# Duality

# Constrained Optimization Problem

Recall the constrained optimization problem:

$$\max_{x \in \mathcal{S}} f(x),$$

subject to

$$h_i(x) \leq 0, i = 1, 2, ..., I,$$
$$g_j(x) = 0, j = 1, 2, ..., J.$$

**Example**

Consider a resource allocation problem:

$$\max_{x} \sum_{r=0}^{2} \log x_r$$
$$x_0 + x_i \leq 1, \qquad \forall i = 1, 2 \quad \text{and} \quad x \geq 0,$$

# Duality

The Lagrangian of this optimization problem is defined to be

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{I} \lambda_i h_i(x) + \sum_{j=1}^{J} \mu_j g_j(x), \qquad \lambda_i \geq 0 \ \forall i.$$

The constants $\lambda_i \geq 0$ and $\mu_j$ are called Lagrange multipliers.

$$L(x, \lambda) = \sum_{r=0}^{2} \log x_r - \sum_{i=1}^{2} \lambda_i (x_0 + x_i - 1).$$

The *Lagrangian dual function* is defined to be

$$D(\lambda, \mu) = \sup_{x \in \mathcal{S}} L(x, \lambda, \mu).$$

# Duality

**Example**

$$D(\lambda) = \max_{x \geq 0} \sum_{r=0}^{2} \log x_r - \sum_{i=1}^{2} \lambda_i (x_0 + x_i - 1)$$

$$\implies \quad x_0^* = \frac{1}{\sum_i \lambda_i}, \quad x_i^* = \frac{1}{\lambda_i} (i \neq 0)$$

# Duality

Let $f^*$ be the maximum of the optimization problem, i.e.,

$$f^* = \max_{x \in \mathcal{S}} f(x).$$

Then, we have the following theorem.

**Theorem**

*$D(\lambda, \mu)$ is a convex function and $D(\lambda, \mu) \geq f^*$.*

**Example**

It is easy to verify that for $\lambda = 1$, we have

$$\begin{aligned} & x_0 = \frac{1}{2}, \quad x_1 = x_2 = 1 \\ \Longrightarrow \quad & D(1) = -\log 2 - 1 \approx -1.7 \geq f^* = -3\log 3 + 2\log 2 \approx -1.9 \end{aligned}$$

# Proof

The convexity comes from a known fact that the pointwise supremum of affine functions is convex.

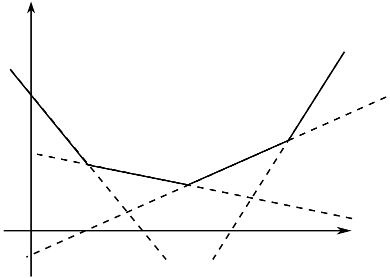

**Figure:** The solid line is the pointwise supremum of the four dashed lines, and is convex.

# Proof

To prove the bound, note that $h_i(x) \leq 0$ and $g_j(x) = 0$ for any feasible $x$, so the following inequality holds for any feasible $x$:

$$L(x, \lambda, \mu) \geq f(x) \Rightarrow \sup_{\substack{x \in \mathcal{S} \\ h(x) \leq 0 \\ g(x) = 0}} L(x, \lambda, \mu) \geq \sup_{\substack{x \in \mathcal{S} \\ h(x) \leq 0 \\ g(x) = 0}} f(x) = f^*.$$

Since removing some constraints of a maximization problem can only result in a larger maximum value, we obtain

$$\sup_{x \in \mathcal{S}} L(x, \lambda, \mu) \geq \sup_{\substack{x \in \mathcal{S} \\ h(x) \leq 0 \\ g(x) = 0}} L(x, \lambda, \mu) \Rightarrow \quad D(\lambda, \mu) = \sup_{x \in \mathcal{S}} L(x, \lambda, \mu) \geq f^*.$$

# Lagrange dual problem

**Lagrange dual problem**

$$\inf_{\lambda \geq 0, \mu} D(\lambda, \mu).$$

- Let $d^*$ be the minimum of the dual problem, i.e.,

$$d^* = \inf_{\lambda \geq 0, \mu} D(\lambda, \mu).$$

- Duality gap: $d^* - f^*$.
- Strong duality: $d^* = f^*$.
- If strong duality holds, then one can solve either the primal problem or the dual problem to obtain $f^*$.

