

Convention: Throughout this discussion a feasible direction \mathbf{d} at a point is by definition taken to be a nonzero vector, although there is no significant harm even if assumed otherwise. Sometimes I may have forgotten to explicitly write it.

Notation: $\nabla^2 f(\mathbf{x}) = H(\mathbf{x})$ is the Hessian matrix of f at \mathbf{x} .

$\nabla f(\mathbf{x})$ is the gradient **row** vector (or the vector of partial derivatives written as a row vector).

Nonlinear Programming

Let f be a real valued function defined on $\Omega \subseteq \mathbb{R}^n$.

The problem is to minimize or maximize $f(\mathbf{x})$ subject to $\mathbf{x} \in \Omega$, where f need not be a linear function.

Throughout the discussion we will assume that $\Omega \subseteq \mathbb{R}^n$, for some n .

Definition 1: A point (or an element) $\mathbf{x}^* \in \Omega$ is called a local minimum (maximum) of f if there exists an $\epsilon > 0$, such that

$\mathbf{x} \in \Omega$ and $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$ implies $f(\mathbf{x}^*) \leq f(\mathbf{x})$ ($f(\mathbf{x}^*) \geq f(\mathbf{x})$).

Definition 2: A point $\mathbf{x}^* \in \Omega$ is called a global minimum (maximum) of f if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ ($f(\mathbf{x}^*) \geq f(\mathbf{x})$) for all $\mathbf{x} \in \Omega$.

Definition 3: A vector $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$ is said to be a feasible direction at $\mathbf{x}^* \in \Omega$, if there exists a $c > 0$ such that for all t , $0 \leq t \leq c$, $\mathbf{x}^* + t\mathbf{d} \in \Omega$.

Example 1: Let $\Omega = \{[x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$.

At $[0, 0]^T$ if $\mathbf{d} = [d_1, d_2]^T$ is a feasible direction then $d_1 \geq 0$ and $d_2 \geq 0$.

At $[0, \frac{1}{2}]^T$ if \mathbf{d} is a feasible direction then $d_1 \geq 0$ but d_2 can be any real number.

At $[\frac{1}{2}, 0]^T$ if \mathbf{d} is a feasible direction then $d_2 \geq 0$ but d_1 can be any real number.

At $[\frac{1}{2}, \frac{1}{2}]^T$ any $\mathbf{d} \in \mathbb{R}^2$ will be a feasible direction.

Remark 1: If \mathbf{x}^* is an interior point of Ω then any $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$ is a feasible direction at \mathbf{x}^* .

First order necessary conditions for a point to be a local minimum

The results obtained in this section is based on **first order approximation** of the function f near the local minimum point \mathbf{x}^* .

Throughout this discussion we will assume $\Omega \subseteq \mathbb{R}^n$ and \mathbf{x}^*, \mathbf{d} are elements of \mathbb{R}^n for some $n \in \mathbb{N}$.

Theorem 1: Let $f : \Omega \rightarrow \mathbb{R}$ be a **continuously differentiable function** (that is, the first order partial derivatives of f exists and are continuous as functions from Ω to \mathbb{R}). If \mathbf{x}^* is a **local minimum point** then for any **feasible direction** \mathbf{d} at \mathbf{x}^* ,

$$\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0,$$

where $\nabla f(\mathbf{x}^*)$, the gradient vector of f at \mathbf{x}^* is written as a row vector (the components of $\nabla f(\mathbf{x}^*)$ are the first order partial derivatives of f at \mathbf{x}^*) and $\mathbf{d} \in \mathbb{R}^n$ is a column vector.

Proof: Let \mathbf{x}^* be a local minimum and let \mathbf{d} be a feasible direction at \mathbf{x}^* .

Let $g(t) = f(\mathbf{x}(t))$, where $\mathbf{x}(t) = \mathbf{x}^* + t\mathbf{d}$.

Since f is continuously differentiable throughout Ω and \mathbf{d} is a **feasible direction** at \mathbf{x}^*

$\lim_{h \rightarrow 0} \frac{f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)}{h}$ exists.

