Static Optimization

Mark Dean

Lecture Notes for Fall 2014 PhD Class - Brown University

1 Lecture 1

We now begin our whirlwind tour of static optimization. The problem that we are interested in is finding the maximum (or minimum) value taken by a real valued function defined on some subset X of $\mathbb{R}^n$. Initially, we are going to to think about the unconstrained problem, whereby we are interested in finding the maximum and minimum on all of X. We are then going to move onto **constrained problems,** in which we are not free to choose any element in the set X, but only elements that satisfy some constraints. We will begin by thinking about *equality* constraints, and then (if we have time) inequality constraints.

1.1 Unconstrained Optimization

We will now set up the general problem of unconstrained optimization. Again, we are going to focus on the case of finding a maximum - the case of finding a minimum is analogous.

Problem 1 *Let $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}$. We wish to find*

$$\max_{x \in X} f(x)$$
$$\text{and } \arg\max_{x \in X} f(x)$$

We will say that x is feasible if $x \in X$, and x^* is an optimal solution if $f(x) \leq f(x^*)\ \forall\ x \in X$

The first thing we are going to do is strengthen the result we had earlier relating the idea of a local maximizer to the derivative

Theorem 2 *Let $X \subset \mathbb{R}$ and $f : X \to \mathbb{R}$ be differentiable*

1. *If $x^* \in int(X)$ is a local maximizer then $f'(x) = 0$. Moreover, if $f''(x)$ exists, then $f''(x) \leq 0$*
2. *If $f'(x^*) = 0$ and $f''(x^*) < 0$ then x^* is a strict local maximizer*

Proof. *We have already shown that if x^* is a local maximizer, then $f'(x) = 0$. Using the second order Taylor approximation, we know that*

$$\begin{aligned} f(x) &= f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2 + o(|x - x^*|^2) \\ &= f(x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2 + o(|x - x^*|^2) \leq f(x^*) \end{aligned}$$

for x close enough to x^*. This implies that

$$\begin{aligned} \frac{1}{2}f''(x^*)(x - x^*)^2 + o(|x - x^*|^2) &\leq 0 \\ &\Rightarrow \frac{1}{2}f''(x^*) \leq -\frac{o(|x - x^*|^2)}{(x - x^*)^2} \end{aligned}$$

But as $\lim_{x \to x^*} \frac{o(|x - x^*|^2)}{(x - x^*)^2} = 0$, this implies that $f''(x^*) \leq 0$

Now assume that $f'(x^*) = 0$ and $f''(x^*) < 0$. As

$$\begin{aligned} f(x) &= f(x^*) + \frac{1}{2}f''(x^*)(x - x^*)^2 + o(|x - x^*|^2) \\ &\Rightarrow f(x) - f(x^*) = \frac{1}{2}f''(x^*)(x - x^*)^2 + o(|x - x^*|^2) \\ &= \left(\frac{1}{2}f''(x^*) + \frac{o(|x - x^*|^2)}{(x - x^*)^2}\right)(x - x^*)^2 \end{aligned}$$

As $\lim_{x \to x^*} \frac{o(|x - x^*|^2)}{(x - x^*)^2} = 0$, we can find some ε such that, for $|x - x^*| < \varepsilon$

$$\left|\frac{o(|x - x^*|^2)}{(x - x^*)^2}\right| \leq \left|\frac{1}{4}f''(x^*)\right|$$

and so

$$f(x) - f(x^*) \leq \frac{1}{4}f''(x^*)(x - x^*)^2 < 0$$

■

We are now going to extend this result in two ways. First, we are going to allow for functions on some arbitrary subset of $\mathbb{R}^n$. Second, we are going to generate a result that allows us to say something about any point in the feasible set. In order to do so we first need to define the concept of a **feasible direction**

Definition 1 *Let $X \subset \mathbb{R}^n$. A vector $d \in \mathbb{R}^n$ is a feasible direction from $x \in X$ if $\exists\ \bar{\alpha} > 0$ such that $x + \alpha d \in X\ \forall\ \alpha \in [0, \bar{\alpha}]$*

Theorem 3 (First Order Necessary Conditions) *Assume f is $\mathcal{C}^2$ on $X \subset \mathbb{R}^n$. if $x^* \in X$ is a local max of f, then*

$$\langle \nabla f(x^*), d \rangle \leq 0$$

for all feasible directions d from x^*

Proof. *Let $d \in \mathbb{R}^n$ be a feasible direction from x^*, and define $g(\alpha) = f(x^* + \alpha.d)$ on $[0, \bar{\alpha}]$. Note that*

$$\begin{aligned} g'(\alpha) &= d_1 \frac{\partial f(x^* + \alpha.d)}{\partial x_1} + .. + d_n \frac{\partial f(x^* + \alpha.d)}{\partial x_n} \\ &= \langle \nabla f(x^* + \alpha.d), d \rangle \end{aligned}$$

Then $g(\alpha) \leq g(0)$ for α sufficiently small. From 1st order Taylor approximation

$$\begin{aligned} g(\alpha) &= g(0) + g'(0)\alpha + o(\alpha) \\ &= g(0) + \langle \nabla f(x^*), d \rangle \alpha + o(\alpha) \leq g(0) \\ &\Rightarrow \langle \nabla f(x^*), d \rangle \leq \frac{o(\alpha)}{\alpha} \end{aligned}$$

and so, by the usual argument $\langle \nabla f(x^*), d \rangle \leq 0$ ■

We can also provide necessary second order conditions

Theorem 4 (Second Order Necessary Conditions) *Assume f is $\mathcal{C}^3$ of X. If x^* is a local maximum, then, for any feasible direction $d \in \mathbb{R}^n$*

1. $\langle \nabla f(x^*), d \rangle \leq 0$

2. $\langle \nabla f(x^*), d\rangle = 0 \Rightarrow d^T H(x^*) d \leq 0$, *where* $H(x^*) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x^*)\right]$

Proof. *(1) we have already proved.*

For (2), using the same function $g(.)$ we can use Taylor's second order approximation to give

$$g(\alpha) = g(0) + g'(0)\alpha + \frac{1}{2}g''(0)\alpha^2 + o(\alpha^2) \leq g(0)$$

If $\langle \nabla f(x^*), d\rangle = 0$ then $g'(0)\alpha = 0$, giving

$$\frac{1}{2}g''(0)\alpha^2 + o(\alpha^2) \leq 0$$

and so, by the usual argument $g''(0) \leq 0$. But $g''(0) = d^T H(x^*) d$ so we are done. ■

Two obvious extensions that we will state and not prove are:

Corollary 1 *If $x^* \in int(X)$ and is a local max, then*

1. $\nabla f(x^*) = 0$
2. $H(x^*)$ *is negative semi - definite*

Just as we have extended the necessary conditions from the $\mathbb{R}$ case to the $\mathbb{R}^n$ case, so we can extend the sufficient conditions. We won't show this, but the technique is to essentially apply the trick involving $g(\alpha)$ above to the proof of the sufficiency condition for the $\mathbb{R}$ case.

Corollary 2 *If $x^* \in int(X)$ such that*

1. $\nabla f(x^*) = 0$
2. $H(x^*)$ *is negative definite*

Then x^* is a strict local minima.