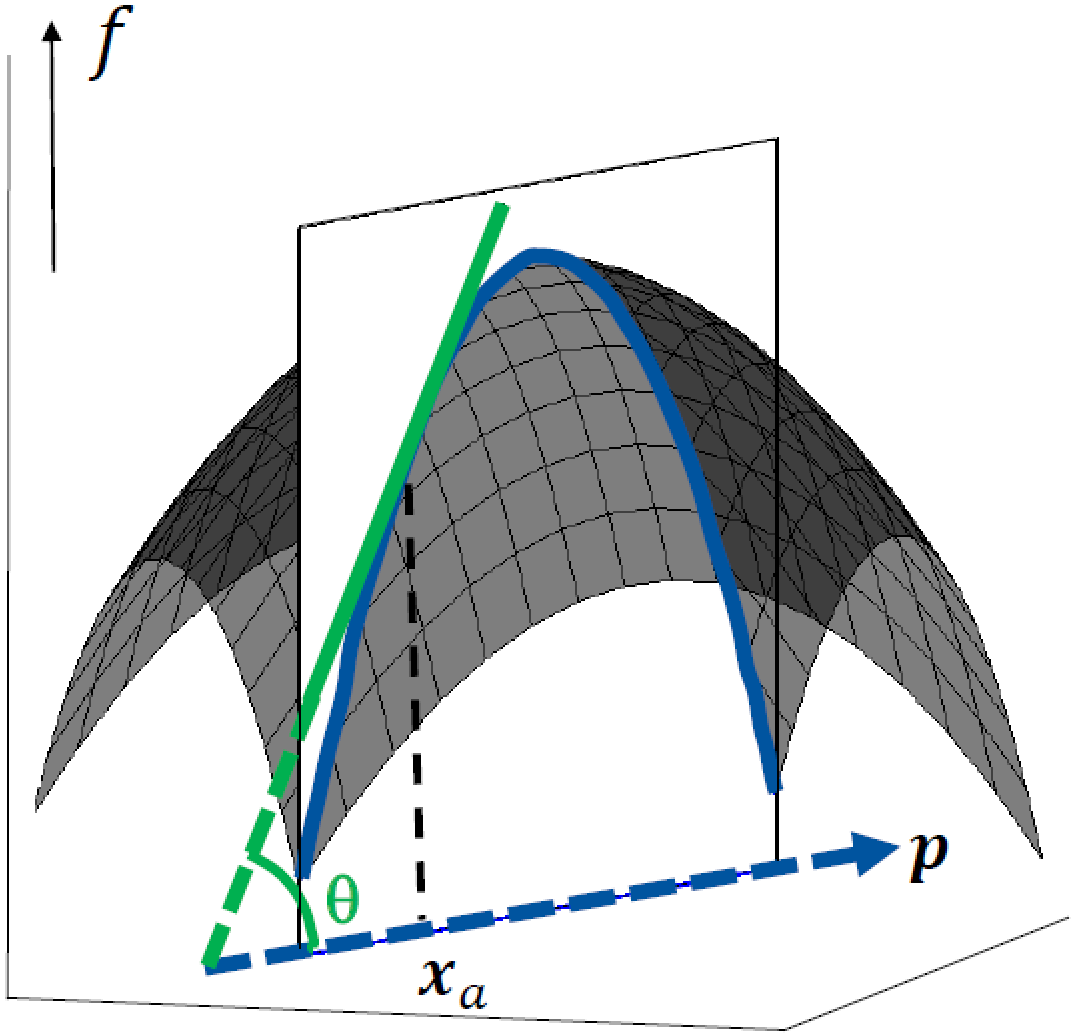
Definición:

Sea $f: D \to R,\ D \subseteq R^n, \boldsymbol{x} \in D$ y $\boldsymbol{p} \in R^n$ con $\|\boldsymbol{p}\| = 1$.

f es diferenciable en el punto $x = x_a$ en la dirección $\boldsymbol{p}$ si el limite,

$$D(f,\boldsymbol{p})|_{\boldsymbol{x}=\boldsymbol{x}_a} = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x}_a + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x}_a)}{\varepsilon} =: \boldsymbol{\nabla}_{\boldsymbol{p}} f(\boldsymbol{x}_a)$$

$D(f,\boldsymbol{p})$ es llamada la derivada direccional de f en la dirección $\boldsymbol{p}$.



