



Institución
Universitaria
Reacreditada en Alta Calidad

Optimización

Optimización sin restricciones

Docente: Cristian Guarnizo Lemus

Somos Innovación Tecnológica con *Sentido Humano*

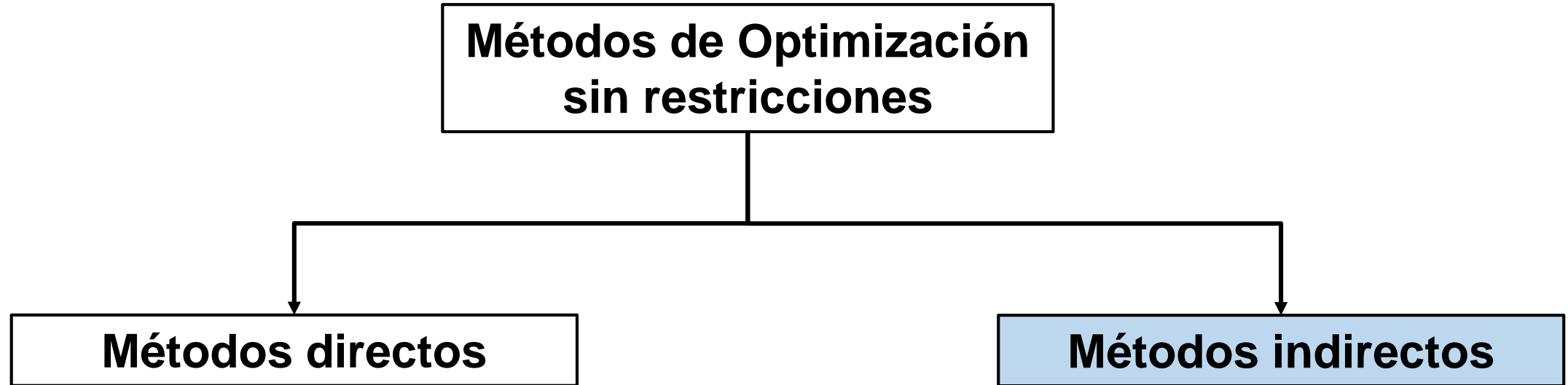


Alcaldía de Medellín

Contenido

1. Métodos de solución para Optimización sin restricciones.
2. Búsqueda en una línea. (Condiciones de Armijo y Wolfe seleccionar paso)
3. Determinación de la dirección (Steepest, Conjugate, Newton)

