as usual, define $\mathcal{F}(P)$ as the set of feasible points (where all the constraints in $P$ are satisfied).

Thus, if $x^* \in \mathcal{F}(P)$ is a local minimizer, there exist (Lagrange) multipliers $(\lambda, \mu) = (\lambda_0, \lambda_1, \ldots, \lambda_r, \mu_1, \ldots, \mu_m)$ such that

$$
\begin{align*}
\lambda_i &\geq 0, \quad i = 0, \ldots, r; \\
(\lambda, \mu) &\neq (0, 0); \\
\lambda_0 f + \sum \lambda_j g_j + \sum \mu_k h_k &= 0 \quad \text{at } x^*.
\end{align*}
$$

Moreover, $A_j g_j(x^*) = 0$ for $j = 1, \ldots, r$.

**Proof.** Consider convex cones:

$$
K = \left\{ v : (\nabla f(x)^T v) < 0, (\nabla g_j(x)^T v) < 0, j \in I(x^*) \right\}
$$

and

$$
L = \left\{ v : (\nabla h_j(x^*)^T v) = 0, j = 1, \ldots, m \right\}
$$

(\text{NB: } K \text{ is open;} L \text{ is linear subspace). Here } I(x^*) \text{ are the indices of active constraints, i.e. } j \in I(x^*) \text{ iff } g_j(x^*) = 0.

First, notice that the implications of the Theorem are satisfied if $\nabla h_j(x^*)j = 1, \ldots, m$ are linearly dependent (just choose $\lambda = 0; \mu$ such that $\sum \lambda_j \nabla h_j(x^*) = 0, v$).

If $\nabla h_j(x^*)$ are linearly independent, implicit function theorem conditions are satisfied, and (after perhaps, reordering indices) we can find $u$ function...
such that the set \{h_j \equiv 0, \ j = 1, ..., m\} is given by
\[ h_j(X_k(x_1, ..., x_{n-m}), ..., X_k(x_1, ..., x_{n-m})) = 0, \]
(i.e. \( h_j(x_1, ..., x_{n-m}, X_k(x_1, ..., x_{n-m}), ..., X_k(x_1, ..., x_{n-m})) = 0 \)).
Now, assume the cones \( K \) and \( L \) have a nontrivial intersection, \( V \in K \cap L, V \not= 0 \). Consider the curve
\[ x_k(t) = x_k^* + tV_k, \quad k = 1, ..., n-m \]
\[ x_k(t) = x_k(\bar{v}^* + t\bar{w}), \quad k = n-m+1, ..., n. \]
(Here \( \bar{w} = (u_1, ..., u_{n-m}) \) for \( u \in \mathbb{R}^n \).) It is immediate to check that \( x(t) \) defines curve
\[ x(t) \in \mathbb{R}^n, \quad x(0) = x^* \] lies in \( \mathbb{P}(P) \) for small \( t > 0 \) (as all active inequalities have decreasing derivative at \( t=0 \), and all equalities \( h_j \) are satisfied by construction). Hence, if \( K \cap L \not= \emptyset \), \( x^* \) is not a minimizer.
Hence, as \( x^* \) is, by assumption,
\[ K \cap L = \emptyset, \] and therefore, there exists a vector \( u \in K^* : (u, \bar{w}) \not= 0 \) for all \( \bar{w} \in L \).
However, any such vector would be a positive combination of vectors \( \nabla f(x^*), \nabla g_j(x^*) \) \( j \in J(x^*) \) (by Dubovitsky-Milyutin theorem) and also, as \( (u, \bar{w}) = 0 \) for all \( \bar{w} \in L \), \( \bar{u} \) is a linear combination of \( \nabla h_j(x^*) \). Hence
\[ 0 = \bar{u} - \bar{w} = \lambda \nabla f + \sum \lambda_j \nabla g_j - \sum \mu_j \nabla h_j. \]  \( \square \)

