Optimización

Optimización con restricciones

Docente: Cristian Guarnizo Lemus

Contenido

Optimización con restricciones

Formulación general:

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

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

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

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

D conjunto anfitrión, $D = R^n$ para tratar las restricciones.

$f: D \rightarrow R$ 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: $\boxed{\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})}$

Recordar: Condiciones para optimalidad

Condiciones de optimalidad para Problemas sin restricciones:

Condiciones necesarias:

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

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

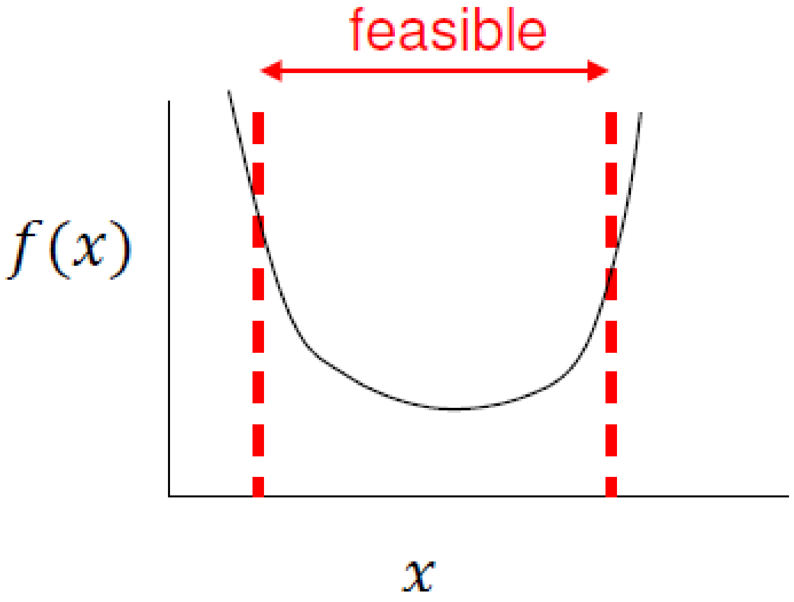
Condiciones suficientes:

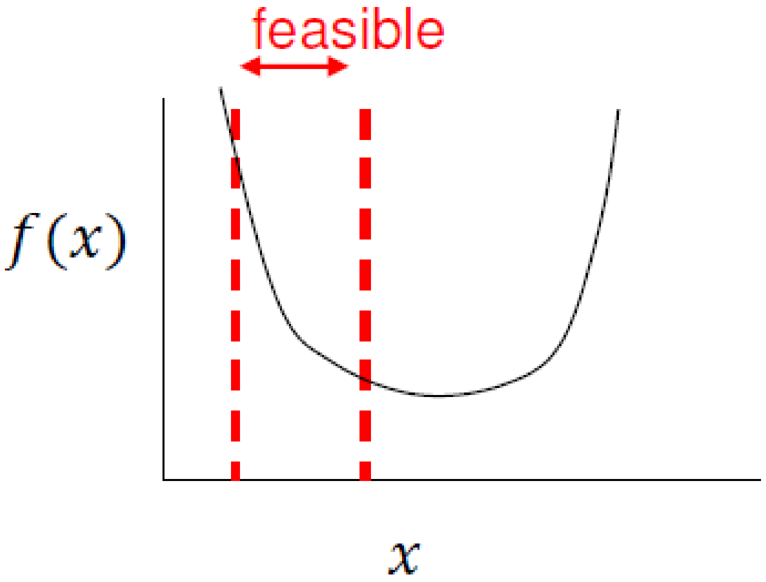
Sea f dos veces diferenciable y sea $\boldsymbol{x}^* \in R^n$, sí

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

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

Sin restricciones y Optimo local restringido





Que condiciones se deben aplicar en este caso?

Contenido

1. Optimización con restricciones.
2. Condiciones de optimalidad para igualdades.
3. Condiciones de optimalidad para desigualdades.
4. Condiciones de optimalidad.

Institución Universitaria
Reacreditada en Alta Calidad

Optimización con restricciones de igualdad

Formulación general:

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

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

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

D conjunto anfitrión, $D = R^n$ para tratar las restricciones.

$f: D \to R$ función objetivo

$c_i: D \to R$ funciones de restricción $\forall i \in E$

E el conjunto índice de las restricciones de igualdad

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

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

Restricción de igualdad: ejemplo

Problema:

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

$$\text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0 \qquad \leftarrow c(\boldsymbol{x})$$

