ORIE 6300 Mathematical Programming I November 9, 2016

Recitation 12

Lecturer: Calvin Wylie *Topic:*

Twice Continuously Differentiable Functions [1]

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at x if $\exists\ v \in \mathbb{R}^n$ and $o : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(y) = f(x) + \langle v, y - x\rangle + o(y - x)\ \forall y \in \mathbb{R}^n,$$

and

$$\lim_{y\to x} \frac{y - x}{\|y - x\|} = 0.$$

Call v the gradient of f and write $\nabla f(x) = v$. If ∇f is continuous say f is C^1 and write $f \in C^1$.

Definition 1 *f is twice differentiable at x if ∇f is continuous at x and there exist $o : \mathbb{R}^n \to \mathbb{R}$ and a linear operator $\nabla^2 f(x)$ such that*

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle\ y - x, \nabla f(x)(y - x)\rangle + o(y - x)\ \forall y \in \mathbb{R}^n,$$

and

$$\lim_{y\to x} \frac{y - x}{\|y - x\|^2} = 0.$$

Write $f \in C^2$.

Theorem 1 *(Taylor's theorem) For any $x, y \in \mathbb{R}^n$,*

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + t(y - x))(y - x)dt.$$

Theorem 2 *$f \in C^2$ is L-Lipschitz differentiable iff $\|\nabla^2 f(x)\| \leq L\ \forall x$.*

Proof: Suppose $\|\nabla^2 f(x)\| \leq L\ \forall x$. Then, by Taylor's theorem,

$$\begin{aligned}
\|\nabla f(x) - \nabla f(y)\| &= \|\int_0^1 \nabla^2 f(x + t(y - x))(y - x)dt\| \\
&\leq \int_0^1 \|\nabla^2 f(x + t(y - x))(y - x)\|dt \\
&\leq \int_0^1 \|\nabla^2 f(x + t(y - x))\|\|y - x\|dt \\
&\leq \int_0^1 L\|y - x)\|dt \\
&= \|y - x\|.
\end{aligned}$$

[1] Based on Nesterov, Yurii. *Introductory lectures on convex optimization: A basic course.*

Now let f be L-Lipschitz differentiable, $s \in \mathbb{R}^n$ and $\alpha > 0$. We have

$$\begin{aligned}
\alpha L\|s\| &\geq \|\nabla f(x+\alpha s) - \nabla f(x)\| \\
&= \|\int_0^1 \nabla^2 f(x+\alpha ts)\alpha s dt\| \\
&= \|\int_0^\alpha \nabla^2 f(x+ws)s dw\|,
\end{aligned}$$

where the last equality follows by making the change of variables $w = \alpha t$. Thus,

$$\frac{1}{\alpha}\|\int_0^\alpha \nabla^2 f(x+ws)s dw\| \leq L\|s\|$$

Taking the limit when $\alpha \to 0$ we obtain

$$\|\nabla^2 f(x)s\| \leq L.$$

Since this is true for every $s \in \mathbb{R}^n$, it follows that $\|\nabla^2 f(x)\| \leq L$. □

Example 1 *Let* $f(x) = \alpha + \langle a, x\rangle$. *Then*

$$f(y) = f(x) + \langle a, y - x\rangle .$$

Thus, $\nabla f(x) \equiv a$ *and* $\nabla^2 f(x) \equiv 0$.

Example 2 *Let* $f(x) = \alpha + \langle a, x\rangle + \frac{1}{2}\langle x, Ax\rangle$, *where* A *is symmetric. We have*

$$\begin{aligned}
f(y) &= \alpha + \langle a, y\rangle + \frac{1}{2}\langle y, Ay\rangle \\
&= f(x) + \langle a + Ax, y - x\rangle + \frac{1}{2}\langle y - x, A(y-x)\rangle .
\end{aligned}$$

Thus, $\nabla f(x) = a + Ax$ *and* $\nabla^2 f(x) \equiv A$.

Theorem 3 *Suppose* x^* *is a local minimizer of* $f \in C^2$. *Then* $\nabla f(x^*) = 0$ *and* $\nabla^2 f(x^*) \succeq 0$.

Proof: Last recitation we proved that $\nabla f(x^*) = 0$, so it remains to prove that $\nabla^2 f(x^*) \succeq 0$. Since x^* is a local minimizer, there exists $r > 0$ such that $f(y) \geq f(x^*)$ $\forall y \in B_r(x^*)$. Moreover, since $\nabla f(x^*) = 0$,

$$f(y) = f(x^*) + \frac{1}{2}\left\langle\, y - x^*, \nabla^2 f(x^*)(y - x^*)\right\rangle + o(y - x^*).$$

Thus,

$$\frac{1}{2}\langle\, y - x^*, \nabla f(x^*)(y - x^*)\rangle + o(y - x^*) \geq 0.$$

Let u be a unit vector and let $y_\epsilon = x^* + \epsilon u$. For ϵ small enough, $y_\epsilon \in B_r(x^*)$. Thus,

$$\frac{1}{2}\left\langle\, y_\epsilon - x^*, \nabla^2 f(x^*)(y_\epsilon - x^*)\right\rangle + o(y_\epsilon - x^*) \geq 0.$$

Divide by $\|y_\epsilon - x^*\|^2$ and let $\epsilon \to 0$ to obtain

$$\left\langle s, \nabla^2 f(x^*)s\right\rangle \geq 0.$$

This finishes the proof. □