

Understanding the Lagrangian in Constrained Optimization

Xiaolei Jiang

School of Control and Computer Engineering, North China Electric Power University, Baoding Hebei, 071003, China

Abstract: A new approach to teaching constrained optimization is proposed, where the Lagrange multipliers in the Lagrangian are viewed as undetermined coefficients, and the Lagrangian are viewed as a weighted sum of multiple objective functions. This perspective facilitates a more intuitive and natural derivation of the Karush-Kuhn-Tucker (KKT) conditions, including the stationary and complementary slackness conditions, assisting students in appreciating their significance in determining the minimizer. This approach helps students understand how the KKT conditions come into being, allowing them to better grasp and master the optimality conditions, ultimately enabling them to apply the theoretical knowledge to real-world problems.

Keywords: Constrained Optimization; Lagrangian; KKT Conditions; Complementary Slackness Conditions.

1. Introduction

Optimization problems are ubiquitous in engineering practice and form the essential knowledge foundation for many artificial intelligence courses, including machine learning. Textbooks on optimization usually introduce the KKT conditions from a geometric perspective [1,2,3], but the approach to introducing complementary slackness conditions can be further explored. Students often find the Lagrangian difficult to understand. We propose an alternative teaching approach by examining the Lagrangian through the lens of undetermined coefficients, that is to say, we think of Lagrange multipliers as adaptive parameters that can be adjusted. This point of view naturally leads to the KKT conditions, including the stationary and complementary slackness conditions. We believe this teaching approach can help students better understand and master the optimality conditions for constrained optimization problems.

The rest of this article is organized as follows. After reviewing the definition of the Lagrangian of a constrained optimization problem, we emphasize the same form shared by both the Lagrangian and many objective functions encountered in machine learning. Section 3 addresses the case in which only equality constraints are present. The general mathematical programming problem which contains both inequality and equality constraints are discussed in Section 4 and 5. We draw our conclusion in Section 6.

2. The Lagrangian of a Constrained Optimization Problem

The general form of a mathematical programming problem, with both equality constraints and inequality constraints, is

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

where $g_1(\mathbf{x}) \leq 0, \dots, g_m(\mathbf{x}) \leq 0$ are inequality constraints, $h_1(\mathbf{x}) = 0, \dots, h_l(\mathbf{x}) = 0$ are equality constraints, $X \subseteq \mathbb{R}^n$, and $f(\mathbf{x})$ is the objective function. It is understood that $f, g_1, \dots, g_m, h_1, \dots, h_l$ are all defined on X .

Since all inequality constraints are in the form of less than or equal to, the constraints of strict inequality, if any, can be formulated into X . The Lagrangian of problem (1) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^l \mu_j h_j(\mathbf{x}), \mathbf{x} \in X, \boldsymbol{\lambda} \in \mathbb{R}_+^m, \boldsymbol{\mu} \in \mathbb{R}^l, \quad (2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_l)$. $\lambda_1, \dots, \lambda_m$ are nonnegative Lagrange multipliers associated with the inequality constraints, and $\mu_1, \dots, \mu_l$ are the Lagrange multipliers associated with the equality constraints. It should be noted that each component of $\boldsymbol{\lambda}$ is nonnegative.

Some practical problems have objective functions that take the same form as the Lagrangian in Equation (2). For example, in optimization problems from machine learning, the loss function often takes the following form:

$$L(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + \lambda_1 g_1(\boldsymbol{\theta}) + \lambda_2 g_2(\boldsymbol{\theta}) \quad (3)$$

where $\boldsymbol{\theta}$ denotes the adaptive parameter vector in a model, $f(\boldsymbol{\theta})$ is the data term, $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$ are two regularization terms, and λ_1 , λ_2 are the associated weight coefficients controlling the relative importance of $f(\mathbf{x})$ with respect to $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$. Usually, the data term comes from likelihood and regularization terms originate from priors. Since the Lagrangian in Equation (2) shares the same form as the objective function in Equation (3), we can view $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_l$ as undetermined coefficients. Specifically, We want to find appropriate Lagrange multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$, such that the minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a minimizer of problem (1). Since the constraints in problem (1) are encapsulated into the Lagrangian, it may be easier to minimize $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ with respect to $\mathbf{x}$ than to solve problem (1) directly.

For the sake of clarity, in the following discussion, we first consider the case with only equality constraints, and then consider the case where both equality and inequality constraints are present.

3. With Only Equality Constraints

If there is no inequality constraint in problem (1), then it degenerates into the constrained optimization problem that students have already learned in their calculus course. Instead of solving some variables in terms of other variables from equality constraints and transforming from constrained

minimization to unconstrained minimization, we can find a minimizer of problem (1) from the minimizer of $f(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$, where $\mu_1^*, \dots, \mu_l^*$ are adaptive parameters that can be adjusted. We start from the following proposition.

