# Optimization

**David S. Hippocampus***
Department of Computer Science
Cranberry-Lemon University
Pittsburgh, PA 15213
hippo@cs.cranberry-lemon.edu

## Abstract

Optimization

## 1 Lagrange Dual

Suppose we want to optimize $x \in \mathbb{R}^n$:

$$\begin{aligned} &\min_x f(x) \\ s.t. \qquad & c_i(x) \leq 0 \\ & h_j(x) = 0 \end{aligned}$$

We introduce the generalized Lagrange function:

$$L(x, \alpha, \beta) = f(x) + \sum_i \alpha_i c_i(x) + \sum_j \beta_j h_j(x) \qquad \alpha_i \geq 0 \tag{1}$$

Then the **primal function**:

$$\theta_P(x) = \max_{\alpha,\beta:\alpha_i \geq 0} L(x, \alpha, \beta) \tag{2}$$

**Important Properties**:

1. if the condition of x does not satisfy, i.e., $c_i(x) > 0$ or $h_j(x) \neq 0$, then we have
$$\theta_P(x) = +\infty$$
2. if x satisfies, we have
$$\theta_P(x) = f(x)$$

We define $p^*$ as the solution of the primal problem:

$$p^* = \min_x \theta_P(x) \tag{3}$$

### 1.1 Dual Problem

We define a function of $\alpha, \beta$:

$$\theta_D(\alpha, \beta) = \min_x L(x, \alpha, \beta) \tag{4}$$

We define the dual problem as:

$$d^* = \max_{\alpha,\beta:\alpha_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha,\beta:\alpha_i \geq 0} \min_x L(x, \alpha, \beta) \tag{5}$$

*

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

### 1.2 Properties

If both $p^*$ and $d^*$ exists, we have:

$$d^* \leq p^* \tag{6}$$

**KKT** condition, if following conditions hold, we have $d^* = p^*$:

1. $f(x)$ and inequality condition $c_i(x)$ both convex;
2. Equality condition $h_i(x)$ is an affine transform;
3. Exist $x$ satisfies $c_i(x) < 0$

**KKT** condition: if $\{x^*, \alpha, \beta\}$ with $x^*$ as a local optimum, we have

1. Stationary:
$$\nabla f(x^*) + \sum_i \alpha_i \nabla c_i(x) + \sum_j \beta_j \nabla h_j(x) = 0 \tag{7}$$
2. Primal feasibility:
$$c_i(x^*) \leq 0 \tag{8}$$
$$h_j(x^*) = 0 \tag{9}$$
3. Dual feasibility:
$$\alpha_i \geq 0 \tag{10}$$
4. Complementary slackness:
$$\alpha_i c_i(x^*) = 0 \tag{11}$$

### 1.3 Application 1: SVM

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{12}$$

$$s.t. \qquad (w^T x_i + b) y_i \geq 1 \tag{13}$$

The function with Lagrange multiplier is:

$$L(w, b; \alpha) = \frac{1}{2}||w||^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i(w x_i + b))$$

Stability:

$$\nabla_w L = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{14}$$

$$\nabla_b L = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{15}$$

Then, make the substitution for $w$, we have:

$$\begin{aligned} L(\alpha) &= \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i - \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j - b(\sum_i \alpha_i y_i) \\ &= \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad s.t. \qquad \sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned}$$

We will maximize the dual subject to the constraint.