Gradiente

Definición:

- La primera derivada de una función continua escalar es llamada gradiente de f en el punto $\boldsymbol{x}$:

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}} \\ \vdots \\ \left.\frac{\partial f}{\partial x_n}\right|_{\boldsymbol{x}} \end{bmatrix}.$$

Observaciones:

- Si $\boldsymbol{x}$ es función de t, se aplica la regla de la cadena:

$$\left.\frac{df}{dt}\right|_{\boldsymbol{x}(t)} = \nabla f(\boldsymbol{x})^{\mathsf{T}} \left.\frac{d\boldsymbol{x}}{dt}\right|_{t} = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}(t)} \left.\frac{\partial x_i}{\partial t}\right|_{t}.$$

- La derivada direccional esta relacionada con el gradiente:

$$D(f(\boldsymbol{x}), \boldsymbol{p}) = \nabla_{\boldsymbol{p}} f(\boldsymbol{x}) = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x} + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x})}{\varepsilon} = \nabla f(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{p}$$

Hessiana (matriz)

Definición:

- La segunda derivada de una función continua escalar diferenciable dos veces es una matriz simétrica llamada Hessiana, $H(\boldsymbol{x})$:

$$\boldsymbol{H}(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \left.\frac{\partial^2 f}{\partial x_1^2}\right|_{\boldsymbol{x}} & \cdots & \left.\frac{\partial^2 f}{\partial x_1 x_n}\right|_{\boldsymbol{x}} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial^2 f}{\partial x_1 x_n}\right|_{\boldsymbol{x}} & \cdots & \left.\frac{\partial^2 f}{\partial x_n^2}\right|_{\boldsymbol{x}} \end{bmatrix}.$$

T

Condiciones suficientes y necesarias

- Condición necesaria: Declaración A es una condición necesaria para la declaración B sí (y solo sí) la falsedad de A garantiza la falsedad de B. En notación matemática: $\text{not}\, A \Longrightarrow \text{not}\, B$.

- Condición suficiente: Declaración A es una condición suficiente para la declaración B si (y solo si) la verdad de A garantiza la verdad de B. En notación matemática: $A \Longrightarrow B$.

T

Condiciones suficientes y necesarias

- Si la declaración A es una condición necesaria para la declaración B, entonces B es una condición suficiente para la declaración A.
  - $\text{not } A \Longrightarrow \text{not } B$ implica $B \Longrightarrow A$

- Si la declaración A es una condición suficiente para la declaración B, entonces B es una condición necesaria para la declaración A.
  - $A \Longrightarrow B$ implica $\text{not } B \Longrightarrow \text{not } A$

- En optimización queremos verificar fácilmente las condiciones de un punto candidato
  - es un optimo local (condición suficiente para optimalidad es suficiente)
  - no es una condición optima (condición necesaria es violada)

T

Idealmente queremos condiciones que son necesarias y suficientes para un optimo local (o mejor aun para global)

Condiciones suficientes y necesarias

- En optimización queremos verificar fácilmente las condiciones de un punto candidato

  - es un optimo local (condición suficiente para optimalidad es suficiente)

  - no es una condición optima (condición necesaria es violada)

Idealmente queremos condiciones que son necesarias y suficientes para un optimo local (o mejor aun para global)

T

Condiciones suficientes y necesarias: ejemplos

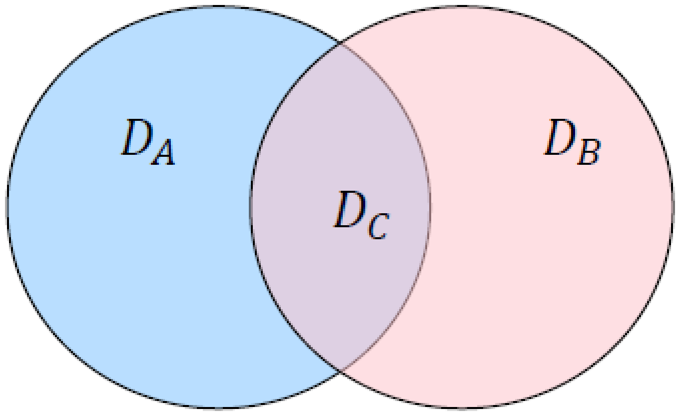
- Ejemplo simple: sea $x \in R$ y $y = x^2$. Declaración A “x es positivo” y declaración B “y es positivo”