Métodos de solución para Optimización sin restricciones



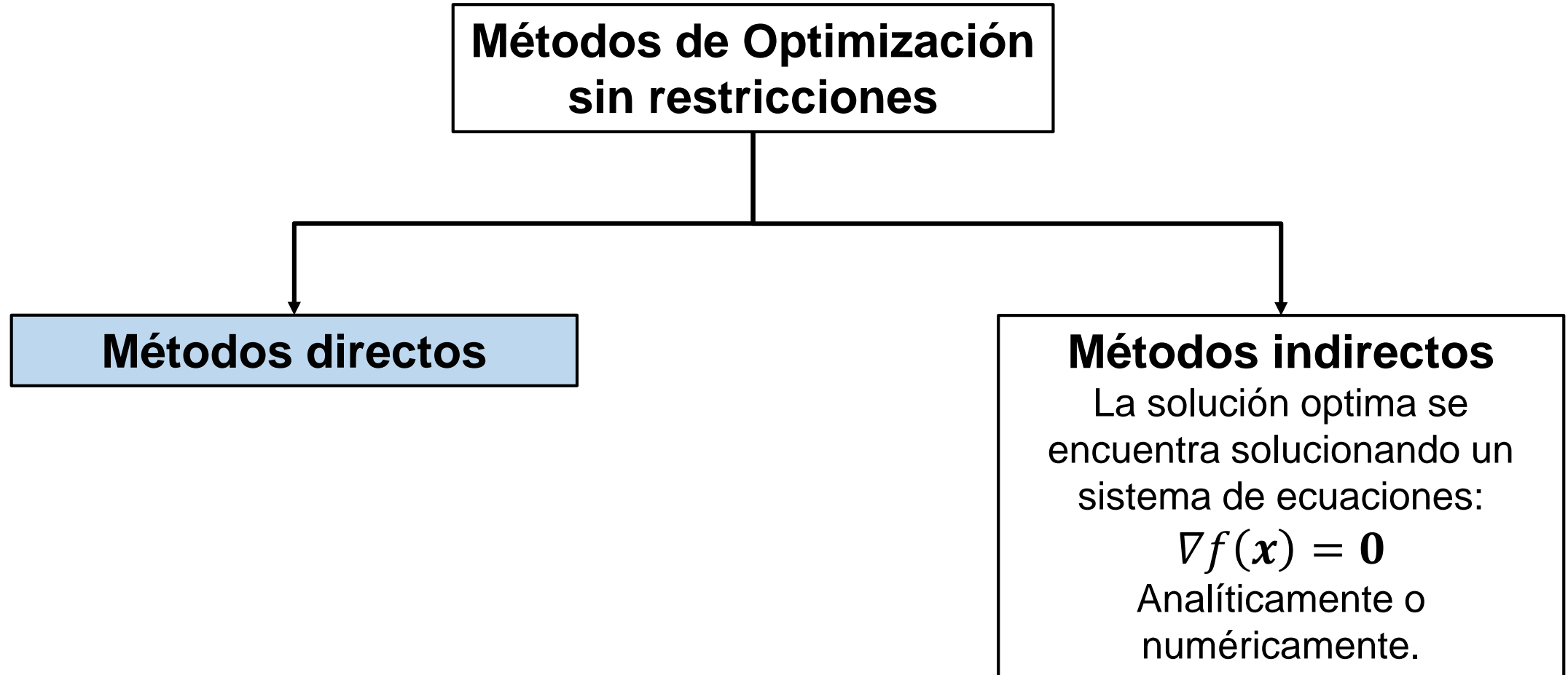
Métodos indirectos - Concepto

- Condiciones necesarias de primer orden

$$\nabla f(\mathbf{x}) = \mathbf{0} \Leftrightarrow \begin{cases} \left. \frac{\partial f}{\partial x_1} \right|_{\mathbf{x}} = 0 = g_1(\mathbf{x}) \\ \left. \frac{\partial f}{\partial x_2} \right|_{\mathbf{x}} = 0 = g_2(\mathbf{x}) \\ \vdots \\ \left. \frac{\partial f}{\partial x_n} \right|_{\mathbf{x}} = 0 = g_n(\mathbf{x}) \end{cases} \Leftrightarrow \begin{matrix} \text{Sistema de ecuaciones no} \\ \text{lineales} \\ \\ \mathbf{g}(\mathbf{x}) = \mathbf{0} \end{matrix}$$

- La solución óptima es encontrada **solucionando el sistema de ecuaciones** analíticamente o numéricamente. (p.e. con el Método de Newton).
- Diferenciar y solucionar un sistema de ecuaciones es **difícil para sistemas complejos**.

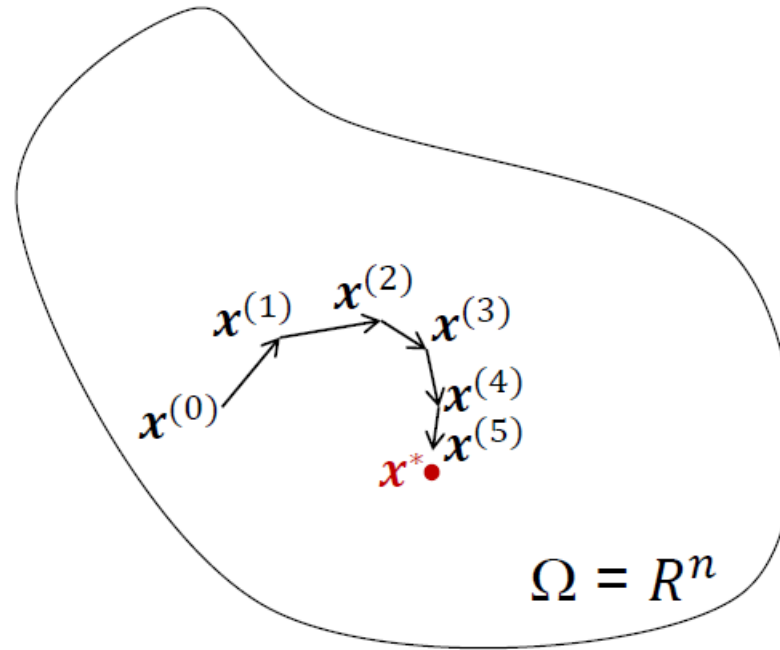
Métodos de solución para Optimización sin restricciones



Métodos directos - Concepto

Idea: Construir una secuencia convergente de $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty}$, que satisfaga las siguientes condiciones:

$$\exists \bar{k} \geq 0: f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)}) \quad \forall k > \bar{k} \quad \text{y} \quad \lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x}^* \in R^n$$



Métodos directos - Concepto

Idea: Construir una secuencia convergente de $\{x^{(k)}\}_{k=1}^{\infty}$, que satisfaga las siguientes condiciones:

$$\exists \bar{k} \geq 0: f(x^{(k+1)}) < f(x^{(k)}) \quad \forall k > \bar{k} \quad \text{y} \quad \lim_{k \rightarrow \infty} x^{(k)} = x^* \in R^n$$

Convergencia:

- **Lineal**: Sí existe una constante $C \in (0,1)$, tal que para un k suficientemente largo:

$$\|x^{(k+1)} - x^{(k)}\| \leq C \|x^{(k)} - x^*\|$$

- **Orden P**: Sí existe una constante $M > 0$, tal que

$$\|x^{(k+1)} - x^{(k)}\| \leq M \|x^{(k)} - x^*\|^p$$

- **Superlineal**: Sí existe una secuencia c_k convergente a cero.

$$\|x^{(k+1)} - x^{(k)}\| \leq c_k \|x^{(k)} - x^*\|$$

Métodos de solución para Optimización sin restricciones

Métodos de Optimización sin restricciones



```
graph TD; A[Métodos de Optimización sin restricciones] --> B[Métodos directos]; A --> C[Métodos indirectos];
```

Métodos directos

La solución optima se encuentra mejorando la función objetivo por medio de **iteraciones descendentes**.