Since $g(h) - g(0) = f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)$, $\lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h}$ also exists and

$$g'(0) = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)}{h} = \nabla f(\mathbf{x}^*)\mathbf{d}.$$

Since $g'(0)$ exists and we look at the first order Taylor's approximation of g around $t = 0$,

$$g(t) = g(0) + tg'(0) + o(t), \text{ where } \lim_{t \rightarrow 0} \frac{o(t)}{t} = 0. \quad (**)$$

If we take $0 < t \leq c$, then $(**)$ gives,

$$f(\mathbf{x}^* + t\mathbf{d}) = f(\mathbf{x}^*) + t\nabla f(\mathbf{x}^*)\mathbf{d} + o(t).$$

Since for t sufficiently small,

$$|t\nabla f(\mathbf{x}^*)\mathbf{d}| \geq |o(t)| \text{ (provided } \nabla f(\mathbf{x}^*)\mathbf{d} \neq 0),$$

if $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$, $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*)$ for all $t > 0$ sufficiently small, which contradicts that \mathbf{x}^* minimizes f locally.

Theorem 2 : Let $f : \Omega \rightarrow \mathbb{R}$ be a **continuously differentiable** function. Let \mathbf{x}^* be an **interior point** of Ω . If \mathbf{x}^* is a **local minimum point** of f then

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

Proof: Follows from **Theorem 1**, by taking $\mathbf{d} = -(\nabla f(\mathbf{x}^*))^T$. Since \mathbf{x}^* is an interior point we know that every $\mathbf{d} \in \mathbb{R}^n, \mathbf{d} \neq \mathbf{0}$ is a feasible direction at \mathbf{x}^* .

Second order necessary conditions for a point to be a local minimum

The following conditions are obtained by considering second order approximation of the function f near the local minimum point \mathbf{x}^* .

Theorem 3: Let $f : \Omega \rightarrow \mathbb{R}$ be **twice continuously differentiable** (that is all the second order partial derivatives of f (given by $\frac{\partial^2 f}{\partial x_j \partial x_i}$) exists and are continuous as functions from Ω to \mathbb{R}).

If \mathbf{x}^* is a **local minimum** of f then for any **feasible direction** \mathbf{d} at \mathbf{x}^*

1. $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$.
2. If $\nabla f(\mathbf{x}^*)\mathbf{d} = 0$, then $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$.

Note: The matrix $\nabla^2 f$ (also denoted by H) is called the **Hessian matrix** of f ,

$(\nabla^2 f(\mathbf{x}^*))_{ij} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^*) = \left[\frac{\partial^2 f}{\partial x_j \partial x_i} \right]_{\mathbf{x}^*}$. Since f **twice continuously differentiable** the **Hessian matrix** $\nabla^2 f$ is a symmetric matrix for all $\mathbf{x} \in \Omega$.

Proof: Let \mathbf{d} be a **feasible direction** at \mathbf{x}^* . That \mathbf{x}^* satisfies condition 1 is already shown in Theorem 1.

As before, take $g(t) = f(\mathbf{x}(t))$, where $\mathbf{x}(t) = \mathbf{x}^* + t\mathbf{d}$.

Since f is **twice continuously differentiable** $g''(0)$ exists. Also the **second order Taylor's approximation** of g around $t = 0$ gives,

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2!}g''(0) + o(t^2), \quad (**)$$

Since $g'(t) = (\nabla f(\mathbf{x}^* + t\mathbf{d}))\mathbf{d} = \sum_i (\frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d})d_i$,

$$g'(t) = \sum_i h_i(t)d_i,$$

$$\text{where } h_i(t) = \frac{\partial f}{\partial x_i}(\mathbf{x}^* + t\mathbf{d}).$$

$$\text{Hence } g''(t) = \sum_i h'_i(t)d_i,$$

$$\text{where } h'_i(t) = ((\nabla \frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d}))\mathbf{d} = \sum_j \frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d})d_j = \sum_j (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^* + t\mathbf{d}))d_j.$$

$$\text{Hence } g''(t) = \sum_i h'_i(t)d_i = \sum_i (\sum_j (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^* + t\mathbf{d}))d_j)d_i,$$

$$\text{Hence } g''(0) = \sum_i (\sum_j (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^*))d_j)d_i = \mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d},$$

where $\nabla^2 f(\mathbf{x}^*) (= H(\mathbf{x}^*))$ is an $n \times n$ matrix whose (i, j) th entry is given by $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}^*)$.

