Lagrangian Multiplier

and also Augmented Lagrangian Multiplier

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : August 2, 2017
Last update : January 26, 2020

Overview

Constrained optimization

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$
$$\text{s.t.} \quad g_i(\mathbf{x}) = 0 \quad i = 1, 2, ..., I$$

- f is called the objective function
- $\mathbf{x} \in \mathbb{R}^k$, it is the variable
- g_i are the equality constraints, there are totally I of them

In words, such problem ask for a solution that minimizes the objective $f(\mathbf{x})$ as long as all the equality constraints $g_i(\mathbf{x}) = 0$ are satisfied.

Such problem can be solved by **the method of Lagrange multiplier**.

The method of Lagrange multiplier

What : it is a way to find the minimizer of a equality constrained problem
How : by the **Lagrangian function**: put the objective f and the constraints g_i all together into a single minimization problem, where g_i are multiplied by a weighting factor λ called **Lagrangian multipliers**

$$L(x, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i g_i(\mathbf{x})$$

- The function $L : \mathbb{R}^k \times \mathbb{R}^I \rightarrow \mathbb{R}$ with dom$L =$dom$f \times \mathbb{R}^I$ is the **Lagrangian**. It is a weighted sum of objective and constraints : the weights of g_i are λ_i and the weight of f is 1
- The variables λ_i are the **Lagrangian multipliers**. The vector variable $\lambda = [\lambda_1 \ldots \lambda_I]$ has dimension I, which is the number of constraints.
- L has two arguments: $\mathbf{x}$ (original variable of f) and λ (newly introduced variable)

Interpretation of the Lagrangian Multiplier

A toy example : *You eat apples. Each apple has 300g of vitamin A and costs 2$. You need 900g of vitamin A. Find the minimum cost of apples for exactly 900g of vitamin A.*

Let x be the number of apples you buy. The problem can be written as

$$\min_x 2x \text{ subject to } 300x = 900.$$

The Lagrangian

$$L(x, \lambda) = 2x + \lambda(300x - 900).$$

The solution is simple : $x = 3$, but let's look at the physical meaning in details.

Interpretation of the Lagrangian Multiplier ... 2

Here is the Dimensional Analysis from Physics :

- x is the number of apples, it has no physical unit
- $2x$ has the unit \$ (dollar) as 2 carries the unit \$
- $300x - 900$ has the unit of gram (amount of mass)

So the expression $2x + \lambda(300x - 900)$ is equivalent to

$$\$ + \lambda \cdot \text{gram}$$

We cannot add dollar to gram. To make the expression shares the same unit (in \$), λ has to be of the unit "dollar per gram".

In general, the unit carried by λ is "cost per unit" : where "cost" is the unit carried by the objective function f and the "unit" is the unit carried by the constraints g.

Interpretation of the Lagrangian Multiplier ... 3

With the "cost per unit" interpretation of λ, then $\lambda(300x - 900)$ means:

- if $x = 3$, $\lambda(300x - 900) = 0$ and this term has no contribution to L
- if $x \neq 3$, says $x = 1$, Then $\lambda(300x - 900) = \lambda(-600)$ and it contributes -600λ to the Lagrangian L

So $\lambda(300x - 900)$ tells the amount of cost (money) of violating the constarints $300x = 900$, and λ tells the cost per amount of the violation.

When the constraint is satisfied $(300x = 900)$, then the term $\lambda(300x - 900)$ is deactivated (the term becomes zero) and it contributes nothing towards L

Hence, $\lambda(300x - 900)$ can be seen as a penalty term that penalizes L if constraint is violated.

In other words, Lagrangian is a kind of penalty method : put the constraint into objective function as penalty to turn a constrained optimization problem into a unconstrained optimization problem

How Lagrangian "helps" to find the solution

Recall that for unconstrained problem $\min_{\mathbf{x}} f(\mathbf{x})$, the first order optimality condition / Fermat's rule tells that $\nabla_{\mathbf{x}} f(\mathbf{x}) = 0$. That is, if $\mathbf{x}$ is a optimal solution to a unconstrained problem, we can just use the optimality condition above.

