A general constrained optimization problem has the form

$$\min_{x} f(x)$$

s.t. \( g_i(x) \leq 0, \ i = 1, \ldots, m \)

\( h_i(x) = 0, \ i = 1, \ldots, p \)

where \( x \in \mathbb{R}^d \). If \( x \) satisfies all the constraints, we say \( x \) is \( \square \). Assume \( f \) is defined on all feasible points.

**Lagrangian Duality**

The \( \square \) function is

$$L(x, \lambda, \nu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

and \( \lambda = [\lambda_1, \ldots, \lambda_m]^T \) and \( \nu = [\nu_1, \ldots, \nu_p]^T \) are called \( \square \) or \( \square \).
The (Lagrange) dual function is

\[ L_D(\lambda, \nu) := \min_x L(x, \lambda, \nu) \]

Note: \( L_D \) is concave, being the point-wise minimum of a family of affine functions.

Then we can define the optimization problem:

\[ \max_{\lambda, \nu: \lambda \geq 0} L_D(\lambda, \nu) \]

why would we constrain \( \lambda \geq 0 \)?

The Primal

Similarly, we could define the function

\[ L_p(x) := \max_{\lambda, \nu: \lambda \geq 0} L(x, \lambda, \nu) \]

and the primal optimization problem

\[ \min_x L_p(x) = \min_x \max_{\lambda, \nu: \lambda \geq 0} L(x, \lambda, \nu) \]

It is like the dual but the min and max are swapped.
Notice that
\[ L_p(x) = \begin{cases} \sum f(x) & \text{if } x \text{ is feasible} \\ -\infty & \text{else} \end{cases} \]

Thus the primal encompasses the original problem (i.e., they have the same solution).

**Weak Duality**

Claim 1: \( \text{d}^* = \max \min L(x, \lambda, \nu) \)
\[ \lambda, \nu : \lambda, \nu \geq 0 \]
\[ \leq \min \max L(x, \lambda, \nu) =: p^* \]
\[ \lambda, \nu : \lambda, \nu \geq 0 \]

Proof: Let \( \hat{x} \) be feasible. Then for any \( \lambda, \nu \) with \( \lambda, \nu \geq 0 \),
\[ L(\hat{x}, \lambda, \nu) = f(\hat{x}) + \sum \lambda_i \varphi_i(\hat{x}) + \sum \nu_i \psi_i(\hat{x}) \leq f(\hat{x}) \]

Hence
\[ L_0(\lambda, \nu) = \min_{x} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f(\hat{x}) \]

This is true for any feasible \( \hat{x} \), so
\[ L_0(\lambda, \nu) \leq \min_{\hat{x}} f(\hat{x}) = p^* \]
\[ \text{\( \hat{x} \) feasible} \]

Taking the max over \( \lambda, \nu : \lambda, \nu \geq 0 \), we have
\[ \text{d}^* = \max_{\lambda, \nu : \lambda, \nu \geq 0} L(\lambda, \nu) \leq p^* \]
The difference $p^* - d^*$ is called the ______.

**Strong Duality**

If $p^* = d^*$, we say **strong duality** holds.

**Theorem** 1. If the primal problem is ______ (f, gi: convex, hi: affine), and a constraint qualification holds, then $p^* = d^*$.

**Examples** of constraint qualifications:

- All $g_i$ are ______.
- $\exists x$ s.t. $h_i(x) = 0 \forall i$, $g_i(x) < 0 \forall i$ (strict feasibility)

**KKT Conditions**

Assume $f, g_i, h_i$ are **differentiable**.

**Necessity**

**Theorem** 2. If $p^* = d^*$, $x^*$ is primal optimal, and $(\lambda^*, \nu^*)$ is dual optimal, then the Kurosh-Kuhn-Tucker conditions hold: 

\[
\begin{align*}
(1) \quad \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0 \\
(2) \quad g_i(x) &\leq 0, \quad i = 1, \ldots, m \\
(3) \quad h_i(x^*) &= 0, \quad i = 1, \ldots, p \\
(4) \quad \lambda_i^* &\geq 0, \quad i = 1, \ldots, m \\
(5) \quad \lambda_i^* g_i(x^*) &= 0, \quad i = 1, \ldots, m \quad \text{(complementary slackness)}
\end{align*}
\]

Proof: To prove the first condition, notice that \( L(x, \lambda^*, \nu^*) \) is minimized by \( x^* \), and thus its gradient wrt \( x \) must be 0 at \( x^* \).

To prove the last condition, write
\[
\begin{align*}
f(x^*) &= L_D(\lambda^*, \nu^*) \\
&= \min_x \left( f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\leq f(x^*)
\end{align*}
\]

and therefore \( \lambda_i^* g_i(x^*) = 0, \quad \forall i \). \( \blacksquare \)
Sufficiency

Theorem: If the primal problem is convex (i.e., $f, g_i$ are convex functions, $h_i$ are affine), and $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy the KKT conditions, then $\tilde{x}$ is primal optimal, $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal, and the duality gap is zero.

Proof: (2), (3) $\Rightarrow$ $\tilde{x}$ is feasible

(4) $\Rightarrow$ $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in $x$

(1) $\Rightarrow$ $\tilde{x}$ is a minimizer of $L(x, \tilde{\lambda}, \tilde{\nu})$. Then

$$L_D(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

$$= f(\tilde{x}) + \sum \tilde{\lambda}_i g_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x})$$

$$= f(\tilde{x})$$

by (5)

Conclusion: If a constrained optimization problem is differentiable, convex, and a constraint qualification holds, then the KKT conditions are necessary and sufficient for primal/dual optimality (with zero duality gap). Thus, the KKT conditions can be used to solve such problems.
Key
A. feasible; Lagrangian; Lagrange multipliers or dual variables
B. dual; primal
C. duality gap; convex; affine

Reference
Convex Optimization, Ch 5
Boyd & Vandenberghe
(available online)