Proposition 1 If $\mathbf{x}^*$ is a minimizer of

$$\begin{aligned} \min f(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}) \\ \text{s.t. } \mathbf{x} \in X \end{aligned} \quad (4)$$

then $\mathbf{x}^*$ is also a minimizer of

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h_j(\mathbf{x}) = h_j(\mathbf{x}^*), j = 1, \dots, l \\ \mathbf{x} \in X \end{aligned}$$

Proof Suppose $\mathbf{x} \in X$ satisfies

$$h_j(\mathbf{x}) = h_j(\mathbf{x}^*), j = 1, \dots, l$$

Since $\mathbf{x}^*$ is a minimizer of problem (4),

$$f(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}^*) \leq f(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}).$$

Hence $f(\mathbf{x}^*) \leq f(\mathbf{x})$.

From Proposition 1 we can immediately have the following corollary.

Corollary 1 If $\mathbf{x}^*$ is a minimizer of

$$\min f(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$$

$$\text{s.t. } \mathbf{x} \in X$$

and

$$h_j(\mathbf{x}^*) = 0, j = 1, \dots, l$$

then $\mathbf{x}^*$ is also a minimizer of

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.t. } h_j(\mathbf{x}) = 0, j = 1, \dots, l \\ \mathbf{x} \in X \end{aligned}$$

To find these undetermined coefficients $\mu_1^*, \dots, \mu_l^*$, We can consider the following scenario. Let $\mathbf{x}^*$ be a minimizer of problem (4). For clarity, we only consider one equality constraint, leaving the others aside for now. Clearly, the larger μ_j^* is, the smaller $h_j(\mathbf{x}^*)$ becomes. Figure 1 provides an example with only one equality constraint, which illustrates how the minimizer of the Lagrangian changes as μ varies. Our goal is to solve problem (4) by way of Corollary 1, so we hope to have $h_j(\mathbf{x}^*) = 0$. If $h_j(\mathbf{x}^*) > 0$, it indicates that μ_j^* is too small; if $h_j(\mathbf{x}^*) < 0$, it indicates that μ_j^* is too large. This adjustment process can be iterative, where we start with an initial guess for $\mu_1^*, \dots, \mu_l^*$, evaluate $h_j(\mathbf{x}^*)$, and then adjust μ_j^* accordingly until the condition $h_j(\mathbf{x}^*) = 0$ is satisfied. Therefore, there is reason to believe that there exists suitable $\mu_1^*, \dots, \mu_l^*$ such that $h_j(\mathbf{x}^*) = 0, j = 1, \dots, l$.

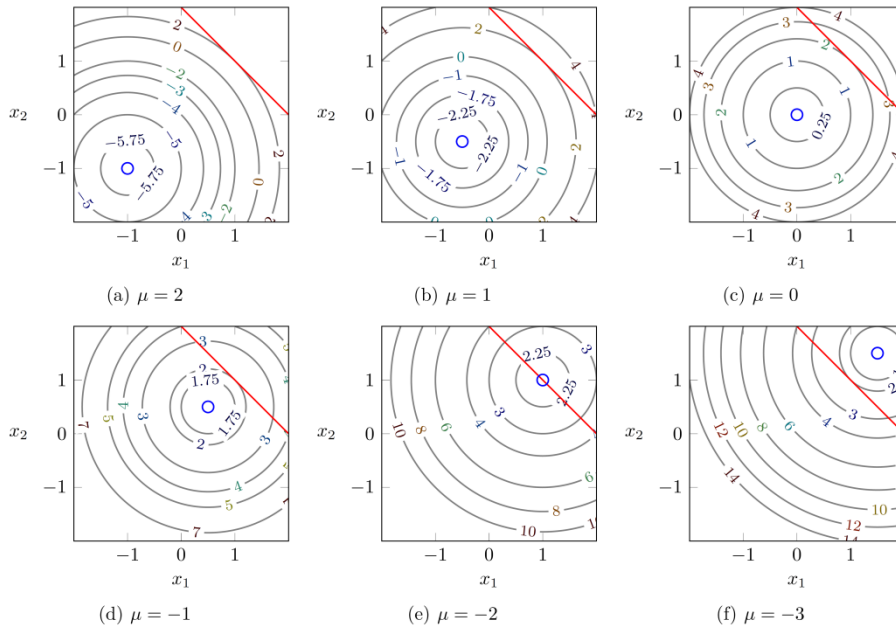


Figure 1. A larger μ leads to a smaller $h(\mathbf{x}^*)$. The objective function is $f(\mathbf{x}) = x_1^2 + x_2^2$, the equality constraint is $h(\mathbf{x}) = x_1 + x_2 - 2 = 0$, and the Lagrangian is $L(x_1, x_2, \mu) = x_1^2 + x_2^2 + \mu(x_1 + x_2 - 2)$. The gray lines represent the level curves of the Lagrangian, the small blue circles indicate the minimizer of the Lagrangian, and the red line represents the equality constraint.