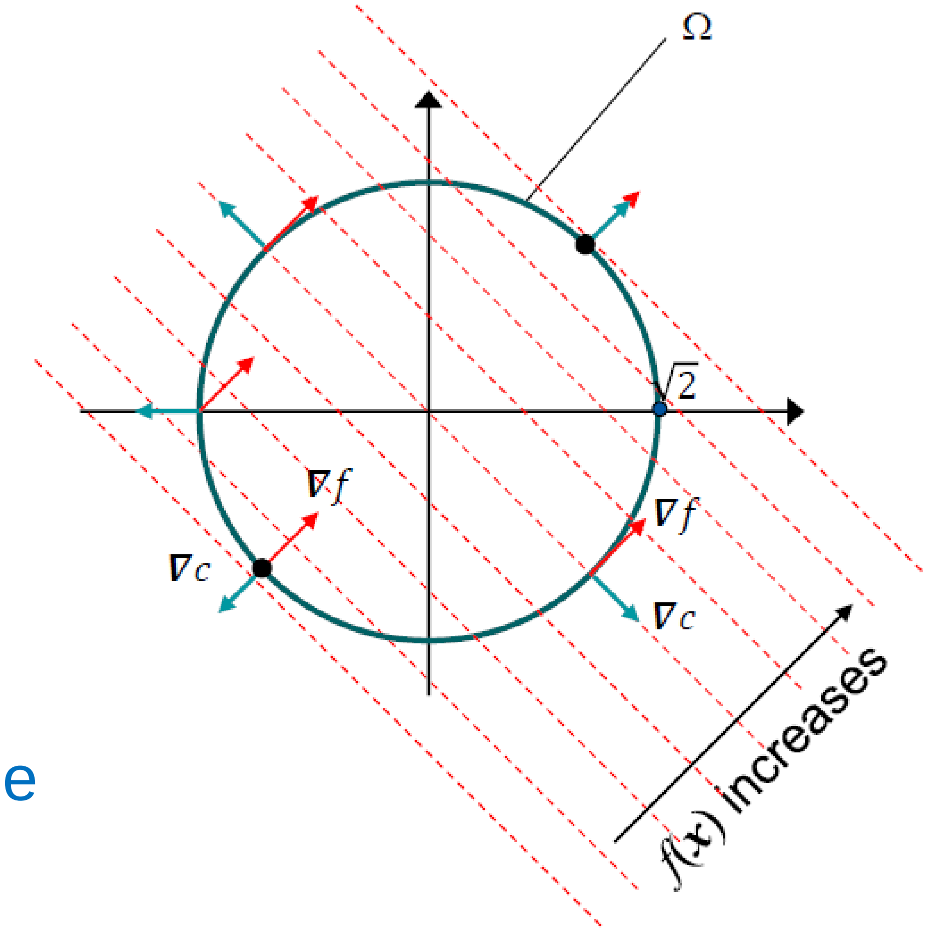
Solución:

$$\boldsymbol{x}^* = (-1, -1)^T$$

$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla c(\boldsymbol{x}^*) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

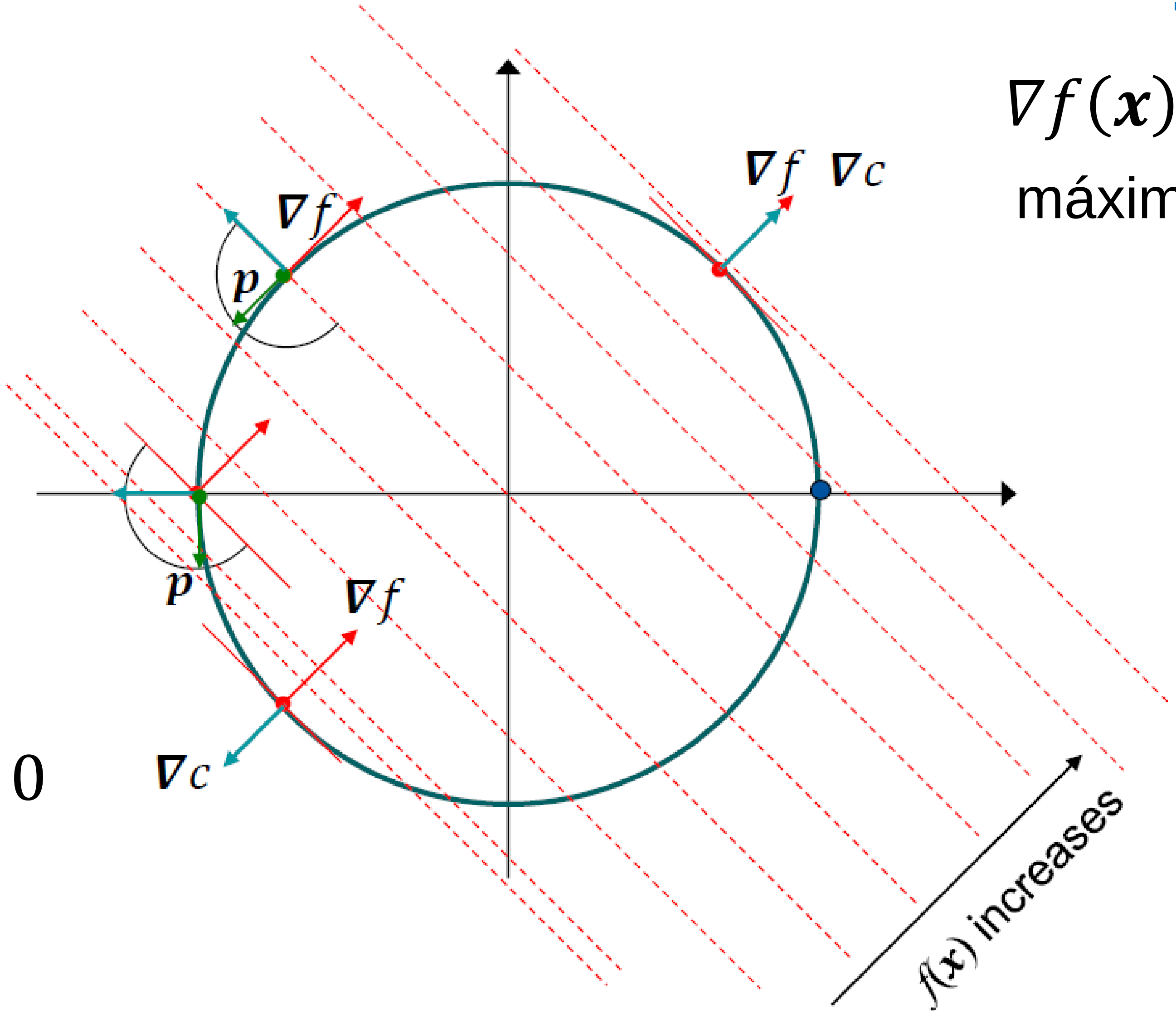
Que condiciones se satisfacen en el mínimo?



