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

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

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 $0 \leq t \leq c$, 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 $0 \leq t \leq a$, for all $i \in I$.

If we take $p = min\{a, c\}$ then clearly $g_i(\mathbf{x}^* + t\mathbf{d}) < 0$ for all $0 \leq t \leq p$, for all $i = 1, \ldots, m$, which implies that $\mathbf{d}$ is a feasible direction at $\mathbf{x}^*$.
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 $G_0 \subseteq D_{\mathbf{x}^*}$, $\quad$ 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 $F_0 \cap D_{\mathbf{x}^*} = \phi$, where $D_{\mathbf{x}^*}$ 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 $G_0 \subseteq D_{\mathbf{x}^*}$ under the assumptions made above, hence $F_0 \cap G_0 = \phi$ is also a necessary condition for $\mathbf{x}^*$ to be a local minimum.

Theorem 7 (**FJ necessary conditions**): Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Consider the problem of minimizing f subject to the conditions $g_i(\mathbf{x}) \leq 0, \ i = 1, ..., m$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i. Let $S = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1, ..., m\}$ 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 non negative 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 $i = 1, ..., m.$ (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 $i = 1, ..., m$, 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** point, or an **FJ** point.

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)
$A\mathbf{y} > \mathbf{0}$ (2)
Proof: System (1) has a solution $\Leftrightarrow$
$\mathbf{u} \geq \mathbf{0}, \quad \mathbf{u}^T A = \mathbf{0}, \quad \sum_i u_i = 1$ has a solution
$\Leftrightarrow$ the system
$\mathbf{u} \geq \mathbf{0}, \quad \begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$ has a solution,
where $\mathbf{e}$ is the column vector with all entries equal to 1.

Farka's lemma gives 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)

Applying Farka's lemma we get exactly one of the following two systems has a solution:

$$\mathbf{u} \geq \mathbf{0}, \qquad \begin{bmatrix} A^T \\ e^T \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} \qquad (1')$$

$$[\mathbf{z}^T] \begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix} \geq \mathbf{0}, \qquad [\mathbf{z}^T] \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} < 0 \qquad (2')$$

If we partition $\mathbf{z}$ as $\mathbf{z}^T = [\mathbf{y}^T, a]$ conformally according to the order of A^T and $\mathbf{e}^T$ (which is a row vector) then system (2') reduces to
$$[\mathbf{y}^T, a] \begin{bmatrix} A^T \\ \mathbf{e}^T \end{bmatrix} \geq \mathbf{0}, \qquad [\mathbf{y}^T, a] \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} < 0. \qquad (2'')$$

On simplification, (2") reduces to
$\mathbf{y}^T A^T \geq (-a)\mathbf{e}^T$, where $a < 0$.
But the system $\mathbf{y}^T A^T \geq (-a)e^T$, $a < 0$ has a solution,
$\Leftrightarrow$ $\mathbf{y}^T A^T > 0$ or $A\mathbf{y} > 0$ has a solution
(For example if $A\mathbf{y} > 0$ has a solution say $A\mathbf{y} = [\frac{1}{2}, 2, 3]^T$ for some $\mathbf{y}$ then $A\mathbf{y} \geq -a[1,1,1]^T$ where $a = -\frac{1}{2}$.
And if $A\mathbf{y} \geq (-a)e$, $a < 0$ has a solution, then clearly for that $\mathbf{y}$, $A\mathbf{y} > \mathbf{0}$).

Hence exactly one of the following two systems has a solution

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

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

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

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

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

$\mathbf{x}^*$ 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 $i = 1, ..., m$. (4)

The above conditions, that is conditions given by (3) or (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 (**KKT necessary conditions**): 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 minimum (as in the previous theorem) then 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 $\nabla g_i(\mathbf{x}^*)$, $i = 1, ..., m$ 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 $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$
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 $[3, 2]^T$, is outside the feasible region, hence we can construct circles of larger and larger radius (starting with radius 0) with center at $[3, 2]^T$, till it cuts the feasible region S. The radius of the smallest such circle centered at $[3, 2]^T$ 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 $u_1 = \frac{2}{6}$, $u_2 = \frac{2}{3}$, 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 $f(x_1, x_2) = -x_1$
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
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^{2} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$.
$\nabla g_i(\mathbf{x}^*)$'s are not LI, also $-\nabla f(\mathbf{x}^*)$ 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 minimum, although FJ conditions are **necessary conditions** for a local minimum.

