More on Lagrange multipliers

CE 377K

April 21, 2015

REVIEW

The standard form for a nonlinear optimization problem is

$$\begin{array}{lrl} \min\limits_{\mathbf{x}} & f(\mathbf{x}) & \\ \text{s.t.} & g_1(\mathbf{x}) & \leq 0 \\ & \vdots & \\ & g_\ell(\mathbf{x}) & \leq 0 \\ & h_1(\mathbf{x}) & = 0 \\ & \vdots & \\ & h_m(\mathbf{x}) & = 0 \end{array}$$

The objective function is to be minimized; all other constraints are of the form $\leq$ or $=$.

What is a convex set?

What is a convex function?

What are some useful facts about convex sets and functions?

What is the method of Lagrange multipliers?

OUTLINE

- Interpretation of Lagrange multipliers
- Solving the transit frequency setting problem with Lagrange multipliers
- Inequality constraints
- Karush-Kuhn-Tucker conditions

MORE PERSPECTIVES ON LAGRANGE MULTIPLIERS

Sensitivity analysis

The numerical value of the Lagrange multiplier is useful in *sensitivity analysis*, and shows how much the objective function would change if the constraint was changed.

Assume that a constraint was changed from $h(\mathbf{x}) = 0$ to $h(\mathbf{x}) = u$, so the optimal solution changes from $\mathbf{x}^*$ to $\mathbf{x}^*(u)$.

The ratio of the difference between $f(\mathbf{x}^*(u)$ and $f(\mathbf{x}^*)$ to the perturbation u is approximately $-\lambda$ when u is small.

$$\frac{df(\mathbf{x}^*)}{du} = -\lambda$$

This is often called a *shadow cost*.

The stationary point of $\min(-x_1 - x_2)$ subject to $x_1^2 + x_2^2 = 1$ was $x_1 = x_2 = \lambda = 1/\sqrt{2}$

If the right-hand side of the constraint was changed slightly (say, to 1.1), $u = 0.1$ so the change in the objective function will be approximately $-0.1/\sqrt{2}$

A geometric interpretation

At a stationary point of the Lagrangian, $\nabla \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = 0$.

This gradient has two parts: the partial derivatives with respect to $\mathbf{x}$ and those with respect to $\boldsymbol{\lambda}$.

The partial derivatives with respect to $\boldsymbol{\lambda}$ give you the original constraints back and ensure the stationary point is feasible.

The partial derivatives with respect to $\mathbf{x}$ give $\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla h(\mathbf{x}) = 0$

In the case of a single equality constraint, this means $\nabla f(\mathbf{x})$ and $\nabla h(\mathbf{x})$ are parallel (or antiparallel).

The stationary point of $\min(-x_1 - x_2)$ subject to $x_1^2 + x_2^2 = 1$ was $x_1 = x_2 = \lambda = 1/\sqrt{2}$

For this function, $\nabla f(\mathbf{x}) = \begin{bmatrix} -1 & -1 \end{bmatrix}$ and $\nabla h(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$

A penalty interpretation

The Lagrangian function is the original function, plus some multiple of the left-hand side of each constraint.

These multipliers can be thought of as "penalties" for violating the constraint. At the optimal solution, $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x})$.

If the penalty is too low or too high, the optimal solution will violate the constraint.

THE TRANSIT FREQUENCY SETTING PROBLEM

Here is a "modified" version without explicit upper and lower limits on the number of buses on each route.

$$\begin{aligned} \min_{\mathbf{n}} \quad & D(\mathbf{n}) = \sum_{r \in R} \frac{d_r T_r}{2 n_r} \\ \text{s.t.} \quad & \sum_{r \in R} n_r = N \end{aligned}$$

Assume there are 3 routes with this problem data:

- Route 1 has demand 1, and the route requires 2 hours to traverse
- Route 2 has demand 8, and the route requires 1 hour to traverse
- Route 3 has demand 6, and the route requires 3 hours to traverse

Furthermore, there are 6 buses to assign to these routes.

The Lagrangian is $\mathcal{L}(\mathbf{n}, \lambda) = \sum_{r \in R} \frac{d_r T_r}{2n_r} + \lambda(\sum_{r \in R} n_r - N)$

The stationary point of the Lagrangian occurs when:

- $\frac{\partial \mathcal{L}}{\partial n_r} = 0$ for all routes r; this means each $n_r = \sqrt{\frac{d_r T_r}{2\lambda}}$
- $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$; this means $\sum_{r \in R} n_r = N$

Substituting the first equations into the second we have

$$\sum_{r \in R} \sqrt{\frac{d_r T_r}{2\lambda}} = N$$

which we can solve for λ.

