Numerical Optimization

Mathematical Background (II)

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Definition

Let $a, b \in \mathbb{R}$. The *closed interval* $[a, b]$ denotes the set, $\{x \in \mathbb{R} : a \leq x \leq b\}$. The set $\{x \in \mathbb{R} : a < x < b\}$ represents the *open interval* (a, b).

Examples:

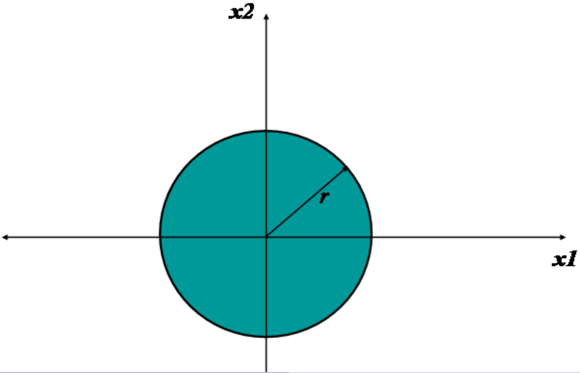
- $[-3, 2]$: a closed interval
- $(1, \infty)$: an open interval

Definition

Let $\mathbf{x}_0 \in \mathbb{R}^n$. A *norm ball* of radius $r > 0$ and centre $\mathbf{x}_0$ is given by $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$ and will be denoted by $B[\mathbf{x}_0, r]$.

Note: We will use $B(\mathbf{x}_0, r)$ to denote $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| < r\}$.

- $B[\mathbf{0}, \mathbf{r}]$



Definition

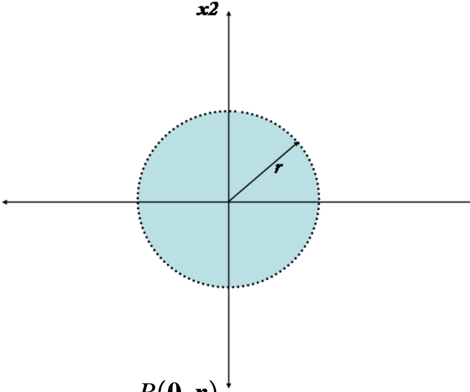
Let $x \in S \subseteq \mathbb{R}^n$. x is called an *interior point* of S if there exists $r > 0$ such that $B[x, r] \subseteq S$. The set of all such points interior to S is called the *interior* of S and is denoted by $\text{int}(S)$.

Definition

A set $S \subseteq \mathbb{R}^n$ is said to be an *open* set if $S = \text{int}(S)$.

Examples:

- $B(\mathbf{0}, \mathbf{r})$
- $(1, 2) \cup (3, 4)$ is an open subset of $\mathbb{R}$



$B(\mathbf{0}, \mathbf{r})$

Definition

A set $S \in \mathbb{R}^n$ is said to be *closed* if its complement in $\mathbb{R}^n$, $\mathbb{R}^n \backslash S = \{x \in \mathbb{R}^n : x \notin S\}$ is open.

Example: $[1,2] \cup [3,4]$ is a closed subset of $\mathbb{R}$

Definition

Let $S \subset \mathbb{R}^n$. $\mathbf{x} \in \mathbb{R}^n$ belongs to the *closure* of S,cl(S) if for each $\epsilon > 0$, $S \cap B[\mathbf{x}, \epsilon] \neq \phi$. The set S is said to be *closed* if $S = \text{cl}(S)$.

Example: Let $S = (1,2] \cup [3,4)$. Then $\text{cl}(S) = [1,2] \cup [3,4]$ and $\text{int}(S) = (1,2) \cup (3,4)$.

Remarks:

- If S is open, then $\text{int}(S) = S$.
- If S is closed, then $\text{cl}(S) = S$.

Definition

The boundary of a set S is defined as $\text{bd}(S) = \text{cl}(S) \backslash \text{int}(S)$.

