

Nonlinear Optimization Lecture 8

Garrick Aden-Buie

Thursday, February 4, 2016

Homework Review

Problem 2.4

Show that the function θ defined by the following optimization problem is concave.

$$\begin{aligned} \theta(u_1, u_2) \quad & \min_{x_1, x_2} x_1(2 - u_1) + x_2(3 - u_2) \\ & \text{s.t. } x_1^2 + x_2^2 \leq 4 \end{aligned}$$

Method 1

Use the four cases where $(2 - u_1) \geq / \leq 0$ and $(3 - u_2) \geq / \leq 0$, and let $x_1(u_1, u_2)$ and $x_2(u_1, u_2)$. Then express θ in terms of u_1, u_2 and show that this function is concave.

Method 2

$$\text{Let } \theta(u) = \min \begin{bmatrix} 2 \\ 3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \text{ s.t. } x \in X.$$

Then $\theta(u) = \min_{x \in X} (a - u)^T x$ where

$$a = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Then consider a linear combination of two points, \bar{u} and \hat{u} .

$$\begin{aligned} \theta(\lambda \bar{u} + (1 - \lambda) \hat{u}) &= \min (a - \lambda \bar{u} - (1 - \lambda) \hat{u})^T x \\ &= \min [\lambda (a - \bar{u})^T x + (1 - \lambda) (a - \hat{u})^T x] \\ &\geq \min [\lambda (a - \bar{u})^T x] + \min [(1 - \lambda) (a - \hat{u})^T x] \\ &= \lambda \theta(\bar{u}) + (1 - \lambda) \theta(\hat{u}) \end{aligned}$$

So $\theta(u)$ is concave.

Method 3

Consider all $x \in \{x^1, x^2, \dots\} =$ all points in $x_1^2 + x_2^2 \leq 4$.

Then $\theta(u_1, u_2) = \min\{f(u; x^1), f(u; x^2), \dots\}$. Note that each $f(u; x^i)$ is linear when x is fixed, so each is a concave (or convex) function. From this perspective, then, $\theta(u) = \min\{\text{concave functions}\}$, so $\theta(u)$ is concave (which draws on a result from a previous homework problem).

Convex / Quasiconvex

Def. $S \subset \mathbb{R}, S \neq \emptyset$ open, $f: S \rightarrow \mathbb{R}, f \in C^1(S)$.

(1) f is *pseudoconvex* on S if

$$[\nabla f(\bar{x})]^T(\hat{x} - \bar{x}) \geq 0 \Rightarrow f(\hat{x}) \geq f(\bar{x}) \quad \forall \hat{x}, \bar{x} \in S$$

(2) f is *strictly pseudoconvex* on S if

$$[\nabla f(\bar{x})]^T(\hat{x} - \bar{x}) > 0 \Rightarrow f(\hat{x}) \geq f(\bar{x}) \quad \forall \hat{x}, \bar{x} \in S, \hat{x} \neq \bar{x}$$

Theorem. f is strictly convex $\Rightarrow f$ strongly quasiconvex.

Optimization

We have been working with convex functions and sets and their properties, and we are now ready to start working with optimization problems.

Unconstrained Problems

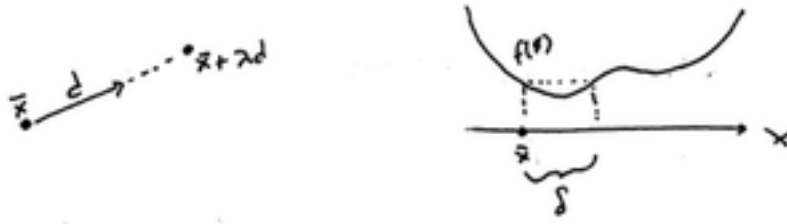
$$\begin{array}{ll} \min & f(x_0) \\ \text{s.t} & x \in \mathbb{R}^n \end{array}$$

Necessary Optimality Conditions

Theorem. $f: \mathbb{R}^n \rightarrow \mathbb{R}$, f is differentiable at \bar{x} . If there is a vector d such that $[\nabla f(\bar{x})]^T d < 0$ then there exists $\delta > 0$ such that

$$f(\bar{x} + \lambda d) < f(\bar{x}) \quad \forall \lambda \in (0, \delta)$$

The idea here is to pick a point \bar{x} and a direction d and then move by step size λ in the direction d , as long as λ is small and no bigger than δ . If you move in this direction, you can decrease the function value, and if the step size is small enough you can decrease the function value. But how do you find the direction d ? By looking at the gradient $\nabla f(\bar{x})$.

Figure 1: Demonstration of step and choice of maximum step size δ .

Remark. d is called a *descent direction* of f at \bar{x} .

Proof.

$$\begin{aligned}
 f(\bar{x} + \lambda d) &= f(\bar{x}) + \lambda [\nabla f(\bar{x})]^T d + \lambda \|d\| \alpha(\bar{x}; \lambda d) \\
 \frac{1}{\lambda} (f(\bar{x} + \lambda d) - f(\bar{x})) &= [\nabla f(\bar{x})]^T d + \|d\| \alpha(\bar{x}; \lambda d) \\
 &= < 0 + \text{arbitrarily small number} \\
 &< 0 \text{ for sufficiently small } \lambda
 \end{aligned}$$

Let's consider $d = -\nabla f(\bar{x})$, then

$$\begin{aligned}
 [\nabla f(\bar{x})]^T d &= -[\nabla f(\bar{x})]^T [\nabla f(\bar{x})] \\
 &= -\|\nabla f(\bar{x})\|^2 \leq 0
 \end{aligned}$$

We want to make sure that we can decrease the function value. In most cases, the value of $-\|\nabla f(x)\|^2$ will be < 0 , then proceed with $d = -\nabla f(\bar{x})$. If $-\|\nabla f(x)\|^2 = 0$, then $\nabla f(\bar{x}) = 0$, so you can stop (global minimum or local minimum, depending on the shape of your function).

The rest of the class will be essentially fine-tuning this statement and building on it.