Theorem 8' (**KKT necessary conditions**): If in addition to the conditions assumed for f and g_i's, as in the previous theorem, if $G_0 \neq \phi$, and $\mathbf{x}^*$ is a local minimum of f then there exists non negative constants u_i, $i \in I$, such that

$\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$. **(**)**

Proof: Since $\mathbf{x}^*$ is a local minimum, it satisfies the FJ conditions, that is
there exists non negative constants, $u_0, u_i, i \in I$, not all zeros such that
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$. **(*)**

Since $G_0 \neq \phi$, $A\mathbf{d} < 0$ has a solution, where the rows of A are given by $\nabla g_i(\mathbf{x}^*)$, $i \in I$.

By Gordon's theorem then
$\mathbf{u} \neq \mathbf{0}$, $\mathbf{u} \geq \mathbf{0}$, $\mathbf{u}^T A = \mathbf{0}$ does not have a solution,
where $\mathbf{u} = [u_1, \ldots, u_k]^T$ (WLOG, $I = \{1, \ldots, k\}$).
Hence in the FJ conditions **(*)**, u_0 cannot be equal to zero.

Example 2: Minimize $f(x_1, x_2) = -x_1$
subject to
$-x_2 - x_1 \leq 0$
$-x_2 \leq 0$
$-x_1 \leq 0$.

$\nabla g_i(\mathbf{x}^*)$'s, $i \in I$ are **LD** at $\mathbf{x}^* = [0, 0]^T$, but $G_0 \neq \phi$.

Remark: If $\nabla g_i(\mathbf{x}^*)$'s, $i \in I$ are **LI** then $G_0 \neq \phi$, but the converse is **not** true.

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

Theorem 9 (**KKT sufficient conditions**): 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 $\mathbf{x}^*$, $F_0 \cap D_{\mathbf{x}^*} = \phi$ then $\mathbf{x}^*$ is a local, hence a global minimum of f.
To show $F_0 \cap D_{\mathbf{x}^*} = \phi$ at $\mathbf{x}^*$.
Let $\mathbf{d} \in D_{\mathbf{x}^*}$, where $D_{\mathbf{x}^*}$ is the set of feasible directions at $\mathbf{x}^*$. If $D_{\mathbf{x}^*} = \phi$ then since S is convex, $\mathbf{x}^*$ should be the only point in S (since for any $\mathbf{y} \in S$, $(\mathbf{y} - \mathbf{x}^*)$ would be a feasible direction at $\mathbf{x}^*$), hence $\mathbf{x}^*$ is the global minimum.
Let $\mathbf{d} \in D_{\mathbf{x}^*} \neq \phi$,
then $g_i(\mathbf{x}^* + t\mathbf{d}) \leq 0 = g_i(\mathbf{x}^*)$ for all $t > 0$ 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 $t > 0$ 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 $\mathbf{x}^*$ is a KKT point, $\nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$ 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 and u_i's are nonnegative, which implies that $\nabla f(\mathbf{x}^*)\mathbf{d} \geq 0$. Since $\mathbf{d} \in D_{\mathbf{x}^*}$ was arbitrary, $F_0 \cap D_{\mathbf{x}^*} = \phi$.

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

Example 1 revisited: Minimize $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$
subject to

$x_1^2 + x_2^2 \leq 5$
$x_1 + 2x_2 \leq 4$
$-x_1 \leq 0$
$-x_2 \leq 0.$

Check that f, g_i for all i are convex functions (by checking that the Hessian matrix of each of the functions f and g_i's are positive semidefinite). We have already seen that $\mathbf{x}^* = [2,1]^T$ is a KKT point. We now conclude from **Theorem 9** that $\mathbf{x}^* = [2,1]^T$ is the **global minimum** of f.

Example 4: Minimize $f(x_1, x_2) = x_1$
subject to
$(x_1 - 1)^2 + (x_2 - 1)^2 \leq 1$
$(x_1 - 1)^2 + (x_2 + 1)^2 \leq 1.$

The feasible region of the above problem consists of only one point $[1,0]^T$ hence it is the **global minimum** of f in this feasible region. Also check that all the functions f, g_i's are convex functions (Hessian matrix positive semi definite). But $[1,0]^T$ is not a KKT point.
Hence even if all the functions f, g_i's are convex, KKT conditions are **not necessary** in general, for a local minimum.

Exercise: Write the KKT conditions for the linear programming problem (LPP):
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 $-x_j \leq 0$ for $j = 1, \ldots, n$.