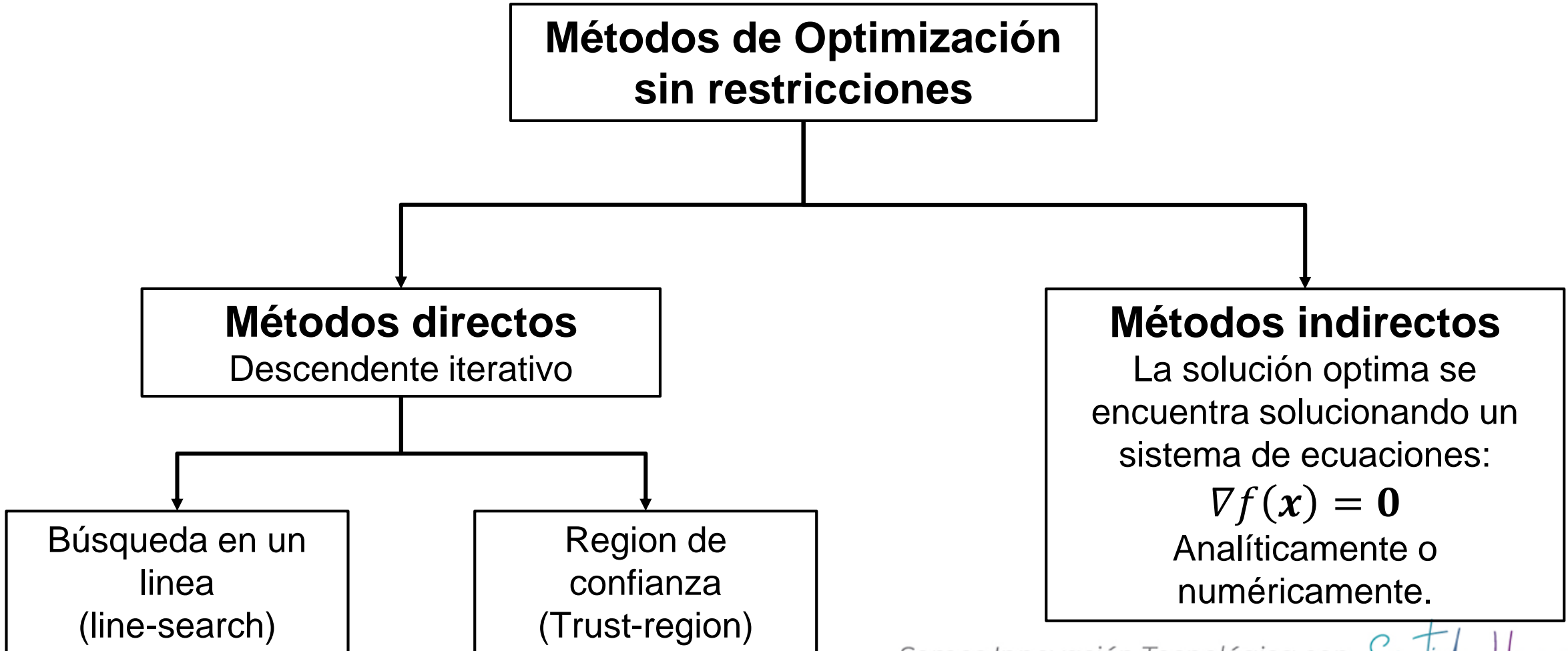
Métodos indirectos

La solución optima se encuentra solucionando un sistema de ecuaciones:

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

Analíticamente o numéricamente.

Métodos de solución para Optimización sin restricciones



Contenido

1. Métodos de solución para Optimización sin restricciones.
2. **Búsqueda en una línea.** (Condiciones de Armijo y Wolfe seleccionar paso)
3. Determinación de la dirección (Steepest, Conjugate, Newton's)

Métodos de solución para Optimización sin restricciones

Métodos de Optimización sin restricciones

Métodos directos Descendente iterativo

Búsqueda en un
línea
(line-search)

Region de
confianza
(Trust-region)

Métodos indirectos

La solución óptima se
encuentra solucionando un
sistema de ecuaciones:

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

Analíticamente o
numéricamente.

Búsqueda en una línea

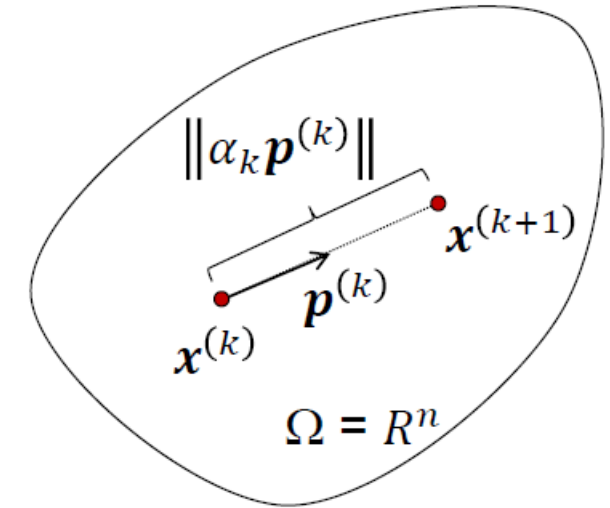
Definición (dirección descendente):

Un vector \mathbf{p} es llamado dirección descendente en $\mathbf{x}^{(k)}$, si $\nabla f(\mathbf{x}^{(k)})^T \mathbf{p} < 0$ se mantiene.

