Optimización

Optimización sin restricciones

Docente: Cristian Guarnizo Lemus

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 de Optimización sin restricciones

- Métodos directos
- Métodos indirectos

Métodos indirectos - Concepto

- Condiciones necesarias de primer orden

$$\nabla f(\boldsymbol{x}) = \mathbf{0} \Leftrightarrow \begin{cases} \left.\dfrac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}} = 0 = g_1(\boldsymbol{x}) \\ \left.\dfrac{\partial f}{\partial x_2}\right|_{\boldsymbol{x}} = 0 = g_2(\boldsymbol{x}) \\ \vdots \\ \left.\dfrac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}} = 0 = g_n(\boldsymbol{x}) \end{cases} \Leftrightarrow$$

Sistema de ecuaciones no lineales

$$\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$$

- La solución optima 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 de Optimización sin restricciones

Métodos directos

Métodos indirectos

La solución optima se encuentra solucionando un sistema de ecuaciones:

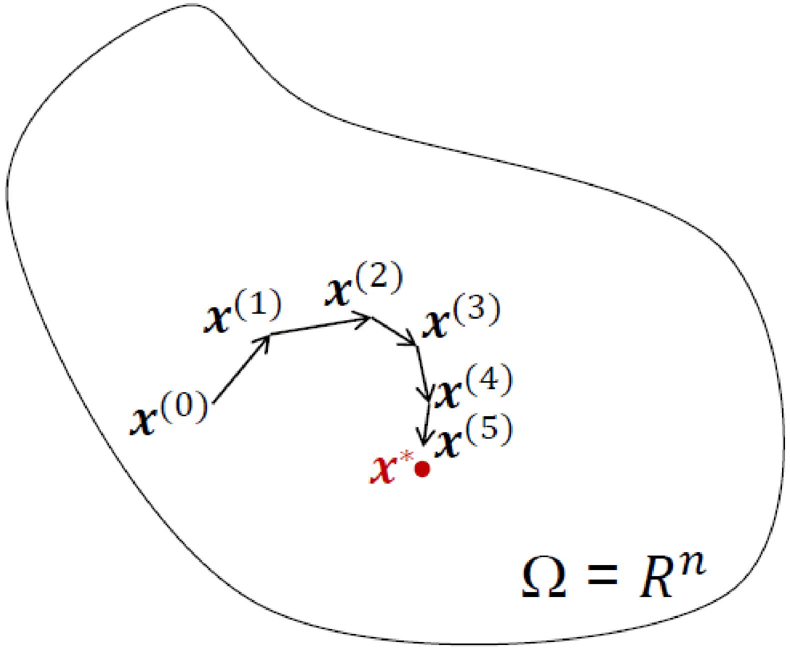
$$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$$

Analíticamente o numéricamente.

Métodos directos - Concepto

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

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



Métodos directos - Concepto

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

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

Convergencia:

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

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

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

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

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

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

Métodos de solución para Optimización sin restricciones

Métodos de Optimización sin restricciones

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:

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

Analíticamente o numéricamente.

Métodos de solución para Optimización sin restricciones

Métodos de Optimización sin restricciones

- **Métodos directos**
  Descendente iterativo
  - Búsqueda en un linea (line-search)
  - Region de confianza (Trust-region)
- **Métodos indirectos**
  La solución optima se encuentra solucionando un sistema de ecuaciones:
  $$\nabla f(\boldsymbol{x}) = \mathbf{0}$$
  Analíticamente o numéricamente.

Institución Universitaria
Reacreditada en Alta Calidad

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 linea (line-search)
  - Region de confianza (Trust-region)
- **Métodos indirectos**
  La solución optima se encuentra solucionando un sistema de ecuaciones:
  $$\nabla f(\boldsymbol{x}) = \mathbf{0}$$
  Analíticamente o numéricamente.

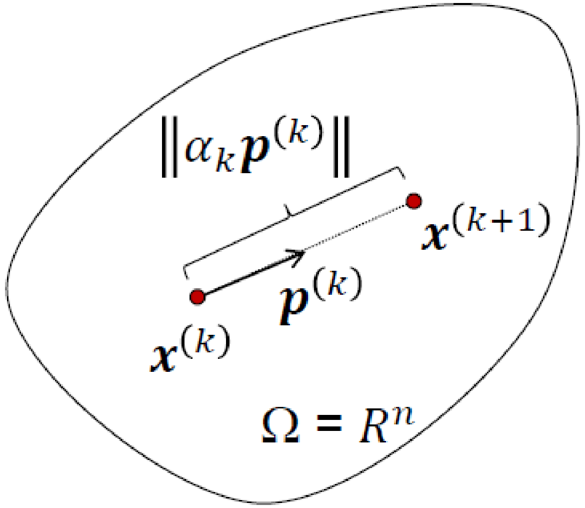
Búsqueda en una línea

Definición (dirección descendente):

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

Algoritmo básico (line-search):

