Optimization

Fourth Class

2020 - 2021

By

Dr. Jawad Mahmoud Jassim

Dept. of Math.

Education College for Pure Sciences

University of Basrah

Iraq

Chapter One

Basic Concepts

Lecture 7

<u>Theorem (8):</u>

If

$\mathbf{1}$: $\boldsymbol{f}$: $\boldsymbol{R^n} \rightarrow \boldsymbol{R}$ $\boldsymbol{has\ continuous\ second\ partial\ derivatives\ in\ an\ open\ convex\ set\ D \subset R^n}$.

2: $\boldsymbol{X^*}$ **is a local minimizer of** $\boldsymbol{f\ over\ D}$.

Then the Hessian matrix $\boldsymbol{G(X^*)}$ **is a positive semi – definite.**

<u>Theorem (9):</u>

If

$\mathbf{1}$: $\boldsymbol{f}$: $\boldsymbol{R^n} \rightarrow \boldsymbol{R}$ $\boldsymbol{has\ continuous\ second\ partial\ derivatives\ in\ an\ open\ convex\ set\ D \subset R^n}$.

2: $\boldsymbol{X^*}$ **is a strict local minimizer of** $\boldsymbol{f\ over\ D}$.

Then the Hessian matrix $\boldsymbol{G(X^*)}$ **is a positive semi – definite at least.**

Theorem (10):

If

1: $\boldsymbol{f}$: $\boldsymbol{R^n \rightarrow R}$ ***has continuous second partial derivatives in an open convex set*** $\boldsymbol{D \subset R^n}$.

2: $\boldsymbol{X^*}$ **is a local maximizer of** $\boldsymbol{f}$ ***over*** $\boldsymbol{D}$.

Then the Hessian matrix $\boldsymbol{G(X^*)}$ **is negative semi – definite.**

Theorem (11):

If

1: $\boldsymbol{f}$: $\boldsymbol{R^n \rightarrow R}$ ***has continuous second partial derivatives in an open convex set*** $\boldsymbol{D \subset R^n}$.

2: $\boldsymbol{X^*}$ **is a strict local maximizer of** $\boldsymbol{f}$ ***over*** $\boldsymbol{D}$.

Then the Hessian matrix $\boldsymbol{G(X^*)}$ **is negative semi – definite at least.**

Note (16):

The critical points for the convex differentiable function are local minimizers or global minimizers.

<u>Theorem (12):</u>

Let $f: D \subset R^n \to R$, where D be a nonempty convex set and X^* is a local minimizer of $f\ over\ D$.

Then:

1: If f is convex the point X^* is also global minimizer.

2: If f is strictly convex the point X^* is unique global minimizer.

<u>Theorem (13):</u>

Let $f: R^n \to R$ be a differentiable convex function. Then the point X^* is a global minimizer of $f\ if\ and\ only\ if\ g(X^*) = 0$, where $g(X^*)$ is the gradient vector of $f\ at\ X^*$.

13: Structure of Optimization Problem

Usually, the optimization method is an iterative one for finding the minimizer of an optimization problem. The basic idea is that, given $an\ initial\ point\ X_0 \epsilon R^n$, one generates an iterate sequence $\{X_k\}$ by means of some iterative rule, such that when $\{X_k\}$ is a finite sequence, the last point is the optimal solution of the problem. When $\{X_k\}$ is infinite, it has a limit point which is the optimal solution of the problem. A typical behavior of an algorithm which is regarded as acceptable is that iterates X_k moves steadily towards the neighborhood of a local minimizer X^* and then rapidly converge to the point X^* when a given convergence rule is satisfied, iterates will be terminated.

In general, the most natural stopping criteria $is\ \|g(X_k)\| \leq\ \varepsilon$, where ε is a prescribed tolerance. If $\|g(X_k)\| \leq\ \varepsilon$ is satisfied it implies that the gradient vector $g(X_k)$ tends to zero and the iterate sequence $\{X_k\}$ converges to a critical point.

Let X_k be k^{th} iterate, d_k k^{th} direction and α_k k^{th} step length factor. Then $X_{k+1} = X_k + \alpha_k d_k$.

Most optimization methods are so – called descent methods in sense that f satisfies at each iteration $f(X_{k+1}) = f(X_k + \alpha_k d_k) < f(X_k)$ in which d_k is a descent direction.

Definition (17): (Descent Direction)

Let $f\colon\ R^n \rightarrow R$ have first partial derivatives $for\ all\ X \epsilon R^n$. If there exists a vector $d \epsilon R^n$ such that $g(X)^T d < 0$, where $g(X)$ is the gradient vector of $f\ at\ X$, then d is called a descent direction.

The basic scheme of optimization methods is as follows:

<u>Algorithm (1): (Basic Scheme Algorithm)</u>

<u>Step 0: (Initial Step)</u>

Given initial point $X_0 \epsilon R^n$ and the tolerance $\varepsilon > 0$.

<u>Step 1: (Termination Criterion)</u>

If $\|g(X_k)\| \leq \varepsilon$, stop.

<u>Step 2: (Finding the Direction)</u>

According to some iterative scheme, find d_k $which\ is\ a\ descent\ direction$.

<u>Step 3: (Line Search)</u>

Determine the step size α_k such that the objective function value decreases, i.e. $f(X_k + \alpha_k d_k) < f(X_k)$.

<u>Step 4: (Loop)</u>

Set $X_{k+1} = X_k + \alpha_k d_k$, $k = k + 1$ and go to step 1.

14: Convergence Rate

Let the iterate sequence $\{X_k\}$ generated by an algorithm converge to X^* in some norm, i.e. $\lim_{k\to\infty}\|X_k - X^*\| = 0$.

If there are real number $\alpha \geq 1$ and a positive constant β which is independent of the *iterative number* k, such that

$$\lim_{k\to\infty}\frac{\|X_{k+1}-X^*\|}{\|X_k-X^*\|^{\alpha}} = \beta.$$

We say that $\{X_k\}$ has $\alpha - order$ of $Q - convergence\ rate$, where $Q - convergence\ rate$ means Quotient – convergence rate. In particular,

1: When $\alpha = 1$ *and* $\beta \in (0,1)$, the sequence $\{X_k\}$ is said to *converge Q – linearly*.

2: When $\alpha = 1$ *and* $\beta = 0$ or $1 < \alpha < 2$ *and* $\beta > 0$, the sequence $\{X_k\}$ *is said to converge Q – super linearly*.

3: *When* $\alpha = 2$, we say that the sequence $\{X_k\}$ has *Q – quadratic converge rate*.

<u>Note (17):</u>

$\boldsymbol{Usually, if\ the\ convergence\ rate\ of\ an\ algorithm\ is\ Q-\ super\ linearly\ or\ Q-\ quadratic\ , we\ say\ that\ it\ has\ rapid\ convergence\ rate}$.

<u>Theorem (14):</u>

If the sequence $\{X_k\}$ $\boldsymbol{converges\ Q-\ super\ linearly\ to\ X^*}$ **, then**

$$\lim_{k \to \infty} \frac{\|X_{k+1}-X^k\|}{\|X_k-X^*\|^\alpha} = 1.$$

<u>Note (18):</u>

In general, the converse of Theorem (14) is not true.