Method of Lagrange multipliers

CE 377K

April 16, 2015

ANNOUNCEMENTS

- HW 3 due today
- Questions about the project?

REVIEW

The standard form for a nonlinear optimization problem is

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

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

What is a convex set?

In one dimension, what is a convex function?

In higher dimensions, what is a convex function?

Outline for today:

1. Finishing up convex functions
2. Relating convex functions and convex sets
3. Defining a convex program
4. Method of Lagrange multipliers (equality constraints only)

For twice-differentiable multidimensional functions, f is convex if any of these equivalent conditions are satisfied:

1. For all x_1 and x_2 in X,

$$f(\lambda x_2 + (1-\lambda)x_1) \leq \lambda f(x_2) + (1-\lambda) f(x_1)$$

2. For all x_1 and x_2 in X,

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1)$$

3. For all x in X, $H(x)$ is positive semidefinite (that is, $y^T H(x) y \geq 0$ for all vectors y).

These conditions can be tedious to check. In this class I will not ask you to apply these definitions directly to multidimensional functions.

However, there are some facts which we can use (even in higher dimensions):

- Any linear function is convex.
- A nonnegative multiple of a convex function is convex.
- The sum of convex functions is convex.
- The composition of convex functions is convex.

Other useful facts:

- For any convex function $g(\mathbf{x})$, the set $\{\mathbf{x} : g(\mathbf{x}) \leq 0\}$ is a convex set.
- For any linear (affine) function $h(\mathbf{x})$, the set $\{\mathbf{x} : h(\mathbf{x}) = 0\}$ is a convex set.
- The intersection of convex sets is convex.

A **convex optimization** problem is any optimization problem where the objective function is a convex function, and the feasible region is a convex set.

From the preceding discussion, a standard-form nonlinear program is a convex program if f is convex, all g are convex, and all h are affine. (These conditions are *sufficient* but not *necessary*.)

Example

Is the transit frequency-setting problem a convex program?

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

1. Is the objective function convex in $\mathbf{n}$?
2. Do the constraints represent a convex set?

METHOD OF LAGRANGE MULTIPLIERS

For a first solution method for nonlinear optimization, assume that we have a standard-form nonlinear program *with only equality constraints*. (We'll tackle inequality constraints next week.)

Assume that f and all h are continuously differentiable functions. This method does *not* require f to be convex or h to be linear, but it is simpler in that case.

The key idea of Lagrange multipliers is that *constraints are what make optimization problems difficult.*

At the start of the class, we found optimal solutions by identifying stationary points of the objective function f.

This worked when there were no constraints... so if we can somehow remove the constraints we can use this technique again.

The idea behind Lagrange multipliers is to move each constraint into the objective function, and then solving it like an unconstrained problem.

$$
\begin{array}{ll}
\min & -x_1 - x_2 \\
\text{s.t.} & x_1^2 + x_2^2 - 1 \quad = 0
\end{array}
$$

What is the solution to this problem?

Multiply the constraint by λ and add it to the objective function to form the **Lagrangian function**:

$$\mathcal{L}(x_1, x_2, \lambda) = -x_1 - x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

The stationary points of this function are the points where all the partial derivatives simultaneously vanish:

- $\frac{\partial \mathcal{L}}{\partial x_1} = -1 + 2\lambda x_1 = 0$
- $\frac{\partial \mathcal{L}}{\partial x_2} = -1 + 2\lambda x_2 = 0$
- $\frac{\partial \mathcal{L}}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0$

Notice that the third equation gives us the original constraint back!

These equations are solved when $x_1 = x_2 = \lambda = 1/\sqrt{2}$.

So, the optimal solution of the original problem is $x_1 = x_2 = 1/\sqrt{2}$.

(We'll leave the interpretation of the Lagrange multiplier for next week.)

If there are multiple equality constraints, introduce a different Lagrange multiplier for each of them.

Example

The feasible region is defined by the intersection of the surfaces $z^2 = x^2 + y^2$ and $x - 2z = 3$. What point in this region is closest to the origin?

$$\begin{array}{lll} \min & x^2 + y^2 + z^2 & \\ \text{s.t.} & x^2 + y^2 - z^2 & = 0 \\ & x - 2z - 3 & = 0 \end{array}$$

The Lagrangian is
$\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = x^2 + y^2 + z^2 + \lambda_1(x^2 + y^2 - z^2) + \lambda_2(x - 2z - 3)$

The partial derivatives are:

- $\frac{\partial \mathcal{L}}{\partial x} = 2(1+\lambda_1)x + \lambda_2$
- $\frac{\partial \mathcal{L}}{\partial y} = 2(1+\lambda_1)y$
- $\frac{\partial \mathcal{L}}{\partial z} = 2(1-\lambda_1)z - 2\lambda_2$
- $\frac{\partial \mathcal{L}}{\partial \lambda_1} = x^2 + y^2 + z^2$
- $\frac{\partial \mathcal{L}}{\partial \lambda_2} = x - 2z - 3$

From the second equation, either $y = 0$ or $\lambda_1 = -1$.

If $\lambda_1 = -1$, the other equations give $\lambda_2 = 0$, $z = 0$, and $x = 3$.

Equation 4 would then require $y^2 + 9 = 0$ which is impossible, so $\lambda = -1$ is not a valid solution.

So, $y = 0$, and the equations reduce to

- $2(1 + \lambda_1)x + \lambda_2 = 0$
- $2(1 - \lambda_1)z - 2\lambda_2 = 0$
- $x^2 - z^2 = 0$
- $x - 2z - 3 = 0$

So, either $z = x$ or $z = -x$.

If $z = x$ we have $x = -3$ and $(x, y, z) = (3, 0, -3)$.

If $z = -x$ we have $x = 1$ and $(x, y, z) = (1, 0, -1)$.

Substitute them into the objective function $x^2 + y^2 + z^2$.

$(3, 0, -3)$ is the *maximum*, and $(1, 0, -1)$ is the *minimum*. The second is the one we want.

This technique should remind you of the stationary point method we used at the start of class.