1. Seleccionar ya dirección descendente, $\boldsymbol{p}^{(k)}$, tal que
$$\nabla f(\boldsymbol{x}^{(k)})^{\mathsf{T}}\boldsymbol{p}^{(k)} < 0$$
2. Determinar el tamaño el paso α_k
3. Calcular $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\, \boldsymbol{p}^{(k)}$



Problemas abiertos:

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

Calculo del paso α_k

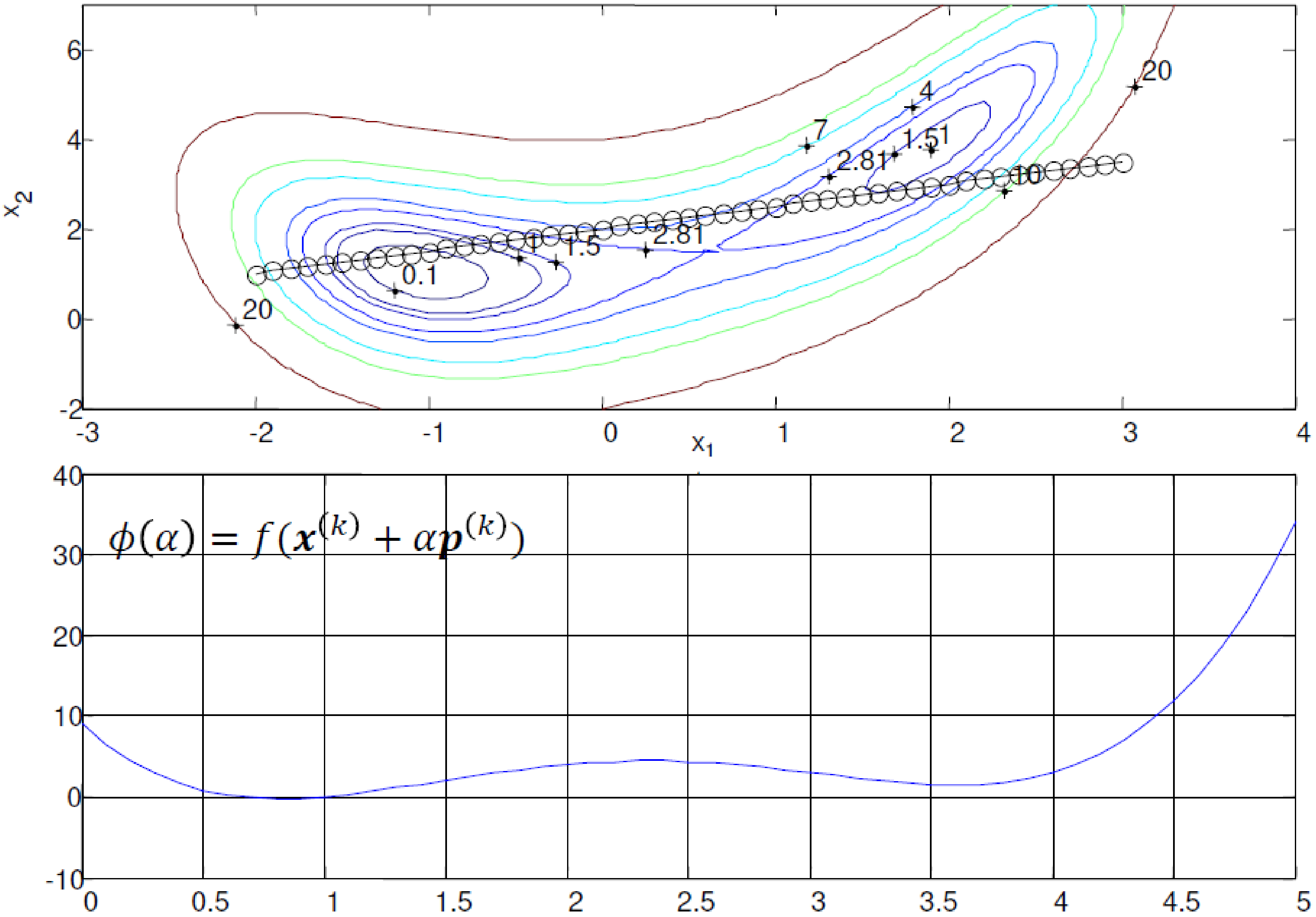
Algoritmo básico (line-search):

1. Definir una función de una dimensión sobre la dirección descendente $\boldsymbol{p}^{(k)}$
$$\phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)})$$
2. Solucionar el problema de minimización de una dimensión
$$\min_{\alpha > 0} \phi(\alpha)$$

Observaciones:

1. De manera ingenua seria ideal minimizer globalmente $\phi(\alpha)$. Generalmente, es muy costoso encontrará esta solucion. No es necesariamente una buena idea buscar en una dimension.
2. Se podria buscar alguna solucion 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 $\nabla f\left(\boldsymbol{x}^{(k+1)}\right)$ se vuelva lo mas posible pequeño con el esfuerzo mínimo.

