Lecture 04 19/09/2019

Proof : $\|Ax\|_2 \le \|A\|_2 \cdot \|x\|_2$

$$\frac{\|Ax\|_2}{\|x\|_2} \le \max_{y \neq 0} \frac{\|Ay\|_2}{\|y\|_2} = \|A\|_2$$

$$\Leftrightarrow$$

$$\|Ax\|_2 \le \|A\|_2 \cdot \|x\|_2$$

proof: Fundamental Theorem of Calculus for multivariate functions

We want to prove that

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \int_0^1 \left(\nabla f(x + t(y-x)) - \nabla f(x)\right)^T (y-x) dt$$

Consider the univariate functional

$$\phi(t) = f(x + t(y-x))$$

The fundamental theorem of calculus asserts

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t) \, dt$$

since $\phi(1) = f(y)$, $\phi(0) = f(x)$ & $\phi'(t) = \nabla f(x + t(y-x))^T (y-x)$

we get

$$f(y) - f(x) = \int_0^1 \nabla f(x + t(y-x))^T (y-x) \, \partial t$$

Lets add and subtract $\nabla f(x)^T (y-x)$

$$f(y) - f(x) = \nabla f(x)^T (y-x) - \nabla f(x)^T (y-x) + \int_0^1 \nabla f(x+t(y-x))^T (y-x) \, \partial t$$

$$= \nabla f(x)^T (y-x) - 1 \cdot \nabla f(x)^T (y-x) + 0 \cdot \nabla f(x)^T (y-x)$$

$$+ \int_0^1 \nabla f(x + t(y-x))^T (y-x) \, \partial t$$

$$= \nabla f(x)^T (y-x) - \left[t \, \nabla f(x)^T (y-x) \right]_0^1$$

$$+ \int_0^1 \nabla f(x + t(y-x))^T (y-x) \, \partial t$$

$$= \nabla f(x)^T (y-x) - \int_0^1 \nabla f(x)^T (y-x) \, \partial t$$

$$+ \int_0^1 \nabla f(x + t(y-x))^T (y-x) \, \partial t$$

$$= \nabla f(x)^T (y-x) + \int_0^1 \overset{\cancel{= \nabla f(x)}}{\nabla f(x + t(y-x))^T (y-x)} - \nabla f(x)^T (y-x) \, \partial t$$

$$= \nabla f(x)^T (y-x) + \int_0^1 (\nabla f(x + t(y-x)) - \nabla f(x))^T (y-x) \, \partial t$$

Definition ; Lipschitz continuity

We say the gradient of f is Lipschitz continuous

if

$$\|\nabla f(y) - \nabla f(x)\|_2 \le L \|y-x\|_2 \quad \forall x,y$$

where L is a constant.

proof: $f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$

We will assume that $\nabla f(x)$ is Lipschitz continuous

$$\|\nabla f(y) - \nabla f(x)\|_2 = L\ \|y-x\|_2$$

Using FToC we get

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \int_0^1 (\nabla f(x+t(y-x)) - \nabla f(x))^T (y-x)\, \partial t$$

Use Cauchy - Schwartz.

$$|x^T y| \le \|x\|_2 \cdot \|y\|_2$$

to get

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\|_2 \cdot \|y-x\|_2\, \partial t$$

use Lipschitz continuity of $\nabla f(x)$

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \int_0^1 L\|x + t(y-x) - x\|_2\, \|y-x\|_2\, \partial t$$

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \int_0^1 Lt \, \|y-x\|^2 \, \partial t$$

$$= f(x) + \nabla f(x)^T (y-x) + L\|y-x\|^2 \int_0^1 t \, \partial t$$

$$= f(x) + \nabla f(x)^T (y-x) + L\|y-x\|^2 \left[\frac{t^2}{2} \right]_0^1$$

$$= f(x) + \nabla f(x)^T (y-x) + \frac{L}{2} \|y-x\|^2$$

Making Gradient Descent slide 14

$$\underset{h}{\text{minimize}} \quad \nabla f(x_0)^T h + \frac{L}{2} \|h\|_2^2$$

~~optimali~~ Convex : $\nabla^2 \left(\nabla f(x_0)^T h + \frac{L}{2} \|h\|_2^2 \right) = L > 0$

First-order optimality $\quad \nabla f(x_0) + L \cdot h = 0$

$$\Leftrightarrow$$

$$h = -\frac{1}{L} \nabla f(x_0)$$

Making Gradient Descent slide 15

Using FToC we get

$$f\left(x_0 - \frac{1}{L} \nabla f(x_0)\right) - f(x_0) \le \nabla f(x_0)^T \left(-\frac{1}{L} \nabla f(x_0) \right) + \frac{L}{2} \left\| -\frac{1}{L} \nabla f(x_0) \right\|_2^2$$

$$= -\frac{1}{L} \|\nabla f(x_0)\|_2^2 + \frac{1}{2L} \|\nabla f(x_0)\|_2^2$$

$$= -\frac{1}{2L} \|\nabla f(x_0)\|_2^2 < 0$$

Line-search complexity : slide 26

We start with a_0 and we half at each iteration. ~~For simplici~~

$k=0$	$k=1$	$k=2$	$k=3$
a_0	$\frac{1}{2} a_0$	$\left(\frac{1}{2}\right)^2 a_0$	$\left(\frac{1}{2}\right)^3 a_0$

~~After~~ At iteration k ~~the~~ we get

$$a_k = \left(\frac{1}{2}\right)^k a_0$$

We proved that for $a \leq \frac{1}{L}$

then $f(x_k - a\nabla f(x_k)) < f(x_k)$

which satisfies the termination condition of line-search. Thus in worst-case line-search will terminate when

$$a_k \leq \frac{1}{L} \Leftrightarrow \left(\frac{1}{2}\right)^k a_0 \leq \frac{1}{L}$$

let's assume for simplicity that

$a_0 = 1$. Then

$$\left(\frac{1}{2}\right)^k \leq \frac{1}{L}$$

$$\Rightarrow$$

$$\log_{1/2}\left(\frac{1}{2}\right)^k \geq \log_{1/2}\frac{1}{L} \quad . \quad \log_{1/2} x \text{ is } \downarrow$$

$$\Leftrightarrow$$

$$k \log_{1/2} 1/2 \geq \log_{1/2}\frac{1}{L}$$

$$\Leftrightarrow$$

$$\begin{aligned} k &\geq \log_{1/2}\frac{1}{L} \\ &= -\left(\frac{1}{\log 2}\right)\cdot \log \frac{1}{L} \\ &= -\frac{1}{\log 2}\left(\log 1 - \log L\right) \\ &= \frac{\log L}{\log 2} \end{aligned}$$

log : natural logarithm

This means that in worst case after 7

$$\cancel{}\ \frac{\log L}{\log 2} \text{ iterations}$$

line-search will terminate.