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 $\Omega \subseteq R^n$.
The problem is to minimize $f(\mathbf{x})$ 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 of f if there exists an $\epsilon > 0$, such that
$\mathbf{x} \in \Omega$ and $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$ implies $f(\mathbf{x}^*) \leq f(\mathbf{x})$.
Definition 2: A point $\mathbf{x}^* \in \Omega$ is called a global minimum of f if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ 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 $f : \Omega \to \mathbb{R}$ be a continuously differentiable function (that is, the first order partial derivatives of f exists and are continuous as functions from $\mathbb{R}^n$ 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 differentiable throughout Ω and $\mathbf{d}$ is a feasible direction at $\mathbf{x}^*$
$\lim_{h \to 0} \frac{f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)}{h}$ exists.
Since $g(h) - g(0) = f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)$, $\lim_{h \to 0} \frac{g(h) - g(0)}{h}$ also exists and
$\lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(\mathbf{x}^* + h\mathbf{d}) - f(\mathbf{x}^*)}{h} = \nabla f(\mathbf{x}^*)\mathbf{d}$.
Hence $g'(0)$ exists and if we look at the first order Taylor's approximation of g around $t = 0$, then
$g(t) = g(0) + tg'(0) + o(t)$, where $o(t)$ is a function of t such that $lim_{t \to 0} \frac{o(t)}{t} = 0$. (**)
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, $|t\nabla f(\mathbf{x}^*)\mathbf{d}| \geq |o(t)|$,
if $\nabla f(\mathbf{x}^*)\mathbf{d} < 0$, $\quad f(\mathbf{x}^*+t\mathbf{d}) < f(\mathbf{x}^*)$ for all t sufficiently small, which contradicts that $\mathbf{x}^*$ minimizes f locally.

Theorem 2 : Let $f : \Omega \to \mathbb{R}$ 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 $f : \Omega \to \mathbb{R}$ be a twice continuously differentiable function (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 $\mathbb{R}^n$ to $\mathbb{R}$).
Because of the assumptions on f our hessian matrix $\nabla^2 f$ for all our discussions is a symmetric matrix for all $\mathbf{x} \in \Omega$. If $\mathbf{x}^*$ is a local minimum 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, is the matrix of the second order partial derivatives of f and the (i,j) th entry of $\nabla^2 f(\mathbf{x}^*)$ is given by $\frac{\partial^2 f}{\partial x_j \partial x_i}(x^*)$ or $\frac{\partial^2 f}{\partial x_j \partial x_i}|_{x^*}$.
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 $\mathbf{x}(t) = \mathbf{x}^* + t\mathbf{d}$.
The second order Taylor's approximation of g around $t = 0$ gives,
$g(t) = g(0) + tg'(0) + \frac{t^2}{2!}g''(0) + o(t^2)$, (**)
Since $g'(t) = (\nabla f(\mathbf{x}^* + t\mathbf{d}))\mathbf{d} = \sum_i (\frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d})d_i$,
$g'(t) = \sum_i h_i(t)d_i$,
where $h_i(t) = \frac{\partial f}{\partial x_i}(\mathbf{x}^* + t\mathbf{d})$.

Hence $g''(t) = \sum_i h_i'(t)d_i$,
where $h_i'(t) = ((\nabla \frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d}))\mathbf{d} = \sum_j \frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i})(\mathbf{x}^* + t\mathbf{d})d_j = \sum_j (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^* + t\mathbf{d}))d_j$.
Hence $g''(t) = \sum_i h_i'(t)d_i = \sum_i(\sum_j(\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}^* + t\mathbf{d}))d_j)d_i$,
Hence $g''(0) = \sum_i(\sum_j(\frac{\partial^2 f}{\partial x_j \partial x_i}(x^*))d_j)d_i = \mathbf{d}^T\nabla^2 f(\mathbf{x}^*)\mathbf{d}$,
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,
$|\frac{t^2}{2!}g''(0)| \geq |o(t^2)|$,
hence if $\nabla f(\mathbf{x}^*) = 0$ and $\mathbf{x}^*$ is a local minimum then it should satisfy the condition
$\mathbf{d}^T\nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0$.