Estrategias de busqueda en una linea

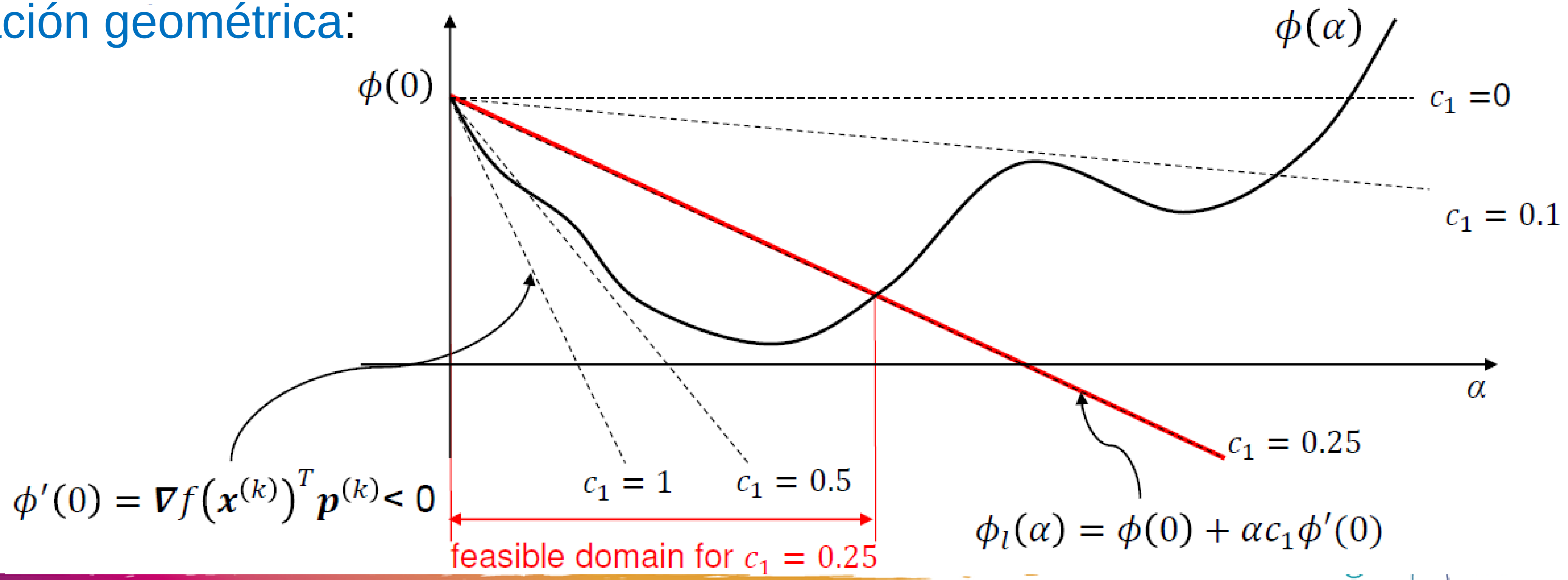


Condición de Armijo

Teorema:

Sea f continuamente diferenciable, $\boldsymbol{p}^{(k)}$ una dirección descendiente, y sea $c_1 \in (0,1)$. Entonces existe un $\alpha > 0$, tal que para $\phi(\alpha) := f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{p}^{(k)})$, la condición $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$ se mantiene.

Interpretación geométrica:



Algoritmo simple de búsqueda en una línea

Observaciones:

1. La selección del tamaño del paso, que cumpla con la condición de Armijo garantiza el descendiente de f.

$$\phi'(0) \leq \nabla f\left(\boldsymbol{x}^{(k)}\right)^{\mathsf{T}} \boldsymbol{p}^{(k)} < 0 (\boldsymbol{p}^{(k)} \text{ es una dirección descendente})$$

$$\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0) \rightarrow \phi(\alpha) < \phi(0)$$

2. La selección de c_1 es crucial:

- Valores grande de c_1 nos llevan a valores pequeños de α. Entonces $\boldsymbol{x}^{(k+1)} \approx \boldsymbol{x}^{(k)}$.
- Un pequeño c_1 resulta potencialmente en una pequeña reducción de f y entonces la convergencia es mas lenta.

Simple line-search algorithm:

choose $\alpha_1 > 0;\ \rho, c_1 \in (0,1)$

set $\alpha = \alpha_1$

repeat $\alpha \leftarrow \rho\alpha$ until $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$

Algoritmo mejorado de búsqueda en una linea

Seleccionar $\alpha_0 > 0$ y $c_1 \in (0,1)$

If $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$ PARAR, else

Encontrar una major solución $\alpha \in (0, \alpha_0)$ por medio de una interpolación cuadrática de los datos disponibles:

$$\alpha_1 = \frac{\phi'(0)\alpha_0^2}{2[\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha]}$$

If $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$ PARAR, else

Encontrar una major solución $\alpha \in (0, \alpha_1)$ por medio de una interpolación cúbica (como?)

Repetir el procedimiento de la interpolación cúbica, hasta que la condición se cumpla.

