Lagrange multipliers

Course of Machine Learning
Master Degree in Computer Science
University of Rome ``Tor Vergata''

Giorgio Gambosi

a.a. 2017-2018

Lagrange method

Widely applied to solve constrained optimization problems, such as

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$h_i(\mathbf{x}) = 0 \qquad i = 1, \dots, l$$

The *lagrangian* of this problem is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i h_i(\mathbf{x})$$

The coefficients λ_i are said *lagrangian multipliers*

Lagrangian optimization

Finding the solutions of

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x}} = \mathbf{0}$$

is a means to identify the values $\mathbf{x}$ which minimize (or maximize) $L(\mathbf{x}, \boldsymbol{\lambda})$.

Also, the solutions of setting to zero the derivates wrt to multipliers

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \mathbf{0}$$

correspond to setting $h_i(\mathbf{x}) = 0$ for all i, that is to satisfy all the constraints.

Hence, requiring that all derivatives are equal to zero is equivalent to solve the original optimization problem wrt $f(\mathbf{x})$, while satisfying all constraints.

Example

Given the following minimization problem,

$$\min_{x_1, x_2} 1 - x_1^2 - x_2^2$$

$$x_1 + x_2 - 1 = 0 \qquad i = 1, \dots, l$$

the corresponding lagrangian is defined

$$L(x_1, x_2, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$$

Setting all derivatives (wrt x_1, x_2, and λ) to zero, we get

$$\frac{\partial}{\partial x_1} L(x_1, x_2, \lambda) = -2x_1 + \lambda = 0$$
$$\frac{\partial}{\partial x_2} L(x_1, x_2, \lambda) = -2x_2 + \lambda = 0$$
$$\frac{\partial}{\partial \lambda} L(x_1, x_2, \lambda) = x_1 + x_2 - 1 = 0$$

which results into $x_1 = 1/2$ and $x_2 = 1/2$ (the solution $\lambda = 1$ also results, but this value is of minor relevance).

In the general definition, not all constraints are defined in terms of equalities. In this case, the following general problem $\mathcal{P}$ is considered:

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$g_i(\mathbf{x}) \leq 0 \qquad i = 1, \ldots, k$$
$$h_i(\mathbf{x}) = 0 \qquad i = 1, \ldots, l$$

Let us introduce the generalized lagrangian

$$L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{l} \alpha_i h_i(\mathbf{x})$$

where the set of lagrangian multipliers (also denoted as dual variables) is given by $\boldsymbol{\alpha} \cup \boldsymbol{\lambda}$, where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_l\}$ and $\boldsymbol{\lambda} = \{\lambda_i, \ldots, \lambda_k\}$).

Primal problem

Let us introduce the maximization problem $\mathcal{P}_p$

$$\theta_p(\mathbf{x}) = \max_{\boldsymbol{\lambda},\boldsymbol{\alpha}:\lambda_i \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$$

For any $\mathbf{x}$,

1. if $\mathbf{x}$ violates a constraint in $\mathcal{P}$ (either $g_i(\mathbf{x}) > 0$ or $h_i(\mathbf{x}) \neq 0$, for some i), then $\theta_p(\mathbf{x})$ can be arbitrarily large.
2. if $\mathbf{x}$ satisfies all constraints in $\mathcal{P}$, then $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) = f(\mathbf{x})$ for all $\boldsymbol{\lambda}, \boldsymbol{\alpha}$ and, as a consequence, $\theta_p(\mathbf{x}) = f(\mathbf{x})$.

Hence,

- if all constraints defined in $\mathcal{P}$ are satisfied, the values of the objective functions of $\mathcal{P}$ and of $\mathcal{P}_p$ are equal (and their optimal values are then equal themselves)
- if some constraint in $\mathcal{P}$ is not satisfied, θ_p has value $+\infty$

Primal problem

The primal optimization problem $\mathcal{P}_1$ is defined as

$$\min_{\mathbf{x}} \theta_p(\mathbf{x}) = \min_{\mathbf{x}} \max_{\boldsymbol{\lambda},\boldsymbol{\alpha}:\lambda_i \geq 0} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$$

with optimal value $p^* = \min_{\mathbf{x}} \theta_p(\mathbf{x})$. The problem has the following properties

1. all feasible solutions of $\mathcal{P}$ are feasible solutions of $\mathcal{P}_1$
2. each unfeasible solution of $\mathcal{P}$ is a feasible solution of $\mathcal{P}_1$, with value $+\infty$ of the objective function

As a consequence

1. all solutions of $\mathcal{P}_1$ are feasible
2. the optimal solution of $\mathcal{P}$ has value p^*

Dual problem

Let us introduce the dual function

$$\theta_d(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$$

and consider the dual problem

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \geq 0} \theta_d(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \geq 0} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha})$$

with optimum value $d^* = \max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \geq 0} \theta_d(\boldsymbol{\lambda}, \boldsymbol{\alpha})$.

In general, for all functions $f(x, y)$ it holds

$$\max_x \min_y f(x, y) \le \min_y \max_x f(x, y)$$

Then,

$$d^* = \max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \ge 0} \theta_d(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \ge 0} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) \le$$
$$= \min_{\mathbf{x}} \max_{\boldsymbol{\lambda}, \boldsymbol{\alpha} : \lambda_i \ge 0} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) = \min_{\mathbf{x}} \theta_p(\mathbf{x}) = p^*$$

Relations between primal and dual

In the case of linear constraints and convex objective function (non negative second derivative), the optimal solutions $\boldsymbol{\lambda}^*$, $\boldsymbol{\alpha}^*$ of $\theta_d(\boldsymbol{\lambda}, \boldsymbol{\alpha})$, and the optimal solution $\mathbf{x}^*$ of $\theta_p(\mathbf{x})$ are such that

$$p^* = \theta_p(\mathbf{x}^*) = \theta_d(\boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*) = d^*$$

and

$$p^* = d^* = L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*)$$

Under the same hypotheses, the optimum values $\boldsymbol{\lambda}^*$, $\boldsymbol{\alpha}^*$, $\mathbf{x}^*$ satisfy the following Karush-Kuhn-Tucker (KKT) conditions

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*) &= \mathbf{0} & & \text{null gradient} \\
g_i(\mathbf{x}^*) &\leq 0 & i = 1, \ldots, k & \quad \text{inequality constraints} \\
h_i(\mathbf{x}^*) &= 0 & i = 1, \ldots, l & \quad \text{equality constraints} \\
\lambda_i^* &\geq 0 & i = 1, \ldots, k & \quad \text{multipliers of inequality constraints} \\
\lambda_i^* g_i(\mathbf{x}^*) &= 0 & i = 1, \ldots, k & \quad \text{complementary slackness}
\end{aligned}
$$

KKT conditions

Condition 1

$$\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*) = \mathbf{0}$$

states that the gradient must be null for the optimum solution.

Conditions 2-3

$$g_i(\mathbf{x}^*) \leq 0 \qquad i = 1, \ldots, k$$

and

$$h_i(\mathbf{x}^*) = 0 \qquad i = 1, \ldots, l$$

are just the original problem constraints.

Condition 4

$$\lambda_i^* \geq 0 \qquad i = 1, \ldots, k$$

for inequality constraints.

KKT conditions

Condition 5

$$\lambda_i^* g_i(\mathbf{x}^*) = 0$$

implies that if $\lambda_i^* > 0$ then $g_i(\mathbf{x}) = 0$, that is the constraint $g_i(\mathbf{x}) \geq 0$ is satisfied at the limit (with equality): in this case, the constraint is said *attivo*.