Algoritmo básico (line-search):

1. Seleccionar ya dirección descendente, $\mathbf{p}^{(k)}$, tal que

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{p}^{(k)} < 0$$
2. Determinar el tamaño el paso α_k
3. Calcular $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$



Problemas abiertos:

1. Determinar la dirección del descendente $\mathbf{p}^{(k)}$?
2. Calcular el tamaño el paso α_k .

Calculo del paso α_k

Algoritmo básico (line-search):

1. Definir una función de una dimensión sobre la dirección descendente $\mathbf{p}^{(k)}$

$$\phi(\alpha) = f(\mathbf{x}^{(k)} + \alpha \mathbf{p}^{(k)})$$

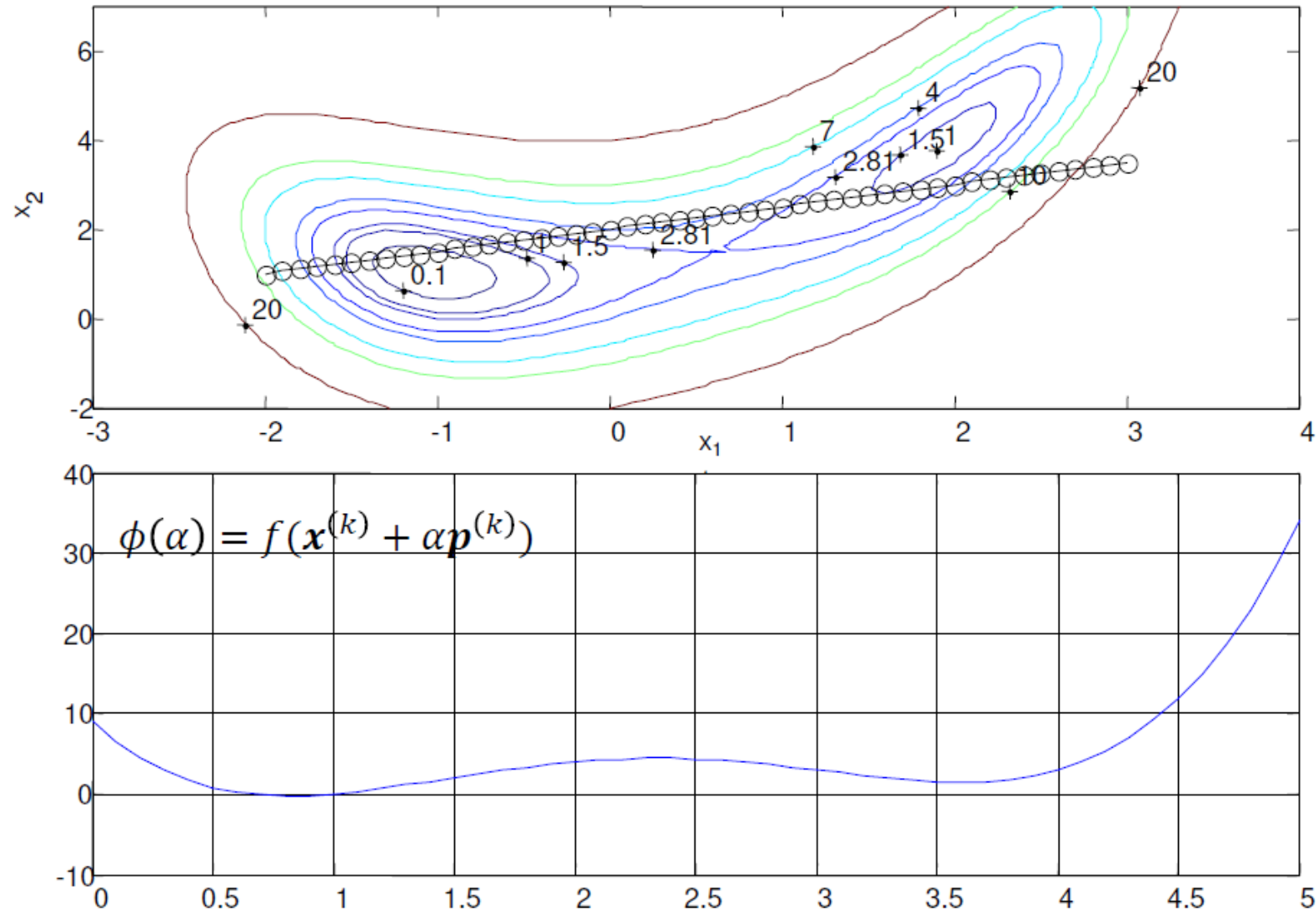
2. Solucionar el problema de minimización de una dimensión

$$\min_{\alpha > 0} \phi(\alpha)$$

Observaciones:

1. De manera ingenua seria ideal minimizar globalmente $\phi(\alpha)$. Generalmente, es muy costoso encontrar esta solución. No es necesariamente una buena idea buscar en una dimensión.
2. Se podría buscar alguna solución local. Pero esto es a menudo costoso (se necesita evaluar la función y/o gradientes en un numero de puntos).
3. Estrategias practicas (también llamadas **LS no exacto**): encontrar α tal que $\nabla f(\mathbf{x}^{(k+1)})$ se vuelva lo mas posible pequeño con el esfuerzo mínimo.