Institución Universitaria
Reacreditada en Alta Calidad

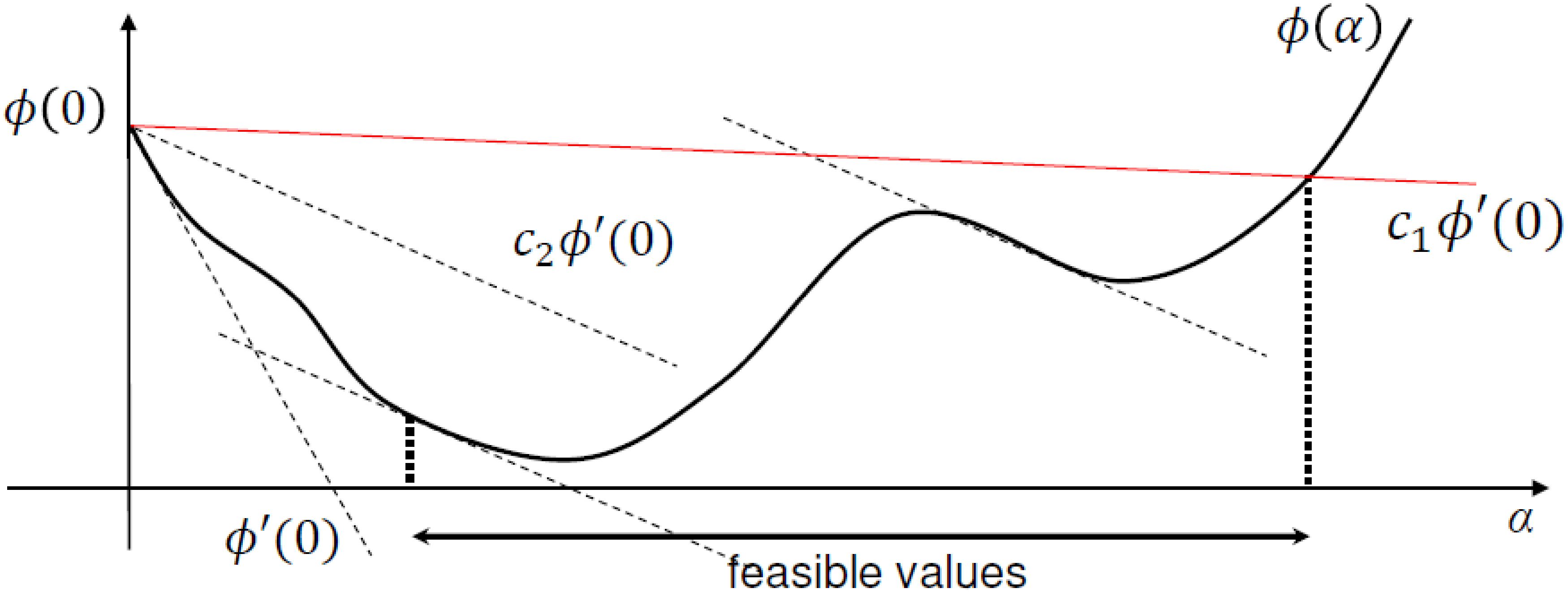
Condición de Wolfe

Teorema:

Sea f continuamente diferenciable, $\boldsymbol{p}^{(k)}$ una dirección descendente, y sea $c_1 \in (0,1)$, $c_2 \in (c_1, 1)$. Entonces existe un $\alpha > 0$, tal que $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$

$\phi'(\alpha) \geq c_2 \phi'(0)$ (condición de pendiente)

Interpretación geométrica:



Las condiciones de Wolfe promueven la convergencia a un punto estacionario.

Institución Universitaria
Reacreditada en Alta Calidad

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)

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.