  - A es suficiente para B.
    Prueba: A verdadero $\Leftrightarrow x > 0 \Rightarrow x^2 > 0 \Rightarrow y > 0 \Leftrightarrow B$ verdadero.
  - A no es necesario para B. Prueba: contra ejemplo
    $x = -1 \Rightarrow y = x^2 = 1$, entonces B es verdadero y A es falso.

T

Condiciones suficientes y necesarias: ejemplos

- Ejemplo con conjuntos: sea $D_A, D_B \subset R^n$ y $D_C = D_A \cap D_B$
  - Declaración A: $x \in D_A$
  - Declaración B: $x \in D_B$
  - Declaración C: $x \in D_C$
  - A es necesario para C, B es necesario para C.
  - C es suficiente para A, C es suficiente para B.
  - (A y B) son ambos necesarios y suficientes para C.



T

Chequeo

- Que funciones son continuas, diferenciables, continuas y diferenciables?
- Como se define la derivada direccional de una función? Como esta la derivada parcial relacionada con la derivada direccional?
- Cual es la definición del gradiente y la Hessiana de una función?

Contenido

1. Definición formal de la optimización
2. Conocimientos matemáticos.
3. Condiciones de optimalidad en problemas suaves sin restricciones.

Optimización sin restricciones

Problema de optimización sin restricciones:
Caso especial para el cual el conjunto factible $\Omega = R^n$

$$\min_{\boldsymbol{x} \in R^n} f(\boldsymbol{x})$$

- $\boldsymbol{x}^*$ es solución local sí $\boldsymbol{x}^* \in R^n$ y un vecindario $N(\boldsymbol{x}^*)$ de $\boldsymbol{x}^*$ existe:

$$f(\boldsymbol{x}^*) \leq f(x), \forall x \in N(\boldsymbol{x}^*)$$

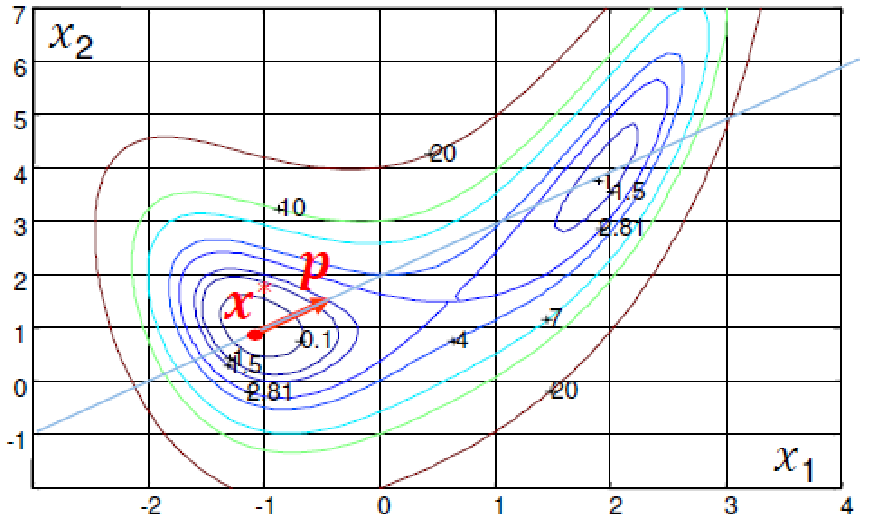
Queremos condiciones fáciles de evaluar
Necesaria: si $\boldsymbol{x}^*$ es optimo entonces las condiciones se satisfacen.
Suficiente: si la condición esta satisfecha entonces $\boldsymbol{x}^*$ es optimo.
Idealmente ambos son necesarios y suficientes!

