Institución Universitaria
Reacreditada en Alta Calidad

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’s)

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:

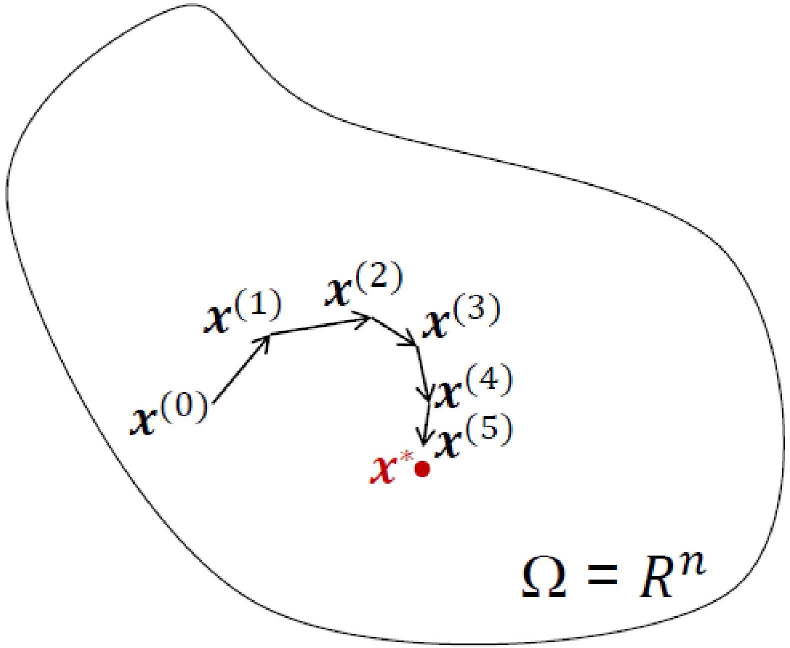
$$\nabla f(\boldsymbol{x}) = \mathbf{0}$$

Analíticamente o numéricamente.

Métodos directos - Concepto

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

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



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:

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

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.

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:
  $$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$$
  Analíticamente o numéricamente.

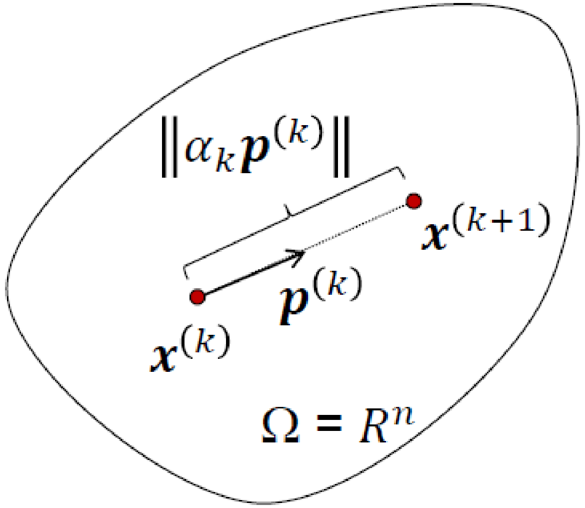
Búsqueda en una línea

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

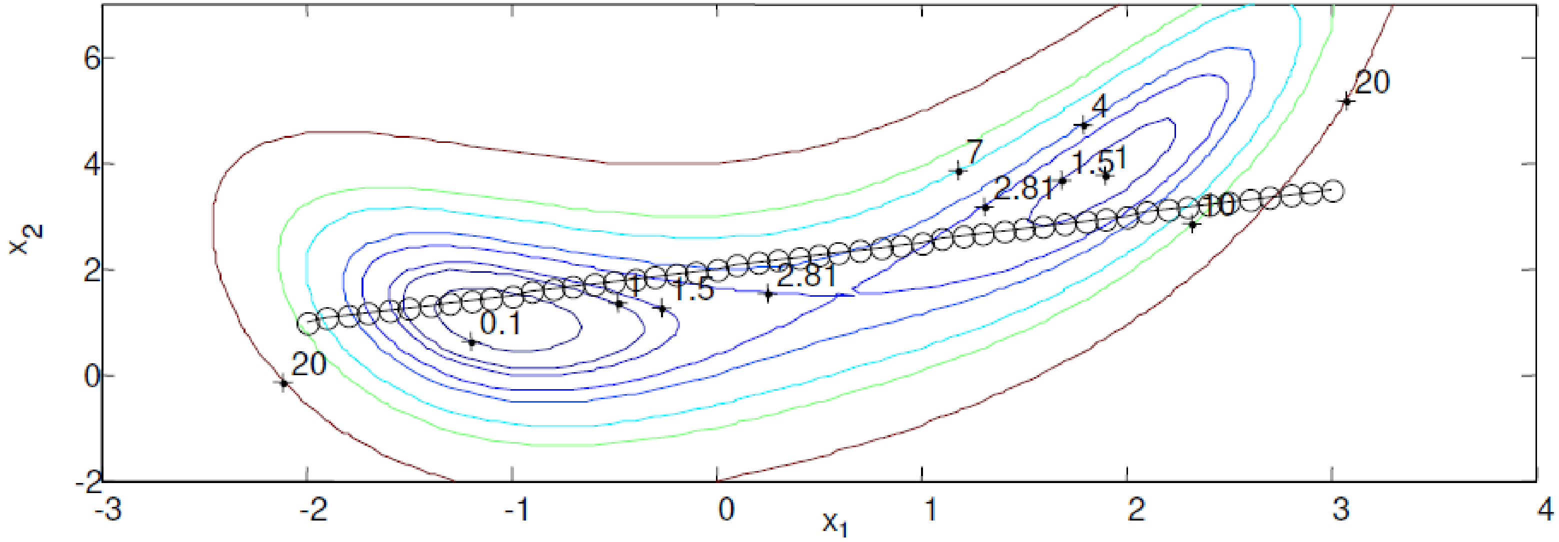
Algoritmo básico (line-search):