Búsqueda en una línea

↓

Determinar la dirección
$\boldsymbol{p}^{(k)}$

↓

Steepest Descent | Direcciones Conjugadas | Newton | Casos Espciales

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

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

Steepest-Descent

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

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

$$\nabla f\left(\boldsymbol{x}^{(k)}\right)^{\mathrm{T}} \boldsymbol{p}^{(k)}$$

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

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

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

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

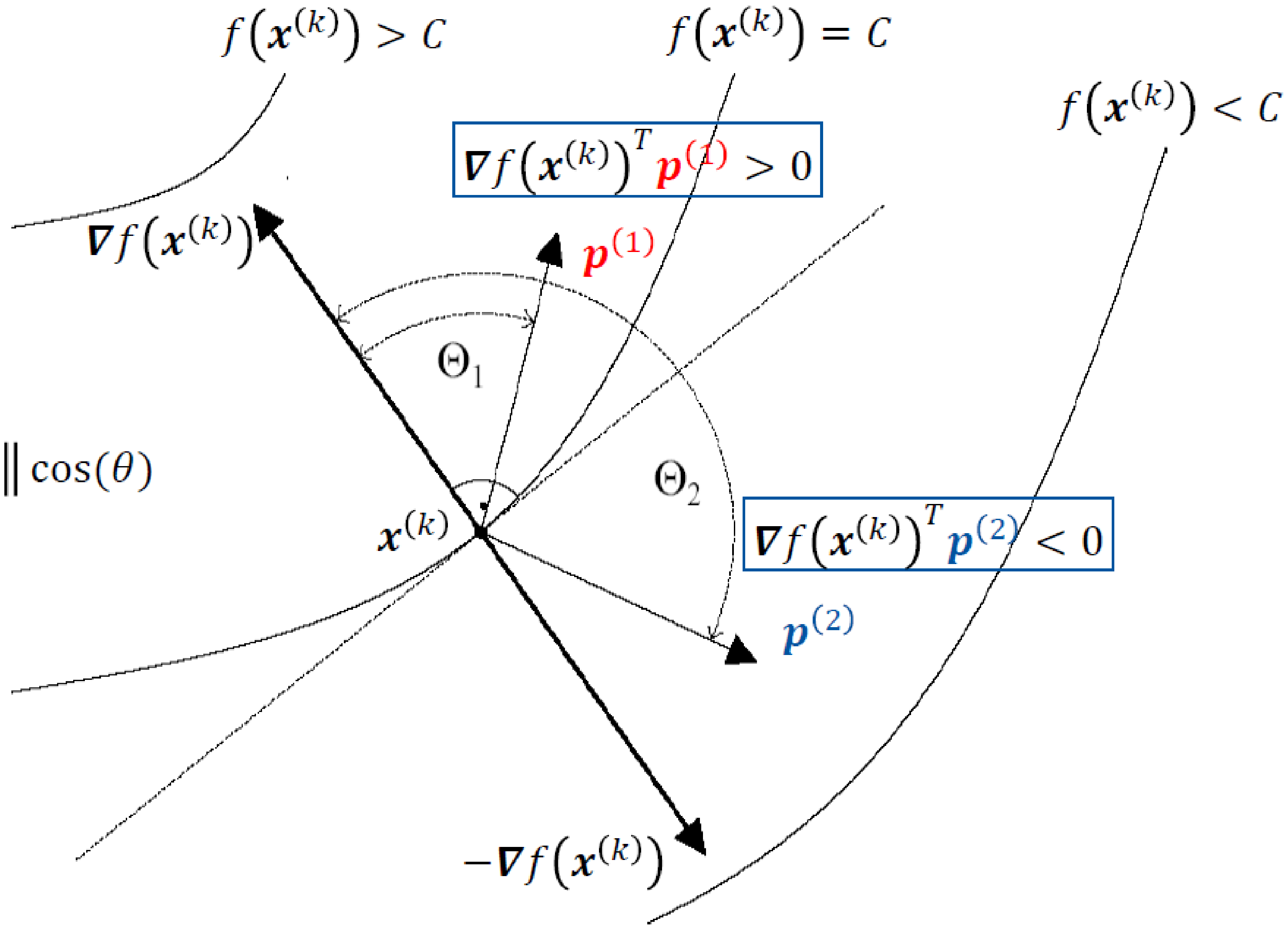
$$\Rightarrow \boldsymbol{p}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) / \left\|\nabla f(\boldsymbol{x}^{(k)})\right\|$$

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

Steepest-Descent

descent direction: $\nabla f\left(\boldsymbol{x}^{(k)}\right)^T \boldsymbol{p}^{(k)} < 0$

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



Steepest-Descent

Algorithm:

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

for $k=0,1,\ldots$

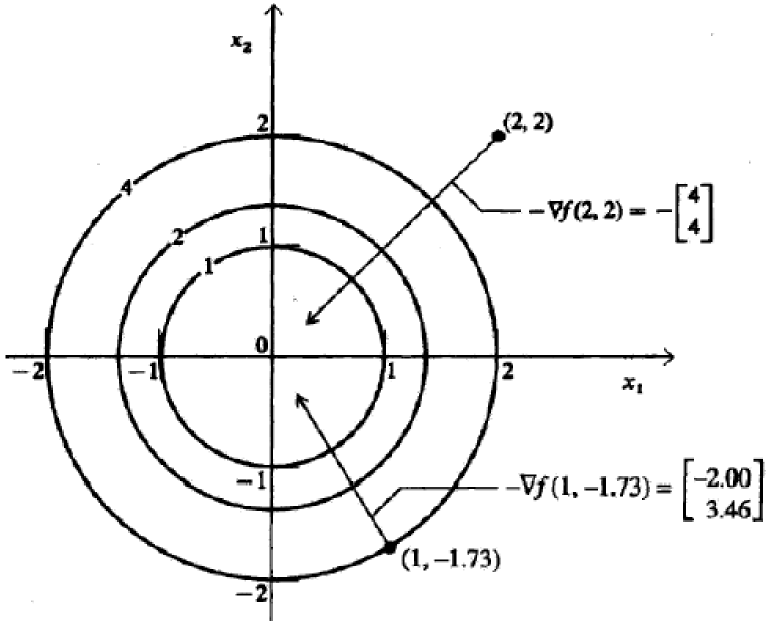
if $\|\nabla f(\boldsymbol{x}^{(k)})\| \leq \varepsilon$ stop, else