Note that because we have assumed f to be **twice continuously differentiable** the matrix $\nabla^2 f(\mathbf{x}^*)$ is a symmetric matrix.

Again from $(**)$ we get,

$$f(\mathbf{x}^* + t\mathbf{d}) = f(\mathbf{x}^*) + t\nabla f(\mathbf{x}^*)\mathbf{d} + \frac{t^2}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} + o(t^2).$$

Since for sufficiently small t ,

$$|\frac{t^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d}| \geq |o(t^2)|, \text{ (provided } \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \neq 0 \text{)}$$

hence if $\nabla f(\mathbf{x}^*) \mathbf{d} = 0$ and \mathbf{x}^* is a **local minimum** then it should satisfy the condition $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$ if $\nabla f(\mathbf{x}^*) \mathbf{d} = 0$.

Theorem 4: Let $f : \Omega \rightarrow \mathbb{R}$ be a twice continuously differentiable function and let \mathbf{x}^* be an interior point of Ω . If \mathbf{x}^* is a local minimum of f then

1. $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
2. $\nabla^2 f(\mathbf{x}^*)$ is positive semidefinite(defined later).

Proof: Follows from the previous theorem and the fact that for an **interior point** of Ω , every nonzero vector $\mathbf{d} \in \mathbb{R}^n$ is a feasible direction.

Hence from **Theorem 3**, $\nabla f(\mathbf{x}^*)(-\nabla f(\mathbf{x}^*))^T \geq 0$, which implies $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Applying the second order conclusion of **Theorem 3**, we get $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$ for all $\mathbf{d} \in \mathbb{R}^n$.

Definition: A real symmetric matrix A is said to be positive semidefinite (negative semidefinite) if $\mathbf{x}^T A \mathbf{x} \geq 0$ ($\mathbf{x}^T A \mathbf{x} \leq 0$) for all $\mathbf{x} \in \mathbb{R}^n$.

Definition: A real symmetric matrix A is said to be positive definite (negative definite) if $\mathbf{x}^T A \mathbf{x} > 0$ ($\mathbf{x}^T A \mathbf{x} < 0$) for all **nonzero** vectors $\mathbf{x} \in \mathbb{R}^n$.

Remark: Note that in general a matrix satisfying the condition $\mathbf{x}^T A \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ need not be symmetric for example $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Theorem: If A is a symmetric, $n \times n$, real matrix then the following statements are equivalent:

1. A is positive semidefinite.
2. All eigenvalues of A are nonnegative.
3. All principal minors of A are nonnegative.

Definition: λ is called an eigenvalue of an $n \times n$ matrix A if there exists an $\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}$ (that is atleast one component of \mathbf{x} is nonzero) such that $A\mathbf{x} = \lambda\mathbf{x}$.

For example the $\mathbf{0}$ matrix has all n eigenvalues equal to 0, the identity matrix I_n has all n eigenvalues equal to 1 and for an upper triangular matrix the diagonal entries are its eigenvalues.

Definition: If A is an $n \times n$ matrix and $\alpha \subseteq \{1, \dots, n\}, \beta \subseteq \{1, \dots, n\}$ then $A[\alpha, \beta]$ is the (sub)matrix obtained from A by deleting all the rows of A which do not belong to α and by deleting all the columns of A which do not belong to β .

If $\alpha = \beta$ then $A[\alpha, \alpha]$ is called a principal submatrix of A and $\det A[\alpha, \alpha]$ is called a principal minor of A .

For example if $\alpha = \beta = \{i\}$ where $i \in \{1, \dots, n\}$ then $A[\alpha, \alpha] = [a_{ii}]$ and $\det A[\alpha, \alpha] = a_{ii}$, the i th diagonal entry.

If $\alpha = \beta = \{i, j\}$ where $i, j \in \{1, \dots, n\}$ and $i < j$ then $A[\alpha, \alpha] = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}$.

If $\alpha = \{1, \dots, n\}$ then $A[\alpha, \alpha] = A$ and $\det A[\alpha, \alpha] = \det(A)$.

Remark: A nonsingular (nonzero determinant) positive semidefinite matrix is positive definite.

