Lagrange multiplier and KKT condition

陳弘軒 Hung-Hsuan Chen
https://www.ncu.edu.tw/~hhchen
中央大學資訊工程學系

Toy example

- Maximize $5x_1x_2$ subject to $2x_1 + x_2 = 100$
- Can we solve the optimization problem by gradient descent/ascent?
  - Probably not, because we have extra constraints

Toy example

- Maximize $5x_1x_2$ subject to $2x_1 + x_2 = 100$

- Sol 1

  $f(x_1, x_2) = 5x_1x_2 = 5x_1(100 - 2x_1)$

  $\rightarrow \dfrac{\partial f(x_1, x_2)}{\partial x_1} = 500 - 20x_1 := 0$

  $\rightarrow (x_1, x_2) = (25, 50)$

  $\rightarrow$ The maximum value of $5x_1x_2$ is 6250

- If the constraints are not complicated, such a method is manageable

Lagrange multiplier

- Lagrange multipliers is a strategy for finding the extreme value of a function ***subject to equality constraints***

- Finding the maximum value of $y = f(\boldsymbol{x})$ subject to $g_i(\boldsymbol{x}) = 0, i = 1,2, \ldots, m$

  - Lagrange function (a.k.a. Lagrangian)

    - $y_\lambda = f(\boldsymbol{x}) + \lambda_1 g_1(\boldsymbol{x}) + \lambda_2 g_2(\boldsymbol{x}) + \cdots + \lambda_m g_m(\boldsymbol{x})$

  - λ_i's are called "Lagrange multipliers"

Solving the problem

- Lagrange function: $\mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$

- Take the derivative of the Lagrange function to every variable (i.e., all the *x*'s and the λ's)

$$\Rightarrow \begin{cases} \dfrac{\partial \mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m)}{\partial x_j} = 0 \ \forall j \\ \dfrac{\partial \mathcal{L}(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m)}{\partial \lambda_i} = 0 \ \forall i \end{cases}$$

- Test each solution set. Whichever gives the greatest (or the smallest) value is the maximum (or minimum) point

Solving the toy example

- Maximize $5x_1x_2$ subject to $2x_1 + x_2 = 100$

$$f(x_1, x_2) = 5x_1x_2, g(x_1, x_2) = 2x_1 + x_2 - 100$$

$$\mathcal{L}(x_1, x_2, \lambda_1) = 5x_1x_2 + \lambda_1(2x_1 + x_2 - 100)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}(x_1, x_2, \lambda_1)}{\partial x_1} = 5x_2 + 2\lambda_1 \coloneqq 0 \\ \dfrac{\partial \mathcal{L}(x_1, x_2, \lambda_1)}{\partial x_2} = 5x_1 + \lambda_1 \coloneqq 0 \\ \dfrac{\partial \mathcal{L}(x_1, x_2, \lambda_1)}{\partial \lambda_1} = 2x_1 + x_2 - 100 \coloneqq 0 \end{cases}$$

$$\Rightarrow (x_1, x_2, \lambda_1) = (25, 50, -125)$$

Generalized Lagrange multiplier

- Lagrange multipliers is generalized to include the ***inequality constraints*** under the Karush–Kuhn–Tucker (KKT) condition

- Standard form problem

Minimize $f(\boldsymbol{x})$ subject to $g_i(\boldsymbol{x}) \leq 0\ (i = 1, \dots, p)$ and $h_j(\boldsymbol{x}) = 0\ (j = 1, \dots, m)$

  - If the task is to maximize $f(\boldsymbol{x})$, transform the problem into minimize $-f(\boldsymbol{x})$

- Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j h_j(\boldsymbol{x})$$

Necessary (not sufficient!) optimal condition (KKT condition)

- If $\boldsymbol{x}^*$ is the optimal solution to the standard form problem, then there exist KKT multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that
  - Lagrangian optimality in $\boldsymbol{x}$

$$\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 ------ (1)$$

  - Primal feasibility

$$g_i(\boldsymbol{x}^*) \leq 0 \ \forall i ------- (2)$$
$$h_j(\boldsymbol{x}^*) = 0 \ \forall j ------- (3)$$

  - **Dual feasibility**

$$\lambda_i \geq 0 \ \forall i --------- (4)$$

  - **Complementary slackness**

$$\lambda_i g_i(\boldsymbol{x}^*) = 0 \ \forall i ------ (5)$$

Example

- Minimize $f(x_1, x_2) = x_1^2 + 2x_2^2$ subject to $x_1 + x_2 \geq 3$ and $x_2 - x_1^2 \geq 1$

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = x_1^2 + 2x_2^2 + \lambda_1(3 - x_1 - x_2) + \lambda_2(1 + x_1^2 - x_2)$$

$$\Rightarrow \begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_1} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial x_2} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_1} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \end{cases} \Rightarrow (x_1, x_2, \lambda_1, \lambda_2) = (-2, 5, 12, -8) \text{ or } (1, 2, 6, 2)$$

$\Rightarrow (-2, 5, 12, -8)$ does not follow the KKT condition (4)
$(\because \lambda_2 = -8 < 0)$
$\Rightarrow \min f(x_1, x_2) = f(1,2) = 1^2 + 2 \cdot 2^2 = 9$

Summary

- When we have constraints on an optimization problem, gradient descent/ascent may not be applicable

- Lagrange multiplier is a useful tool to solve the optimization problem with equality and inequality constraints