CONSTRAINED OPTIMIZATION

A constrained optimization problem has the form

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, n \end{aligned}$$

where $x \in \mathbb{R}^d$. If x satisfies all the constraints, it is said to be feasible. Assume f is defined at all feasible points.

The Lagrangian

The Lagrangian function is

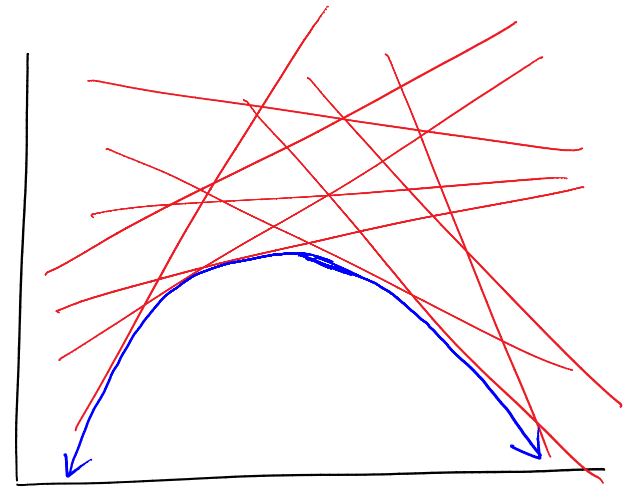
$$L(x, \lambda, \nu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{n} \nu_j h_j(x)$$

and $\lambda = [\lambda_1, \dots, \lambda_m]^T$ and $\nu = [\nu_1, \dots, \nu_n]^T$ are called Lagrange multipliers or dual variables.

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



The dual optimization problem is

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

Similarly, the primal function is

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

and the primal optimization problem is

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

Notice that

$$L_P(x) = \begin{cases} f(x) & \text{if } x \text{ is feasible} \\ \infty & \text{otherwise.} \end{cases}$$

Therefore, the primal problem and the original problem have the same solution, yet the primal problem is unconstrained.

Weak Duality

Denote the optimal objective function values of the primal and dual

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

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

Weak duality refers to the following fact, which always holds.

Proposition $d^* \leq p^*$

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

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

Hence

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

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

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

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

$$d^* = \max_{\lambda, \nu : \lambda_i \ge 0} L_D(\lambda, \nu) \le p^*. \qquad \square$$

Strong Duality

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

The original unconstrained optimization problem is said to be convex if f and $g_1, \dots, g_m$ are convex functions and $h_1, \dots, h_n$ are affine.

We state the following without proof.

Theorem If the original problem is convex and a constraint qualification holds, then $p^* = d^*$.

Examples of constraint qualifications

- All g_i are affine
- (strict feasibility) $\exists\, x$ s.t. $h_j(x) = 0 \ \forall j$ and $g_i(x) < 0 \ \forall i$.

$g_i(x) = 0 \quad \forall i.$

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

From now on, assume $f, g_1, \dots, g_m, h_1, \dots, h_n$ are differentiable. For unconstrained optimization, we know $\nabla f(x^*) = 0$ is necessary for x^* to be a global minimizer, and sufficient if f is additionally convex. The following two results generalize these properties to constrained optimization.

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

(1) $\nabla f(x^*) + \sum_i \lambda_i^* \nabla g_i(x^*) + \sum_j \nu_j^* \nabla h_j(x^*) = 0$

(2) $g_i(x^*) \le 0 \quad \forall i$

(3) $h_j(x^*) = 0 \quad \forall j$

(4) $\lambda_i^* \ge 0 \quad \forall i$

(5) $\lambda_i^* g_i(x^*) = 0 \quad \forall i$ (complimentary slackness)

Proof (2)-(3) hold since x^* is feasible. (4) holds by definition of the dual problem. To prove (5) and (1):

$$
\begin{aligned}
f(x^*) &= L_D(\lambda^*, \nu^*) && \text{[by strong duality]} \\
&= \min_x \left(f(x) + \sum \lambda_i^* g_i(x) + \sum \nu_j^* h_j(x) \right) \\
&\le f(x^*) + \sum \lambda_i^* g_i(x^*) + \sum \nu_j^* h_j(x^*) \\
&\le f(x^*) && \text{[by (2)-(4)]}
\end{aligned}
$$

and therefore the two inequalities are equalities. Equality of the last two lines implies $\lambda_i^* g_i(x^*) = 0 \ \forall i$. 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 (Sufficiency) If the original problem is convex 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 strong duality holds.

Proof By (2) and (3), $\tilde{x}$ is feasible. By (4), $L(x, \tilde{\lambda}, \tilde{\nu})$

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

$$
\begin{aligned}
L_D(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\
&= f(\tilde{x}) + \underbrace{\sum \tilde{\lambda}_i g_i(\tilde{x}) + \sum \tilde{\nu}_j h_j(\tilde{x})}_{= 0 \text{ by } (5) \text{ and } (3)} \\
&= f(\tilde{x}).
\end{aligned}
$$

Therefore $p^* = d^*$ and the result follows. $\square$

In conclusion, if a constrained optimization problem is differentiable and convex, 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 $\tilde{x}$ is primal optimal, $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal, and strong duality holds, then $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ is a saddle point of L, i.e.,

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

for all $x \in \mathbb{R}^d$, $\lambda \in \mathbb{R}^m$ with $\lambda_i \geq 0$, and $\nu \in \mathbb{R}^n$.

The proof is left as an exercise.