In the following examples there is a slight abuse of notation. Instead of writing $f([x_1, x_2]^T)$, to avoid cumbersome notation, I have written it as $f(x_1, x_2)$.

Example 1 revisited: Consider the following problem:

Minimize $f(x_1, x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2$

subject to $x_1 \geq 0, x_2 \geq 0$, hence $\Omega = \{[x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$.

To search for points which might be local minimum points of f we initially search for points in Ω which satisfy first order necessary conditions for a local minimum. The points which obviously satisfy first order necessary conditions for a local minimum are the points at which the gradient vector vanishes.

Hence we try to find \mathbf{x} at which $\nabla f(\mathbf{x}) = 0$.

$\nabla f(\mathbf{x}) = (3x_1^2 - 2x_1 x_2, -x_1^2 + 4x_2) = [0, 0]$ has two solutions $x_1 = 0, x_2 = 0$ and $x_1 = 6, x_2 = 9$.

Here $[6, 9]^T$ an interior point of the feasible region $\Omega = \{[x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$ satisfies the first order necessary conditions for a local minimum.

But since $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{pmatrix}$.

At $\mathbf{x}^* = [6, 9]^T$, $\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 18 & -12 \\ -12 & 4 \end{pmatrix}$ is not positive semidefinite.

Hence $\mathbf{x}^* = [6, 9]^T$ is not a local minimum point of f .

Hence the first order necessary conditions are **necessary** but **not sufficient** for a point to be a local minimum.

At $\mathbf{x}^* = [0, 0]^T$ a $\mathbf{d} \neq \mathbf{0}$ is a feasible direction if and only if $d_1 \geq 0$ and $d_2 \geq 0$.

Since $\nabla f(\mathbf{x}^*) = \mathbf{0}$, $\nabla f(\mathbf{x}^*)\mathbf{d} = 0$ for all \mathbf{d} .

Since $\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$, $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} = 4d_2^2 \geq 0$ for all \mathbf{d} .

Since $[0, 0]^T$ satisfies both the first and second order necessary conditions, we can only say that $\mathbf{x}^* = [0, 0]^T$ can be a candidate for local minimum, but since these are only necessary conditions we cannot conclude from previous calculations that $[0, 0]^T$ is indeed a local minimum of f in Ω .

Exercise: Check that $[0, 0]^T$ is a local minimum point of f in **Example 1**.

Solution: Note that $f(x_1, x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2 = x_1^2(x_1 - x_2) + 2x_2^2$ can take values < 0 only when either $x_2 = 0$ and $x_1 < 0$ or when $x_2 \neq 0$ and $x_2 > x_1$. But $[-1, 0]^T$ is not a feasible direction at $\mathbf{x}^* = [0, 0]^T$ for $\Omega = \{[x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$.

For $x_2 \neq 0$ and $x_2 > x_1$ sufficiently small, that is $|x_1|, |x_2| < 1$, clearly $x_1^2(x_1 - x_2) + 2x_2^2 \geq 0$. Hence $[0, 0]^T$ is a local minimum of f .

It is quite clear that for any f , of the form $f(x_1, x_2) = x_1^3 - x_1^2 x_2 + cx_2^2, c > 0$, $\mathbf{x}^* = [0, 0]^T$ will be a local minimum point of f for the domain given in **Example 1**.

However if we take $c = 0$ in the above expression, then $f(x_1, x_2) = x_1^3 - x_1^2 x_2$, then one can check that $\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, is a positive semidefinite matrix, but $\mathbf{x}^* = [0, 0]^T$ is not a local optimum point.

Hence the **first order** and the **second order** necessary conditions are **necessary** but **not sufficient** for a point \mathbf{x}^* to be a **local minimum**.

Exercise: Will $[0, 0]^T$ be a local minimum point of f given in **Example 1** if $\Omega = \mathbb{R}^2$?

Example 2 : Consider the following problem:

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

subject to $x_1 \geq 0, x_2 \geq 0$.

f is a twice continuously differentiable function.

At $[\frac{1}{2}, 0]^T$, $\frac{\partial f}{\partial x_1} = 2x_1 - 1 + x_2 = 0$

$$\frac{\partial f}{\partial x_2} = 1 + x_1 = \frac{3}{2}.$$

If \mathbf{d} is a feasible direction at $[\frac{1}{2}, 0]^T$, then d_2 has to be nonnegative.