Restricción de igualdad: ejemplo

$$\nabla c(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{p} = 0$$
$$\nabla f(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{p} < 0$$

No mínimo



$$\nabla f(\boldsymbol{x}) + \lambda\nabla c(\boldsymbol{x}) = 0$$

máximo

$$\nabla f(\boldsymbol{x}) + \lambda\nabla c(\boldsymbol{x}) = 0$$

mínimo

Restricción de igualdad: Condiciones generales de optimalidad

- Defina la función Lagrangiana: $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda c(\boldsymbol{x})$
- Puntos estacionarios de L (mas precisamente puntos de ensilladura) corresponden a puntos estacionarios del problema con restricciones

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

$$\text{s.t.} \quad c(\boldsymbol{x}) = 0$$

- Condiciones de optimalidad necesaria de primer orden:

$$\boldsymbol{\nabla} L(\boldsymbol{x}, \lambda) = \begin{bmatrix} \boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) \\ \nabla_{\lambda} L(\boldsymbol{x}, \lambda) \end{bmatrix} = \mathbf{0} \quad \begin{cases} \boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = \boldsymbol{\nabla} f(\boldsymbol{x}) + \lambda \boldsymbol{\nabla} c(\boldsymbol{x}) = \mathbf{0} & \text{(Estacionario)} \\ \nabla_{\lambda} L(\boldsymbol{x}, \lambda) = c(\boldsymbol{x}) = 0 & \text{(factibilidad primaria)} \end{cases}$$

- Múltiples restricciones: multiplicador de Lagrange λ_i para cada restricción c_i, sumados sobre i: $\nabla f(\boldsymbol{x}) + \sum_i \lambda_i \nabla c_i(\boldsymbol{x}) = \mathbf{0}$ e imponen la factibilidad primaria $\forall i$: $c_i(\boldsymbol{x}) = 0$.

Restricción de igualdad: aplicación de condiciones

Problema:

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

$$\text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0 \qquad \leftarrow c(\boldsymbol{x})$$

Solución:

Función Lagrangiana: $L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda c(\boldsymbol{x}) = (x_1 + x_2) + \lambda(x_1^2 + x_2^2 - 2)$

Condiciones de Optimalidad necesarias de primer orden:

$$\begin{matrix} \nabla_{\boldsymbol{x}} L(\boldsymbol{x},\lambda) = 0 \\ \nabla_{\lambda} L(\boldsymbol{x},\lambda) = 0 \end{matrix} \Leftrightarrow \begin{cases} \nabla f(\boldsymbol{x}) + \lambda \nabla c(\boldsymbol{x}) = \boldsymbol{0} \\ c(\boldsymbol{x}) = 0 \end{cases} \Leftrightarrow \begin{cases} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ x_1^2 + x_2^2 = 2 \end{cases}$$

Dos soluciones $\begin{cases} \min\text{: } \lambda^* = 0.5,\ \boldsymbol{x}^* = (-1,-1) \\ \max\text{: } \lambda^* = -0.5,\ \boldsymbol{x}^* = (1,1) \end{cases}$

Signo del multiplicador de Lagrange

- En general no se puede concluir sí λ es negativo o positivo para el mínimo. El signo de c, de ∇c y λ es arbitrario

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

$$\text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0 \qquad \leftarrow c(\boldsymbol{x})$$

Es equivalente a

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

$$\text{s.t.} \quad -x_1^2 - x_2^2 + 2 = 0 \qquad \leftarrow c(\boldsymbol{x})$$

pero tienen signos contrarios para c y ∇c, y para λ también.

Contenido

1. Optimización con restricciones.
2. Condiciones de optimalidad para igualdades.
3. Condiciones de optimalidad para desigualdades.
4. Condiciones de optimalidad.

Optimización con restricciones de desigualdad

Formulación general:

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

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

$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^\top \in D$ un vector (punto n-dimensional)

D conjunto anfitrión, $D = R^n$ para tratar las restricciones.

