GD II

Monday, January 13, 2020 5:34 PM

Yulan

- general f, ∇f is L-Lipschitz $\|\nabla^2 f\|_{op} \le L$

GD gives x: $\|\nabla f(x)\| \le \varepsilon$

in at most $(f(x^0) - f(x^*)) \cdot \frac{2L}{\varepsilon^2}$ steps.

- convex f, x: $f(x) \le f(x^*) + \varepsilon$

in at most $\frac{2LR^2}{\varepsilon}$ steps.

$\left(R = \max\limits_{x:\ f(x) \le f(x^0)} \|x^0 - x^*\|_2\right)$

<u>Today</u>: f is <u>strongly</u> convex.

$\|\nabla^2 f\|_{op} \ge \mu$

<u>Thm</u>. GD gives x s.t. $f(x) \le f^* + \varepsilon$

in $O\left(\frac{L}{\mu} \log \frac{f(x^0) - f^*}{\varepsilon}\right)$ steps.

$\frac{1}{\varepsilon} \longrightarrow \log \frac{1}{\varepsilon}$!!

$$\frac{1}{\varepsilon} \longrightarrow \log \frac{1}{\varepsilon} \quad !!$$

<u>Pf.</u> – $f(x^{k+1}) = f\left(x^k - \frac{1}{L} \nabla f(x^k)\right)$

$$\leq f(x^k) - \frac{1}{2L} \|\nabla f(x^k)\|^2 .$$

– $f(x^*) = f(x^k) + \nabla f(x^k)(x^* - x^k)$

$$+ \frac{1}{2}(x^* - x^k)^T \nabla^2 f(z)(x^* - x^k)$$

$$\geq f(x^k) + \underbrace{\nabla f(x^k) . \Delta + \frac{1}{2} \mu . \Delta^2}_{\substack{\text{minimized by} \\ \Delta = -\frac{\nabla f(x^k)}{\mu}}}$$

$$\geq f(x^k) - \frac{\|\nabla f(x^k)\|^2}{2\mu}$$

$$2\mu\left(f(x^k) - f(x^*)\right) \leq \|\nabla f(x^k)\|^2$$

$$\therefore f(x^{k+1}) - f^* \leq f(x^k) - f^* - \frac{1}{2L} . 2\mu\left(f(x^k) - f^*\right)$$

$$\leq \left(1 - \frac{\mu}{L}\right)\left(f(x^k) - f^*\right)$$

$$\le \left(1-\frac{\mu}{L}\right)\left(f(x^k)-f^*\right)$$

$$\Rightarrow \quad f(x^k)-f^* \le \left(1-\frac{\mu}{L}\right)^k \cdot \left(f(x^0)-f^*\right).$$

GD comes from a simple, continuous algorithm:

$$dx_t = -\nabla f(x_t)\,dt$$

$$df(x_t) = \nabla f(x_t)\cdot dx_t = -\|\nabla f(x_t)\|^2\,dt$$

$$\le -2\mu\left(f(x_t)-f^*\right)dt$$

$$\Rightarrow \quad \boxed{f(x_t) \le e^{-2\mu t}\left(f(x_t)-f^*\right)}$$

General technique:

- find a simple, continuous process (algorithm) that converges to desired solution.
- Show it is fast in continuous time.
- [illegible] time: try to maintain efficiency.

- slow [illegible]
- Discretize time; try to maintain efficiency.

Some calculus

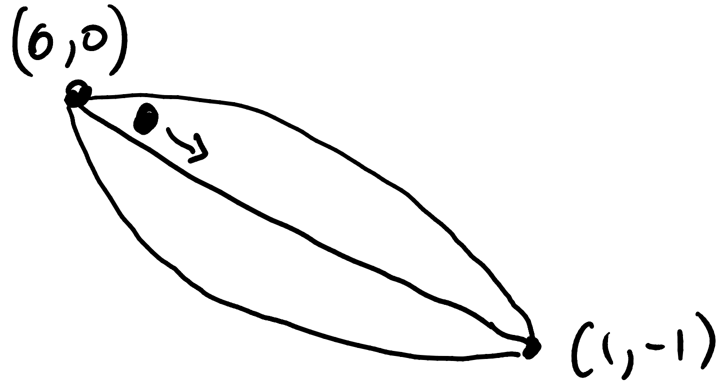
$$f : X \to \mathbb{R}^n$$

$$Df(x)[h] = \frac{d}{dt}\Big|_{t=0} f(x + t h) \quad \text{"directional derivative"}$$

Often easier, more compact and less mistake-prone.