Theorem 4: Let $f : \Omega \to \mathbb{R}$ 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, every nonzero vector $\mathbf{d} \in \mathbb{R}^n$ is a feasible direction.

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 $x \in \mathbb{R}^n$.

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 rows of A which do not belong to α and by deleting all the columns which do not correspond to β.
If $\alpha = \beta$ then $A[\alpha, \alpha]$ is called a principal submatrix of A and $detA[\alpha, \alpha]$ 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 $detA[\alpha, \alpha] = a_{ii}$ the i th diagonal entry.
If $\alpha = \beta = \{i, j\}$ where $i, j \in \{1, \dots, n\}$ and $i < j$ then $A[\alpha, \alpha] = \begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix}$. If $\alpha = \{1, \dots, n\}$ then $A[\alpha, \alpha] = A$ and $detA[\alpha, \alpha] = det(A)$.

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 : Consider the following problem:

Minimize $f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1x_2$
subject to $x_1 \geq 0$, $x_2 \geq 0$.
One can easily check that f has a global minimum at $x_1 = \frac{1}{2}, x_2 = 0$.
Also 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.

Example 2: Consider the following problem:

Minimize $f(x_1, x_2) = x_1^3 - x_1^2 x_2 + 2x_2^2$
subject to $x_1 \geq 0$, $x_2 \geq 0$.
$\nabla f(x) = (3x_1^2 - 2x_1x_2, -x_1^2 + 4x_2) = [0, 0]$ 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.
But since $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{pmatrix}$.
At $\mathbf{x}^* = [6, 9]^T$, $\qquad \nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 18 & -12 \\ -12 & 4 \end{pmatrix}$ $\qquad$ 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 nonzero direction $\mathbf{d}$ is feasible if and only if $d_1 \geq 0$ and $d_2 \geq 0$.
Since $\nabla f(\mathbf{x}^*) = \mathbf{0}$, $\quad \nabla f(\mathbf{x}^*)\mathbf{d} = 0$ for all $\mathbf{d}$.
Since $\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$, $\quad \mathbf{d}^T\nabla^2 f(\mathbf{x}^*)\mathbf{d} = 4d_1^2 \geq 0$ 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.

Exercise: Check that $[0, 0]^T$ is a local minimum point of f in **Example 2**.
Solution: Note that $f(x_1, x_2) = x_1^3 - x_1^2x_2 + 2x_2^2 = x_1^2(x_1 - x_2) + 2x_2^2$ can take values < 0 only when $x_2 > x_1$, but then for $|x_1|, |x_2| < 1$, clearly $x_1^2(x_1 - x_2) + 2x_2^2 \geq 0$, hence $[0, 0]^T$ is a local minimum.

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

However if we take $c = 0$ in the above expression, then $f(x_1, x_2) = x_1^3 - x_1^2x_2$, 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.

Sufficient conditions for a local minima :

Theorem 4 : Let $f : \Omega \to \mathbb{R}$ 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.$\nabla^2 f(\mathbf{x}^*)$ 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
$f(\mathbf{x}^* + t\mathbf{d}) = f(\mathbf{x}^*) + \frac{t^2}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} + o(t^2)$, for all $\mathbf{d} \in \mathbb{R}^n$ and all $t > 0$ sufficiently small.
For t small, $|\frac{t^2}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d}| \geq |o(t^2)|$,
hence $\mathbf{x}^*$ is a local minimum point if $\nabla^2 f(\mathbf{x}^*)$ is positive definite.

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 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: A function f defined on a convex set Ω is said to be concave if $-f$ is convex.

Theorem 1 : If f is a convex function defined on Ω (a convex set), then the set $S = \{\mathbf{x} : f(\mathbf{x}) \leq c\}$ 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 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 $f : \Omega \to \mathbb{R}$ 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 $\alpha \to 0$ 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

$\alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \geq f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})$.

Since the above condition is difficult to verify and check, let me state the following result without proof.

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.
Revisiting Example 2 : 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.

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 and the positive semidefinite matrix in the previous theorem can be appropriately replaced by a negative semidefinite matrix.

Theorem 4: Let $f : \Omega \to \mathbb{R}$ be a continuously differentiable function. If f is convex on Ω, then $\mathbf{x}^*$ is a local minimum point of f if and only if for all feasible direction $\mathbf{d}$ 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 local minimum of f (in fact a global minimum of f).