$f: D \rightarrow R$ función objetivo

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

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:

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

Restricción de deigualdad: ejemplo

Problema:

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

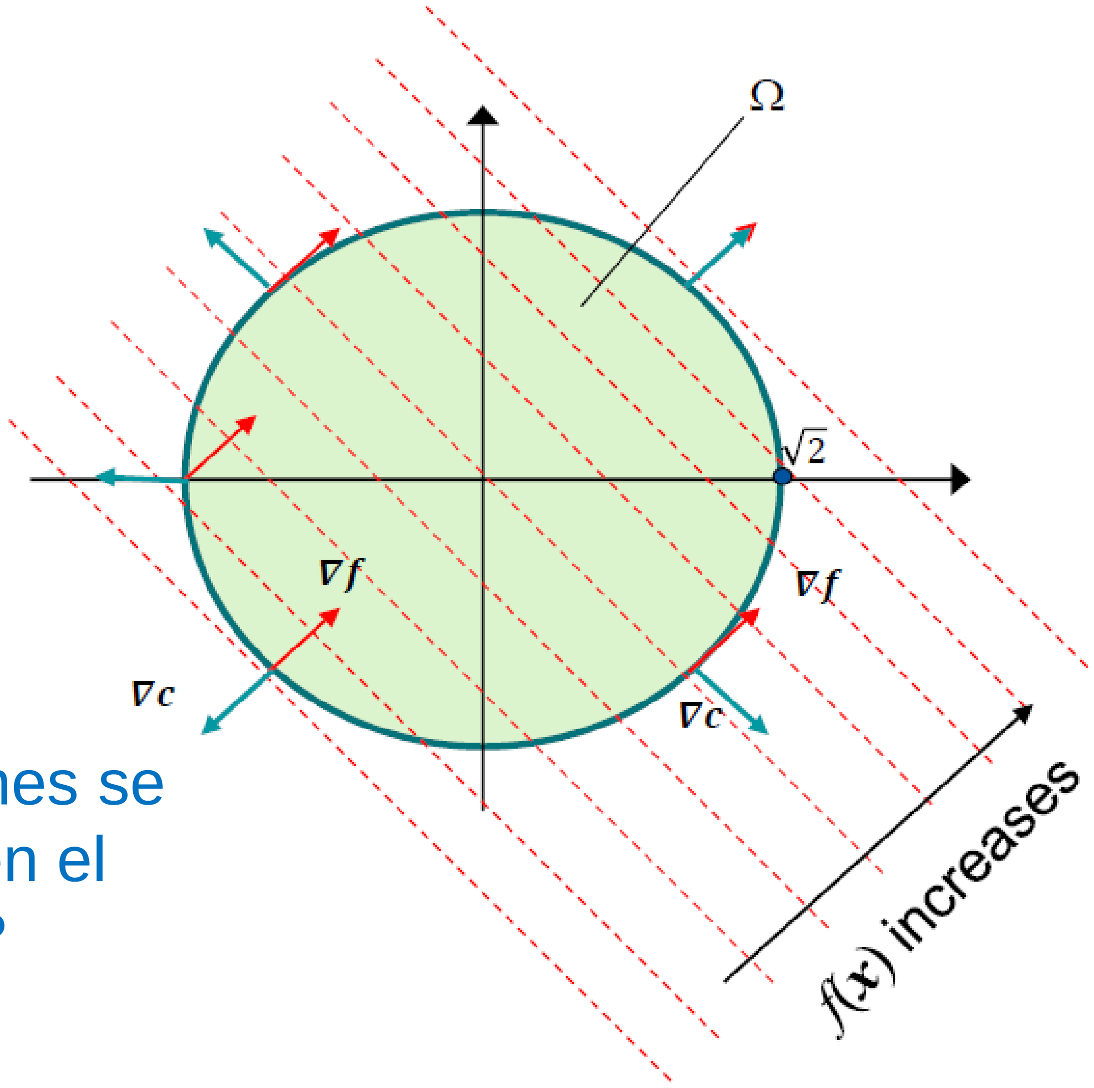
$$\text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0 \qquad \leftarrow c(\boldsymbol{x})$$

Solución:

$$\boldsymbol{x}^* = (-1, -1)^T$$

$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla c(\boldsymbol{x}^*) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$



Que condiciones se satisfacen en el mínimo?