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

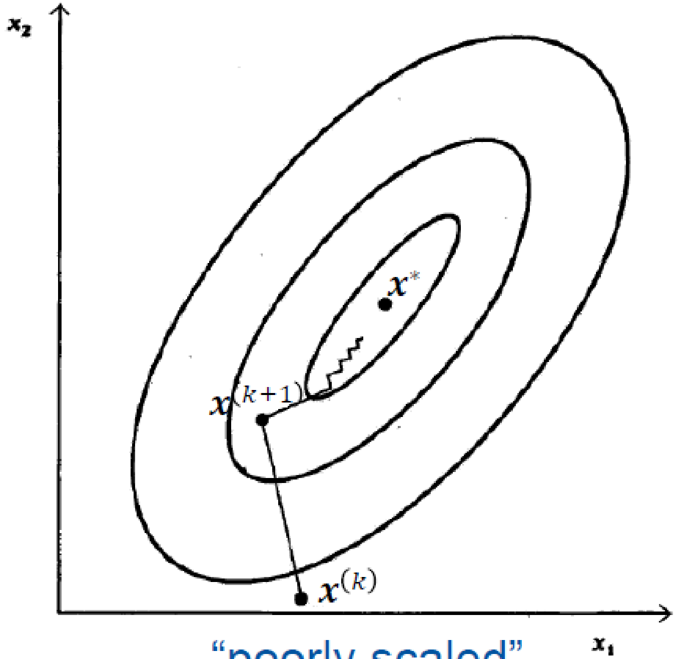
determine the step length α_k (e.g. using the Armijo rule)

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

end for



"well scaled"



"poorly scaled"

Directions become perpendicular

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.

Búsqueda en una línea

↓

Determinar la dirección
$\boldsymbol{p}^{(k)}$

Steepest Descent | Direcciones Conjugadas | Newton | Casos Espciales

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

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

Dirección del descendente de Newton

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

$$m\left(\boldsymbol{x}^{(k+1)}\right) = f\left(\boldsymbol{x}^{(k)}\right) + \nabla f\left(\boldsymbol{x}^{(k)}\right)^{\mathsf{T}}\left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right) + \left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right)^{\mathsf{T}}\nabla^2 f\left(\boldsymbol{x}^{(k)}\right)\left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right)$$

Primera condición necesaria para m

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

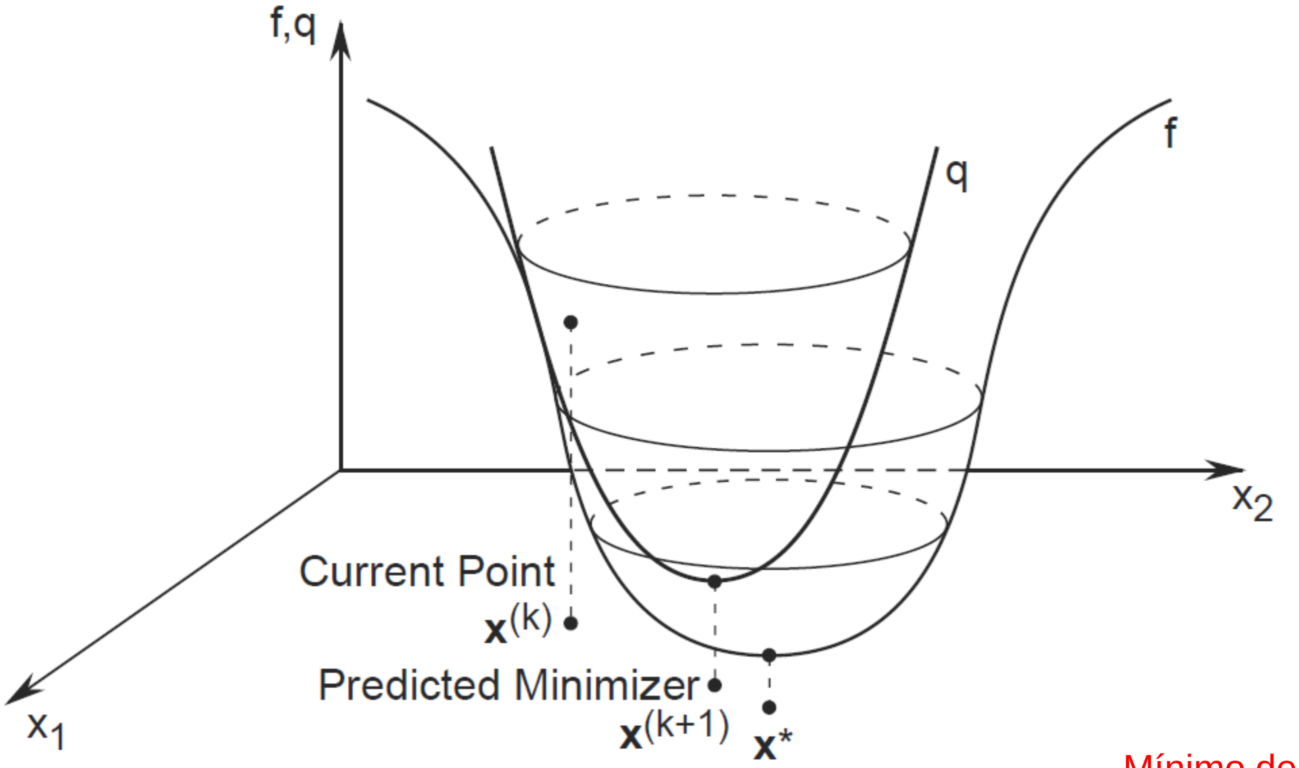
$$\Rightarrow \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left[\nabla^2 f\left(\boldsymbol{x}^{(k)}\right)\right]^{-1}\nabla f\left(\boldsymbol{x}^{(k)}\right)$$