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 $\nabla g_j(\mathbf{x}^*) = -\mathbf{e}_j^T$ for the n non negativity constraints,
where $\mathbf{a}_i^T$ for $i = 1, \ldots, m$, is the i th row of A, and $\mathbf{e}_j^T$ for $i = 1, \ldots, n$, is the j th row of I_n, respectively.
Since all the functions f and g_i's are continuously differentiable, $\mathbf{x}^*$ is a KKT point
$\Leftrightarrow$ there exists u_i, u_i', v_j's non negative such that
$\mathbf{c}^T + \sum_{i=1}^m u_i\mathbf{a}_i^T + \sum_{i=1}^m u_i'(-\mathbf{a}_i^T) + \sum_{j=1}^n v_j(-\mathbf{e}_j^T) = \mathbf{0}$ (1),
$u_i(A\mathbf{x} - \mathbf{b})_i = 0$ for all $i = 1, ..., m$ (2),
$u_i'(-A\mathbf{x} + \mathbf{b})_i = 0$ for all $i = 1, ..., m$ (3),
and $v_jx_j = 0$ for all $j = 1, ..., n$ (4).

(1), (2), (3) and (4) is true
$\Leftrightarrow$
there exists vectors $\mathbf{v}$ and $\mathbf{y}$ such that $\mathbf{c}^T - \mathbf{y}^TA - \mathbf{v}^T = \mathbf{0}$, or $\mathbf{c}^T - \mathbf{y}^TA = \mathbf{v}^T$ (1')
where $\mathbf{v} = [v_1, \ldots, v_n]^T$ is a non negative vector, $\mathbf{y}$ is unrestricted, with $(\mathbf{y})_i = u_i' - u_i$ and $(A\mathbf{x} - \mathbf{b})_iy_i = 0$ for all $i = 1, ..., m$ (2')
$v_jx_j = 0$ for all $j = 1, ..., n.$ (3')
The above conditions are equivalent to the **dual feasibility** and the **complementary slackness conditions** for the above LPP as follows:
$\mathbf{y}^TA \leq \mathbf{c}^T$,
and $(\mathbf{c}^T - \mathbf{y}^TA)_jx_j = 0$ for all $j = 1, 2, ..., n$.

Conclusion: Hence from the above 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, $A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}$.

The following gives **FJ necessary conditions** where the feasible region has constraints of the type $h(\mathbf{x}) = 0$, where $h : \mathbb{R}^n \to \mathbb{R}$.

Theorem 10: (**FJ necessary conditions**) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a **continuously differentiable** function.
Consider the problem **P** of minimizing f subject to the conditions
$g_i(\mathbf{x}) \leq 0, i = 1, ..., m$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i,
$h_j(\mathbf{x}) = 0, j = 1, ..., l$, where $h_j : \mathbb{R}^n \to \mathbb{R}$ for all j.
Let $S = Fea(P)$ and $\mathbf{x}^* \in S$.
For all $i \in I^*$, g_i's are **continuous** at $\mathbf{x}^*$,
for all $i \in I$, g_i's are **continuously differentiable** at $\mathbf{x}^*$,
and for all j, h_j's are **continuously differentiable** at $\mathbf{x}^*$.

Then if $\mathbf{x}^*$ is a **local minimum** of f over S
there exists $u_0, u_i, i \in I$, **non negative** constants, and v_j constants (**unrestricted in sign**), not all zeros, such that

$$u_0 \nabla f(\mathbf{x}^*) + \sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{l} v_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}. \qquad \textbf{(*)}$$

Example 5: Minimize $f(x_1, x_2) = x_1 + x_2$
subject to
$x_2 - x_1 = 0$.

Check that no point of the above problem satisfies the **FJ** conditions (as in Theorem 10). However every point in the feasible region is an **FJ** point (as in Theorem 7) if you convert the equality into two inequalities of the form, $x_2 - x_1 \leq 0$ and $-x_2 + x_1 \leq 0$.
Since
$0(\nabla f(\mathbf{x}^*)) + 1(\nabla g_1(\mathbf{x}^*)) + 1(\nabla g_1(\mathbf{x}^*)) = \mathbf{0}$,
where $g_1(\mathbf{x}) = x_2 - x_1 = -g_2(\mathbf{x})$.

Hence in a feasible region which has a constraint of the form $h(\mathbf{x}) = 0$, every point $\mathbf{x}$ is an **FJ** point (as in **Theorem 7**), but $\mathbf{x}$ may not satisfy the **FJ** conditions of **Theorem 10**.