Estrategias de búsqueda en una línea



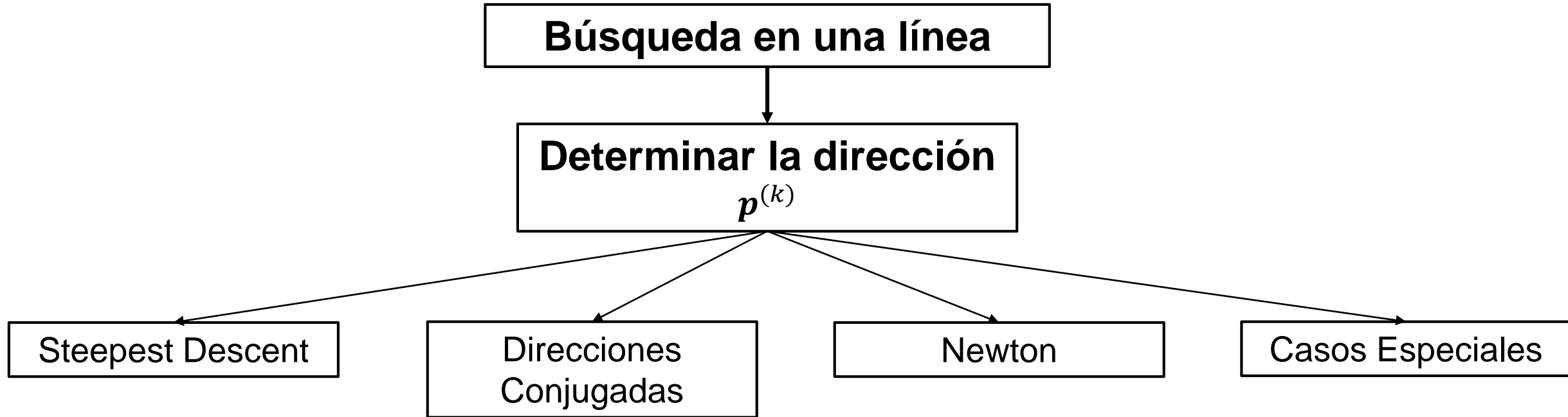


Contenido

1. Métodos de solución para Optimización sin restricciones.
2. Búsqueda en una línea.
3. Determinación de la dirección (Steepest, Newton, Quasi-Newton)

Determinación de la dirección

Los métodos de búsqueda en una línea difieren el uno del otro con respecto a la determinación de la dirección del descendente y el tamaño del paso.



Muchos métodos de gradiente usan matrices definidas positivas $\mathbf{D}^{(k)}$ y calculan

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \mathbf{D}^{(k)} \nabla f(\mathbf{x}^{(k)})$$

Steepest-Descent

Series de Taylor: $f(\mathbf{x}^{(k)} + \alpha \mathbf{p}^{(k)}) = f(\mathbf{x}^{(k)}) + \alpha \nabla f(\mathbf{x}^{(k)})^T \mathbf{p}^{(k)} + O(\alpha^2)$

La razón del cambio de f en $\mathbf{x}^{(k)}$ en la dirección $\mathbf{p}^{(k)}$ es el coeficiente en expresión lineal

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{p}^{(k)}$$

La dirección unitaria con el mayor cambio es la solución del siguiente problema

$$\min_{\mathbf{p}^{(k)} \in \mathbb{R}^n} \nabla f(\mathbf{x}^{(k)})^T \mathbf{p}^{(k)} \quad \text{s. t. } \|\mathbf{p}^{(k)}\| = 1$$

Note que $\nabla f(\mathbf{x}^{(k)})^T \mathbf{p}^{(k)} = \|\nabla f(\mathbf{x}^{(k)})\| \|\mathbf{p}^{(k)}\| \cos(\theta)$

La solución del problema se alcanza para $\cos(\theta) = -1 \Rightarrow \theta = \pi$

$$\Rightarrow \mathbf{p}^{(k)} = -\nabla f(\mathbf{x}^{(k)}) / \|\nabla f(\mathbf{x}^{(k)})\|$$

La selección de $\mathbf{D}^{(k)}$ es la matriz identidad \mathbf{I} .

Steepest-Descent

Algorithm:

choose $\mathbf{x}^{(0)}$

for $k=0, 1, \dots$

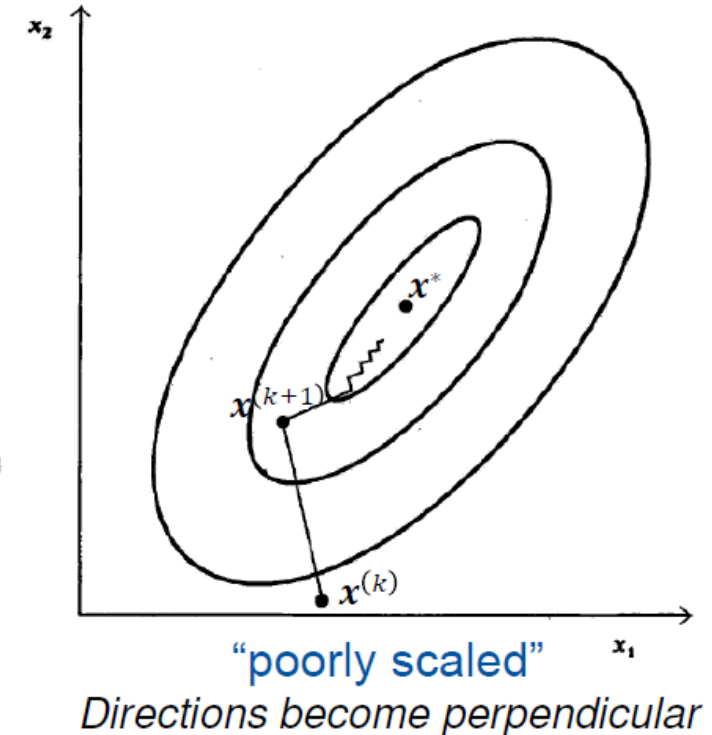
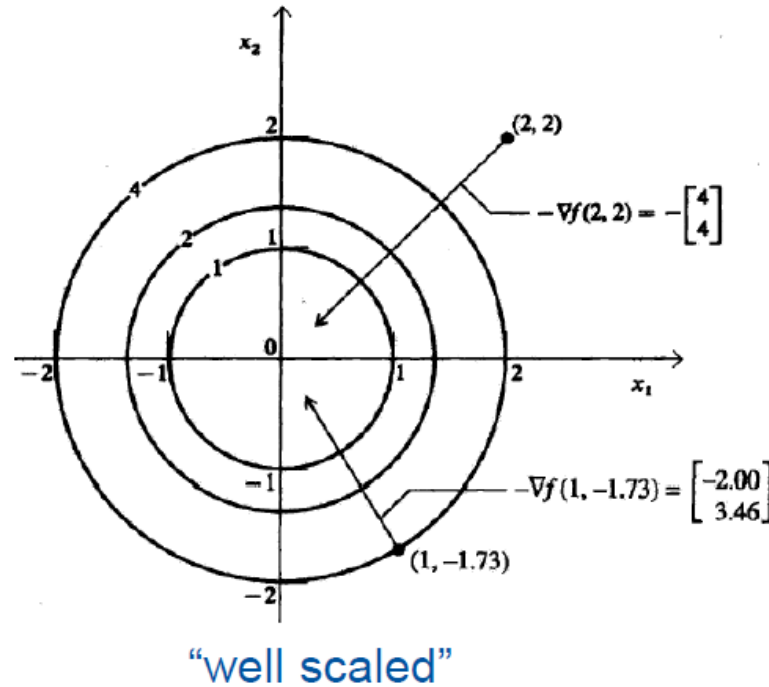
if $\|\nabla f(\mathbf{x}^{(k)})\| \leq \varepsilon$ stop, else

set $\mathbf{p}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$

determine the step length α_k (e.g.
using the Armijo rule)

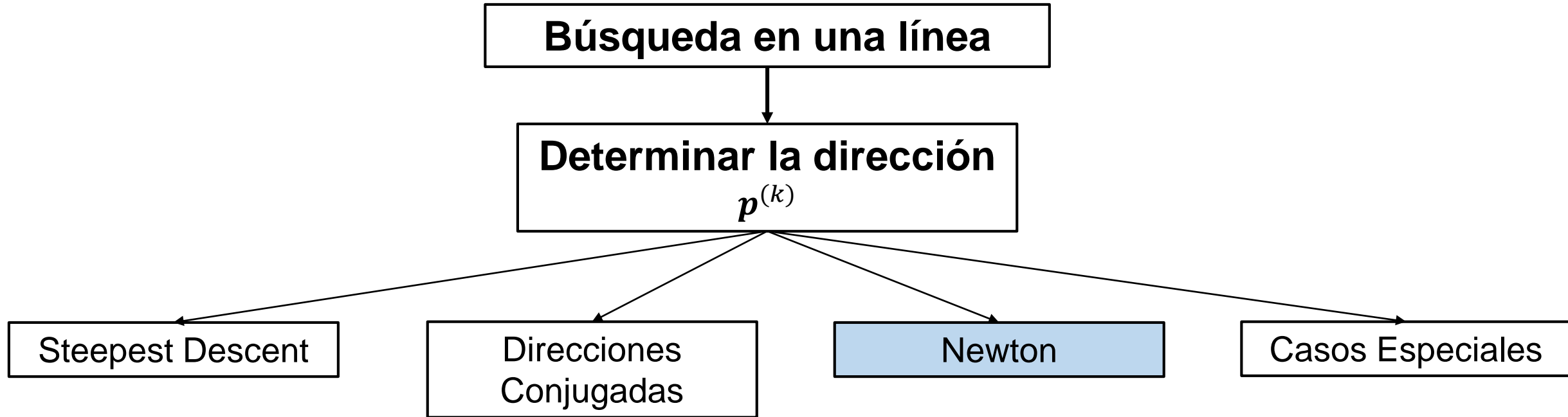
set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{p}^{(k)}$

end for



Determinación de la dirección

Los métodos de búsqueda en una línea difieren el uno del otro con respecto a la determinación de la dirección del descendente y el tamaño del paso.