Definition

A set $S \subset \mathbb{R}^n$ is said to be bounded if there exists R $(0 < R < \infty)$ and $\mathbf{x} \in \mathbb{R}^n$, such that $S \subset B(\mathbf{x}, R)$.

Examples:

- $(1, 2] \cup [3, 100)$: a bounded set
- $[0, \infty)$: not a bounded set

Definition

A set S in $\mathbb{R}^n$ is said to be *compact* if it is closed and bounded.

Example:

- $[0, 100] \cup [1000, 10000]$

Sequences

- $S \subseteq \mathbb{R}^n$
- $\{\mathbf{x}^k\}$: A sequence of points belonging to S

Definition

A sequence $\{\mathbf{x}^k\}$ *converges* to $\mathbf{x}^*$, if for any given $\epsilon > 0$, there is a positive integer K such that

$$\|\mathbf{x}^k - \mathbf{x}^*\| \leq \epsilon, \quad \forall\, k \geq K.$$

We write this as $\mathbf{x}^k \rightarrow \mathbf{x}^*$ or $\lim_{k\rightarrow\infty} \mathbf{x}^k = \mathbf{x}^*$.

Definition

A sequence $\{\mathbf{x}^k\}$ is called a *Cauchy sequence* if, for any given $\epsilon > 0$, there is a positive integer K such that $\|\mathbf{x}^k - \mathbf{x}^m\| \leq \epsilon$ for all $k, m \geq K$.

Sequences

Examples:

- The sequence $\{\mathbf{x}^k\}$ where $\mathbf{x}^k = (1 + 2^{-k}, 1/k)^T$ converges to $(1, 0)^T$.
- The sequence $\{x^k\}$ where $x^k = (-1)^k$ does not converge.

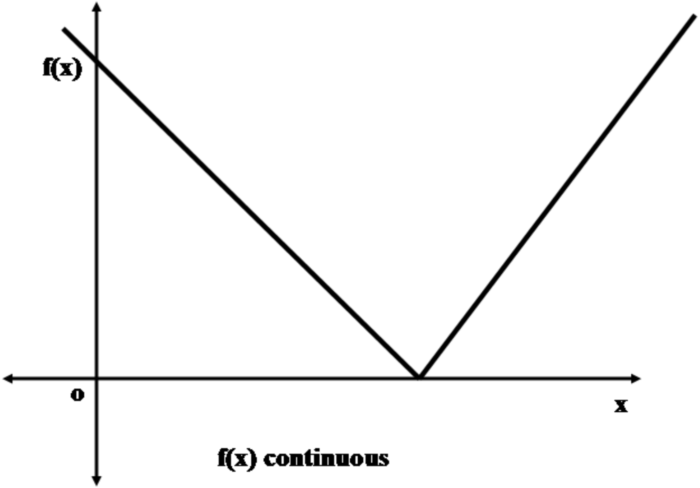
Continuous Functions

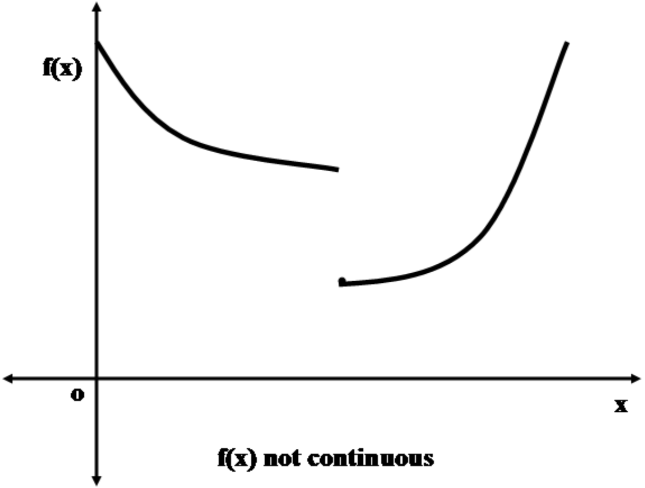
Definition

Let $S \subseteq \mathbb{R}^n$. A function $f : S \to \mathbb{R}$ is said to be *continuous* at $\bar{\mathbf{x}} \in S$ if for any given $\epsilon > 0$ there exists a $\delta > 0$ such that $\mathbf{x} \in S$ and $\|\mathbf{x} - \bar{\mathbf{x}}\| < \delta$ implies that $|f(\mathbf{x}) - f(\bar{\mathbf{x}})| < \epsilon$.