It is worth noting that the iterative adjustment process described above is not the actual method used to find the undetermined coefficients. It is merely intended to facilitate understanding the existence of these coefficients. To find these undetermined coefficients, assuming that $f(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$ is differentiable, then we have

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* \nabla h_j(\mathbf{x}^*) = \\ h_j(\mathbf{x}^*) = 0, \quad j = 1, \dots, l \end{aligned}$$

This accords with the Lagrange multiplier method.

4. With Both Equality and Inequality Constraints: Simple Case

Inequality constraints are more complex to handle than equality constraints. We start with a simple case in this section and consider the general case in the next section. Proposition 1 can be extended to cases involving inequality constraints as follows.

Proposition 2 Suppose $\lambda^* \in \mathbb{R}_+^m$ and $\mu^* \in \mathbb{R}^l$. If $\mathbf{x}^*$ is a minimizer of

$$\min L(\mathbf{x}, \lambda^*, \mu^*)$$

$$\text{s.t. } \mathbf{x} \in X$$

then $\mathbf{x}^*$ is also a minimizer of

$$\min f(\mathbf{x})$$

$$\text{s.t. } g_i(\mathbf{x}) \leq g_i(\mathbf{x}^*), i = 1, \dots, m$$

$$h_j(\mathbf{x}) = h_j(\mathbf{x}^*), j = 1, \dots, l$$

$$\mathbf{x} \in X$$

Proof To make it easier for students to understand the correctness of this proposition, two proofs are provided below: one is a direct proof, and the other is a proof by contradiction.

Method 1 (direct proof): For all $\mathbf{x} \in X$ satisfying

$$g_i(\mathbf{x}) \leq g_i(\mathbf{x}^*), i = 1, \dots, m$$

$$h_j(\mathbf{x}) = h_j(\mathbf{x}^*), j = 1, \dots, l$$

we have

$$f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}^*)$$

$$= L(\mathbf{x}^*, \lambda^*, \mu^*)$$

$$\leq L(\mathbf{x}, \lambda^*, \mu^*)$$

$$= f(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$$

$$\leq f(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}^*).$$

Hence $f(\mathbf{x}^*) \leq f(\mathbf{x})$.

Method 2 (proof by contradiction): Suppose there exists $\mathbf{x} \in X$ such that

$$g_i(\mathbf{x}) \leq g_i(\mathbf{x}^*), i = 1, \dots, m$$

$$h_j(\mathbf{x}) = h_j(\mathbf{x}^*), j = 1, \dots, l$$

and $f(\mathbf{x}) < f(\mathbf{x}^*)$, then

$$L(\mathbf{x}, \lambda^*, \mu^*)$$

$$= f(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$$

$$< f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}^*)$$

$$= L(\mathbf{x}^*, \lambda^*, \mu^*).$$

This leads to a contradiction, and therefore, the conclusion holds.

It should be noted that the Lagrange multipliers for inequality constraints is nonnegative, and this condition is indispensable when proving Proposition 2. From Proposition 2 the following corollary can be derived.

Corollary 2 Suppose $\lambda^* \in \mathbb{R}_+^m$ and $\mu^* \in \mathbb{R}^l$. If $\mathbf{x}^*$ is a minimizer of

$$\min L(\mathbf{x}, \lambda^*, \mu^*)$$

$$\text{s.t. } \mathbf{x} \in X$$

and

$$g_i(\mathbf{x}^*) = 0, i = 1, \dots, m$$

$$h_j(\mathbf{x}^*) = 0, j = 1, \dots, l$$

then $\mathbf{x}^*$ is also a minimizer of problem (1).

So if we can find appropriate Lagrange multipliers λ^* and μ^* such that not only equality constraints are satisfied, but also the equality holds for all inequality constraints, then $\mathbf{x}^*$ is also a minimizer of problem (1). However, the restrictions on the minimizers of problem (1) found in this way are somewhat too strong, because for these minimizers, the inequality constraints are all active constraints. In the next section, we will discuss the general case, where it is not

required that inequality constraints must be active at minimizers.