Lagrange's multipliers(brief overview, optional reading):

Let us consider the problem of minimizing, $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$
subject to
$g(x_1, x_2) = 5$,
where $g(x_1, x_2) = x_1^2 + x_2^2$.
Then we consider circles of bigger and bigger radius centered at $[3, 2]^T$ till it touches the circle

$x_1^2 + x_2^2 = 5$, say at the point $\mathbf{x}^*$, where $(x_1 - 3)^2 + (x_2 - 2)^2 = c$.
At the point $\mathbf{x}^*$, the tangent to the circle $(x_1 - 3)^2 + (x_2 - 2)^2 = c$, is same as the tangent to the curve $g(x_1, x_2) = 5$ (which is also a circle in this example) hence $\nabla f(\mathbf{x}^*)$ is parallel to $\nabla g(\mathbf{x}^*)$ (assuming $\nabla g(\mathbf{x}^*) \neq \mathbf{0}$), since both are orthogonal to the tangent and we have
$\nabla f(\mathbf{x}^*) = \lambda \nabla g(\mathbf{x}^*)$ (for some $\lambda \in \mathbb{R}$).
This λ is called Lagrange's multiplier.

Similarly if we consider minimizing, $f(x_1, x_2, x_3) = (x_1 - 3)^2 + (x_2 - 2)^2 + (x_3 - 1)^2$ subject to
$g_1(x_1, x_2, x_3) = 5$,
$g_2(x_1, x_2, x_3) = 10$.

Consider the curve $C = \{\mathbf{x} \in \mathbb{R}^3 : g_1(x_1, x_2, x_3) = 5, g_2(x_1, x_2, x_3) = 10\}$.
Let us assume $[3, 2, 1]^T$ is not in the curve C. Then we consider spheres of bigger and bigger radius centered at $[3, 2, 1]^T$ till a sphere touches the curve C at a point, say $\mathbf{x}^*$, where $(x_1 - 3)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 = a$ (for some a).
If $r(t)$ is a curve contained in the sphere $(x_1 - 3)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 = a$, passing through the point $\mathbf{x}^*$ then tangent vector of $r(t)$ at $\mathbf{x}^*$(with direction $r'(t)$), is an element of the tangent plane of the sphere $(x_1 - 3)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 = a$, at $\mathbf{x}^*$ (discussed in class), and for some $r(t)$ its tangent at $\mathbf{x}^*$ will be same as the tangent to the curve C at $\mathbf{x}^*$. Or in other words, a tangent vector say $\mathbf{y}$ (taken to be nonzero) to the curve C at $\mathbf{x}^*$ belongs to the tangent plane of the sphere $(x_1 - 3)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 = a$, at $\mathbf{x}^*$.
Hence $\mathbf{y}$ is orthogonal to $\nabla f(\mathbf{x}^*)$ (discussed in class).
Since $\mathbf{y}$ also belongs to both the tangent planes of the surfaces $g_1(x_1, x_2, x_3) = 5$, and $g_2(x_1, x_2, x_3) = 10$, at $\mathbf{x}^*$, $\mathbf{y}$ is also orthogonal to both, $\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$.
So $\mathbf{y}$ (nonzero vector) is orthogonal to each of $\nabla f(\mathbf{x}^*)$, $\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$.
If in addition $\{\nabla g_1(\mathbf{x}^*), \nabla g_2(\mathbf{x}^*)\}$ is assumed to be LI, then since a nonzero vector in $\mathbb{R}^3$ cannot be orthogonal to 3 LI vectors, hence $\{\nabla f(\mathbf{x}^*), \nabla g_1(\mathbf{x}^*), \nabla g_2(\mathbf{x}^*)\}$ is LD.
Since $\{\nabla g_1(\mathbf{x}^*), \nabla g_2(\mathbf{x}^*)\}$ is LI, $\nabla f(\mathbf{x}^*) = \lambda_1 \nabla g_1(\mathbf{x}^*) + \lambda_2 \nabla g_2(\mathbf{x}^*)$ for some $\lambda_1, \lambda_2 \in \mathbb{R}$.
The λ_i s are called Lagrange's multipliers.

The above portion on Lagrange Multipliers is **optional reading**, you can leave it if you are not interested.