Condiciones necesarias de primer orden

Teorema (Condiciones necesarias de primer orden):

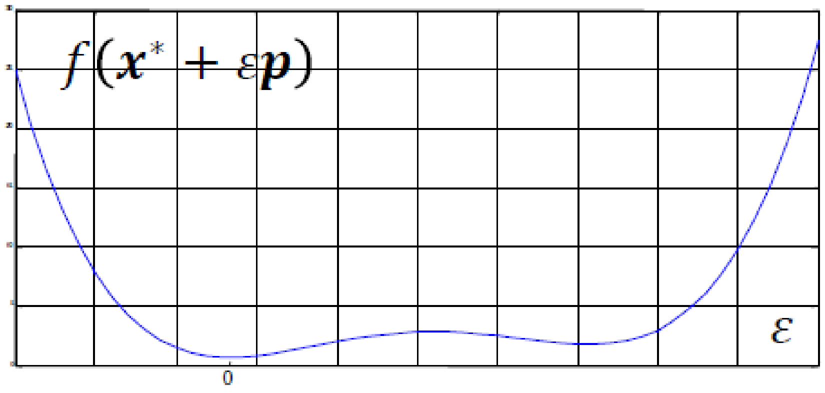
Sea f continuamente diferenciable y sea $\boldsymbol{x}^* \in R^n$ un minimizador local de f, entonces

$$\boxed{\nabla f(\boldsymbol{x}^*) = \mathbf{0}}$$



Prueba:

Como $\boldsymbol{x}^*$ es un minimizador local de f, para cada $\boldsymbol{p} \in R^n$, existe un $\tau > 0$, tal que $f(\boldsymbol{x}^* + \varepsilon \boldsymbol{p}) \geq f(\boldsymbol{x}^*) \forall \varepsilon \in [0, \tau]$



By the definition of the directional derivative:

$$\nabla_{\boldsymbol{p}} f(\boldsymbol{x}^*) = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x}^* + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x}^*)}{\varepsilon} = \nabla f(\boldsymbol{x}^*)^T \boldsymbol{p} \geq 0 \quad (1)$$

The special choice, $\boldsymbol{p} = -\nabla f(\boldsymbol{x}^*)$, leads to

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{p} = -\nabla f(\boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) = -\|\nabla f(\boldsymbol{x}^*)\|^2 \leq 0 \text{ (norm property)} \quad (2)$$

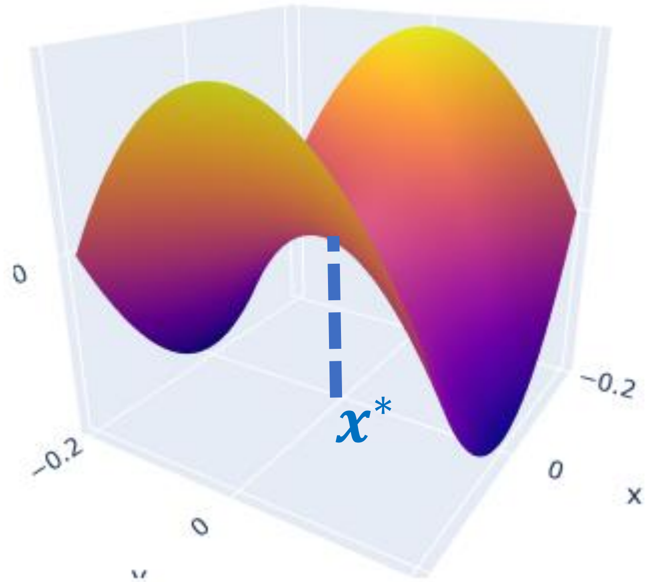
(1) and (2) $\Rightarrow \nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

Puntos estacionarios

- Sea f continuamente diferenciable y $\boldsymbol{x} \in R^n$. Si $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ se mantiene, entonces $\boldsymbol{x}^*$ es llamado un punto estacionario de f.
- Esta condición es necesaria, pero no una condición suficiente para un mínimo local.
- Ejemplo: $f(x) = -x^2$ posee un solo punto estacionario en $x^* = 0$, debido que $\nabla f(x^*) = -2x^* = 0$. Este punto no es un mínimo sino un máximo global.