Now, L is a unconstrained formulation, the same approach applies : to find the optimal solution $\mathbf{x}$ to the Lagrangian, solve $\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0$. As L is a weighted sum of objective f and constraints g_i, so ∇L contains the information of gradients of both f and g_i :

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = \nabla_x f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i \nabla_{\mathbf{x}} g_i(\mathbf{x})$$

- If there is no constraint ($g = 0$), $\nabla_{\mathbf{x}} L = \nabla_{\mathbf{x}} f$: the gradient of L is just the gradient of f.
- If constraints exist ($g \neq 0$), the terms $\nabla_{\mathbf{x}} g_i$ modify the gradient of f and move the solution of $\nabla_{\mathbf{x}} L = 0$ to the feasible region.

* Note that there is no assumption on the convexity of f and g_i, so generally the roots of $\nabla_{\mathbf{x}} L = 0$ are only local minimizer of the problem.

Solving for the minimizer

- Recall $\mathbf{x} \in \mathbb{R}^k$ and $\lambda \in \mathbb{R}^I$, so $L(\mathbf{x}; \lambda)$ has totally $k + I$ variables.
- The equation $\nabla_{\mathbf{x}} L = 0$ only provides k equations.
- To solve a system of $k + I$ variables with k equations, we need I more equations.
- Such I equations come from the gradient of L with respect to λ : $\nabla_\lambda L = 0$, where $\dfrac{\partial L}{\partial \lambda_i} = g_i(\mathbf{x})$

The solution process

The Lagrangian approach converted the problem from

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g_i(\mathbf{x}) = 0,\ i = 1, 2, ..., I$$

to a problem of solving system of equations in the form

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0 \\ \nabla_{\lambda} L(\mathbf{x}, \lambda) = 0 \end{cases}$$

Note that in general, solving this system of equations can be hard : either the equations are highly non-linear, or I is very large, or k is very large. Specific techniques are developed for these situations.

Augmented Lagrangian

The original constrained problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g_i(\mathbf{x}) = 0,\ i = 1, 2, ..., I$$

Lagrangian

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i g_i(\mathbf{x})$$

Method of Lagrangian Multiplier

$$\min_{\mathbf{x}} L(\mathbf{x}, \lambda)$$

Penalty Method

$$\min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i \Phi(g_i(\mathbf{x}))$$

Augmented Lagrangian

$$L_\rho(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i g_i(\mathbf{x}) + \frac{\rho}{2} \Phi(g_i(\mathbf{x}))$$

here $\Phi(\cdot) = \cdot^2$

What is augmented Lagrangian

Augmented Lagrangian = Lagrangian + a quadratic term.
Example : the original constrained problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

Lagrangian

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \langle \lambda, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle$$

Augmented Lagrangian

$$\begin{aligned} L_\rho(\mathbf{x}, \lambda) &= L(\mathbf{x}, \lambda) + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \\ &= f(\mathbf{x}) + \langle \lambda, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \end{aligned}$$

What's the purpose of adding the quadratic term : smoothing. Adding the quadratic term, *under a properly selected* ρ, can improve the convexity of the L. That is, it is the quadratic term making solving $\min L_p$ easier than solving $\min L$, which comes with a trade off – we need to tune ρ.

Last page – summary

Problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g_i(\mathbf{x}) = 0,\ i = 1, 2, ..., I$$

- Lagrangian $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{I} \lambda_i g_i(\mathbf{x})$
- The interpretation of λ : cost per unit violation
- Lagrangian approach of solving the equality constrained problem as a system of equations
$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda) = 0 \\ \nabla_{\lambda} L(\mathbf{x}, \lambda) = 0 \end{cases}$$

Not discussed

- Convergence issues of method of Lagrangian multiplier
- Algorithm for method of augmented Lagrangian
- Convergence issues of method of augmented Lagrangian

End of document