A general constrained optimization problem has the form

\[
\min_{x} f(x)
\]

\[
\text{s.t. } g_i(x) \leq 0, \quad i = 1, \ldots, m
\]

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

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

**Lagrangian Duality**

The 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 \( \Box \) or \( \Box \).
The (Lagrange) dual function is

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

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

(\textbf{3}) Then we can define the optimization problem:

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

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

\underline{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 + 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), yet it is unconstrained.

Weak Duality

Claim: \( d^* := \max_{\lambda, v: \lambda \geq 0} \min_{x} L(x, \lambda, v) \)

\[ \leq \min_{x} \max_{\lambda, v: \lambda \geq 0} L(x, \lambda, v) =: p^* \]

Proof: Let \( x \) be feasible. Then for any \( \lambda, v \) with \( \lambda \geq 0 \),

\[ L(x, \lambda, v) = f(x) + \sum \lambda_i g_i(x) + \sum \nu_i h_i(x) \leq f(x) \]

Hence

\[ L_0(\lambda, v) = \min_{x} L(x, \lambda, v) \leq L(x, \lambda, v) \leq f(x). \]

This is true for any feasible \( x \), so

\[ L_0(\lambda, v) \leq \min_{\hat{x}} f(\hat{x}) = p^* \]

Taking the max over \( \lambda, v: \lambda \geq 0 \), we have

\[ d^* = \max_{\lambda, v: \lambda \geq 0} L_0(\lambda, v) \leq p^* \]
Strong Duality

If \( p^* = d^* \), we say **strong duality** holds.

Theorem 1: If the primal problem is \( f, g_i; \text{convex}, h_i \text{ affine} \), and a constraint qualification holds, then \( p^* = d^* \).

Examples of constraint qualifications:

- All \( g_i \) are ___
- \( \exists x \text{ s.t. } h_i(x) = 0 \quad \forall i, \quad g_i(x) < 0 \quad \forall i \) (strict feasibility)

KKT Conditions

Assume \( f, g_i, h_i \) are differentiable.

Necessity

Theorem 1: If \( p^* = d^* \), \( x^* \) is primal optimal, and \( (\lambda^*, \nu^*) \) is dual optimal, then the KKT conditions hold.
\[
KKT \text{ conditions}
\]

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

Proof: (2)-(3) hold because \( x^* \) must be feasible, and (4) holds by def. of the dual problem. To prove (5),

\[
f(x^*) = L_b(\lambda^*, \nu^*) \quad \text{[by strong duality]}
\]

\[
= \min_x \left( f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)
\]

\[
\leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*)
\]

\[
\leq f(x^*) \quad \text{[by (2)-(4)]}
\]

and therefore the two inequalities are equalities. Equality of the last two lines implies \( \lambda_i^* g_i(x^*) = 0 \ \forall i \) which is (5). Equality of the 2nd and 3rd lines implies \( x^* \) is a minimizer of \( L(x, \lambda^*, \nu^*) \) w.r.t. \( x \). Therefore \( \nabla_x L(x^*, \lambda^*, \nu^*) = 0 \), which is (1). \( \square \)
**Theorem**  If the original problem is convex (i.e., \( f, g_i \) are convex functions, \( h_i \) are affine), and \( \bar{x}, \bar{\lambda}, \bar{\nu} \) satisfy the KKT conditions, then \( \bar{x} \) is primal optimal, \( (\bar{\lambda}, \bar{\nu}) \) is dual optimal, and the duality gap is zero.

**Proof:** (2), (3) \( \Rightarrow \) \( \bar{x} \) is feasible

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

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

\[
L(\bar{x}, \bar{\lambda}, \bar{\nu}) = L (\bar{x}, \bar{\lambda}, \bar{\nu})
\]

\[
= f(\bar{x}) + \sum \bar{\lambda}_i \, g_i(\bar{x}) + \sum \bar{\nu}_i \, h_i(\bar{x})
\]

\[
= f(\bar{x}) \quad \text{by (5)}
\]

**Conclusion**  If a constrained optimization problem is differentiable, and has convex objective function and constraint sets, 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.
Saddle Point Property

If $\bar{x}$ and $(\bar{\lambda}, \bar{\nu})$ are primal/dual optimal w/ zero duality gap, they are a saddle point of $L$, i.e.

$$L(\bar{x}, \lambda, \nu) \leq L(\bar{x}, \bar{\lambda}, \bar{\nu}) \leq L(x, \lambda, \nu)$$

for all $x, (\lambda, \nu)$.

Justification of this fact is left as an exercise.

Key

A. feasible; Lagrangian; Langrange multipliers or dual variables
B. dual; primal C. duality gap; convex; affine

Reference

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