Hence $\nabla f(\mathbf{x})|_{[\frac{1}{2}, 0]^T} \mathbf{d} = \frac{3}{2}d_2 \geq 0$ for any feasible direction \mathbf{d} .

Hence the first order necessary conditions for $\mathbf{x}^* = [\frac{1}{2}, 0]^T$ to be a locally minimum point is satisfied.

Also if $\nabla f(\mathbf{x}^*)\mathbf{d} = 0$, then $d_2 = 0$, and for all such \mathbf{d}

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} = 2d_1^2 \geq 0.$$

Hence the second order necessary conditions for \mathbf{x}^* to be locally minimum is also satisfied.

One can easily check that f has a global minimum at $x_1 = \frac{1}{2}, x_2 = 0$.

Sufficient conditions for a local minima :

Theorem 4 : Let $f : \Omega \rightarrow \mathbb{R}$ be a **twice continuously differentiable** function. Let \mathbf{x}^* be an interior point of Ω . If \mathbf{x}^* satisfies the following conditions

$$1. \nabla f(\mathbf{x}^*) = \mathbf{0}$$

$$2. \nabla^2 f(\mathbf{x}^*) \text{ is positive definite,}$$

then \mathbf{x}^* is a **local minimum point** of f .

Proof: Since \mathbf{x}^* is an interior point of Ω , if $\nabla f(\mathbf{x}^*) = \mathbf{0}$, then by Taylor's second order approximation formula for f near \mathbf{x}^* , we get

$$f(\mathbf{x}^* + t\mathbf{d}) = f(\mathbf{x}^*) + \frac{t^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(t^2), \text{ for all } \mathbf{d} \in \mathbb{R}^n \text{ and all } t > 0 \text{ sufficiently small.}$$

For t small, $|\frac{t^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d}| > |o(t^2)|$ (if $\mathbf{d} \neq \mathbf{0}$).

Hence $f(\mathbf{x}^* + t\mathbf{d}) > f(\mathbf{x}^*)$ for all t sufficiently small. The above proof is not enough to conclude that \mathbf{x}^* is a **local minimum point** of f .

The actual proof is given below but you may skip it if you find it difficult.

Optional reading:

$$f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \frac{1}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\|\mathbf{d}\|^2).$$

Since $\nabla^2 f(\mathbf{x}^*)$ is real, symmetric (By spectral theorem for real symmetric matrices), there exists a real matrix U , orthogonal (that is, $U^T U = U U^T = I$) such that

$$\nabla^2 f(\mathbf{x}^*) = U^T \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} U, \text{ where } \lambda_i \text{'s are the eigenvalues of } \nabla^2 f(\mathbf{x}^*).$$

$$\Rightarrow \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} = \mathbf{y}^T \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} \mathbf{y}, \text{ where } \mathbf{y} = U\mathbf{d}.$$

$$\Rightarrow \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} = \sum_i \lambda_i y_i^2, \text{ where } y_i \text{ s the } i \text{ th component of } \mathbf{y}. \text{ (multiply and check).}$$

$$\Rightarrow \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq \lambda_{\min} \sum_i y_i^2 = \lambda_{\min} \mathbf{y}^T \mathbf{y} = \lambda_{\min} \mathbf{d}^T \mathbf{d}$$

$$(\mathbf{y}^T \mathbf{y} = \mathbf{d}^T U^T U \mathbf{d} = \mathbf{d}^T \mathbf{d} = \|\mathbf{d}\|^2 \text{ since } U^T U = I).$$

$$\text{Hence } f(\mathbf{x}^* + \mathbf{d}) \geq f(\mathbf{x}^*) + \lambda_{\min} \|\mathbf{d}\|^2 + o(\|\mathbf{d}\|^2).$$

For $\|\mathbf{d}\| > 0$ sufficiently small we have

$$|\lambda_{\min} \|\mathbf{d}\|^2| > |o(\|\mathbf{d}\|^2)| \text{ hence } f(\mathbf{x}^* + \mathbf{d}) > f(\mathbf{x}^*) \text{ for all } \|\mathbf{d}\| > 0, \text{ sufficiently small.}$$

Remark: Since **maximizing** f is same as **minimizing** $-f$, all the previous theorems have corresponding analogues for a **maximization problem** with some obvious changes. For example \leq conditions in the results are replaced by \geq conditions and with positive semidefinite (or positive

definite) matrices in the results are appropriately replaced by negative semidefinite matrices (or negative definite matrices).

