Gradient Descent

Wednesday, January 8, 2025 6:38 AM

Goal: $\min_{x \in \mathbb{R}^n} f(x)$ f is differentiable

GD: family of algorithms $\longrightarrow$ find x s.t. $\nabla f(x) \approx 0$.

Continuous time: Start at X_0

$$\frac{d X_t}{dt} = -\nabla f(X_t)$$ move opposite to gradient.

When $\frac{dX_t}{dt} = 0$, $\nabla f(X_t) = 0$.

<u>Algo</u> $X^{(0)}$, step-size h. While $\|\nabla f(X^{(k)})\| > \varepsilon$:

$$\left[\; X^{(k+1)} = X^{(k)} - h \nabla f(X^{(k)}) \right.$$

When it stops, $\|\nabla f(X^{(k)})\| \le \varepsilon$. So, $X^{(k)}$ has nearly optimal f in some small region around it. But $\nabla f(X^{(k)})$ could change rapidly, making the region very small.

So we assume ∇f is Lipschitz.

[illegible] g is L-Lipschitz if $|g(y) - g(x)| \le L \|y - x\|_2$

So we ...

Def. $\forall x, y$ g is L-Lipschitz if $|g(y) - g(x)| \leq L \|y - x\|_2$

Lemma. The following are equivalent:

(1) ∇f is L-Lipschitz

(2) $-LI \preceq \nabla^2 f(x) \preceq LI$

(3) $\| f(y) - f(x) - \langle \nabla f(x), y - x \rangle \|_2 \leq \frac{L}{2} \|y - x\|_2^2$

Pf. $\| \nabla^2 f(x)[v] \| = \| \lim_{t \to 0} \frac{\nabla f(x + tv) - \nabla f(x)}{t} \|$

(1) ⇒ (2). $\leq L \|v\|$

(2) ⇒ (3): $f(y) = f(x) + \int_{t=0}^{1} \langle \nabla f(x + t(y-x)), y - x \rangle \, dt$

$= f(x) + \langle \nabla f(x), y - x \rangle + \int_{t=0}^{1} \int_{s=0}^{t} \nabla^2 f(x + s(y-x)) [y-x, y-x] \, ds \, dt$

So

$\| f(y) - f(x) - \langle \nabla f(x), y - x \rangle \|_2 \leq L \|y - x\|^2 \cdot \int_{t=0}^{1} \int_{s=0}^{t} ds \, dt$

$= \frac{L}{2} \|y - x\|^2$

(3) ⇒ (1).

Let $h(x) = f(x) + \frac{L}{2} \|x\|^2$ $\quad \nabla h(x) = \nabla f(x) + Lx$

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

Then $h(y) - h(x) = f(y) - f(x) + \frac{L}{2}(\|y\| \ldots$

$$\geq \langle \nabla f(x), y - x \rangle - \frac{L}{2}\|y-x\|^2 + \frac{L}{2}(\|y\|^2 - \|x\|^2)$$

$$= \langle \nabla f(x), y - x \rangle + \langle Lx, y - x \rangle$$

$$= \langle \nabla h(x), y - x \rangle$$

$\therefore \nabla^2 h(x) \succcurlyeq 0$
by convexity

i.e. $\nabla^2 f(x) + LI \succcurlyeq 0.$

llly, $h(x) = -f(x) + \frac{L}{2}\|x\|^2$ shows $\nabla^2 f(x) \preccurlyeq LI$

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

Pf. $f(x - \frac{1}{L}\nabla f(x)) = f(x) + \langle \nabla f(x), -\frac{1}{L}\nabla f(x) \rangle$

for some $z \in [x - \frac{1}{L}\nabla f(x), x]$ $\quad + \frac{1}{2}\left(-\frac{1}{L}\nabla f(x)\right)^T \nabla^2 f(z) \left(-\frac{1}{L}\nabla f(x)\right)$

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

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

Thm1 For general f, GD with $h = \frac{1}{L}$ achieves $\|\nabla f(x)\| \leq \varepsilon$ in at most $\frac{2L}{\varepsilon^2}(f(x^0) - f(x^*))$ steps.