Note:

- The function f is said to be continuous on $A \subset \mathbb{R}^n$ if it is continuous at each point of A.
- When we say that f is continuous, we mean that f is continuous on its domain.
- $\mathcal{C}$: Class of all continuous functions


f(x)
0
x
f(x) continuous


f(x)
0
x
f(x) not continuous

Gradient

Definition

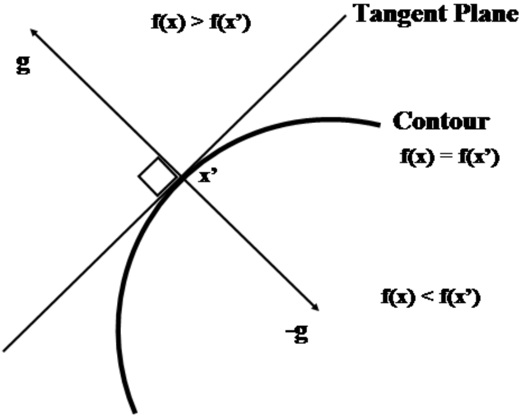
A continuous function, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, is said to be continuously differentiable at $\mathbf{x} \in \mathbb{R}^n$, if $\frac{\partial f}{\partial x_i}(\mathbf{x})$ exists and is continuous, $i = 1, \ldots, n$.

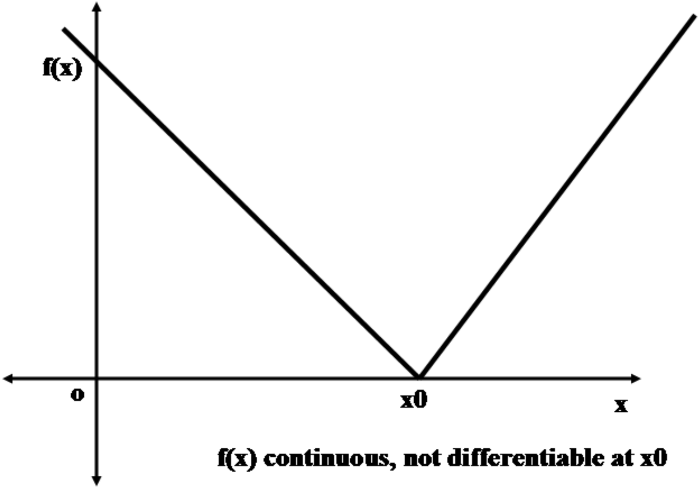
- $\mathcal{C}^1$: Class of functions whose first partial derivatives are continuous
- Assumption: $f \in \mathcal{C}^1$

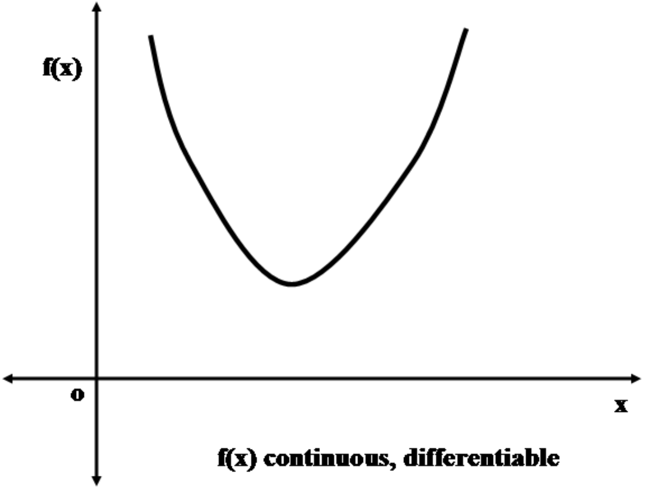
Definition

We define the *gradient* of f at $\mathbf{x}$ to be the vector

$$\mathbf{g}(\mathbf{x}) \equiv \nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)^T$$