Theorem 11: (**KKT necessary conditions**) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a **continuously differentiable** function.
Consider the problem $\mathbf{P}$ of minimizing f subject to the conditions
$g_i(\mathbf{x}) \leq 0, i = 1, ..., m$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ for all i,
$h_j(\mathbf{x}) = 0, j = 1, ..., l$, where $h_j : \mathbb{R}^n \to \mathbb{R}$ for all j.
Let $S = Fea(P)$ and $\mathbf{x}^* \in S$.
For all $i \in I^*$, g_i's are **continuous** at $\mathbf{x}^*$,
for all $i \in I$, g_i's are **continuously differentiable** at $\mathbf{x}^*$,
and for all j, h_j's are **continuously differentiable** at $\mathbf{x}^*$.
Let $\{\nabla g_i(\mathbf{x}^*),\ \nabla h_j(\mathbf{x}^*),\ i \in I,\ j \in \{1, \ldots, l\}\}$ be **LI**. Then if $\mathbf{x}^*$ is a **local minimum** of f over S there exists $u_i, i \in I$, **non negative** constants, and v_j constants (**unrestricted in sign**) such that

$$\nabla f(\mathbf{x}^*) + \textstyle\sum_{i \in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{l} v_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}. \qquad \textbf{(*)}$$

Proof: If $\mathbf{x}^*$ is a **local minimum** of f over $S = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m, h_j(\mathbf{x}) = 0, j = 1, ..., l\}$ then $\mathbf{x}^*$ is a **local minimum** of f over S_0, where $S_0 = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i \in$

$I^*, g_i(\mathbf{x}) = 0, i \in I, h_j(\mathbf{x}) = 0, j = 1, ..., l\}$ (note that $\mathbf{x}^* \in S_0$).
Then there exists Lagrange's multipliers u_i's and v_j's such that

$\nabla f(\mathbf{x}^*) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{l} v_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$.
If we can show that all the u_i's are non negative in the above expression then we are done and we get **(*)**, but the proof in that line is lengthy and beyond the scope of this course.
So we rather look for alternate ways to prove that above by using the results we already have on KKT conditions for inequality constraints.

Aliter: Let us write $h_j(\mathbf{x}) = 0$, for all $j = 1, ..., l$ as $h_j(\mathbf{x}) \leq 0$ and $-h_j(\mathbf{x}) \leq 0$.
Since $\mathbf{x}^*$ is a local minimum from the **FJ** necessary conditions, there exists u_0, u_i, y_j, y_j's, **non negative** constants, not all zeros such that
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{l} y_j \nabla h_j(\mathbf{x}^*) - \sum_{j=1}^{l} y_j' \nabla h_j(\mathbf{x}^*) = \mathbf{0}$ **(**)**

The above system can be rewritten as:
$u_0 \nabla f(\mathbf{x}^*) + \sum_{i\in I} u_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{l} v_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$ **(**)**
where $v_j = y_j - y_j'$, is unrestricted in sign.

Lastly, since $\{\nabla g_i(\mathbf{x}^*), i \in I, \nabla h_j(\mathbf{x}^*), j \in \{1, \ldots, l\}\}$ is an LI set, hence in **(**)**, $u_0 \neq 0$.
If we divide **(**)** throughout by u_0 which is positive, we get **(*)**.

Example 6: Minimize $f(x_1, x_2) = (x_1 - 1)^2 + x_2$
subject to
$x_2 - x_1 = 1$
$x_1 + x_2 \leq 2$.

Check that $\mathbf{x}^* = [\frac{1}{2}, \frac{3}{2}]^T$ is a **KKT** point (as in Theorem 11).

$\nabla f(\mathbf{x}^*) + 0(\nabla g_1(\mathbf{x}^*)) + (-1)\nabla h_1(\mathbf{x}^*) = 0$, where $h_1(\mathbf{x}) = x_2 - x_1$ and $g_1(\mathbf{x}) = x_1 + x_2$.

Example 7: Minimize $f(x_1, x_2) = x_1^3 - x_1 + x_2$,
subject to
$(x_1 - 1)^2 + 2x_2 \leq 1$,
$x_1 - x_2 = 0$.

Check that at $\mathbf{x}^* = [0, 0]^T$, the first constraint is also binding. But $\nabla g_1(\mathbf{x}^*)) = 2[-1, 1]^T$, and $\nabla h_1(\mathbf{x}^*) = [1, -1]^T$, which implies $\{\nabla g_1(\mathbf{x}^*)), \nabla h_1(\mathbf{x}^*)\}$ is LD, so Theorem 11 is not applicable.
But check that $\mathbf{x}^*$ satisfies the KKT conditions as given in **(3)**(immediately after Theorem 7).
Since f, g_1, g_2, g_3 are all convex functions, where $g_2(x_1, x_2) = x_1 - x_2$ and $g_3(x_1, x_2) = -x_1 + x_2$, (obtained by writing $h_1(x_1, x_2) = 0$ as two inequalities), $[0, 0]^T$ is the global minimum of f in the given feasible region.