Definition 4: A real valued function f defined on a **convex set** $\Omega \subseteq \mathbb{R}^n$ is said to be a **convex function** on Ω if for all $\mathbf{x}, \mathbf{y} \in \Omega$ and all $0 \leq \alpha \leq 1$,
 $f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$.

Definition 5: An $f : \Omega \rightarrow \mathbb{R}$ is said to be a concave function if $-f$ is a convex function.

Theorem 1 : If f is a **convex function** on Ω (a convex set), then the set $S = \{\mathbf{x} \in \Omega : f(\mathbf{x}) \leq c\}$ is a convex set (for all real c).

Proof: Exercise.

Theorem 2: Let f be a **continuously differentiable** function defined on a convex set, $\Omega \subseteq \mathbb{R}^n$, then f is **convex** on Ω if and only if $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \Omega$.

Proof: Let $f : \Omega \rightarrow \mathbb{R}$ be a convex function. Then for all $\mathbf{x}, \mathbf{y} \in \Omega$ and all $0 \leq \alpha \leq 1$,
 $f(\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}) \leq \alpha f(\mathbf{y}) + (1 - \alpha)f(\mathbf{x})$.

For all $\alpha > 0$, sufficiently small,

$$\frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha} \leq f(\mathbf{y}) - f(\mathbf{x})$$

 Letting $\alpha \rightarrow 0$ we get
 $\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$.

To show the converse,
 let $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \Omega$. (**)
 Fix $\mathbf{x}, \mathbf{y} \in \Omega$, and let \mathbf{z} be a point in between and on the straight line segment joining \mathbf{x} and \mathbf{y} .
 That is $\mathbf{z} = \alpha \mathbf{x} + (1 - \alpha)\mathbf{y}$ for some $0 \leq \alpha \leq 1$. From (**) we get
 $f(\mathbf{x}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{x} - \mathbf{z})$ and
 $f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})(\mathbf{y} - \mathbf{z})$.
 By multiplying the first equation by α , the second by $(1 - \alpha)$ and adding the two equations we get,
 $\alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y} - \mathbf{z})$.
 Since $\mathbf{z} = \alpha \mathbf{x} + (1 - \alpha)\mathbf{y}$ we get
 $\alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}) \geq f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y})$.
 But the above condition is difficult to verify. The following result gives an easy way to the convexity of a function provided that the function satisfies certain conditions.

Theorem 3 : Let f be a **twice continuously differentiable** function on a convex set Ω (let Ω be such that it has atleast one interior point), then f is convex on Ω if and only if for all $\mathbf{x} \in \Omega$, $\nabla^2 f(\mathbf{x})$ is **positive semidefinite**.

Proof: For those interested in the proof, refer to Luenberger(Page 196, Third Edition) or Bazaara Sherali Shetty (page 92, Second edition).

Revisiting Example 1 : Let $f(x_1, x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2$ be defined on $\Omega = \{[x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$.

Since $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{pmatrix}$, at $x_1 = 1, x_2 = 3$, $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 0 & -2 \\ -2 & 4 \end{pmatrix}$
 is clearly not positive semidefinite. Hence f is not a convex function on Ω .

Theorem 4: Let $f : \Omega \rightarrow \mathbb{R}$ be a **continuously differentiable** function. If f is **convex** on Ω , then \mathbf{x}^* is a **global minimum** of f if and only if for all **feasible directions** \mathbf{d} at \mathbf{x}^* $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$.

Proof: Since the **only if** part is already shown before, we have to only show the **if** part.

Let $\mathbf{x}^* \in \Omega$ satisfy $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$ for all feasible \mathbf{d} at \mathbf{x}^* .

Since f is convex $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \Omega$. (1)

Hence for all $\mathbf{y} \in \Omega$, $f(\mathbf{y}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*)$.