Pf. f decreases by at least $\frac{\varepsilon^2}{2L}$ in each step.

What if f is convex?

Then $\nabla f(x) = 0 \Rightarrow x$ is a global min.

Thm. f convex, GD with $h = \frac{1}{L}$ gives

$$f(x^{(k)}) - f(x^*) \le \frac{2L}{k+4} \| x^{(0)} - x^* \|_2^2$$

Pf. $f(x^{(k+1)}) = f(x^{(k)} - \frac{1}{L} \nabla f(x^{(k)})) \le f(x^{(k)}) - \frac{1}{2L} \| \nabla f(x^{(k)}) \|_2^2$

Fact. $|f(y) - f(x)| \le \|\nabla f(x)\|_2 \|y - x\|_2$ for f convex

Pf.

$$f(x) - f(y) \le - \langle \nabla f(x), y - x \rangle$$

$$f(y) - f(x) \le \|\nabla f(x)\|_2 \|y - x\|_2$$

$$\underbrace{f(x^{(k+1)}) - f(x^*)}_{\varepsilon_{k+1}} \le \underbrace{f(x^{(k)}) - f(x^*)}_{\varepsilon_k} - \frac{1}{2L} \frac{(f(x^{(k)}) - f(x^*))^2}{\|x^{(k)} - x^*\|^2}$$

Let $R = \max_{x:\, f(x) \le f(x^{(0)})} \|x - x^*\|$

$$\varepsilon_{k+1} \le \varepsilon_k - \frac{1}{2L}\left(\frac{\varepsilon_k}{R}\right)^2$$

$$\frac{1}{\varepsilon_{k+1}} - \frac{1}{\varepsilon_k} = \frac{\varepsilon_k - \varepsilon_{k+1}}{\varepsilon_{k+1}\varepsilon_k} \ge \frac{1}{\varepsilon_k^2} \cdot \frac{\varepsilon_k^2}{2LR^2} \ge \frac{1}{2LR^2}.$$

$$\varepsilon_0 = f(x^{(0)}) - f(x^*) \le \langle \nabla f(x^*), x^{(0)} - x^* \rangle + \frac{L}{2}\|x^{(0)} - x^*\|_2^2$$

$$\le \frac{L}{2}R^2$$

$$\therefore \frac{1}{\varepsilon_k} \ge \frac{2}{LR^2} + \frac{k}{2LR^2} \ge \frac{k+4}{2LR^2} \Rightarrow \varepsilon_k \le \frac{2LR^2}{k+4}.$$

To complete the proof we will shows that

$$\|x^{(k+1)} - x^*\| \le \|x^{(k)} - x^*\|$$

Why? $\|x^{(k+1)} - x^*\|^2 = \|x^{(k)} - h\nabla f(x^{(k)}) - x^*\|^2$

$$= \|x^{(k)} - x^*\|^2 + h^2\|\nabla f(x^{(k)})\|^2 - 2h\langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle$$

$$\nabla f(x^{(k)}) = \int_{t=0}^{1} \left(\nabla^2 f(x^* + t(x^{(k)} - x^*))\right)(x^{(k)} - x^*)\, dt$$

$$= \underbrace{\int_0^1 \nabla^2 f(x^* + t(x^{(k)} - x^*))\, dt}_{H \succeq 0}\,(x^{(k)} - x^*)$$

$H^2 \preceq \|H\|\, H$

$H \succcurlyeq 0$

$H^2 \preccurlyeq \|H\|_{op} H$

i.e. $H \succcurlyeq \frac{1}{L} H^2$

So $\langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle \geq \frac{1}{L} (x^{(k)} - x^*)^T H^2 (x^{(k)} - x^*)$

$$= \frac{1}{L} \| H (x^{(k)} - x^*) \|_2^2$$

$$= \frac{1}{L} \| \nabla f(x^{(k)}) \|_2^2$$

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

$\leq \|x^{(k)} - x^*\|^2$ for $h \leq \frac{2}{L}$.