Muchos métodos de gradiente usan matrices definidas positivas $\mathbf{D}^{(k)}$ y calculan

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \mathbf{D}^{(k)} \nabla f(\mathbf{x}^{(k)})$$

Dirección del descendente de Newton

Aproximación cuadrática de f en $\mathbf{x}^{(k+1)}$:

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

Primera **condición necesaria** para m

$$0 = \nabla m(\mathbf{x}^{(k+1)}) = \nabla f(\mathbf{x}^{(k)}) + \nabla^2 f(\mathbf{x}^{(k)}) (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

$$\Rightarrow \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - [\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)})$$

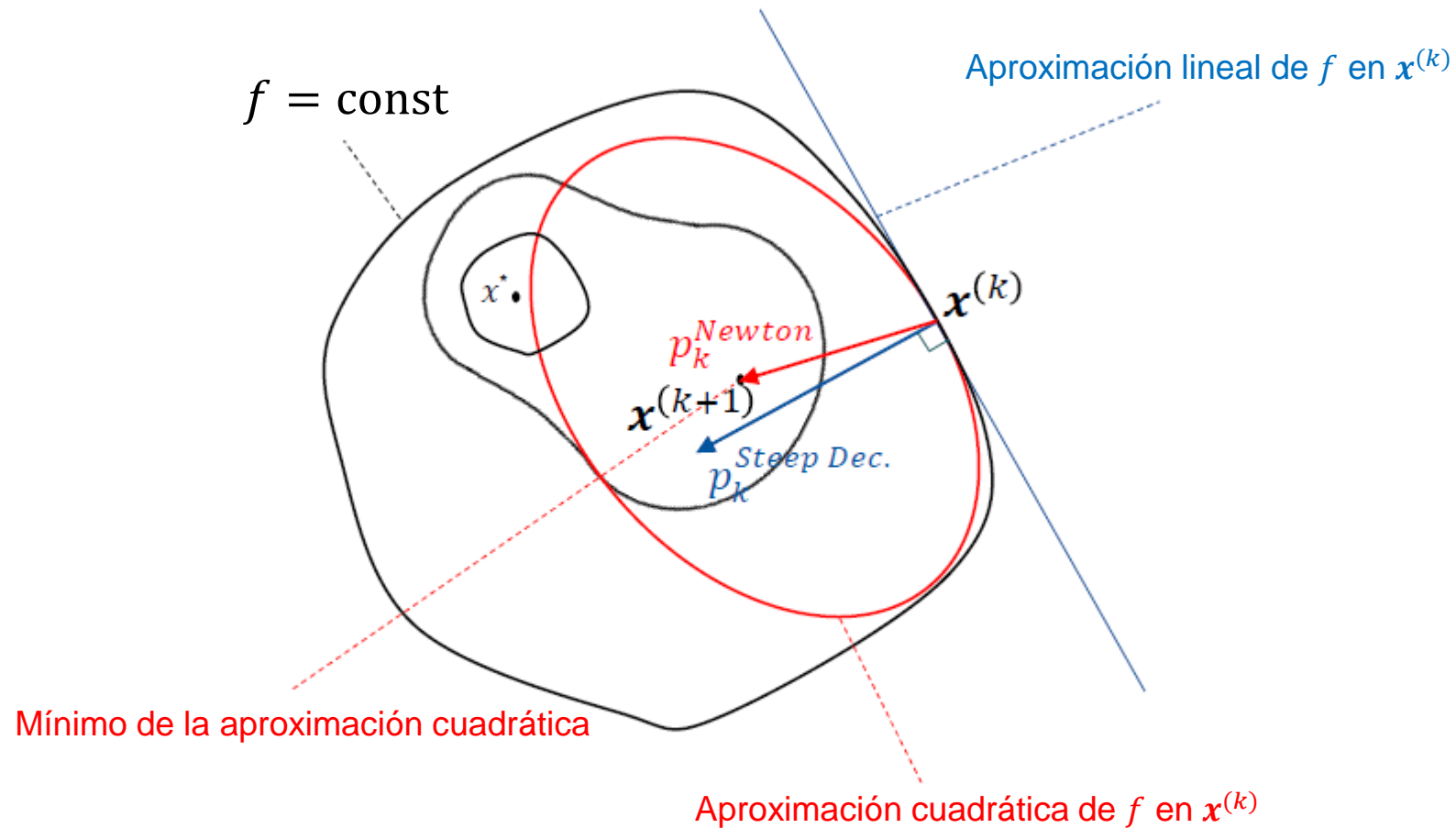


$$\Rightarrow \mathbf{p}^{(k)} = -[\nabla^2 f(\mathbf{x}^{(k)})]^{-1} \nabla f(\mathbf{x}^{(k)})$$

$$\Rightarrow \alpha_k = 1$$

Se selecciona $\mathbf{D}^{(k)}$ como la inversa de la Hessiana.

Dirección del descendente de Newton



Método de Newton

Algoritmo:

Choose $x^{(0)}$

FOR $k = 0, 1, 2, \dots$

IF $\|\nabla f(x^{(k)})\| \leq \varepsilon$ **STOP, ELSE**

SET $p^{(k)} = -[\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})$

SET $x^{(k+1)} = x^{(k)} + p^{(k)}$

END_FOR

Ventajas:

- Localmente tiene convergencia cuadrática, si $x^{(k)}$ es cercano a x^*
- Si f es cuadrático, el algoritmo converge en **una iteración**.

Desventajas:

- Es muy costoso para muchas variables, por el cálculo de la 2da derivada y la inversa.