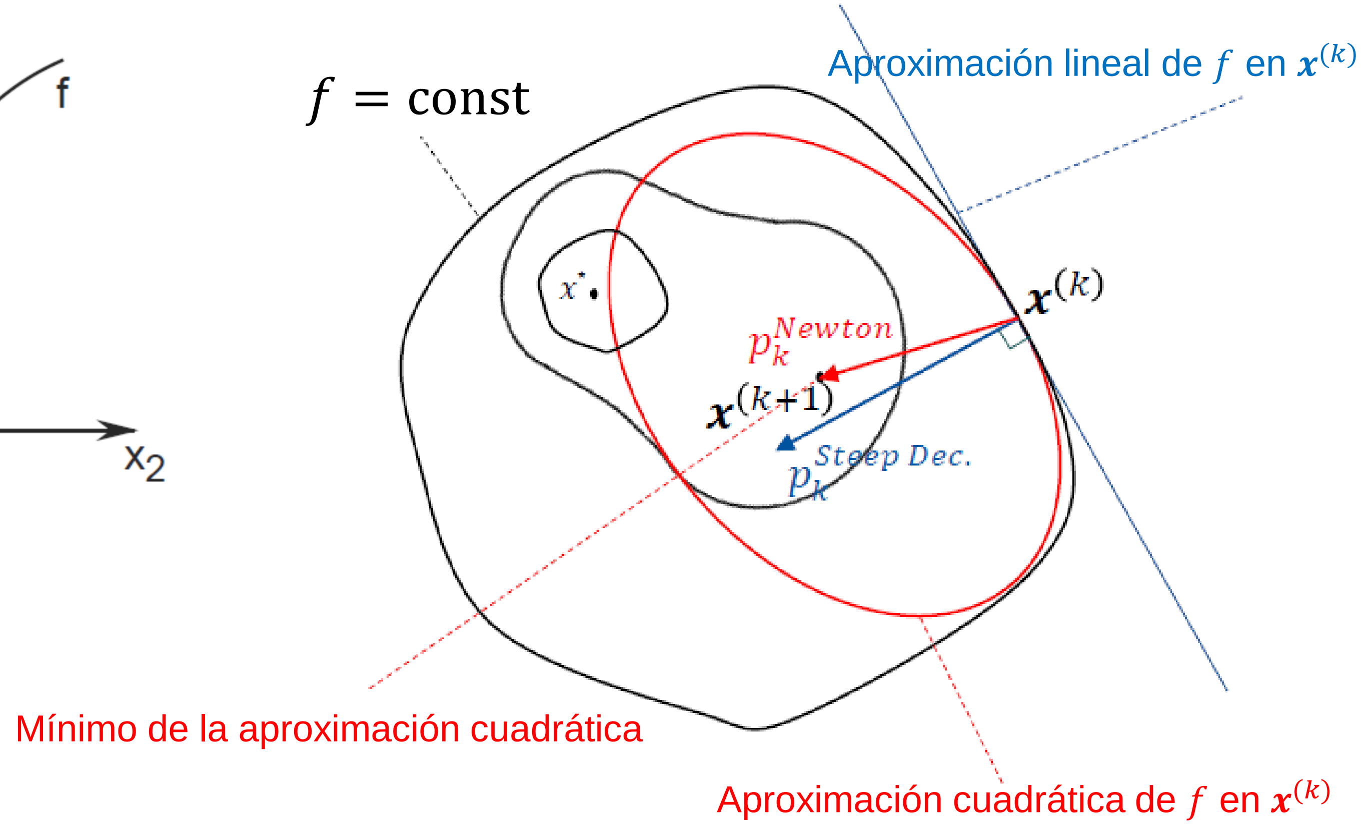
$$\Longrightarrow \boldsymbol{p}^{(k)} = -\left[\nabla^2 f\left(\boldsymbol{x}^{(k)}\right)\right]^{-1}\nabla f\left(\boldsymbol{x}^{(k)}\right)$$

$$\Longrightarrow \alpha_k = 1$$

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

Dirección del descendente de Newton





https://suzyahyah.github.io/calculus/optimization/2018/04/06/Taylor-Series-Newtons-Method.html

Método de Newton

Algoritmo:

Choose $x^{(0)}$

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

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

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

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

END_FOR

Ventajas:

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

Desventajas:

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

Método Quasi-Newton

Definiciones: $f^{(k)} = f(\boldsymbol{x}^{(k)})$, $\boldsymbol{g}^{(k)} := \nabla f\left(\boldsymbol{x}^{(k)}\right)$, $\boldsymbol{H}^{(k)} := \nabla^2 f\left(\boldsymbol{x}^{(k)}\right)$ y $\boldsymbol{B}^k := (\boldsymbol{H}^{(k)})^{-1}$

Del método de Newton:

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

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

Idea:

El sistema lineal $\boldsymbol{H}^{(k)}\boldsymbol{p}^{(k)} = -\boldsymbol{g}^{(k)}$ 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.