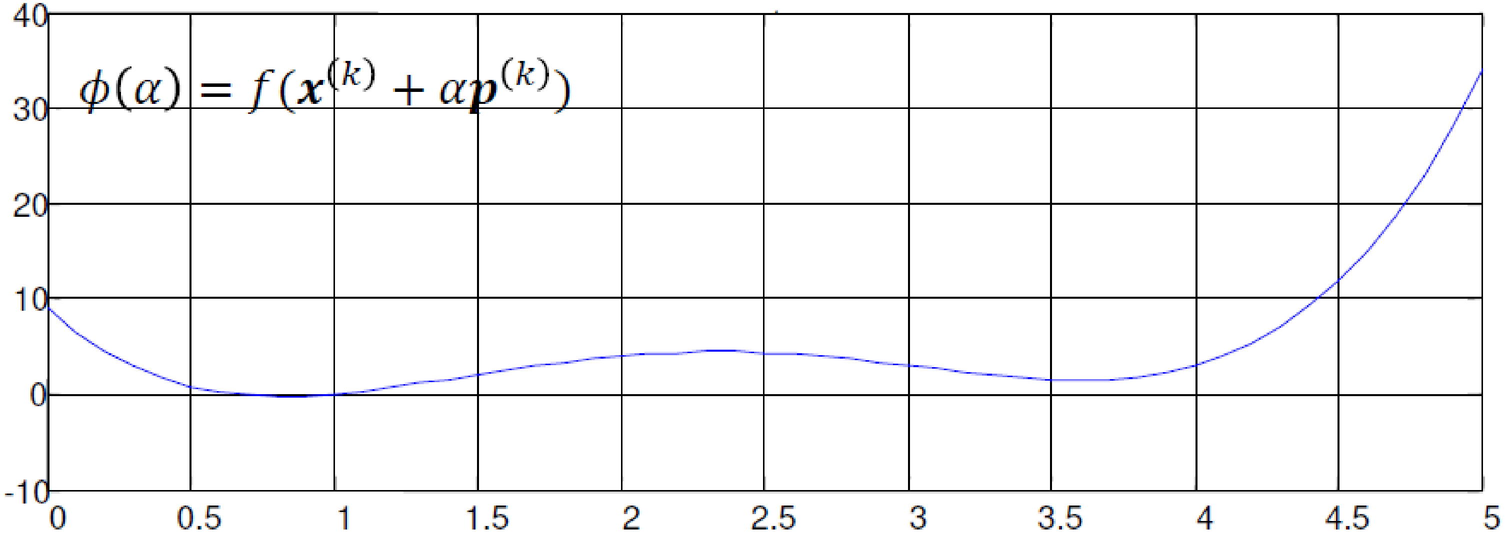
1. Definir una función de una dimensión sobre la dirección descendente $\boldsymbol{p}^{(k)}$
$$\phi(\alpha) = \nabla f(\boldsymbol{x}^{(k)} + \boldsymbol{p}^{(k)})$$
2. Solucionar el problema de minimización de una dimensión
$$\min_{\alpha>0} \phi(\alpha)$$