Punto de ensilladura

- Un punto estacionario no tiene que ser un mínimo o un máximo. Este punto estacionario es llamado un punto de ensilladura (saddle point).
- Ejemplo: el gradiente de $f(\boldsymbol{x}) = x_1^2 - x_2^2$ es $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^{\mathsf{T}}$. Entonces, $\boldsymbol{x}^* = \mathbf{0}$ es el único punto estacionario. Como f es curvada positiva en la dirección x_1 y negativamente curvada en la dirección x_2, $\boldsymbol{x}^*$ es un punto de ensilladura.



Condiciones necesarias de segundo orden

Teorema (condiciones necesarias de segundo orden):
Sea f dos veces diferenciable y sea $\boldsymbol{x}^* \in R^n$ un minimizador local de f, entonces

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$,
2. $\nabla^2 f(\boldsymbol{x}^*)$ es semidefinida positiva.

Estas condiciones son solamente necesarias pero no suficientes

- El único punto estacionario de $f(x) = x^3$ es $x^* = 0$, con $\nabla f(0) = 0$, $\nabla^2 f(0) = 0$. Se cumplen ambas condiciones. $x^* = 0$ no es un mínimo local pero si es un punto de ensilladura.
- El único punto estacionario de $f(x) = -x^4$ es $x^* = 0$, con $\nabla f(0) = 0$, $\nabla^2 f(0) = 0$. Se satisfacen las dos condiciones. Pero $x^* = 0$ no es un minimo local sino un máximo local.

Condiciones necesarias de segundo orden: demostración

Sea f dos veces diferenciable y sea $\boldsymbol{x}^* \in R^n$ un minimizador local de f.

Asumir $\nabla^2 f(\boldsymbol{x}^*)$ no es semidefinida positiva.

Entonces, $\exists \boldsymbol{p} \in R^n$: $\boldsymbol{p}^{\mathsf{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{p} < 0$

La expansión de Taylor en $\boldsymbol{x}^*$ nos da

$$f(\boldsymbol{x}^* + \epsilon \boldsymbol{p}) = f(\boldsymbol{x}^*) + \epsilon \nabla f(\boldsymbol{x}^*)^{\mathsf{T}} \boldsymbol{p} + \frac{1}{2} \epsilon^2 \boldsymbol{p}^{\mathsf{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{p} + O(\epsilon^3).$$

$\boldsymbol{x}^*$ es un mínimo local y entonces la condición necesaria de primer orden

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

Para ϵ lo suficientemente pequeño, $O(\epsilon^2)$ domina sobre $O(\epsilon^3)$. Debido que $\boldsymbol{p}^{\mathsf{T}} \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{p} < 0 \rightarrow f(\boldsymbol{x}^* + \epsilon \boldsymbol{p}) < f(\boldsymbol{x}^*)$

$\boldsymbol{x}^*$ no es un mínimo local.