Corollary 4: Let $f : \Omega \to R$ be a continuously differentiable function. If f is convex on Ω and $\mathbf{x}^*$ an interior point of Ω, then $\mathbf{x}^*$ is a local 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 2**.

Theorem 5: Let f be a convex function defined on a convex set Ω, 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 not true.
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 convex set, the minimum may be attained at an interior point of Ω.
The following theorem however gives a similar result when maximizing a convex function over a convex set.
I will just state the result without proof.

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. (For proof refer, Luenberger).

FJ conditions and Karush Kuhn Tucker (KKT) conditions in constrained optimization problems:

Consider the following nonlinear programming problem (P) of the form,
Minimize $f(\mathbf{x})$
subject to $g_i(\mathbf{x}) \leq \mathbf{0}$, for $i = 1, \ldots, m$, $\quad \mathbf{x} \in \Omega, (\Omega \text{ open })$ or $\mathbf{x} \in \mathbb{R}^n$.
Assume all the functions f,$g_i : \mathbb{R}^n \to \mathbb{R}$. As the name suggests the function f and the constraint functions g_i may not be linear functions.

The feasible region now is given by $S = Fea(P) = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, \text{ for } i = 1, 2, .., m\}$.
Since the problem is again of optimizing a function, we are again looking for feasible directions at $\mathbf{x}^* \in S$ such that starting from $\mathbf{x}^*$ if we move along $\mathbf{d}$, we will get points of the feasible region with better values of the objective function than that obtained at $\mathbf{x}^*$.

Note that for this problem the set D of feasible directions at $\mathbf{x}^*$ is given by,
$D = \{\mathbf{d} \in \mathbb{R}^n : g_i(\mathbf{x}^* + t\mathbf{d}) \leq 0, \text{ for all } i = 1, 2, .., m, \text{and for all } 0 \leq t \leq c, \text{ for some } c > 0\}$.

If I is the set of indices which corresponds to the constraints binding at $\mathbf{x}^* \in S$ then,
$I = \{i \in \{1, \ldots, m\} : g_i(\mathbf{x}^*) = 0\}$. Let $I^* = \{i \in \{1, \ldots, m\} : g_i(\mathbf{x}^*) < 0\}$.

For all $i \in I^*$, we assume that g_i is continuous at $\mathbf{x}^*$.
Then note that for each $i \in I^*$ there exists $c_i > 0$ such that
$g_i(\mathbf{x}^* + t\mathbf{d}) < 0$ for all $0 \leq t \leq c_i$,
since $g_i(\mathbf{x}^*) < 0$ and g_i is continuous at $\mathbf{x}^*$.
By taking $c = min_{i \in I^*}\{c_i\}$, we can get a $c > 0$ (which depends on $\mathbf{d}$) such that
$g_i(\mathbf{x}^* + t\mathbf{d}) < 0$ for all $0 \leq t \leq c$ and for all $i \in I^*$.

Since $g_i(\mathbf{x}^*) = 0$, for $i \in I$, if g_i's are continuously differentiable at $\mathbf{x}^*$ and $\mathbf{d}$ satisfies the condition, $\nabla g_i(\mathbf{x}^*)\mathbf{d} < 0$ for all $i \in I$,
then from Taylor's formula of first order approximation applied to the g_i's,
we get that for each $i \in I$ there exists an $a_i > 0$ such that
$g_i(\mathbf{x}^* + t\mathbf{d}) < 0$ for all $0 \leq t \leq a_i$.
Again by taking $a = min_{i \in I}\{a_i\} > 0$, we get that $g_i(\mathbf{x}^* + t\mathbf{d}) < 0$ for all $0 \leq t \leq a$ and for all $i \in I$.

