KKT conditions, Descent methods

CE 377K

April 23, 2015

REVIEW

What is the method of Lagrange multipliers?

What are three interpretations?

OUTLINE

- Karush-Kuhn-Tucker conditions
- Descent methods for convex optimization

INEQUALITY CONSTRAINTS

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 $x \leq 0$.

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

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 $x \leq 0$ to $x \leq 1$, the objective function can be reduced by approximately 10.

DESCENT METHODS

Assume that we are given a convex optimization problem.

A *descent* method is an iterative algorithm consisting of the following steps:

1. Choose an initial feasible solution $\mathbf{x} \leftarrow \mathbf{x}^{\mathbf{0}}$.
2. Identify a feasible "target" solution $\mathbf{x}^*$ in a "downhill direction."
3. Choose a step size $\lambda \in [0, 1]$ and set $\mathbf{x} \leftarrow \lambda \mathbf{x}^* + (1 - \lambda)\mathbf{x}$
4. Test for termination, and return to step 2 if we need to improve further.

Hiking analogy

The two key questions are:

1. How do I choose the target solution?
2. How do I choose λ

You might also ask how to choose the initial solution $\mathbf{x^0}$. For a convex optimization problem, it won't matter too much, and if we answer the previous two questions correctly we will converge to the global optimum from any starting point.

Today and next class, I'll give a few different ways these steps can be performed:

For choosing the target $\mathbf{x}^*$, I will show you the **conditional gradient** and **gradient projection** methods.

For choosing the step size λ, I will show you the **method of successive averages**, **limited minimization rule**, and **Armijo rule**.

For a convex optimization problem, any combination of these will converge to the global optimum from any starting point.

Target selection

Both of the target selection rules are based on the gradient at the current point, $\nabla f(\mathbf{x})$.

Remember that the gradient is a vector pointing in the direction of steepest *ascent*. Since our standard form is a minimization problem, we will want to move in the opposite direction as the gradient.

However, just using the gradient as a search direction creates problems, because it fails to account for the constraints.

The conditional gradient and gradient projection methods are designed to provide reasonable target points while accounting for constraints.

Conditional gradient

In the conditional gradient rule, we assume that the slope of the objective is the *same* throughout the feasible region. (This is not true, but gives us a direction to move in.)

This is equivalent to replacing the objective function with its linear approximation at the current point $\mathbf{x}$.

We then solve the optimization problem for the approximate objective function $f(\mathbf{x}^*) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x}^* - \mathbf{x})$ with the same constraints, and use the optimal $\mathbf{x}^*$ value for the target.

If all of the constraints are linear, the conditional gradient method is nothing more than solving a linear program.

Example

Transit frequency setting problem with just 2 routes, same data as before:

Minimize $1/n_1 + 4/n_2$ subject to $n_1 + n_2 = 6$, $n_1 \geq 0$, $n_2 \geq 0$.

Gradient projection

In the gradient projection method, the target is found by calculating the point $\mathbf{x} - s\nabla f(\mathbf{x})$ (where s is another step size), and then calculating the *projection* of that point onto the feasible region.

The **projection** of a point $\mathbf{x}$ onto a set X is the point in X closest to $\mathbf{x}$, and is denoted by $\Pi_X \mathbf{x}$.

Examples of projection

Project the point $(5, 10)$ onto the set defined by $0 \leq x_1 \leq 7$, $0 \leq x_2 \leq 7$.

Project the point $(4, 2)$ onto the set defined by $x_1 + x_2 = 5$

Minimize $1/n_1 + 4/n_2$ subject to $n_1 + n_2 = 6$, $n_1 \geq 0$, $n_2 \geq 0$.