Tema 6. Optimización numérica.

Computación Numérica

Antonio Palacio

Departamento de Matemáticas
Universidad de Oviedo

palacioantonio@uniovi.es

Curso 2021-2022

Contenidos I

Contenidos

Preliminares

Motivación

Dada una función

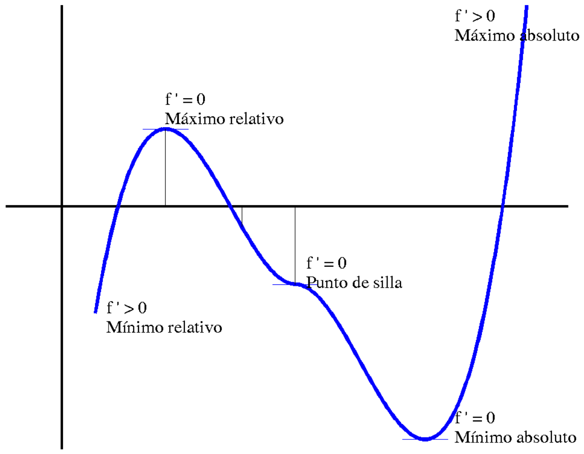
$$f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$$

nos planteamos el problema de encontrar los máximos y/o los mínimos de f en U. Cuando U es un conjunto abierto, se suele utilizar el término **Optimización sin restricciones** para este tipo de problemas.

Nota 6.1

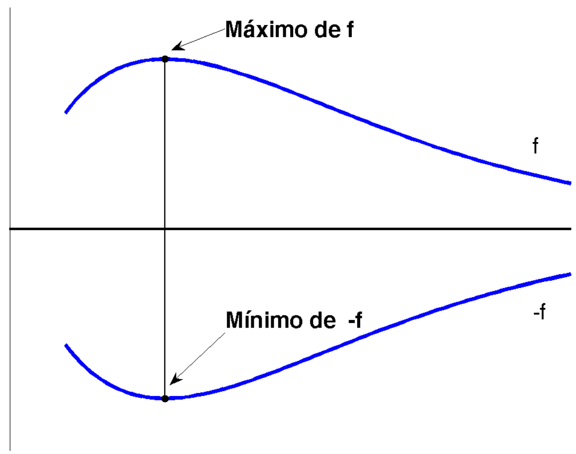
Un punto $x^* \in U$ se dice que es un ***mínimo relativo*** *de f si existe un entorno V de x^* tal que $f(x) \geq f(x^*)$ para todo $x \in V \cap U$.*
Un punto se denomina ***máximo relativo*** *si $f(x) \leq f(x^*)$ para todo $x \in V \cap U$.*
El máximo y el mínimo se dice que son ***absolutos*** *si $V = U$. Utilizaremos el término extremo para referirnos a los máximos y mínimos.*

Preliminares



Preliminares

Teniendo en cuenta que todo máximo de f es un mínimo para $-f$, limitaremos nuestro estudio al cálculo de mínimos.



Preliminares

Teorema 6.1 (Teorema de Weierstrass)

Sea $K \subset U$. Si K es compacto (cerrado y acotado) y f es continua en K entonces f alcanza un mínimo absoluto y un máximo absoluto en K.

Teorema 6.2

Sea $\mathbf{x}^*$ punto interior de U y supongamos que $f \in C^1(U)$.

- ***Condición necesaria de extremo**: Si x^* es un extremo relativo de f entonces:*

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

- ***Condición suficiente de mínimo**: Supongamos ahora que $f \in C^2(U)$ y que $\nabla f(\mathbf{x}^*) = 0$. Si se verifica que*

$$\mathbf{y}^t\, \nabla^2 f(\mathbf{x}^*)\, \mathbf{y} > 0 \text{ para todo } y \neq 0$$

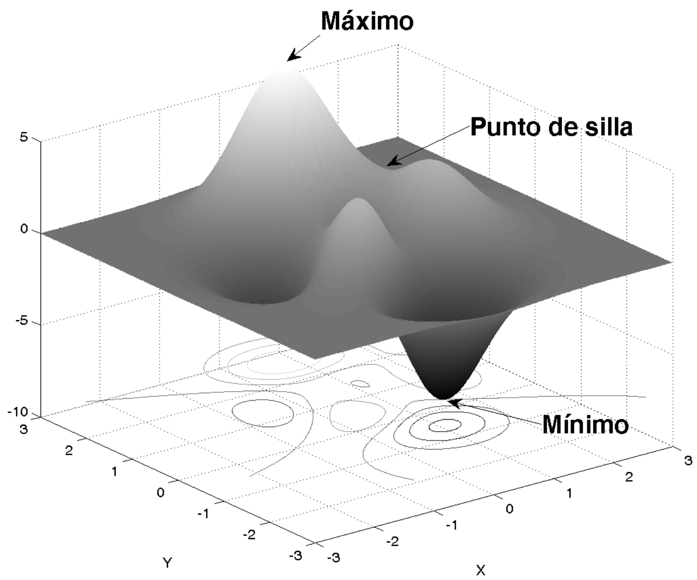
entonces x^* es un mínimo de f.