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

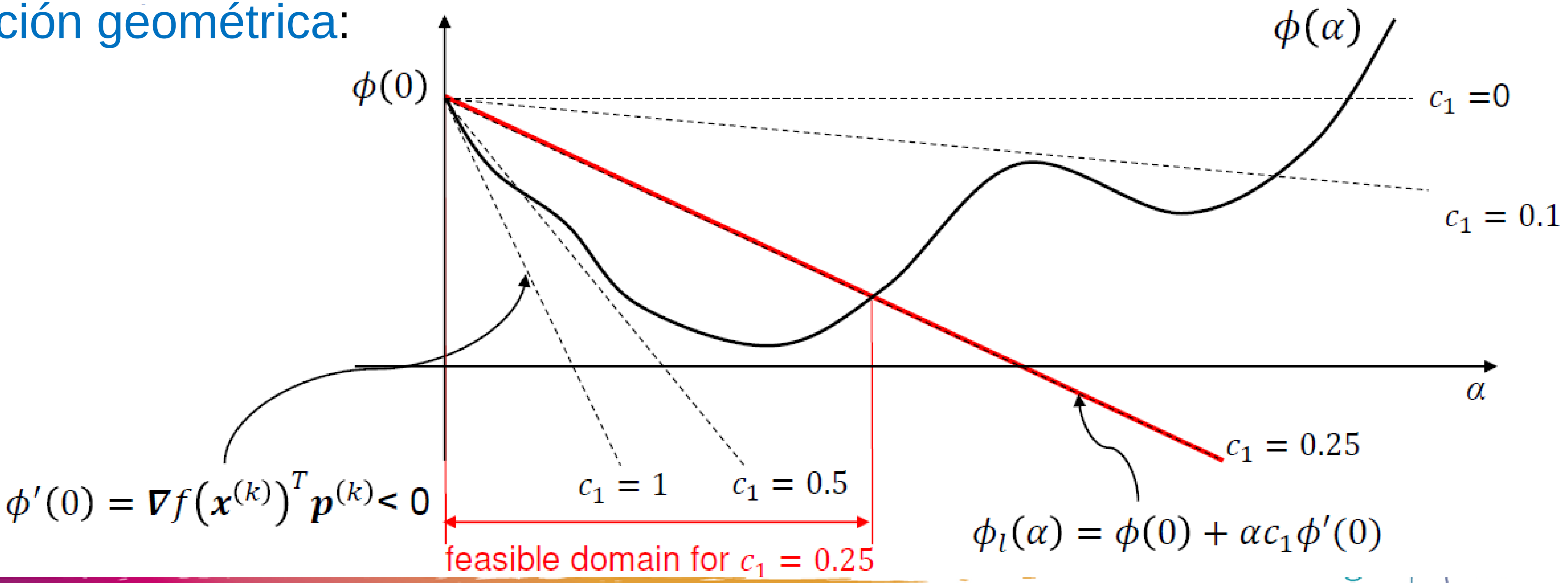




Condición de Armijo

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 linea

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.