Métodos Quasi-Newton

La matriz $\boldsymbol{B}$ debe ser simétrica y positiva definida.

Method	$B_{k+1} =$	$H_{k+1} = B_{k+1}^{-1} =$
BFGS	$B_k + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \Delta x_k} - \frac{B_k \Delta x_k (B_k \Delta x_k)^{\mathrm{T}}}{\Delta x_k^{\mathrm{T}} B_k \, \Delta x_k}$	$\left(I - \frac{\Delta x_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \Delta x_k}\right) H_k \left(I - \frac{y_k \Delta x_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \Delta x_k}\right) + \frac{\Delta x_k \Delta x_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \, \Delta x_k}$
Broyden	$B_k + \frac{y_k - B_k \Delta x_k}{\Delta x_k^{\mathrm{T}} \, \Delta x_k} \, \Delta x_k^{\mathrm{T}}$	$H_k + \frac{(\Delta x_k - H_k y_k) \Delta x_k^{\mathrm{T}} H_k}{\Delta x_k^{\mathrm{T}} H_k \, y_k}$
Broyden family	$(1 - \varphi_k) B_{k+1}^{\mathrm{BFGS}} + \varphi_k B_{k+1}^{\mathrm{DFP}}, \quad \varphi \in [0, 1]$	
DFP	$\left(I - \frac{y_k \, \Delta x_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \, \Delta x_k}\right) B_k \left(I - \frac{\Delta x_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \, \Delta x_k}\right) + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \, \Delta x_k}$	$H_k + \frac{\Delta x_k \Delta x_k^{\mathrm{T}}}{\Delta x_k^{\mathrm{T}} \, y_k} - \frac{H_k y_k y_k^{\mathrm{T}} H_k}{y_k^{\mathrm{T}} H_k y_k}$
SR1	$B_k + \frac{(y_k - B_k \, \Delta x_k)(y_k - B_k \, \Delta x_k)^{\mathrm{T}}}{(y_k - B_k \, \Delta x_k)^{\mathrm{T}} \, \Delta x_k}$	$H_k + \frac{(\Delta x_k - H_k y_k)(\Delta x_k - H_k y_k)^{\mathrm{T}}}{(\Delta x_k - H_k y_k)^{\mathrm{T}} y_k}$

https://en.wikipedia.org/wiki/Quasi-Newton_method

Chequeo

- Cual es la diferencia entre los métodos directos y los indirectos?
- Cual es la dirección que se emplea en el método Steepest Descent?
- Cual es el calcula de la actualización del nuevo punto a evaluar en el método de Newton?
- Por que aparecen modificaciones al método de Newton?

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

Método Quasi-Newton (2)

Idea: reducir la compejidad simplificando el calculo de $\boldsymbol{H}^{(k)}$ (Davidon):

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

Metodo:

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

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

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

Método Quasi-Newton (3)

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

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

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

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

Método Quasi-Newton (3)

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

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

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

$$\min_{\boldsymbol{B}} \left\|\boldsymbol{B} - \boldsymbol{B}^{(k)}\right\|_W \quad \text{s.t.} \quad \boldsymbol{B}^T = \boldsymbol{B}, \quad \boldsymbol{B}\boldsymbol{s}^{(k)} = \boldsymbol{y}^{(k)}$$

$\leftarrow$ weighted Frobenius-Norm

$\|A\|_W = \left\|W^{1/2}AW^{1/2}\right\|_F$, for any W s.t. $Wy_k = s_k$

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

$$\Rightarrow \boldsymbol{B}^{(k+1)} = \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{y}^{(k)}\boldsymbol{s}^{(k)^T}\right)\boldsymbol{B}^{(k)}\left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{s}^{(k)}\boldsymbol{y}^{(k)^T}\right) + \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{y}^{(k)}\boldsymbol{y}^{(k)^T}$$

$\rightarrow$ DFP formula

$$\Rightarrow \boldsymbol{B}^{(k+1)^{-1}} = \left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{s}^{(k)}\boldsymbol{y}^{(k)^T}\right)\boldsymbol{B}^{(k)^{-1}}\left(\boldsymbol{I} - \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{y}^{(k)}\boldsymbol{s}^{(k)^T}\right) + \frac{1}{\boldsymbol{y}^{(k)^T}\boldsymbol{s}^{(k)}}\boldsymbol{s}^{(k)}\boldsymbol{s}^{(k)^T}$$

$\rightarrow$ BFGS formula