Preliminares

Observación 6.1

- *Los puntos que verifican la condición $\nabla f(\mathbf{x}) = 0$ reciben el nombre de **puntos estacionarios o críticos**. Pueden existir puntos estacionarios que no son extremos y suelen denominarse puntos de silla.*
- *La matriz $H(\mathbf{x}^*) = \nabla^2 f(\mathbf{x}^*) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}^*)\right)$ se denomina hessiano. La condición suficiente se puede expresar diciendo que el Hessiano es definido positivo.*
- *La condición suficiente de máximo es que el hessiano sea definido negativo.*
- *Observe que en dimensión 1 el hessiano está definido por $f''(x^*)$ y ser definido positivo significa que $f''(x^*) > 0$.*

Preliminares



Contenidos

Optimización en una variable

Introducción

Dada una función
$$g\ :\ [a,b] \longrightarrow \mathbb{R}$$
nos planteamos el problema de hallar sus mínimos en $[a,b]$.
Cuando g es derivable, la condición necesaria de extremo nos proporciona un procedimiento para obtener los mínimos en (a,b) ya que estos deben encontrarse entre las soluciones de la ecuación:
$$g'(x) = 0$$
La resolución de esta ecuación puede abordarse mediante los métodos numéricos vistos en el *Tema 2* (*bisección, Newton-Raphson, secante,...*). Sin embargo, en algunas ocasiones el uso de g' resulta muy complejo lo que hace necesario disponer de métodos que no requieran su cálculo. A continuación describiremos uno de estos procedimientos alternativos que consiste en reducir iterativamente la longitud del intervalo que contiene al mínimo mediante el análisis de los valores de g en ciertos puntos.

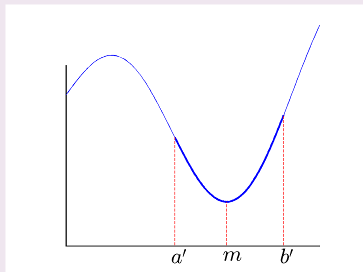
Optimización en una variable

Procedimiento

Supondremos que g es una función continua que posee un único mínimo en $x = m$, punto interior de su dominio. Por la continuidad, existe un intervalo $[a',b']$ con $m \in [a',b']$ y verificando que:

g **estrictamente decreciente en** $[a',m]$

g **estrictamente creciente en** $[m,b']$



Optimización en una variable

Metodología

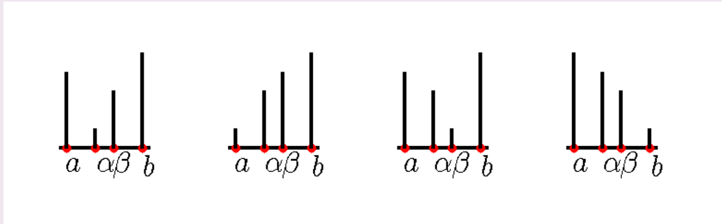
A partir de estas propiedades se puede diseñar un método similar al método de bisección consistente en reducir iterativamente la longitud del intervalo inicial $[a', b']$ (que denotaremos en lo que sigue $[a,b]$ para simplificar las notaciones) que contiene a m.

- Se eligen dos puntos α y β tal que $a < \alpha < \beta < b$.
- Se reduce la longitud del intervalo que contiene al mínimo evaluando g en α y β:
  1. Si $g(\alpha) < g(\beta) \Rightarrow m \in [a, \beta]$
  2. Si $g(\alpha) \geq g(\beta) \Rightarrow m \in [\alpha, b]$
- Se utiliza como aproximación de m el punto medio del intervalo resultante.

Optimización en una variable

Observación 6.1

- *La reducción del intervalo puede ser mejorada en algunas ocasiones. Por ejemplo, si en el caso 1 además se cumple que $g(\alpha) > g(a)$ podemos concluir que $m \in [a, \alpha]$. En el caso 2 se pueden plantear razonamientos análogos. Desde el punto de vista de la programación, esta mejora del intervalo no representa una ventaja significativa en la eficiencia del método.*
- *Eligiendo adecuadamente los puntos α y β se garantiza que la sucesión de puntos medios de los intervalos converge al mínimo. Por ejemplo, una elección muy sencilla sería $\alpha = (b-a)/3$ y $\beta = 2(b-a)/3$. Un elección no tan simple pero mas eficiente desde el punto de vista numérico es el método de la sección de oro.*



Optimización en una variable

Método de la sección de oro

Se reduce el tamaño del intervalo en cada iteración en una razón fija dada por $\Phi - 1$, siendo Φ el **número de oro (o razón áurea)**:

$$\Phi = \frac{\sqrt{5}+1}{2} \approx 1{,}6180339887 \qquad \text{y} \qquad \frac{1}{\Phi} = \Phi - 1 \approx 0{,}6180339887$$

- Datos iniciales: $[a_1, b_1] = [a, b]$, $\quad \Delta = (\Phi - 1)(b_1 - a_1)$.
- Se eligen $\alpha = b_1 - \Delta$, $\quad \beta = a_1 + \Delta$.
- Se aplica el método para obtener un nuevo intervalo $[a_2, b_2]$ cuya longitud es Δ y que debe coincidir con $[a_1, a_1 + \Delta]$ o con $[b_1 - \Delta, b_1]$.