Restricción de desigualdad: derivación de condiciones

El punto actual factible $\boldsymbol{x}$ no es optimo, se puede encontrar un $\boldsymbol{p}$, tal que

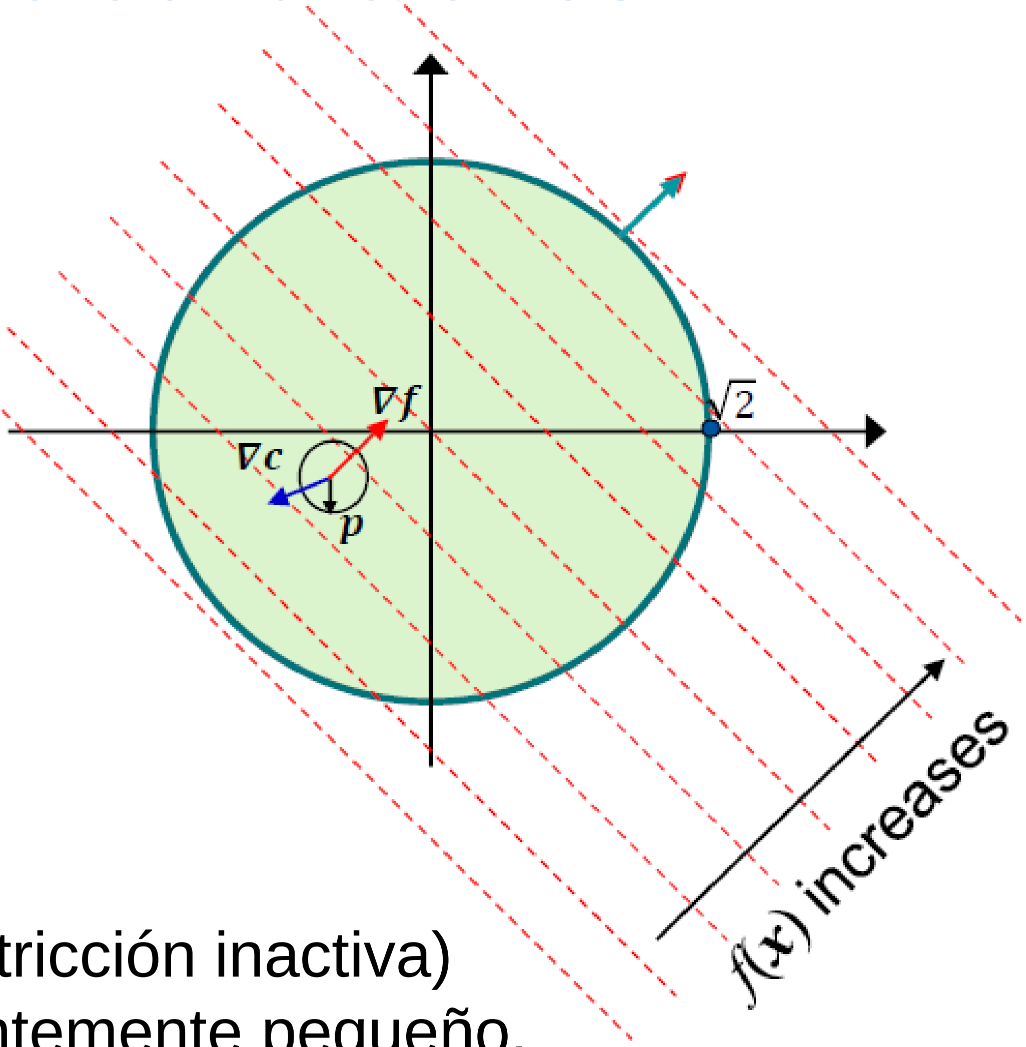
- Se retiene la factibilidad.
- f se puede disminuir.

Una disminución de f se logra solo en la dirección descendente $\boldsymbol{p}$ tal que $\nabla f(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{p} < 0 \quad (1)$

La restricción de desigualdad debe mantener la factibilidad, esto es:

$$0 \geq c(\boldsymbol{x}+\boldsymbol{p}) \approx c(\boldsymbol{x}) + \nabla c(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{p}$$

$$\Rightarrow c(\boldsymbol{x}) + \nabla c(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{p} \leq 0 \qquad (2)$$



Dos casos:

1. Punto cae en el interior del disco $c(\boldsymbol{x}) < 0$ (restricción inactiva)
   $\Rightarrow$ (2) siempre se satisface para un $\|\boldsymbol{p}\|$ suficientemente pequeño.
   (1) también se satisface a menos que $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$.

Restricción de desigualdad: derivación de condiciones

Dos casos:

1. Punto cae en el interior del disco $c(\boldsymbol{x}) < 0$ (restricción inactiva)

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

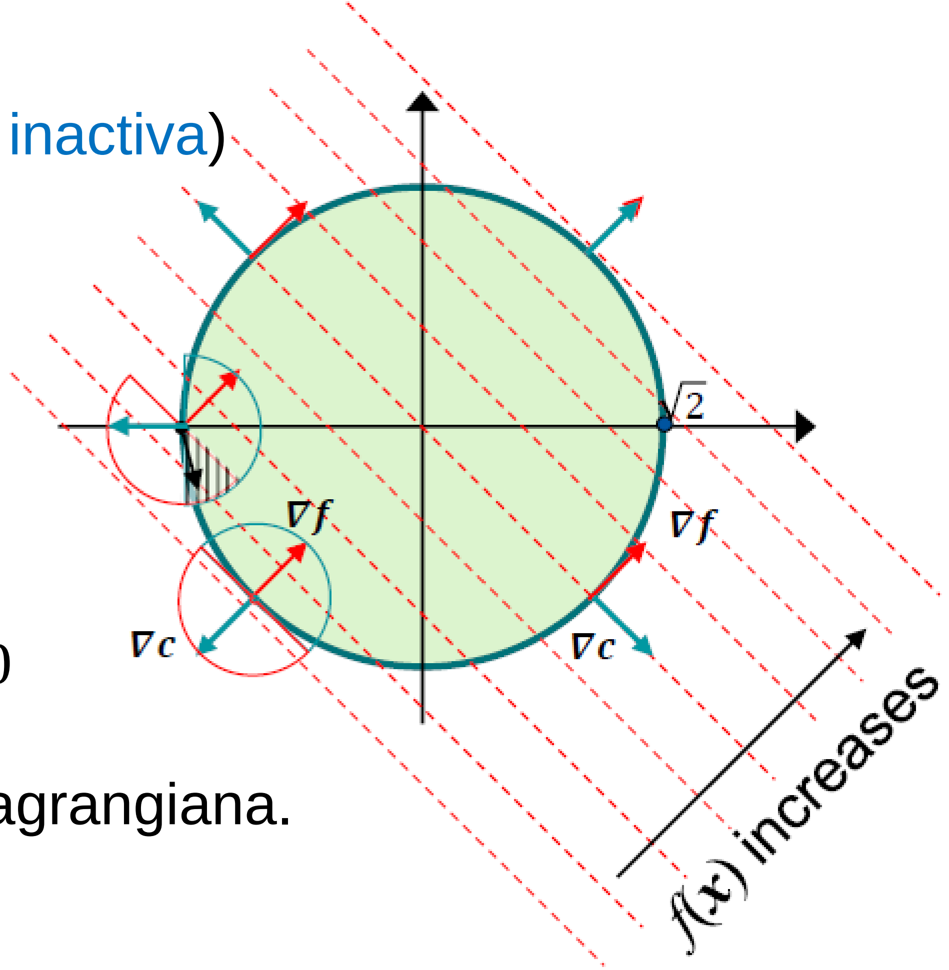
2. Punto cae en el circulo $c(\boldsymbol{x}) = 0$ (restricción activa).

   Dirección descendente $\boldsymbol{p}$: $\nabla f(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{p} < 0$ (1)

   Factibilidad: $c(\boldsymbol{x}) + \nabla c(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{p} \leq 0$

   $\Longrightarrow\ \nabla c(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{p} \leq 0$ (2)

   $\Longrightarrow$ No se puede descender sí $\nabla f(\boldsymbol{x}) + \lambda \nabla c(\boldsymbol{x}) = \boldsymbol{0}$, $\lambda \geq 0$



Ambos casos se pueden caracterizar usando la función Lagrangiana.

Restricción de igualdad: Condiciones de optimalidad

Ejemplo:

$$\min_{\boldsymbol{x}} \quad x_1 + x_2 \qquad \leftarrow f(\boldsymbol{x})$$

$$\text{s.t.} \; x_1^2 + x_2^2 - 2 \leq 0 \qquad \leftarrow c(\boldsymbol{x})$$

Solución:

Función Lagrangiana: $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda c(\boldsymbol{x}) = (x_1 + x_2) + \lambda(x_1^2 + x_2^2 - 2)$

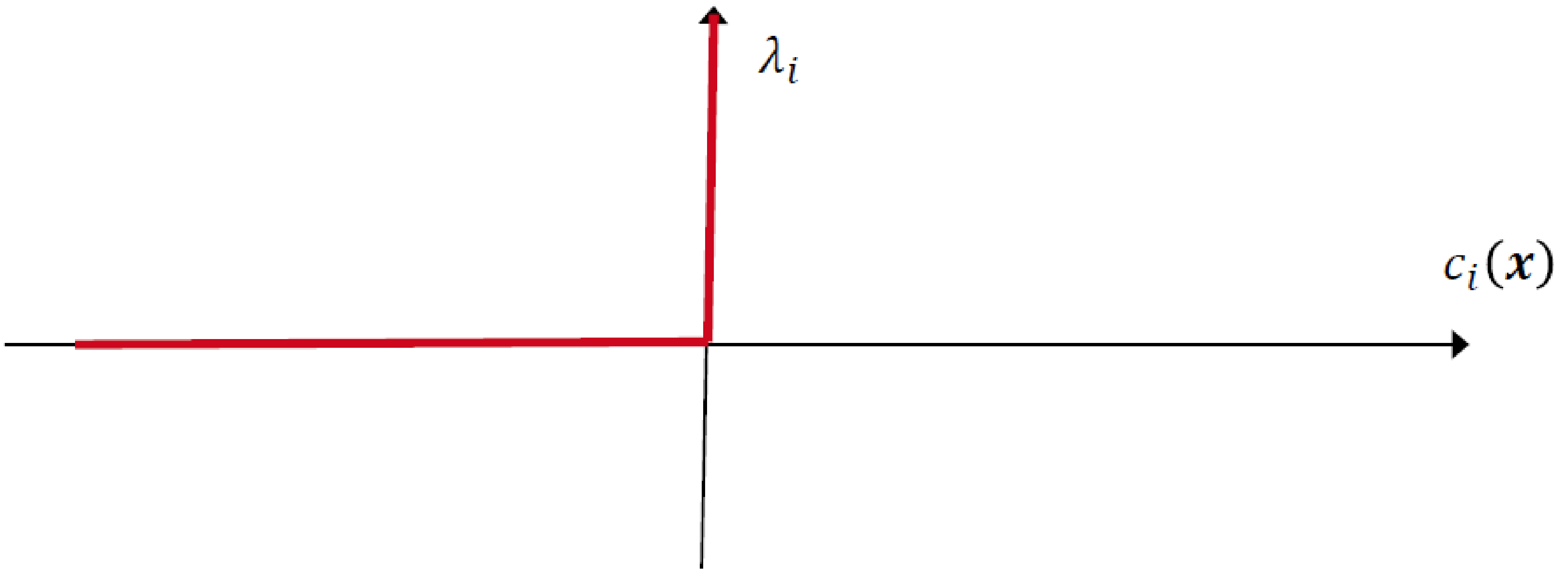
Condiciones de optimalidad de primer orden:

$$\begin{cases} \nabla f(\boldsymbol{x}) + \lambda \nabla c(\boldsymbol{x}) = \boldsymbol{0} & \text{(Estacionario)} \\ c(x) \leq 0 & \text{(factibilidad primaria)} \\ \lambda c(x) = 0 & \text{(Condición complementaria)} \\ \lambda \geq 0 & \text{(factibilidad dual)} \end{cases}$$

- Restricciones múltiples: multiplicador de Lagrange λ_i para cada restricción c_i, sumando sobre i: $\nabla f(\boldsymbol{x}) + \sum_i \lambda_i \nabla c_i(\boldsymbol{x}) = \boldsymbol{0}$ e imponen otras condiciones $\forall i$: $c_i(\boldsymbol{x}) \leq 0, \lambda_i c_i(\boldsymbol{x}) = 0, \lambda_i \geq 0$

(Complementary slackness) Holgura complementaria de desigualdades

- La holgura complementaria es engañosa solamente $\lambda_i c_i(\boldsymbol{x}) = 0, i \in I$
- Junto con la factibilidad primaria y dual, CS es equivalente a una declaración if-then-else
  - Restricción inactiva: $c_i(\boldsymbol{x}) < 0 \implies \lambda_i = 0$
  - Restricción activa: $c_i(\boldsymbol{x}) = 0 \implies \lambda_i \geq 0$
  - Conjunto activo $A(\boldsymbol{x})$: conjunto de desigualdades activas, $A(\boldsymbol{x}) = \{i \in I | c_i(\boldsymbol{x}) = 0\}$
- En algunos casos CS no es una restricción suave: conjunto factible es una esquina sin interior!



Contenido

Optimización con restricciones

Formulación general:

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

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

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

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

D conjunto anfitrión, $D = R^n$ para tratar las restricciones.

$f: D \to R$ función objetivo

$c_i: D \to R$ 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: $\boxed{\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})}$

Definiciones: condiciones de primer orden

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

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

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

Definición de conjunto activo:

El conjunto activo $A(\boldsymbol{x})$: conjunto de desigualdades activas y todas las igualdades

$$A(\boldsymbol{x}) = E \cup \{i \in I | c_i(\boldsymbol{x}) = 0\}$$

Función Lagrangiana:

$$L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \sum_{i\in E} \lambda_i c_i(\boldsymbol{x}) + \sum_{i\in I} \lambda_i c_i(\boldsymbol{x})$$

$$= f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{c}(\boldsymbol{x})$$

Condiciones de primer orden necesarias: Condiciones KKT

Teorema de Karush-Kuhn-Tucker (KKT):

Sea $\boldsymbol{x}^* \in R^n$ un minimizador local. Asuma que $\nabla c_i(\boldsymbol{x}^*), i \in A(\boldsymbol{x}^*)$ son linealmente independientes. Entonces, existen multiplicadores de Lagrange, $\lambda_i^*, i \in E \cup I$:

(1) $\nabla f(\boldsymbol{x}^*) + \sum_{i \in E \cup I} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = \boldsymbol{0}$ ← Estacionariedad

(2) $c_i(\boldsymbol{x}^*) = 0, \ \forall i \in E$ ← Factibilidad primaria

(3) $c_i(\boldsymbol{x}^*) \leq 0, \ \forall i \in I$ ← Factibilidad primaria

(4) $\lambda_i^* \geq 0, \ \forall i \in I$ ← Factibilidad dual

(5) $\lambda_i^* c_i(\boldsymbol{x}^*) = 0, \ \forall i \in I$ ← Holgura complementaria (CS)

Observaciones:

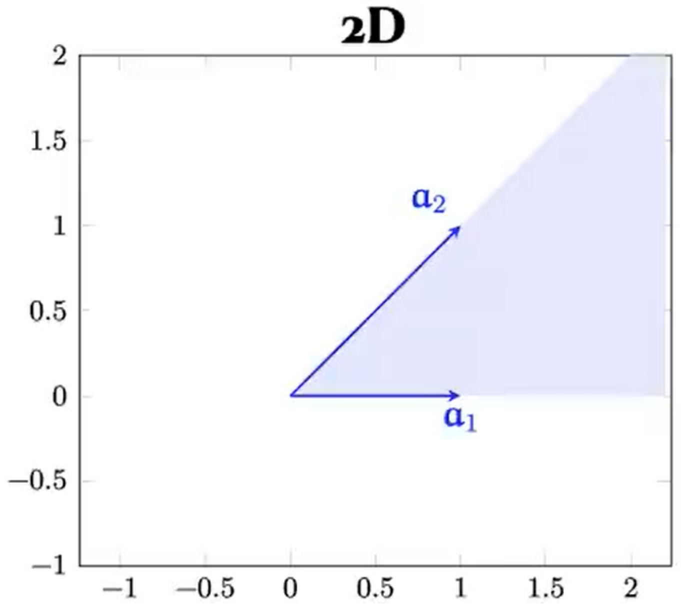
- De (5) y las restricciones inactivas ($c_i(\boldsymbol{x}^*) < 0$) los multiplicadores de Lagrange son $\lambda_i^* = 0$. Entonces, (1) se puede escribir como $\boldsymbol{0} = \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*)$
- (5) se satisface también para las igualdades debido que $c_i(\boldsymbol{x}^*) = 0$ dado (2)
- (1) implica que $\nabla f(\boldsymbol{x}^*)$ es una combinación lineal de $\nabla c_i(\boldsymbol{x}^*)$, $i \in A(\boldsymbol{x}^*)$. Geometría: $\nabla f(\boldsymbol{x}^*)$ cae en el cono de $\nabla c_i(\boldsymbol{x}^*)$.
- Existen condiciones suficientes de segundo orden.

Cono convexo - revisar

Convex cone For any given real $m \times n$ matrix A, denote the columns of A by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$, then $A\boldsymbol{x} = \boldsymbol{b}$ means $\boldsymbol{b}$ is a linear combination of the columns of A, that is,

$$\boldsymbol{b} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n.$$

When we insist that all the coefficients $x_i \geq 0$, then the set of all such linear combinations is a **convex cone** of the vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$.



$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Condiciones de optimalidad para desigualdades ≥

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0,\ i \in E \\ & \boxed{c_i(\boldsymbol{x}) \leq 0,\ i \in I} \\ & E = \{1, \dots, n_E\} \\ & I = \{n_E + 1, \dots, n_E + n_I\} \end{aligned}$$

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0,\ i \in E \\ & \boxed{c_i(\boldsymbol{x}) \geq 0,\ i \in I} \\ & E = \{1, \dots, n_E\} \\ & I = \{n_E + 1, \dots, n_E + n_I\} \end{aligned}$$

Funciones Lagrangianas

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}) &= f(\boldsymbol{x}) + \sum_{i \in E \cup I} \lambda_i c_i(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{c}(\boldsymbol{x}) \end{aligned}$$

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}) &= f(\boldsymbol{x}) - \sum_{i \in E \cup I} \lambda_i c_i(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \boldsymbol{c}(\boldsymbol{x}) \end{aligned}$$

El teorema KKT es valido para ambas, solo una diferencia

(3) $c_i(\boldsymbol{x}^*) \leq 0, \forall i \in I$

(3) $c_i(\boldsymbol{x}^*) \geq 0, \forall i \in I$

Multiplicadores de Lagrange y la sensibilidad

- Restricciones activas: $i \in A(\boldsymbol{x}^*)$

Estacionariedad: $\mathbf{0} = \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i \in A(\boldsymbol{x}^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*)$:

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \approx (\boldsymbol{x} - \boldsymbol{x}^*)^{\mathsf{T}} \nabla f(\boldsymbol{x}^*) = -\sum_{i \in A(\boldsymbol{x}^*)} \lambda_i^* (\boldsymbol{x} - \boldsymbol{x}^*)^{\mathsf{T}} \nabla c_i(\boldsymbol{x}^*) \approx -\sum_{i \in A(\boldsymbol{x}^*)} \lambda_i^* [c_i(\boldsymbol{x}) - c_i(\boldsymbol{x}^*)]$$

Entonces, $\delta f = -\sum_{i \in A(\boldsymbol{x}^*)} \lambda_i^* \delta c_i$

- Perturbar una restricción activa tiene impacto en el valor optimo del objetivo! (a menos que por coincidencia $\lambda_i^* = 0$).
- Perturbar una restricciones inactiva no tiene impacto en el valor optimo del objetivo.
- Análisis de sensibilidad: que tan importante es cada restricción.

Chequeo

- Defina la función Lagrangiana. Por qué se introduce?
- Que son las condiciones complementarias? Que información adicional provee sobre la solución?
- Que significa conjunto activo?
- Puedes pensar en un problema donde los gradientes de las restricciones sean linealmente dependientes?

Referencias

- Basado en el curso “Applied Numerical Optimization” por el profesor Alexander Mitsos.
- Nocedal J. Wright S. J. Numerical Optimization, 2nd Edition, Springer, 2006.

¡Gracias!

Institución Universitaria
Reacreditada en Alta Calidad

Somos Innovación Tecnológica con Sentido Humano

Alcaldía de Medellín