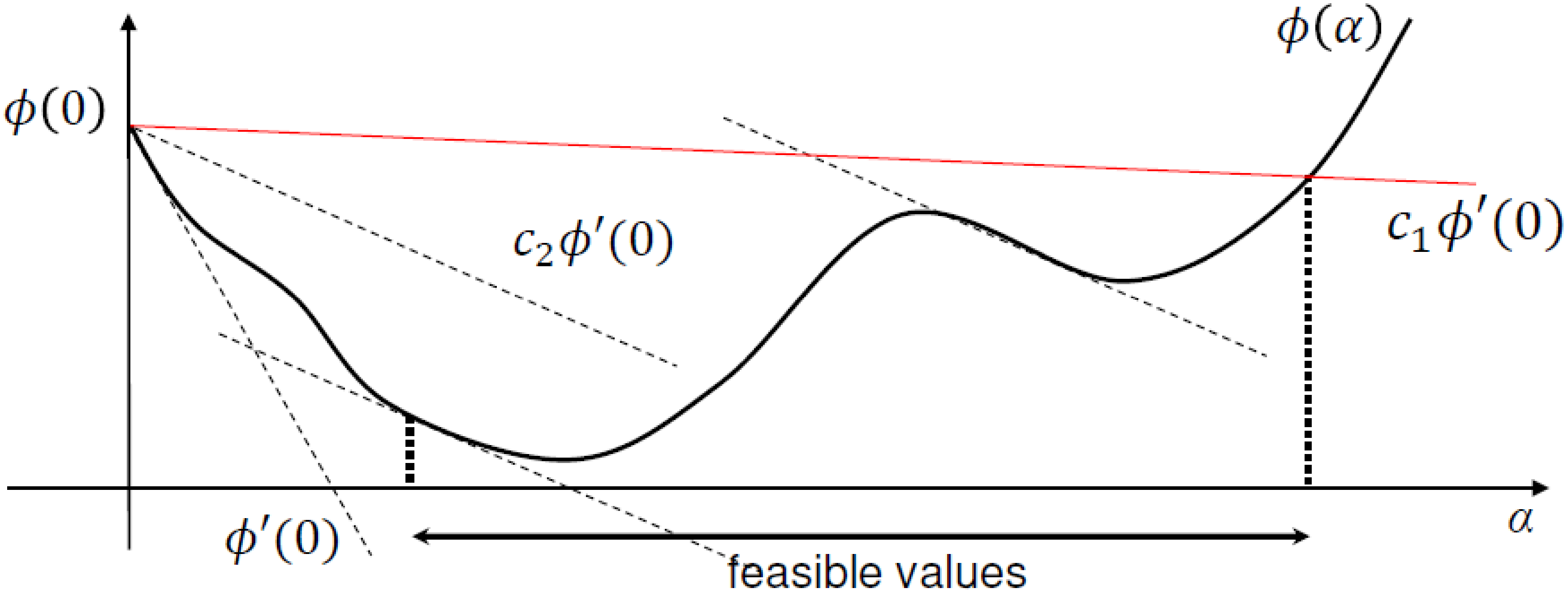
Condición de Wolfe

Teorema:

Sea f continuamente diferenciable, $\boldsymbol{p}^{(k)}$ una dirección descendiente, 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.

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 descendiente 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

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

Steepest-Descent

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

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

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

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

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

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

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

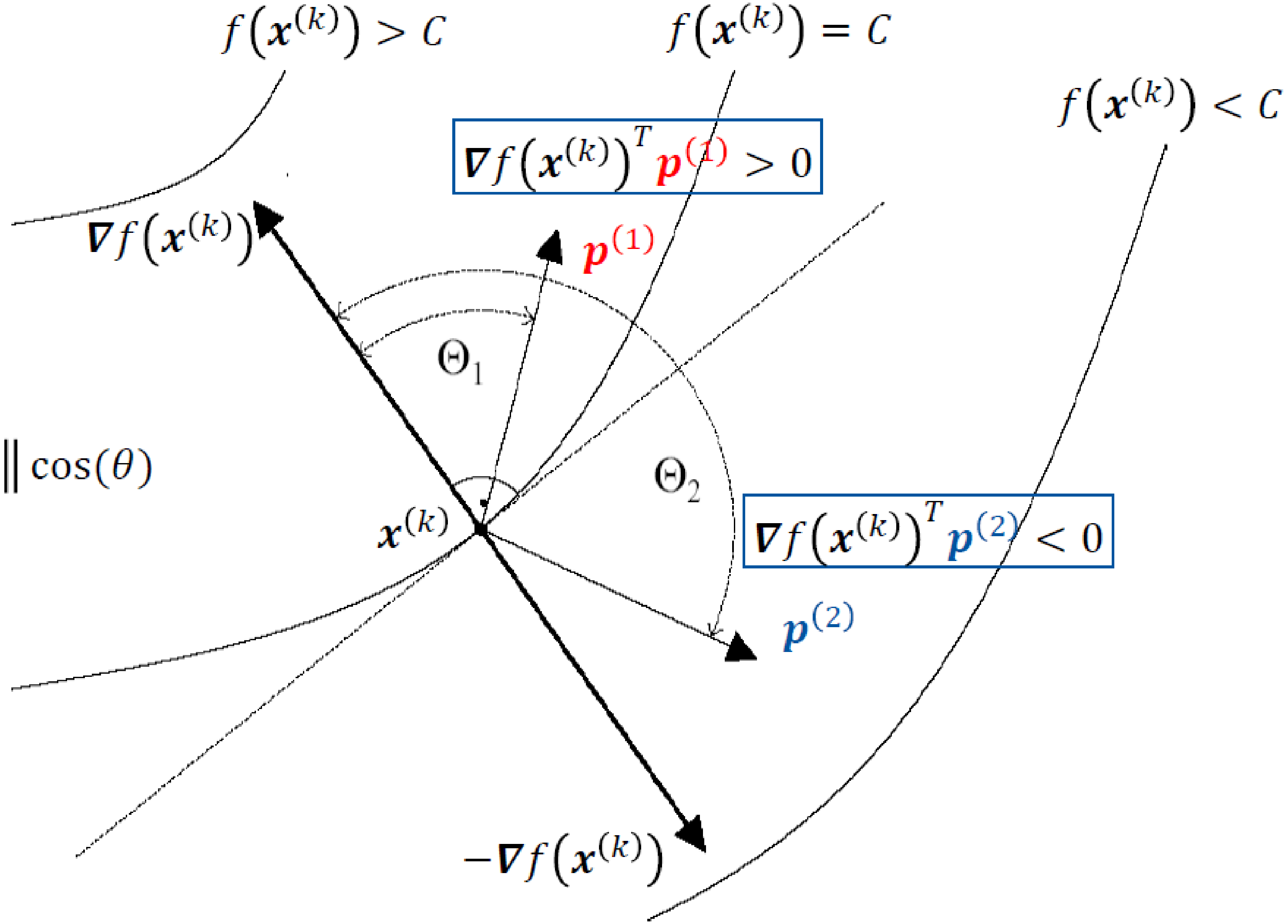
$$\Rightarrow \boldsymbol{p}^{(k)} = -\boldsymbol{\nabla} f(\boldsymbol{x}^{(k)}) / \left\|\boldsymbol{\nabla} f(\boldsymbol{x}^{(k)})\right\|$$