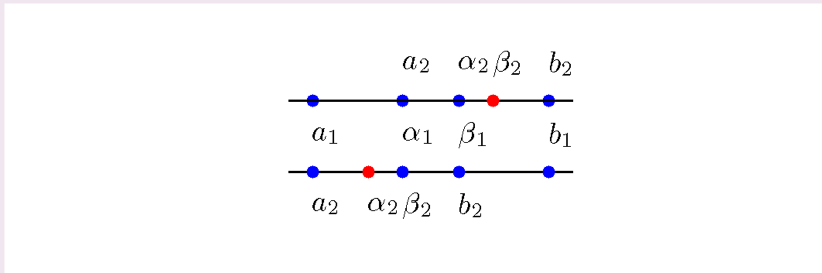
Del proceso anterior se deduce que $(b_n - a_n) = (\Phi - 1)^{n-1}(b - a)$.
Puesto que $0 < \Phi - 1 < 1$ se tiene que $\lim_{n\to\infty}(b_n - a_n) = 0$, lo que garantiza la convergencia del método.

Observación 6.1

La obtención de $[a_n, b_n]$ requiere realizar $n-1$ iteraciones.

Nota 6.2

El uso de la razón áurea hace que $\beta_2 = \alpha_1$ (si $[a_2, b_2] = [a_1, \beta_1]$) o bien $\alpha_2 = \beta_1$ (si $[a_2, b_2] = [\alpha_1, b_1]$).



Problema 6.1

Calcule una aproximación del mínimo de $g(x) = \dfrac{\operatorname{sen} x}{x}$, $x \in [\pi, 2\pi]$ aplicando dos iteraciones del método de la sección de oro. ¿Cuántas iteraciones del método garantizan que el valor que hace mínima la función está en un intervalo de longitud menor que una centésima?

Contenidos

Optimización en $\mathbb{R}^n$ sin restricciones

Introducción

Se considera ahora una función

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

que supondremos de clase dos. Un procedimiento para buscar los mínimos de f consiste en utilizar la condición necesaria de extremo y plantear la resolución del sistema:

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

Este sistema, que posiblemente sea no lineal, puede resolverse mediante las técnicas desarrolladas en el tema 3.
El objetivo de esta sección es presentar varios métodos iterativos que permitan evitar la resolución del sistema. Estos métodos adoptan la forma:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \mathbf{d}^k \tag{1}$$

siendo la elección del parámetro λ_k y del vector $\mathbf{d^k}$ característica de cada método.

Método de máxima pendiente

Método de máxima pendiente

Es conocido que el vector $\nabla f(\mathbf{x})$ determina la dirección y sentido en el que la función f crece más rápidamente en torno a $\mathbf{x}$. Si, partiendo de $\mathbf{x}^k$, queremos obtener un punto en el que la función tome un valor menor que $f(\mathbf{x}^k)$, en la fórmula (1) es natural elegir un vector $\mathbf{d}^k$ que verifique

$$\nabla f(\mathbf{x}^k) \cdot \mathbf{d}^k < 0$$

es decir, una *dirección de descenso*. Cuando se elige $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$, se obtiene el denominado **método de máxima pendiente**:

- Dato Inicial: $\mathbf{x}^0$
- Dado $\mathbf{x}^k$ se define $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$.
- Se calcula λ_k por $f(\mathbf{x}^k + \lambda_k \mathbf{d}^k) = \min\limits_{\lambda \geq 0} f(\mathbf{x}^k + \lambda \mathbf{d}^k)$
- Se define $\mathbf{x}^{k+1} = \mathbf{x}^k + \lambda_k \mathbf{d}^k$
- Test de Parada: $\|\nabla f(\mathbf{x}^{k+1})\| \leq Tol$

Métodos de máxima pendiente

Observación 6.2

- *En el proceso anterior, el escalar λ_k no se puede calcular, en general, de forma exacta. Suele utilizarse algún método de optimización para funciones de una variable.*
- *El método de máxima pendiente tiene, en general, convergencia de orden 1.*

Problema 6.2

Realice una iteración del método de máxima pendiente para calcular

$$\min_{(x,y)\in\mathbb{R}^2} \frac{x^4 + y^4 + 2xy}{4}$$

partiendo del punto $\mathbf{x}^0 = (1,\ 0)$. *Utilice un paso del método de la sección de oro para resolver* $\min_{\lambda\geq 0} f(\mathbf{x}^k + \lambda \mathbf{d}^k)$ *considerando que* $\lambda \in [0,1]$.