CONSTRAINED OPTIMIZATION

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,

(A) we say x is ________. Assume f is defined on all feasible points.

Lagrangian Duality

The ____________ function is

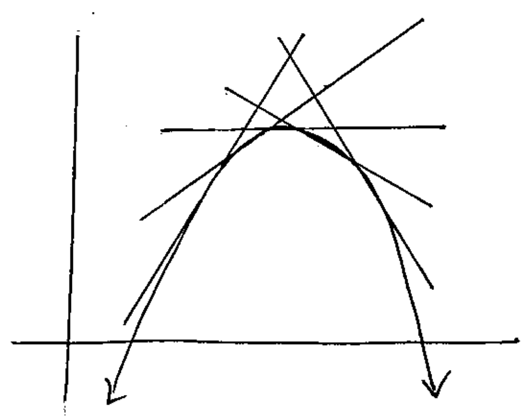
$$L(x, \lambda, \nu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

and $\lambda = [\lambda_1, \ldots, \lambda_m]^T$ and $\nu = [\nu_1, \ldots, \nu_p]^T$ are called __________ ________ or ______ ________.

The (Lagrange) dual function is

$$L_D(\lambda, \nu) := \min_x \; L(x, \lambda, \nu)$$

<u>Note</u> L_D is concave, being the **point**-wise minimum of a family of affine functions



3 Then we can define the ____ <u>optimization problem</u>:

$$\max_{\lambda, \nu : \lambda_i \geq 0} \; L_D(\lambda, \nu)$$

Why would we constrain $\lambda_i \geq 0$?

The Primal

Similarly, we could define the ______ function

$$L_P(x) := \max_{\lambda, \nu :\, \lambda_i \geq 0} \; L(x, \lambda, \nu)$$

and the primal optimization problem

$$\min_x \; L_P(x) = \min_x \; \max_{\lambda, \nu :\, \lambda_i \geq 0} \; L(x, \lambda, \nu)$$

It is like the dual but the min + max are swapped.

Notice that

$$L_p(x) = \begin{cases} f(x) & \text{if } x \text{ is feasible} \\ \infty & \text{else} \end{cases}$$

Thus the primal encompasses the original problem (i.e., they have the same solution), yet it is <u>unconstrained</u>.

Weak Duality

<u>Claim</u> $d^* := \max_{\lambda,\nu : \lambda_i \geq 0} \min_x L(x, \lambda, \nu)$

$$\leq \min_x \max_{\lambda,\nu : \lambda_i \geq 0} L(x, \lambda, \nu) =: p^*$$

Proof: Let $\tilde{x}$ be feasible. Then for any λ, ν with $\lambda_i \geq 0$,

$$L(\tilde{x}, \lambda, \nu) = f(\tilde{x}) + \sum \lambda_i g_i(\tilde{x}) + \sum \nu_i h_i(\tilde{x}) \leq f(\tilde{x})$$

Hence

$$L_D(\lambda, \nu) = \min_x L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f(\tilde{x}).$$

This is true for any feasible $\tilde{x}$, so

$$L_D(\lambda, \nu) \leq \min_{\substack{\tilde{x} \\ \tilde{x} \text{ feasible}}} f(\tilde{x}) = p^*$$

Taking the max over $\lambda, \nu : \lambda_i \geq 0$, we have

$$d^* = \max_{\lambda,\nu : \lambda_i \geq 0} L_D(\lambda, \nu) \leq p^*$$

$\blacksquare$

(c) The difference $p^* - d^*$ is called the ______ ______.

Strong Duality

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

Theorem | If the original problem is _____ (f, g_i convex, h_i 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)

$\boxed{\text{KKT Conditions}}$

Assume f, g_i, h_i are differentiable.

Necessity

Theorem | If $p^* = d^*$, x^* is primal optimal, and (λ^*, ν^*) is dual optimal, then the Karesh-Kuhn-Tucker conditions hold:

KKT conditions

(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) $g_i(x^*) \le 0, \quad i = 1, \ldots, m$

(3) $h_i(x^*) = 0, \quad i = 1, \ldots, p$

(4) $\lambda_i^* \ge 0, \quad i = 1, \ldots, m$

(5) $\lambda_i g_i(x^*) = 0, \quad i = 1, \ldots, m$ (complementary slackness)

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

$$
\begin{aligned}
f(x^*) &= L_D(\lambda^*, \nu^*) && \text{[by strong duality]} \\
&= \min_x \left(f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\le f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\le f(x^*) && \text{[by (2)-(4)]}
\end{aligned}
$$

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). ∎

Sufficiency

Theorem If the original 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

$$
\begin{aligned}
L_D(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\
&= f(\tilde{x}) + \underbrace{\sum \tilde{\lambda}_i g_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x})}_{= 0 \text{ by } (5)} \\
&= f(\tilde{x})
\end{aligned}
$$

$\blacksquare$

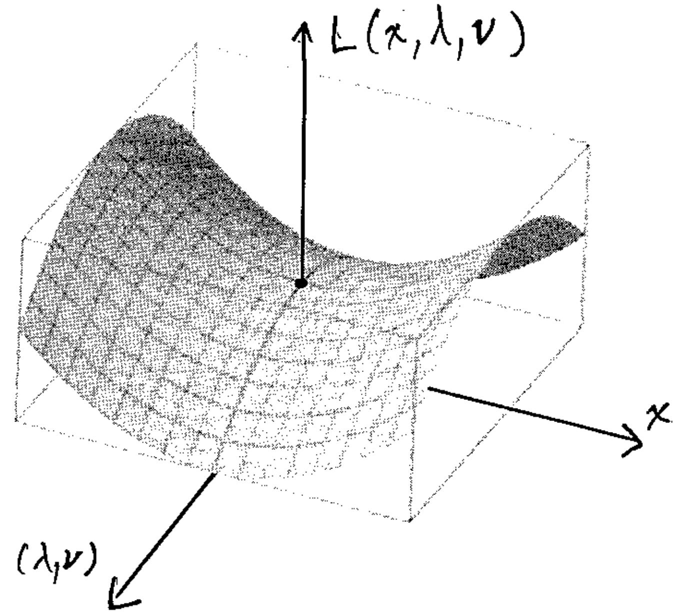
Conclusion If a constrained optimization problem is differentiable, and has convex objective function and constraint sets, then the KKT conditions are necessary and sufficient for primal / dual optimality (with zero duality gap). Thus, the KKT conditions can be used to solve such problems.

Saddle Point Property If $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal / dual optimal w/ zero duality gap, they are a saddle point of L, ie.

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

for all $x \in \mathbb{R}^d$, $\lambda \in \mathbb{R}_+^m$, $\nu \in \mathbb{R}^p$

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)