From the above discussion it is clear that if
for all $i \in I^*$, g_i's are assumed to be continuous at $\mathbf{x}^*$
and for all $i \in I$, if g_i's are assumed to be continuously differentiable at $\mathbf{x}^*$ then
$G_0 \subseteq D$,
where $G_0 = \{\mathbf{d} \in \mathbb{R}^n : \nabla g_i(\mathbf{x}^*)\mathbf{d} < 0 \text{ for all } i \in I\}$.
From the first order necessary conditions for a local minimum (discussed earlier) it is already seen that if $\mathbf{x}^*$ is a local minimum then it should satisfy the condition that $F_0 \cap D = \phi$, where D is the set of feasible directions at $\mathbf{x}^*$ and $F_0 = \{\mathbf{d} \in \mathbb{R}^n : \nabla f(\mathbf{x}^*)\mathbf{d} < 0\}$.
Since $G_0 \subseteq D$ under the assumptions made above, hence necessary conditions for $\mathbf{x}^*$ to be a local minimum is also given by $F_0 \cap G_0 = \phi$.

Theorem 7: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Consider the problem of minimizing f subject to the conditions $g_i(\mathbf{x}) \leq 0, i = 1, 2, .., m$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i. Let $S = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1, 2, .., m\}$ and $\mathbf{x}^* \in S$.
For all $i \in I^*$, g_i's are assumed to be continuous at $\mathbf{x}^*$ and for all $i \in I$, g_i's are assumed to be continuously differentiable at $\mathbf{x}^*$.
Then if $\mathbf{x}^*$ is a local minimum of f over S
there exists nonnegative constants, $u_0, u_i, i \in I$, not all zeros such that
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$. (1)

Proof: If $\mathbf{x}^*$ is a local minimum point of f then $F_0 \cap G_0 = \phi$, hence the following system
$\nabla f(\mathbf{x}^*)\mathbf{d} < 0$ and $\nabla g_i(\mathbf{x}^*)\mathbf{d} < 0$ for $i \in I$
does not have a solution.
That is, the system $A\mathbf{d} < 0$ does not have a solution,

where the rows of A are given by $\nabla f(\mathbf{x}^*)$ and $\nabla g_i(\mathbf{x}^*)$, $i \in I$.
From a theorem of alternative (called Gordon's theorem, proof given at the end) we get that the system
$\mathbf{u} \neq \mathbf{0}$, $\mathbf{u} \geq \mathbf{0}$, $\mathbf{u}^T A = \mathbf{0}$ has a solution.
Hence the components of $\mathbf{u}$ satisfy condition (1).

If all the g_i's for $i = 1, \ldots, m$ are continuously differentiable at $\mathbf{x}^*$, then the above conditions reduces to
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$, and $u_i g_i(\mathbf{x}^*) = 0$ for all $i = 1, 2, ..., m$. (2)
Here $\mathbf{x}^* \in S$ is called the primal feasibility condition, the condition given in (1) together with non-negativity of the u_i's, is called the dual feasibility condition.
$u_i g_i(\mathbf{x}^*) = 0$ for all $i = 1, 2, ..., m$, is called the complementary slackness condition.
All the conditions taken together are called the **FJ (Fritz John)** conditions and the point $(\mathbf{x}^*, \mathbf{u})$ (or $\mathbf{x}^*$) is called a **Fritz John**, or an **FJ** point.

$\mathbf{x}^*$ is said to satisfy **KKT** condition if there exists nonnegative constants u_i, $i \in I$, such that
$\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$. (3)
If all the g_i's for $i = 1, \ldots, m$ are continuously differentiable at $\mathbf{x}^*$ then the above condition reduces to
$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$ and $u_i g_i(\mathbf{x}^*) = 0$ for all $i = 1, 2, ..., m$. (4)
The above conditions, that is conditions given by (3) and (4) are called **KKT (Karush, Kuhn, Tucker)** conditions.
Any $(\mathbf{x}^*, \mathbf{u})$ (or $\mathbf{x}^*$) which satisfies the **Karush Kuhn Tucker (KKT)** conditions is called a **KKT** point.

Theorem 8: If in addition to the conditions assumed for f and g_i's, as in the previous theorem, $\nabla g_i(\mathbf{x}^*)$'s for $i \in I$ are assumed to be linearly independent (as vectors) where $\mathbf{x}^*$ is a local minima (as in the previous theorem) then it can be easily shown that there exists nonnegative constants u_i, $i \in I$, such that
$\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$. (3)

Proof: Follows from the previous theorem and the fact that if $\nabla g_i(\mathbf{x}^*)$, $i = 1, 2, .., m$ are LI then the system
$\mathbf{u}' \neq \mathbf{0}$, $\mathbf{u}' \geq \mathbf{0}$, $\mathbf{u}'^T A' = \mathbf{0}$ does not have a solution,
where A' is the matrix whose rows are given by $\nabla g_i(\mathbf{x}^*)$, $i \in I$.
Hence in the FJ conditions which $\mathbf{x}^*$ must satisfy since it is local minima, u_0 cannot be zero, hence one can divide equation (1) by $u_0 > 0$ to get condition (3).