Método Quasi-Newton

Definiciones: $f^{(k)} = f(\mathbf{x}^{(k)})$, $\mathbf{g}^{(k)} := \nabla f(\mathbf{x}^{(k)})$ y $\mathbf{H}^{(k)} := \nabla^2 f(\mathbf{x}^{(k)})$

Del método de Newton:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{p}^{(k)}$$

$$[\nabla^2 f(\mathbf{x}^{(k)})]\mathbf{p}^{(k)} = -\nabla f(\mathbf{x}^{(k)}) \Rightarrow \mathbf{H}^{(k)}\mathbf{p}^{(k)} = -\mathbf{g}^{(k)}$$

Idea:

El sistema lineal $\mathbf{H}^{(k)}\mathbf{p}^{(k)} = -\mathbf{g}^{(k)}$ se puede solucionar aproximadamente por un método iterativo.

Observaciones:

Factorizaciones como LU o Cholesky – costo computacional alto.

Ocurren errores grandes para problemas mal condicionados.

No se requiere la solución exacta.

Método Quasi-Newton (2)

Idea: reducir la complejidad simplificando el cálculo de $H^{(k)}$ (Davidon):

- Reemplazar $H^{(k)}$ por una aproximación $B^{(k)}$.
- En vez de calcular $B^{(k)}$, buscamos una simple actualización usando la información de las últimas iteraciones.

Método:

- Considerar la aproximación cuadrática de f en $x^{(k)}$, $m^{(k)}(p) = f^{(k)} + [g^{(k)}]^T p + \frac{1}{2} p^T B^{(k)} p$.
- Condición de optimalidad de primer orden: $p^{(k)} = -B^{(k)^{-1}} g^{(k)}$
- Por convexidad es necesaria y suficiente para la minimización de $m^{(k)}(p)$.
- Construir la aproximación cuadrática en $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$

$$m^{(k+1)}(p) = f^{(k+1)} + [g^{(k+1)}]^T p + \frac{1}{2} p^T B^{(k+1)} p$$

- Que condiciones debe satisfacer $B^{(k+1)}$?

Método Quasi-Newton (3)

Condiciones de $B^{(k+1)}$:

1. Gradiente de $m^{(k+1)}$ en $x^{(k+1)}$ debe ser igual al gradiente de f .

$\nabla m^{(k+1)}(p) = g^{(k+1)} + B^{(k+1)}p$	
At $x = x^{(k+1)}$, $p = 0$ Queremos $\nabla m^{(k+1)}(0) = g^{(k+1)}$	At $x = x^{(k)}$, $p = -\alpha_k p^{(k)}$ Queremos $\nabla m^{(k+1)}(-\alpha_k p^{(k)}) = g^{(k)}$
Satisface automáticamente	$\Rightarrow g^{(k+1)} - \alpha_k B^{(k+1)} p^{(k)} = g^{(k)}$ $\Rightarrow B^{(k+1)} \underbrace{\alpha_k p^{(k)}}_{=x^{(k+1)} - x^{(k)}} = g^{(k+1)} - g^{(k)}$ $\Rightarrow B^{(k+1)} s^{(k)} = y^{(k)}, \text{ donde } s^{(k)} = x^{(k+1)} - x^{(k)} \text{ y } y^{(k)} = g^{(k+1)} - g^{(k)}$

2. Debido que $B^{(k+1)}$ es simétrica definida positiva: $s^{(k)T} B^{(k+1)} s^{(k)} > 0, \forall s^{(k)} \neq 0 \Rightarrow s^{(k)T} y^{(k)} > 0$

Método Quasi-Newton (3)

Condiciones de $\mathbf{B}^{(k+1)}$:

$\mathbf{B}^{(k+1)} \mathbf{s}^{(k)} = \mathbf{y}^{(k)}$ gives many solutions for $\mathbf{B}^{(k+1)}$

- Unique solution: $\mathbf{B}^{(k+1)}$ should be close to $\mathbf{B}^{(k)}$

$$\min_{\mathbf{B}} \|\mathbf{B} - \mathbf{B}^{(k)}\|_W \quad \leftarrow \text{weighted Frobenius-Norm}$$

$$\text{s. t. } \mathbf{B}^T = \mathbf{B}$$

$$\|A\|_W = \|W^{1/2} A W^{1/2}\|_F, \text{ for any } W \text{ s.t. } W \mathbf{y}_k = \mathbf{s}_k$$

$$\mathbf{B} \mathbf{s}^{(k)} = \mathbf{y}^{(k)}$$

$$\|\cdot\|_F^2: R^{n \times n} \rightarrow R_{\geq 0}, \|C\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2$$

$$\Rightarrow \mathbf{B}^{(k+1)} = \left(\mathbf{I} - \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{y}^{(k)} \mathbf{s}^{(k)T} \right) \mathbf{B}^{(k)} \left(\mathbf{I} - \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{s}^{(k)} \mathbf{y}^{(k)T} \right) + \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{y}^{(k)} \mathbf{y}^{(k)T} \rightarrow \text{DFP formula}$$

$$\Rightarrow \mathbf{B}^{(k+1)-1} = \left(\mathbf{I} - \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{s}^{(k)} \mathbf{y}^{(k)T} \right) \mathbf{B}^{(k)-1} \left(\mathbf{I} - \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{y}^{(k)} \mathbf{s}^{(k)T} \right) + \frac{1}{\mathbf{y}^{(k)T} \mathbf{s}^{(k)}} \mathbf{s}^{(k)} \mathbf{s}^{(k)T} \rightarrow \text{BFGS formula}$$

Referencias

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



Institución
Universitaria
Reacreditada en Alta Calidad

¡Gracias!

Somos Innovación Tecnológica con *Sentido Humano*



Alcaldía de Medellín