Condiciones suficiente para optimalidad

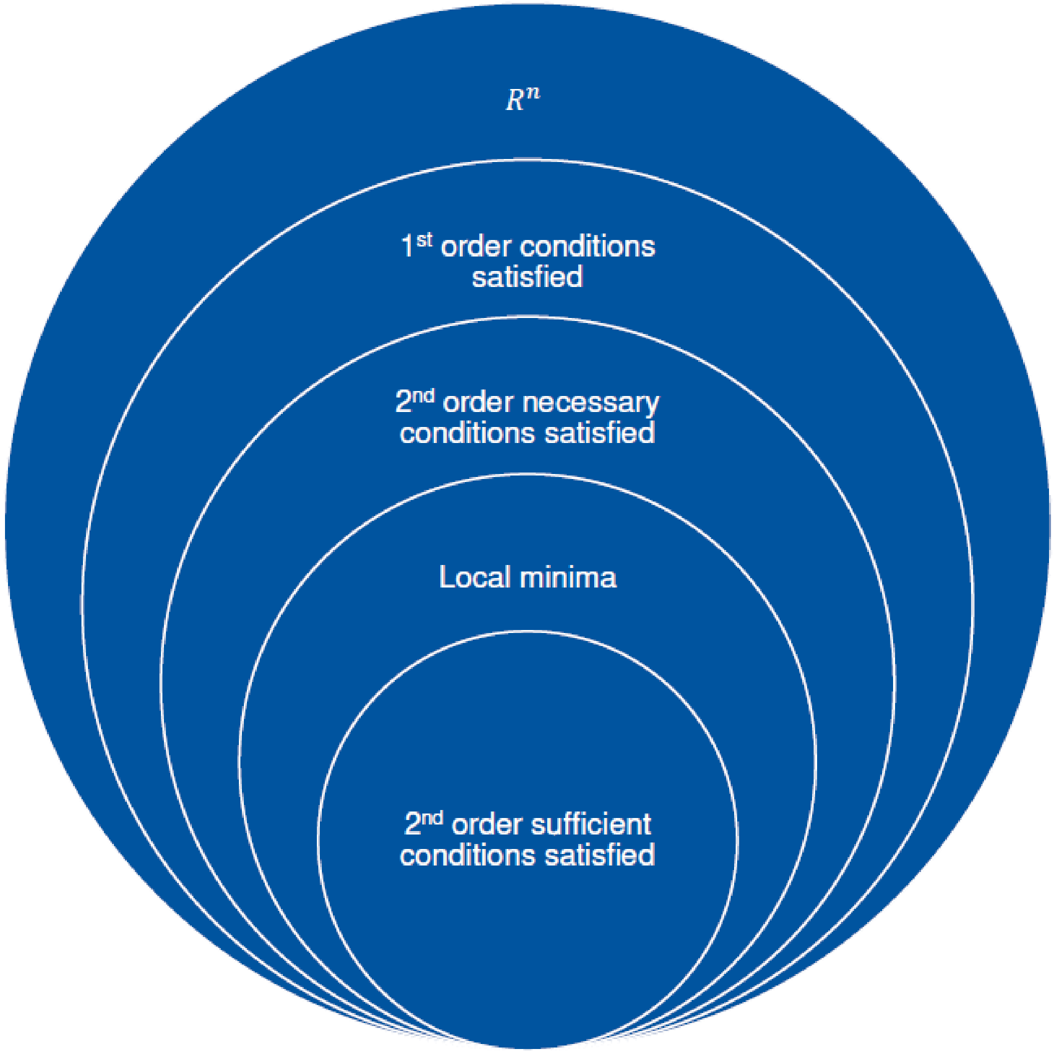
Teorema (condiciones suficientes para optimalidad):
Sea f dos veces diferenciable y sea $\boldsymbol{x}^* \in R^n$, sí

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$,
2. $\nabla^2 f(\boldsymbol{x}^*)$ es definida positiva.

Entonces $\boldsymbol{x}^*$ es un minimizador local estricto de f.
Observación:

- $f(x) = x^4$ tiene en $x = 0$ su (único) mínimo global estricto. Además, $\nabla f(0) = 0$ y $\nabla^2 f(0) = 0$ se mantienen, entonces la segunda condición del teorema se viola.
- Por eso, las condiciones mencionadas en el teorema son suficientes pero no necesarias.

Condiciones de optimalidad para problemas suaves



- Condiciones de optimalidad están en un punto, no para todo R^n.
- Todos los conjuntos mostrados son subconjuntos.
- Las condiciones necesarias de primer orden excluyen los puntos no estacionarios.
- Las condiciones necesarias de segundo orden excluyen algunos puntos y algunos máximos locales, pero no todos.

Institución Universitaria
Reacreditada en Alta Calidad

Chequeo

- Que es un punto estacionario? Existen diferentes tipos de puntos estacionarios?
- Cuales son las condiciones necesarias de primer orden para problemas suaves sin restricciones?
- Cuales son las condiciones necesarias de segundo orden para problemas suaves sin restricciones?
- Cuales son las condiciones suficientes de segundo orden para problemas suaves sin restricciones?

Institución
Universitaria
Reacreditada en Alta Calidad

¡Gracias!

Somos Innovación Tecnológica con Sentido Humano

Alcaldía de Medellín