Remark: Hence from **Theorem 8** it is clear that if $(\mathbf{x}^*, \mathbf{u})$ (or $\mathbf{x}^*$) is an **FJ** point and if $\nabla g_i(\mathbf{x}^*)$'s for $i \in I$ are LI then $(\mathbf{x}^*, \mathbf{u})$ (or $\mathbf{x}^*$) is a **KKT** point.

Remark: If $\mathbf{x}^*$ satisfies the **Karush Kuhn Tucker (KKT)** conditions then it necessarily satisfies the **FJ** conditions.
Remark: KKT conditions basically says that under the assumptions of the theorem, if $\mathbf{x}^*$ is a local minimum, then $-\nabla f(\mathbf{x}^*)$ lies in the cone generated by the $\nabla g_i(\mathbf{x}^*)$'s, $i \in I$.

Example 1: Minimize $(x_1 - 3)^2 + (x_2 - 2)^2$
subject to
$x_1^2 + x_2^2 \leq 5$.

$x_1 + 2x_2 \leq 4.$
$-x_1 \leq 0$
$-x_2 \leq 0.$
By inspection one can see that f takes its minimum value at $[2,1]^T$.
Solution given by a student **Romel**: The point $[2,3]^T$, is outside the feasible region, hence we can construct circles of larger and larger radius (starting with radius 0) with center at $[2,3]^T$, till it cuts the feasible region S. The radius of the smallest such circle centered at $[2,3]^T$ which intersects S will give the optimal value, and the point where this circle cuts S will give optimal solutions.
Here $g_1(\mathbf{x}) = x_1^2 + x_2^2 - 5$, $g_2(\mathbf{x}) = x_1 + 2x_2 - 4$, $g_3(\mathbf{x}) = -x_1$ and $g_4(\mathbf{x}) = -x_2$.
At $\mathbf{x}^* = [2,1]^T$ the binding constraints are g_1 and g_2.
$\nabla g_1(\mathbf{x}^*) = [4,2]$ and $\nabla g_2(\mathbf{x}^*) = [1,2]$ and $\nabla f(\mathbf{x}^*) = [-2,-2]$.
$\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$ are linearly independent.
Take $u_1 = \frac{1}{6}$, $u_2 = \frac{1}{3}$, then
$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{2} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$.
Hence $[2,1]^T$ satisfies the KKT condition.

Example 2: Minimize $-x_1$.
subject to
$x_2 - (1-x_1)^3 \leq 0$
$-x_2 \leq 0$
It is clear that $\mathbf{x}^* = [1,0]^T$ is a local minimum.
At $[1,0]^T$ both the constraints are binding.
$\nabla f(\mathbf{x}^*) = [-1,0]$, $\nabla g_1(\mathbf{x}^*) = [0,1]$ and $\nabla g_2(\mathbf{x}^*) = [0,-1]$.
Take $u_0 = 0, u_1 = 1, u_2 = 1$, then
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^{2} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$.
But since $\nabla g_i(\mathbf{x}^*)$'s are not LI, $-\nabla f(\mathbf{x}^*)$ does not lie in the cone generated by the $\nabla g_i(\mathbf{x}^*)$'s, $i = 1,2$. Hence $[1,0]^T$ is an FJ point but **not** a KKT point.
Hence the above example shows that the KKT condition is not a necessary condition for a local minima, although FJ conditions are necessary conditions for a local minima.

The following theorem shows that if f and the g_i's are (in addition to conditions already assumed) assumed to be convex functions, then the KKT conditions become a sufficient condition for a local minima (although not necessary).