5. With Both Equality and Inequality Constraints: General Case

The Lagrange multipliers of equality constraints can be found by iterative adjustments, as envisioned in Section 3. Similar to μ_j^* , increasing λ_i^* makes $g_i(\mathbf{x}^*)$ smaller. Suppose we want find a minimizer of problem (1) with the help of Corollary 2. If $g_i(\mathbf{x}^*) > 0$, it indicates that λ_i^* is too small; if $g_i(\mathbf{x}^*) < 0$, it indicates that λ_i^* is too large. Since $\lambda_i^* \geq 0$ is one of the conditions used in Corollary 2, λ_i^* cannot be adjusted freely like μ_j^* . If $g_i(\mathbf{x}^*) < 0$ and $\lambda_i^* = 0$, this means that after removing the term $\lambda_i^* g_i(\mathbf{x}^*)$ corresponding to the inequality constraint $g_i(\mathbf{x}^*) \leq 0$ from $L(\mathbf{x}, \lambda^*, \mu^*)$, the minimizer $\mathbf{x}^*$ satisfies $g_i(\mathbf{x}^*) \leq 0$, so $\mathbf{x}^*$ is a minimizer of problem (1) and the constraint $g_i(\mathbf{x}^*) \leq 0$ is inactive. Similarly, if $g_i(\mathbf{x}^*) = 0$ and $\lambda_i^* = 0$, then $\mathbf{x}^*$ is a minimizer of problem (1) and the constraint $g_i(\mathbf{x}^*) \leq 0$ is active. If $g_i(\mathbf{x}^*) > 0$ and $\lambda_i^* = 0$, λ_i^* should be increased. For an appropriate $\lambda_i^* > 0$, we expect $g_i(\mathbf{x}^*) = 0$, and in this case $\mathbf{x}^*$ is a minimizer of problem (1) and the constraint $g_i(\mathbf{x}^*) \leq 0$ is active. We have discussed three cases for an inequality constraint:

$$(1) \lambda_i^* = 0, g_i(\mathbf{x}^*) < 0,$$

$$(2) \lambda_i^* = 0, g_i(\mathbf{x}^*) = 0,$$

$$(3) \lambda_i^* > 0, g_i(\mathbf{x}^*) = 0.$$

In all three cases, $\mathbf{x}^*$ is a minimizer of problem (1). We can combine these three cases into $g_i(\mathbf{x}^*) \leq 0$ and $\lambda_i^* g_i(\mathbf{x}^*) = 0$. So the complementary slackness condition can be naturally derived from the perspective of undetermined coefficients. The above analysis leads us to the following proposition.

Proposition 3 Suppose $\lambda^* \in \mathbb{R}_+^m$ and $\mu^* \in \mathbb{R}^l$. If $\mathbf{x}^*$ is a minimizer of

$$\min L(\mathbf{x}, \lambda^*, \mu^*)$$

$$\text{s.t. } \mathbf{x} \in X$$

and

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, \dots, m$$

$$g_i(\mathbf{x}^*) \leq 0, i = 1, \dots, m$$

$$h_j(\mathbf{x}^*) = 0, j = 1, \dots, l$$

then $\mathbf{x}^*$ is also a minimizer of problem (1).

Proof Suppose $\mathbf{x}$ is a feasible point of problem (1), then we have

$$f(\mathbf{x}^*)$$

$$= f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x}^*)$$

$$= L(\mathbf{x}^*, \lambda^*, \mu^*)$$

$$\leq L(\mathbf{x}, \lambda^*, \mu^*)$$

$$= f(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}) + \sum_{j=1}^l \mu_j^* h_j(\mathbf{x})$$

$$\leq f(\mathbf{x}).$$

To find these undetermined coefficients, assuming that $L(\mathbf{x}, \lambda^*, \mu^*)$ is differentiable, then we have

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^l \mu_j^* \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

$$\begin{aligned}\lambda_i^* g_i(\mathbf{x}^*) &= 0, & i &= 1, \dots, m \\ g_i(\mathbf{x}^*) &\leq 0, & i &= 1, \dots, m \\ h_j(\mathbf{x}^*) &= 0, & j &= 1, \dots, l\end{aligned}$$

These are just KKT conditions.

6. Conclusion

We introduce a new teaching approach to constrained optimization by considering the Lagrangian from the perspective of a weighted sum of several objective functions. By viewing Lagrange multipliers as undetermined coefficients, students are able to develop a clearer and more intuitive understanding of the optimality conditions associated with constrained optimization problems. This method not only simplifies the process of deriving the stationary and complementary slackness conditions but also provides students with a more structured approach to understanding how the different components of the

optimization problem are interconnected. Ultimately, this approach helps students grasp the KKT conditions more effectively, deepening their understanding of optimization theory and enhancing both their theoretical knowledge and analytical skills.

Acknowledgments

The author thanks his colleges and students for their kindness and support.

References

- [1] Amir Beck (2023). Introduction to nonlinear optimization: theory, algorithms, and applications with python and matlab (2nd edition), SIAM.
- [2] Jorge Nocedal, Stephen J. Wright (2006). Numerical optimization (2nd edition). Springer.
- [3] Mordecai Avriel (2003). Nolinear programming: analysis and methods. Dover.