Example from physics

Brachistochrone problem. What is the curve along which a particle travels a given distance <u>fastest</u> assuming constant gravity (and no friction)?



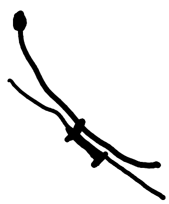
[illegible] in $(x, u(x))$

Thm. From $(0,0)$ to $(1,-1)$, opt curve is $(x, u(x))$

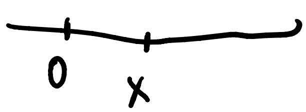
$$1 + u'(x)^2 + 2u(x)u''(x) = 0.$$

Not $u''(x) = 0$ (line) !!

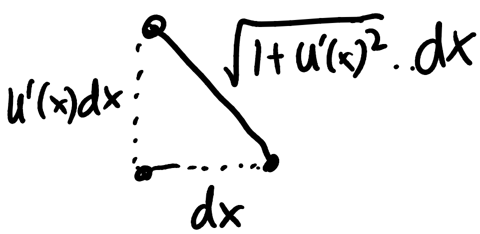
Pf. $T(u) =$ length of u



$$= \int_0^1 \frac{ds(x)}{v(x)} \qquad v(x): \text{velocity.} \quad v(0) = 0.$$



$$= \int_0^1 \frac{\sqrt{1+u'(x)^2}}{v(x)}\, dx$$



By conservation of energy,

$$\frac{1}{2} m v(x)^2 = -mg\,u(x)$$

$$v(x) = \sqrt{-2g\,u(x)} \qquad (u(x) \le 0)$$

So $$T(u) = \int_0^1 \frac{\sqrt{1+u'(x)^2}}{\sqrt{-2g\,u(x)}}\, dx.$$

Next since u is a shortest curve, $\forall$ variations h, Next since u is a shortest curve, $\forall$ variations h,

$$DT(u)[h] = 0$$

$$\frac{d}{dt}\Big|_{t=0} T(u+th)$$

$$DT(u)[h] = \int_0^1 -\frac{1}{2}\frac{\sqrt{1+u'^2}}{\sqrt{-2g}\cdot u^{3/2}}\frac{d}{dt}(u+th)\,dx + \int_0^1 \frac{1}{2}\cdot\frac{2u'}{\sqrt{-2gu}\cdot\sqrt{1+u'^2}}\cdot\frac{d}{dt}\left((u+th)'\right)dx$$

$$= \int_0^1 -\frac{1}{2}\frac{\sqrt{1+u'^2}}{\sqrt{-2gu}\cdot u}\cdot h(x)\,dx + \underbrace{\int_0^1 \frac{u'\cdot h'(x)}{\sqrt{-2gu}\,\sqrt{1+u'^2}}\,dx}$$

want to replace h' with h. So integrate by parts:

$$\left[\frac{u'\ h(x)}{\sqrt{-2gu}\,\sqrt{1+u'^2}}\right]_0^1 - \int_0^1 \frac{h(x)\cdot u''}{\sqrt{-2gu}\cdot\sqrt{1+u'^2}}\,dx$$

$$-\int \frac{h(x)\cdot u'\cdot u'}{-2\,\sqrt{-2gu}\,u\,\sqrt{1+u'^2}}\,dx - \int \frac{h(x)\cdot u'\cdot 2u'\cdot u''}{-2\cdot\sqrt{-2gu}\,(1+u'^2)^{3/2}}\,dx$$

$$\int -2\sqrt{-2gu}\;u\;\sqrt{1+u'^2} \qquad \int -2.\sqrt{-2gu}\,(1+u'^2) -$$

$h(0) = h(1) = 0$.

So

$$0 = DT(u)[h] = \int_0^1 \Big(\frac{-1}{2}(1+u'^2)^2 - u''.u\,(1+u'^2) + \frac{1}{2}u'^2(1+u'^2) + u'^2.u''.u \Big) h(x)\,dx = 0$$

$\forall h$, ~~so~~ integrand $= 0$.

$$0 = -1 - 2u'^2 - \cancel{u'^4} - 2uu'' \cancel{-2uu''u'^2} + u'^2 + \cancel{u'^4} \cancel{+2uu''u'^2}$$

or $1 + u'^2 + 2uu'' = 0$