Theorem 9: Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and continuously differentiable. Consider the problem of minimizing f subject to the conditions $g_i(\mathbf{x}) \leq 0, i = 1,\ldots,m$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i. Let $S = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1,\ldots,m\}$ and $\mathbf{x}^* \in S$. For all $i \in I^*$, we assume that g_i is continuous at $\mathbf{x}^*$ and for all $i \in I$, g_i's are assumed to be continuously differentiable at $\mathbf{x}^*$. Let all the g_i's be convex functions, so that $S = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1,2,\ldots,m\}$ is convex. Then $\mathbf{x}^*$ is a global minimum of f over S if there exists nonnegative constants, $u_i, i \in I$ such that
$\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$.   (1)
Proof: Since this problem now becomes the problem of minimizing a convex function over a convex set hence if at $\mathbf{x}^*$, $F_0 \cap D = \phi$ then $\mathbf{x}^*$ is a local minimum, hence a global minimum of f.
To show $F_0 \cap D = \phi$ at $\mathbf{x}^*$.
Let $\mathbf{d} \in D$, where D is the set of feasible directions at $\mathbf{x}^*$. If $D = \phi$ then since S is convex, $\mathbf{x}^*$ should be the only point in S (since for any $\mathbf{y} \in S$, $(\mathbf{y} - \mathbf{x}^*)$ should be a feasible direction at $\mathbf{x}^*$), hence $\mathbf{x}^*$ is the global minimum.
Hence let $\mathbf{d} \in D \neq \phi$,
then $g_i(\mathbf{x}^* + t\mathbf{d}) \leq 0 = g_i(\mathbf{x}^*)$ for all t sufficiently small and for all $i \in I$.   (**)

But since g_i's are convex functions and continuously differentiable for $i \in I$,
$g_i(\mathbf{x}^* + t\mathbf{d}) \geq g_i(\mathbf{x}^*) + t\nabla g_i(\mathbf{x}^*)\mathbf{d}$ for all t sufficiently small and for all $i \in I$,
which together with (**) implies $\nabla g_i(\mathbf{x}^*)\mathbf{d} \leq 0$ for all $i \in I$.
Since $\nabla f(\mathbf{x}^*) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$ has a solution,
$\nabla f(\mathbf{x}^*)\mathbf{d} + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}^*)\mathbf{d} = \mathbf{0}$ has a solution, which implies that $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$. Since $\mathbf{d} \in D$ was arbitrary, $F_0 \cap D = \phi$.

Note that for a nonconvex function even if $\mathbf{x}^*$ satisfies the above conditions it may not be a local minimizer. Check this for the following example by taking $\mathbf{x}^* = (0, 0)$.
Example 3: $f(x) = -x^2$ for $x \leq 0$
$= x^2$ for $x \geq 0$.
In fact any x^* for which $\nabla f(x^*) = 0$ and x^* not a local minima will provide an example.

Exercise: Write the KKT conditions for the linear programming problem
Min $\mathbf{c}^T\mathbf{x}$
subject to, $A_{m\times n}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$.

Solution: Note that we have to write the constraints as :
$(A\mathbf{x} - \mathbf{b})_i \leq 0$ for $i = 1, \ldots, m$, $(-A\mathbf{x} + \mathbf{b})_i \leq 0$ for $i = 1, \ldots, m$ and $-x_i \leq 0$ for $i = 1, \ldots, n$.