Using the given data for this problem, the equation simplifies to

$$\frac{1}{\sqrt{\lambda}} + \frac{2}{\sqrt{\lambda}} + \frac{3}{\sqrt{\lambda}} = 6$$

which is solved when $\lambda = 1$.

If the functions and values are not so nice, you can use Newton's method or the bisection method to solve for λ.

Substituting $\lambda = 1$ into each equation, we find that the optimal solution is $n_1 = 1$, $n_2 = 2$, and $n_3 = 3$.

The interpretation of the Lagrange multiplier $\lambda = 1$ is that (at the margin), adding one more bus to the fleet will reduce total waiting time by approximately one hour if allocated optimally.

The requirement $n_r = \sqrt{\frac{d_r T_r}{2\lambda}}$ can also be interpreted in saying that the marginal impact of an additional bus on each route must be equal at the optimal solution.

INEQUALITY CONSTRAINTS

The theory of Lagrange multipliers dates to the 18th century; techniques for handling inequality constraints are more recent.

This theory is generalized in the *Karush-Kuhn-Tucker* conditions, which accounts for both inequality and equality constraints.

Karush first came up with this idea in his 1939 MS thesis.

Kuhn and Tucker independently came up with the idea in 1951.

An equivalent way of phrasing the Lagrange multiplier technique is :

At an optimal solution $\mathbf{x}^*$ to the problem $\min_{\mathbf{x}} f(\mathbf{x})$ subject to $h_i(\mathbf{x}) = 0$ for $i \in \{1, \ldots, m\}$, we have

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla h(\mathbf{x}) = 0$$

for some λ_i, and furthermore $h_i(\mathbf{x}) = 0$ for all $i \in \{1, \ldots, m\}$.

The Karush-Kuhn Tucker conditions are as follows:

At an optimal solution $\mathbf{x}^*$ to the problem $\min_{\mathbf{x}} f(\mathbf{x})$ subject to $h_i(\mathbf{x}) = 0$ for $i \in \{1, \ldots, m\}$ and $g_j(\mathbf{x}) \leq 0$ for $j \in \{1, \ldots, \ell\}$ we have

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla h(\mathbf{x}) + \sum_{j=1}^{\ell} \mu_j \nabla g(\mathbf{x}) = 0$$

for some λ_i and **nonnegative** μ_j, and furthermore $h_i(\mathbf{x}) = 0$ for all $i \in \{1, \ldots, m\}$, $g_j(\mathbf{x}) \leq 0$ for $j \in \{1, \ldots, \ell\}$, **and** $\mu_j = 0$ for any inactive constraint.

Unpacking the KKT conditions:

- A multiplier μ_j is introduced for each inequality constraint, just like a λ_i is introduced for each equality.
- We distinguish between an *active* and an *inactive* inequality constraint. The constraint $g_j(\mathbf{x}) \leq 0$ is *active* if $g_j(\mathbf{x}) = 0$ and *inactive* if $g_j(\mathbf{x}) < 0$.
- The multiplier for each μ_j must be nonnegative, and zero for each inactive constraint.

The second and third points warrant further explanation.

If a constraint is inactive at the optimum solution, it is essentially irrelevant, and changing the right-hand side by a small amount will not affect the optimal solution at all.

Therefore $\mu_j = 0$ for any inactive constraint.

Increasing the right-hand side of the constrant $g_j(\mathbf{x})$ can *only improve the optimal value* of the objective function. So μ_j cannot be negative.

(Why is this not true for equality constraints?)

A simple example

Minimize $f(x) = (x+5)^2$ subject to $x \leq 0$.

The optimal solution is clearly $x = -5$. The inequality constraint is active, so $\mu = 0$.

Here $\nabla f(x) = 2(x+5)$ and $\nabla g(x) = 1$; if we set $x = -5$ and $\mu = 0$, then

$$\nabla f(x) + \mu \nabla g(x) = 0$$

so this is (potentially) the optimal solution.

A simple example

Minimize $f(x) = (x-5)^2$ subject to $x \le 0$.

The optimal solution is now $x = 0$. The inequality constraint is active, so $\mu \ge 0$.

Here $\nabla f(x) = 2(x-5)$ and $\nabla g(x) = 1$; if we set $x = 0$ then

$$\nabla f(x) + \mu \nabla g(x) = 0$$

is true when $\mu = 10$.

The interpretation: by changing the constraint from $x \le 0$ to $x \le 1$, the objective function can be reduced by approximately 10.