Lemma 2.1 (Descent Lemma). *Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L, then*

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

Remark: If assuming $\|\nabla^2 f\| \le L \Rightarrow |\lambda(\nabla^2 f)| = \sigma(\nabla^2 f) \le L$

Minmax Principle $\Rightarrow -L \le \dfrac{h^T \nabla^2 f h}{h^T h} \le L$

Taylor Thm $f(\mathbf{y}) = f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{1}{2}(\mathbf{x}-\mathbf{y})^T\nabla^2 f(\mathbf{z})(\mathbf{x}-\mathbf{y})$

implies

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

Remark: If ∇f is cont., Strong Convexity

$$\Leftrightarrow \quad f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\|\mathbf{x}-\mathbf{y}\|^2$$

Lemma 2.4 (Sufficient Decrease Lemma). *Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L, then the gradient descent method* (2.1) *satisfies*

$$f(\mathbf{x}) - f(\mathbf{x}-\eta\nabla f(\mathbf{x})) \ge \eta(1-\frac{L}{2}\eta)\|\nabla f(\mathbf{x})\|^2, \quad \forall \mathbf{x}, \forall \eta > 0.$$

Proof $y = x - \eta \nabla f(x)$ in Descent Lemma

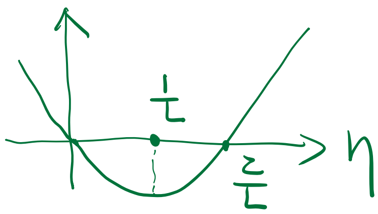
$$f(\mathbf{y}) \le f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|^2.$$

GD $x_{k+1} = x_k - \eta \nabla f(x_k)$

$$f(x_{k+1}) - f(x_k) \le -\eta\left(1 - \frac{L}{2}\eta\right)\|\nabla f(x_k)\|^2$$

I. To have $f(x_{k+1}) < f(x_k)$, we need

$$-\eta\left(1-\frac{L}{2}\eta\right) = \frac{L}{2}\left(\eta^2 - \frac{2}{L}\eta\right) = \frac{L}{2}\left(\eta - \frac{1}{L}\right)^2 - \frac{L}{2} < 0$$



$\Rightarrow \quad 0 < \eta < \frac{2}{L}$

Stability

① GD $x_{k+1} = x_k - \eta \nabla f(x_k)$ with $\eta > 0$ is numerically stable if $\eta < \frac{2}{L}$

$f(x_*) \le f(x_{k+1}) < f(x_k)$ if global minimum $f(x_*)$ exists.

② In practice it's hard to have exact L.

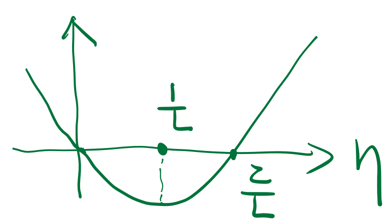
Assume ∇f is L-continuous, then

GD is stable for any $\eta \in (0, \frac{2}{L})$ with unknown L

$\Rightarrow$ GD is stable for small enough η.

II. "Best" constant step size is to minimize $-\eta\left(1 - \frac{L}{2}\eta\right)$

$$\eta = \frac{1}{L}$$



$$\Rightarrow f(x_{k+1}) - f(x_k) \le -\eta\left(1 - \frac{L}{2}\eta\right)\|\nabla f(x_k)\|^2$$

$\min_{\eta} f(x_k - \eta \nabla f(x_k))$ $\qquad \leq -\frac{L}{2}\|\nabla f(x_k)\|^2$

"Best" only in the sense of minimizing $-\eta(1-\frac{L}{2}\eta)$

III. Convergence of Constant Step Size $\eta \in (0, \frac{2}{L})$

$$f(x_{k+1}) - f(x_k) \leq -\eta(1-\frac{L}{2}\eta)\|\nabla f(x_k)\|^2$$

$$\eta \in (0, \frac{2}{L}) \Rightarrow \omega = \eta(1-\frac{L}{2}\eta) > 0$$

$$f(x_k) - f(x_{k+1}) \geq \omega \|\nabla f(x_k)\|^2$$

① Sum it for $k=0,1,2,\ldots$

$$\sum_{k=0}^{\infty} [f(x_k) - f(x_{k+1})] \geq \omega \sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2$$

② $\{f(x_k)\}$ is a decreasing sequence, thus it is also bounded $(f(x_*) \leq f(x_k) < f(x_0))$

[Completeness Theorem of Real numbers]

monotone bounded sequence has a limit.