**Definition**
\[ L : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} : \text{Lagrange function}. \]
\[ L(x, \lambda, \mu) = \lambda_0 \cdot f(x) + \sum \lambda_j \cdot g_j(x) + \sum \mu_j \cdot h_j(x) \]
Examples. Find vector closest to $b \in \mathbb{R}^n$, subject to bound on its $l_1$ norm: 
\[ \frac{1}{2} \| x - b \|_2^2 \rightarrow \min, \quad \| x \|_1 = \sum |x_i| \leq c. \]

Making it smooth: for any binary vector $\epsilon \in \{-1, 1\}^n$,
\[ (\epsilon, x) \leq c. \]

Lagrange multipliers:
\[ \lambda_\epsilon, \epsilon \in \mathcal{I}_2^n \]
\[ L(x, \lambda) = \frac{1}{2} \| x - b \|_2^2 + \sum \lambda_\epsilon (\epsilon, x) - c. \]

Necessary conditions for optimality:
\[ \begin{cases} \text{for all } \epsilon: (\epsilon, x) = c, \quad \lambda_\epsilon = 0, \\ \lambda_0 (x - b) + \sum \lambda_\epsilon \epsilon = 0. \end{cases} \]

**•** $y \rightarrow \min$ sub $b$
\[ y - 2x^2 \leq 0 \]
\[ ax^2 - y \leq 0 \]

Lagrange function:
\[ \lambda_0 y + \lambda_1 (y - 2x^2) + \lambda_2 (ax^2 - y), \]

Necessary conditions for minimizer:
\[ \lambda_0 (0, 1) + \lambda_1 (-2x, 1) + \lambda_2 (2x, -1) = 0, \quad \lambda_j \geq 0 \ldots \]
\[ \lambda_1 = 2\lambda_2, \quad \lambda_0 + \lambda_1 = \lambda_2 \]
\[ \therefore x = 0, \quad y = 0. \]

**•** $x + y \rightarrow \min$ sub $b$
\[ y - 2x^2 \leq 0 \]
\[ ax^2 - y \leq 0 \]

\[ \lambda_0 (1, 1) + \lambda_1 (-6x^2, 1) + \lambda_2 (3x^2, -1) = 0 \]
Useful view at Lagrange multipliers: penalty function.

Define \( g^+ = \max(g, 0) \)

Consider \( F = f(x) + \sum g^+_j (x)^2 + \sum h_j(x)^2 \). \( \text{(strict)} \)

Then \( x^* \) is a global minimizer in some ball \( B^o(x^*) \). The value of \( x^* \) is \( f(x^*) \), and if \( x^* \) is a local minimizer, then there is a global minimizer in \( B^o(x^*) \cap F(P) \), as \( F(x) > f(x^*) = F(x^*) \) for \( x \in \partial B^o(x^*) \).

Let \( x_k \) be an unconstrained minimizer for \( F \) in \( B^o(x^*) \). As \( \inf_{x \in B^o(x^*)} f \leq F(x_k) \leq f(x^*) \), the terms \( g_j(x_k) \) and \( h_j(x_k) \) tend to 0. Take \( \bar{x} \) to be a limit of \( x_k \)'s along a subsequence. Then \( \bar{x} \in B^o(x^*) \cap F(P) \), and \( f(\bar{x}) \geq f(x^*) \). So, also \( |\bar{x} - x^*|^2 = 0 \), i.e. there is a sequence \( x_k \to x^* \).

At \( x_k \),

\[ 0 = \nabla F(x_k) = \nabla f + \sum k^+_g \nabla g_j(x_k) + \sum k_h \nabla h_j(x_k) + \nabla f(x_k) + f(x_k) \]

We can rescale these coefficients so that they sit on the sphere, so that

\[ 0 = \lambda^+_g \nabla f(x_k) + \sum \lambda_{j,k} \nabla g_j(x_k) + \sum \mu_{j,k} \nabla h_j(x_k) + f(x_k) \]

Taking limits, we get the desired.