Note that $\nabla f(\mathbf{x}^*) = \mathbf{c}^T$, $\nabla g_i(\mathbf{x}^*) = \mathbf{a}_i^T$ for the first m constraints, $\nabla g_i(\mathbf{x}^*) = -\mathbf{a}_i^T$ for the next m constraints and $\nabla g_i(\mathbf{x}^*) = -\mathbf{e}_i^T$ for the nonnegativity constraints,
where $\mathbf{a}_i^T$ for $i = 1, \ldots, m$, is the i th row of A, and $\mathbf{e}_i^T$ for $i = 1, \ldots, n$, is the i th row of I_n, respectively.
Since all the functions f and g_i's are continuously differentiable, the KKT conditions reduces to
$\mathbf{c}^T + \sum_{i=1}^m u_i\mathbf{a}_i^T + \sum_{i=1}^m u_i'(-\mathbf{a}_i^T) + \sum_{i=1}^n v_i(-\mathbf{e}_i^T) = \mathbf{0}$ (1)
and $u_i(A\mathbf{x} - \mathbf{b})_i = 0$ for all $i = 1, 2, ..., m$ (2)
$u_i'(-A\mathbf{x} + \mathbf{b})_i = 0$ for all $i = 1, 2, ..., m$ (3)
$v_i x_i = 0$ for all $i = 1, 2, ..., n$ (4)
where all u_i, u_i', v_i are nonnegative.
Conditions (1), (2), (3) and (4) can be rewritten as:
$\mathbf{c}^T - \mathbf{y}^T A - \mathbf{v}^T = \mathbf{0}$, or $\mathbf{c}^T - \mathbf{y}^T A = \mathbf{v}^T$
where $\mathbf{v} = [v_1, \ldots, v_n]^T$ is a nonnegative vector, $y_i = u_i' - u_i$ and $\mathbf{y} = [y_1, \ldots, y_m]^T$.
$(A\mathbf{x} - \mathbf{b})_i y_i = 0$ for all $i = 1, 2, ..., m$
and $v_i x_i = 0$ for all $i = 1, 2, ..., n$.
The above conditions reduces to the dual feasibility and the complementary slackness conditions in a LPP as follows:
$\mathbf{y}^T A \leq \mathbf{c}^T$,
$(A\mathbf{x} - \mathbf{b})_i y_i = 0$ for all $i = 1, 2, ..., m$,
and $(\mathbf{c}^T - \mathbf{y}^T A)_i x_i = 0$ for all $i = 1, 2, ..., n$.

Conclusion: Hence from the above derivation we can conclude that $\mathbf{x}^* \in Fea(P)$ is optimal for (P) if and only if $\mathbf{x}^*$ is a KKT point of (P).

Exercise: What are the FJ points of the above problem?
Exercise: Write the KKT conditions for the linear programming problem
Min $\mathbf{c}^T\mathbf{x}$
subject to, $A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}$.

Gordon's Theorem: Exactly one of the following two systems has a solution:
$\mathbf{u} \neq \mathbf{0}$, $\mathbf{u} \geq \mathbf{0}$, $\mathbf{u}^T A = \mathbf{0}$ (1)
$\mathbf{y}^T A > \mathbf{0}$ (2)

Proof: System (1) has a solution if and only if
$\mathbf{u} \geq \mathbf{0}$, $\mathbf{u}^T A = \mathbf{0}$, $\sum_i u_i = 1$ has a solution,
that is the system
$\mathbf{u} \geq \mathbf{0}$, $\mathbf{u}^T \begin{bmatrix} A \\ e^T \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$ has a solution,
where e is the vector with all entries equal to 1.
But we already know (Farka's lemma) that exactly one of the following two systems has a solution
$A\mathbf{x} = \mathbf{b}$ $\mathbf{x} \geq \mathbf{0}$ (1)
$\mathbf{y}^T A \geq \mathbf{0}$, $\mathbf{y}^T \mathbf{b} < 0$ (2)
Hence by using this lemma in the previous systems we get that exactly one of the following two systems has a solution

$$\mathbf{u} \geq \mathbf{0}, \qquad \mathbf{u}^T \begin{bmatrix} A \\ e^T \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \qquad (1)$$

$$[\mathbf{y}^T, a] \begin{bmatrix} A \\ e^T \end{bmatrix} \geq \mathbf{0}, \qquad [\mathbf{y}^T, a] \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} < 0 \qquad (2)$$

But system (2) reduces to $\mathbf{y}^T A > (-a)e^T$, where $a < 0$.
But $\mathbf{y}^T A > (-a)e^T$, $a < 0$ has a solution, if and only if $\mathbf{y}^T A > 0$ has a solution.
Hence exactly one of the following two systems has a solution

$$\mathbf{u} \geq \mathbf{0}, \qquad \mathbf{u}^T \begin{bmatrix} A \\ e^T \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \qquad (1)$$

$$\mathbf{y}^T A > 0 \qquad (2)$$

Or exactly one of the following two systems has a solution:

$$\mathbf{u} \neq \mathbf{0}, \qquad \mathbf{u} \geq \mathbf{0}, \qquad \mathbf{u}^T A = \mathbf{0} \qquad (1)$$

$$\mathbf{y}^T A > \mathbf{0} \qquad (2)$$