f(x) > f(x')
Tangent Plane
g
Contour
f(x) = f(x')
x'
f(x) < f(x')
-g


f(x)
o
x0
x
f(x) continuous, not differentiable at x0


f(x)
0
x
f(x) continuous, differentiable

Directional Derivatives

Definition

Let $S \subset \mathbb{R}^n$ be an open set and $f : S \to \mathbb{R}$, f continuously differentiable in S. Then, for any $\mathbf{x} \in S$ and any nonzero $\mathbf{d} \in \mathbb{R}^n$, the *directional derivative* of f at $\mathbf{x}$ in the direction of $\mathbf{d}$, defined by

$$\frac{\partial f}{\partial d}(\mathbf{x}) \equiv \lim_{\epsilon \to 0} \frac{f(\mathbf{x} + \epsilon \mathbf{d}) - f(\mathbf{x})}{\epsilon}$$

exists and equals $\nabla f(\mathbf{x})^T \mathbf{d}$.

Define $\phi : \mathbb{R} \to \mathbb{R}$ as $\phi(t) = f(\mathbf{x} + t\mathbf{d})$.

$$\frac{d\phi}{dt}(\alpha) = \nabla f(\mathbf{x} + \alpha \mathbf{d})^T \mathbf{d}$$

Substituting $\alpha = 0$ gives

$$\frac{\partial f}{\partial d}(\mathbf{x}) = \nabla f(\mathbf{x})^T \mathbf{d}$$

Hessian

Definition

A continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *twice continuously differentiable* at $\mathbf{x} \in \mathbb{R}^n$, if $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ exists and is continuous.

- $\mathcal{C}^2$: Class of twice continuously differentiable functions

Definition

Let $f \in \mathcal{C}^2$. We define the *Hessian* of f at $\mathbf{x}$ to be the matrix

$$H(\mathbf{x}) \equiv \nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Hessian

$$H(\mathbf{x}) \equiv \nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Note: The Hessian matrix is symmetric.

Definition

Let $S \subset \mathbb{R}^n$ be an open set and $f : S \to \mathbb{R}$, f twice continuously differentiable in S. Then, for any $\mathbf{x} \in S$ and any nonzero $\mathbf{d} \in \mathbb{R}^n$, the *second directional derivative* of f at $\mathbf{x}$ in the direction of $\mathbf{d}$ equals $\mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d}$.

Example

- Consider the Rosenbrock function,

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

The gradient of f at $\mathbf{x} = (x_1, x_2)^T$ is

$$\mathbf{g}(\mathbf{x}) \equiv \nabla f(\mathbf{x}) = \begin{pmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$

The Hessian of f at $\mathbf{x} = (x_1, x_2)^T$ is

$$H(\mathbf{x}) \equiv \nabla^2(\mathbf{x}) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

Example

- Consider the function,

$$f(x_1, x_2) = x_1 e^{(-x_1^2 - x_2^2)}$$

The gradient of f at $\mathbf{x} = (x_1, x_2)^T$ is

$$\mathbf{g}(\mathbf{x}) \equiv \nabla f(\mathbf{x}) = \begin{pmatrix} (1 - 2x_1^2)e^{(-x_1^2 - x_2^2)} \\ -2x_1 x_2 e^{(-x_1^2 - x_2^2)} \end{pmatrix}$$

The Hessian of f at $\mathbf{x} = (x_1, x_2)^T$ is

$$H(\mathbf{x}) \equiv \nabla^2(\mathbf{x}) = \begin{pmatrix} (4x_1^3 - 6x_1)e^{(-x_1^2 - x_2^2)} & -2x_2(1 - 2x_1^2)e^{(-x_1^2 - x_2^2)} \\ -2x_2(1 - 2x_1^2)e^{(-x_1^2 - x_2^2)} & (4x_1 x_2^2 - 2x_1)e^{(-x_1^2 - x_2^2)} \end{pmatrix}$$

Taylor Series

$\mathcal{C}^\infty$: Class of all functions for which the derivative of any order is continuous.

Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^\infty$.

Let x^0 be the point about which we write the Taylor series.

$$f(x) = f(x^0) + f'(x^0)(x - x^0) + \frac{1}{2}f''(x^0)(x - x^0)^2 + \ldots$$

Suppose we use only $f'(x^0)$. Then $f(x)$ at x^0 can be approximated by

$$f(x) \approx f(x^0) + f'(x^0)(x - x^0).$$

Similarly, using $f'(x^0)$ and $f''(x^0)$, then the quadratic approximation of f at x^0 is

$$f(x) \approx f(x^0) + f'(x^0)(x - x^0) + \frac{1}{2}f''(x^0)(x - x^0)^2.$$

Truncated Taylor Series (First Order)
Let $f : \mathbb{R}^n \rightarrow \mathbb{R}, f \in \mathcal{C}^1, \mathbf{x}^0 \in \mathbb{R}^n$.
Then, for every $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\bar{\mathbf{x}}) \cdot (\mathbf{x} - \mathbf{x}^0)$$

where $\bar{\mathbf{x}}$ is some point that lies on the line segment joining $\mathbf{x}$ and $\mathbf{x}^0$; $\bar{\mathbf{x}}$ depends on $\mathbf{x}$, $\mathbf{x}^0$ and f.

Truncated Taylor Series (Second Order)
Let $f : \mathbb{R}^n \rightarrow \mathbb{R}, f \in \mathcal{C}^2, \mathbf{x}^0 \in \mathbb{R}^n$.
Then, for every $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0) \cdot (\mathbf{x} - \mathbf{x}^0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)^T \nabla^2 f(\bar{\mathbf{x}})(\mathbf{x} - \mathbf{x}^0)$$

where $\bar{\mathbf{x}}$ is some point that lies on the line segment joining $\mathbf{x}$ and $\mathbf{x}^0$; $\bar{\mathbf{x}}$ depends on $\mathbf{x}$, $\mathbf{x}^0$ and f.

Proofs of Theorems

- $A \Rightarrow B$
  - If A is true, then B is true.
  - *Direct Proof* : Assume A and derive B.
  - *Proof by contradiction* : Assume “not B" and derive “not A"
- $A \Longleftrightarrow B$
  - A if and only if B
  - B is a necessary and sufficient condition for A.
  - We must prove $A \Rightarrow B$ and $B \Rightarrow A$.

Proof by Mathematical Induction

Induction Principle: Let $N = \{1, 2, \ldots\}$ denote the set of natural numbers and let $M \subset N$. If the following properties hold:

1. 1 is in M, and
2. if n is in M, then $n + 1$ is in M,

then, $M = N$.

Example: Define $S_n = 1 + 2 + \ldots + n, \quad n = 1, 2, 3, \ldots$

Claim: $S_n = \frac{n(n+1)}{2}, \quad n = 1, 2, 3, \ldots$

Let M denote the set of natural numbers for which the above claim is true.

If $n = 1$, $S_1 = 1 = \frac{1 \times 2}{2}$. Hence 1 is in M.

Now, assume that n is in M and consider S_{n+1},

$S_{n+1} = S_n + (n + 1) = \frac{n(n+1)}{2} + (n + 1) = \frac{(n+1)(n+2)}{2}$.

So, $n + 1$ is in M.

From the induction principle, $M = N$ and hence the claim is proved.

Some References

- Thomas G. B. and Finney R. L., Calculus and Analytic Geometry, Addison Wesley, 1995.
- Royden H. L., Real Analysis, Prentice Hall, 1988.