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

Steepest-Descent

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

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



Steepest-Descent

Algorithm:

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

for $k = 0, 1, \ldots$

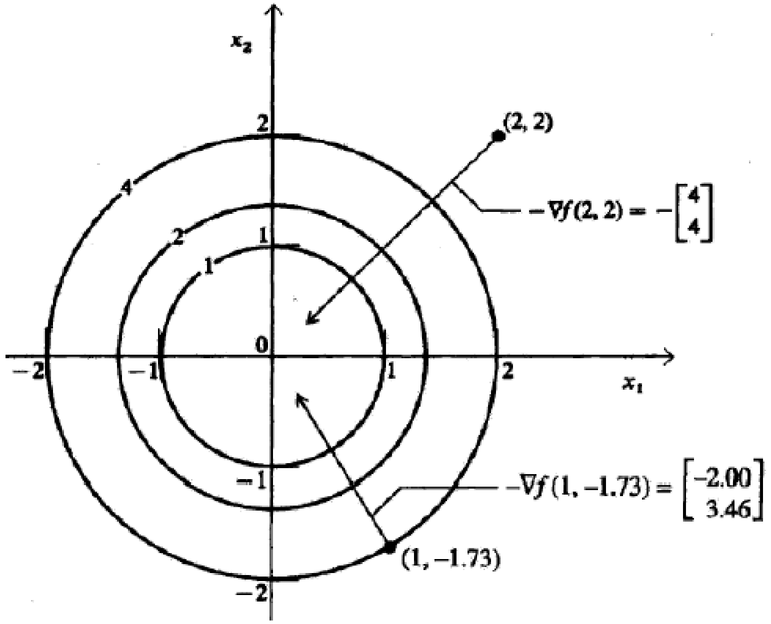
  if $\left\|\nabla f\left(\boldsymbol{x}^{(k)}\right)\right\| \leq \varepsilon$ stop, else

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

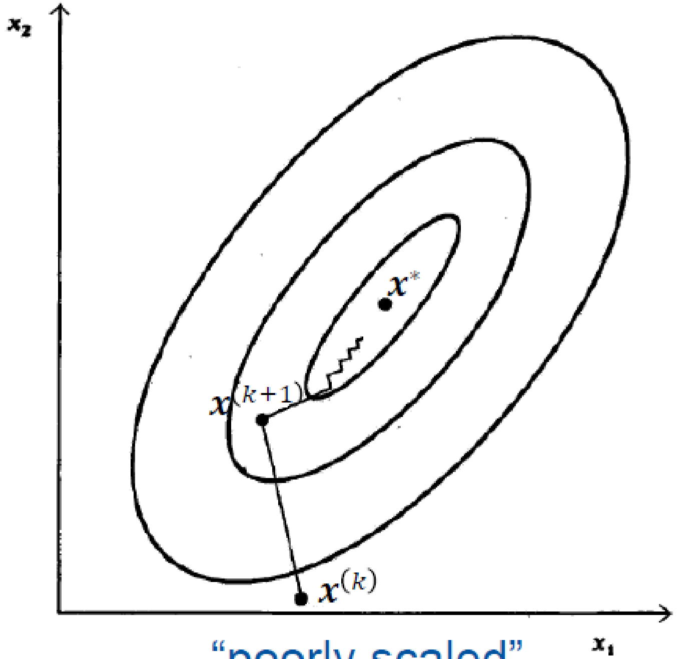
   determine the step length α_k (e.g. using the Armijo rule)

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

end for



"well scaled"



"poorly scaled"

Directions become perpendicular

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