# Slater's Condition

Recall the constrained optimization problem:

$$\max_{x \in \mathcal{S}} f(x)$$

$$\text{subject to:} \quad h_i(x) \leq 0, i = 1, 2, ..., I, \text{ and } g_j(x) = 0, j = 1, 2, ..., J.$$

**Slater's condition**

Consider the constrained optimization problem. Strong duality holds if the following conditions are true:

- $f(x)$ is a concave function and $h_i(x)$ are convex functions;
- $g_j(x)$ are affine functions;
- There exists an $x$ that belongs to the relative interior of $\mathcal{S}$ such that $h_i(x) < 0$ for all $i$ and $g_j(x) = 0$ for all $j$.

### Example

Consider the resource allocation problem:

$$\max_{x} \sum_{r=0}^{L} \log x_r$$

$$x_0 + x_l \le 1, \qquad \forall l = 1, \ldots, L \quad \text{and} \quad x \ge 0.$$

Slater's condition is satisfied by considering $x_r = 0.1$.

# Karush–Kuhn–Tucker (KKT) conditions

## Karush–Kuhn–Tucker (KKT) conditions

Consider the constrained optimization problem. Assume that $f$ and $h_i$ $(i = 1, 2, \ldots, l)$ are differentiable functions and that Slater's conditions are satisfied. Let $x^*$ be a feasible point, i.e., a point that satisfies all the constraints. Such an $x^*$ is a global maximizer for the optimization problem if and only if there exist constants $\lambda_i^* \geq 0$ and $\mu_j^*$ such that

$$\frac{\partial f}{\partial x_k}(x^*) - \sum_i \lambda_i^* \frac{\partial h_i}{\partial x_k}(x^*) + \sum_j \mu_j^* \frac{\partial g_j}{\partial x_k}(x^*) = 0, \quad \forall k, \tag{1}$$

$$\lambda_i^* h_i(x^*) = 0, \quad \forall i. \tag{2}$$

Further, (1) and (2) are also necessary and sufficient conditions for $(\lambda^*, \mu^*)$ to be a global minimizer of the Lagrange dual problem. If $f$ is strictly concave, then $x^*$ is also the unique global maximizer.

# Karush–Kuhn–Tucker (KKT) conditions

Consider the Lagrangian

$$L(x, \lambda, \mu) = f(x) - \sum_i \lambda_i h_i(x) + \sum_j \mu_j g_j(x).$$

- Condition (1) is the first-order necessary condition for the maximization problem $\max_{x \in \mathcal{S}} L(x, \lambda^*, \mu^*)$.
- Complementary slackness: When strong duality holds, we have

$$f(x^*) = f(x^*) - \sum_i \lambda_i^* h_i(x^*) + \sum_j \mu_j^* g_j(x^*),$$

which results in condition (2) since $g_j(x^*) = 0 \ \forall j$, and $\lambda_i^* \geq 0$ and $h_i(x^*) \leq 0 \ \forall i$.

# Karush–Kuhn–Tucker (KKT) conditions

$$\max_{x} \sum_{r=0}^{L} \log x_r$$

$$x_0 + x_l \leq 1, \qquad \forall l = 1, \dots, L \quad \text{and} \quad x \geq 0.$$

$$\frac{1}{x_0^*} - \sum_{l=1}^{L} \lambda_l^* = 0 \quad \text{and} \quad \frac{1}{x_r^*} - \lambda_r^* = 0, \;\; r = 1, \cdots, L$$

$$\lambda_l^*(x_0 + x_l^* - 1) = 0.$$

$$x_0^* = \frac{1}{L+1}, \; x_r^* = \frac{L}{L+1}, \; \lambda_l^* = \frac{L+1}{L}$$

# Reference


- R. Srikant and L. Ying. Chapter 2.1.2, Communication networks: an optimization, control, and stochastic networks perspective. Cambridge University Press, 2013.
- Efroni Yonathan, Shie Mannor, and Matteo Pirotta. "Exploration-exploitation in constrained MDPs." arXiv preprint arXiv:2003.02189, 2020.
- H. Wei, X. Liu, and L. Ying. "Triple-Q: A Model-Free Algorithm for Constrained Reinforcement Learning with Sublinear Regret and Zero Constraint Violation", AISTATS 2022.