Since Ω is convex, $\mathbf{x} = \alpha\mathbf{y} + (1 - \alpha)\mathbf{x}^* = \mathbf{x}^* + \alpha(\mathbf{y} - \mathbf{x}^*)$ belongs to Ω , for all $0 \leq \alpha \leq 1$, hence $(\mathbf{y} - \mathbf{x}^*)$ is a feasible direction at \mathbf{x}^* and $\nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0$.

Hence $f(\mathbf{y}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq f(\mathbf{x}^*)$.

Since $\mathbf{y} \in \Omega$ was arbitrary, \mathbf{x}^* is a global minimum of f .

Corollary 4: Let $f : \Omega \rightarrow \mathbb{R}$ be a **continuously differentiable** function. If f is **convex** on Ω and \mathbf{x}^* an **interior point** of Ω , then \mathbf{x}^* is a **global minimum** for f if and only if

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

Proof: Follows from the previous result.

That the above result is not necessarily true if f is not convex as you have already seen in **Example 1**.

Remark : Since **minimizing** f is same as **maximizing** $-f$, all the previous theorems for **minimizing a convex function** have corresponding analogues for **maximizing a concave function**.

Theorem 5: Let $f : \Omega \rightarrow \mathbb{R}$ be a **convex function**, then the following statements are true.

1. Let S be the collection of all \mathbf{x} 's where f attains its **minimum value** (that is, the set of all optimal solutions of f for a minimization problem). Then S is a **convex set**, or in other words the set $S = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{y}) \text{ for all } \mathbf{y} \in \Omega\}$, is a **convex set**.
2. If \mathbf{x}^* is a **local minimum** point of f then it is also a **global minimum** point of f .

Proof:

1. If f does not have a minimum then the above result is vacuously true.

If f takes a minimum value then let $a = \min_{\mathbf{x} \in \Omega} \{f(\mathbf{x})\}$.

Since f is a convex function, $S_1 = \{\mathbf{x} : f(\mathbf{x}) \leq a\}$ is convex, by Theorem 1.

Note that $S_1 = S$.

2. To show that a local minimum of f is a global minimum of f .

If not, then let \mathbf{x}^* be a local minimum point of f and let there exist a $\mathbf{y} \in \Omega$ such that $f(\mathbf{y}) < f(\mathbf{x}^*)$.

Join \mathbf{x}^* and \mathbf{y} by a straight line.

Since Ω is a convex set, the straight line segment joining \mathbf{x}^* and \mathbf{y} lies entirely in Ω .

Since f is a convex function, for all $0 < \alpha \leq 1$,

$f((1 - \alpha)\mathbf{x}^* + \alpha\mathbf{y}) \leq (1 - \alpha)f(\mathbf{x}^*) + \alpha f(\mathbf{y}) < f(\mathbf{x}^*)$.

This contradicts that \mathbf{x}^* is a local minimum point of f .

Remark: A natural question would be whether the conclusions of Theorem 5 holds good when **maximizing a convex function**. The answer however is a **NO**.

Take $f(x) = x^2$, $-1 \leq x \leq 2$.

From previous discussions however it is clear that the above result is true if you are **maximizing** $-f$ or **maximizing** a **concave** function.

Remark: We had seen while **minimizing** or **maximizing** a **linear** function over a **polyhedral** set, the extremum was attained in at least one **extreme point**. An extreme point of a polyhedral set, (more generally convex set) is one which cannot be written as a strict convex combination of two distinct points of that set).

But in the problem of **minimizing** a **convex** function over a **polyhedral** set, the only minimum may be attained at an **interior point** of Ω .

For example take $f(x) = x^2$, $-1 \leq x \leq 1$, the only minimum of this function on the interval $[-1, 1]$ which is a convex set is attained at $x = 0$ which is an interior point of $[-1, 1]$.

However when maximizing a convex function over a closed and bounded convex set the maximum is attained at an extreme point as the following theorem will illustrate.

Theorem 6: Let f be a **convex** function defined on a **closed** and **bounded convex** set Ω (so it has atleast one extreme point), then there exists an **extreme point** of Ω , where f takes its **maximum** value. (In case you are interested to know the proof refer to Luenberger page 198, Third Edition).