So $\lim_{k\to\infty} f(x_k)$ exists (doesn't imply $\lim_{k\to\infty} x_k$ exists)

$$LHS = f(x_0) - \lim_{k\to\infty} f(x_k)$$

$$\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \leq \frac{1}{\omega}\left[f(x_0) - \lim_{k\to\infty} f(x_k)\right]$$

$g_n = \sum_{k=0}^{n} \|\nabla f(x_k)\|^2$ is $\nearrow$ and bounded

The series $\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2$ converges

$$\Rightarrow \lim_{k\to\infty} \|\nabla f(x_k)\| = 0$$

(doesn't imply $\lim_{k\to\infty} x_k$ exists)

(3) Let $g_N = \min_{0 \leq k \leq N} \|\nabla f(x_k)\|$, then

$$f(x_k) - f(x_{k+1}) \geq \omega \|\nabla f(x_k)\|^2$$

$$\Rightarrow \sum_{k=0}^{N} \|\nabla f(x_k)\|^2 \leq \frac{1}{\omega}\left[f(x_0) - f(x_{N+1})\right]$$

$$\leq \frac{1}{\omega}\left[f(x_0) - f(x_*)\right]$$

$$(N+1)\, g_N^2 \leq \sum_{k=0}^{N} \|\nabla f(x_k)\|^2$$

$$\Rightarrow g_N \leq \frac{1}{\sqrt{N+1}} \sqrt{\frac{1}{\omega}\left[f(x_0) - f(x_*)\right]}$$

<u>Theorem</u> Assume ∇f is L-continuous.

Assume $f(x) \geq f(x_*)$, $\forall x \in \mathbb{R}^n$

Then for $x_{k+1} = x_k - \eta \nabla f(x_k)$

where $\eta \in (0, \frac{2}{L})$ is a contant:

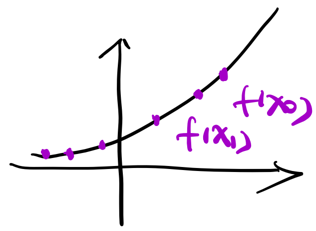
① $f(x_{k+1}) - f(x_k) \le -\eta(1-\frac{L}{2}\eta)\|\nabla f(x_k)\|^2$

$\omega = \eta(1-\frac{L}{2}\eta)$

② $\lim\limits_{x \to \infty} \|\nabla f(x_k)\| = 0$

③ $\min\limits_{0 \le k \le N} \|\nabla f(x_k)\| \le \frac{1}{\sqrt{N+1}} \sqrt{\frac{1}{\omega}\left[f(x_0) - f(x_*)\right]}$

Example: GD for $f(x) = e^x$



III. Convergence of $\{x_k\}$ for convex $f(x)$

Theorem If ∇f is L-continuous and $f(x)$ is convex:

(a) $f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$

(b) $\|\nabla f(x) - \nabla f(y)\|^2 \le L \langle \nabla f(x) - \nabla f(y), x - y \rangle$

Remark (a) can be written as

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$$

Compare it with

∇f is L-Cont. $\Rightarrow$ $f(\mathbf{y}) \le f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2$

Strong Convexity $\Rightarrow$ $f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2$

Pick any x_0, then for fixed x_0

Proof: Define $\phi(x) = f(x) - \langle \nabla f(x_0), x \rangle \Rightarrow \nabla\phi(x_0) =$

Then $\nabla\phi(x) = \nabla f(x) - \nabla f(x_0)$

$\nabla\phi(y) = \nabla f(y) - \nabla f(x_0)$

$\Rightarrow \|\nabla\phi(x) - \nabla\phi(y)\| = \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$

$\Rightarrow \nabla\phi(x)$ is L-continuous

Descent Lemma on $\phi(x)$:

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

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

$\langle a, b \rangle \leq \|a\| \cdot \|b\|$

$r = \|x-y\|$

$$\min_{x\in\mathbb{R}^n} \phi(x) \leq \min_{x\in\mathbb{R}^n}\left[\phi(y) + \|\nabla\phi(y)\| \cdot \|y-x\| + \frac{L}{2}\|x-y\|^2\right]$$

$\phi(x_0)$

$$= \min_{r\geq 0}\left[\phi(y) + r\|\nabla\phi(y)\| + \frac{L}{2}r^2\right]$$

$$= \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|^2$$

$r = -\frac{1}{L}\|\nabla\phi\|$

$\phi(x) = f(x) - \langle \nabla f(x_0), x \rangle \Rightarrow \phi(x)$ is convex

linear function satisfies

$g(\lambda x+(1-\lambda)y)\leq \lambda g(x)+(1-\lambda)g(y)$

Sum of two convex functions is still convex.

$$\left.\begin{array}{l}\nabla\phi(x_0)=0\\ \phi \text{ is convex}\end{array}\right\}\Rightarrow x_0 \text{ minimizes } \phi(x)$$

So $\phi(x_0)\leq \phi(y)-\frac{1}{2L}\|\nabla\phi(y)\|^2$

$$f(x_0)-\langle\nabla f(x_0),x_0\rangle \leq f(y)-\langle\nabla f(x_0),y\rangle-\frac{1}{2L}\|\nabla f(y)-\nabla f(x_0)\|^2$$

$$\underset{(a)}{\Rightarrow} f(x)+\langle\nabla f(x),y-x\rangle+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2\leq f(y) \quad ①$$

$$f(y)+\langle\nabla f(y),x-y\rangle+\frac{1}{2L}\|\nabla f(y)-\nabla f(x)\|^2\leq f(x) \quad ②$$

Add two inequalities $\Rightarrow$

(b) $\|\nabla f(x)-\nabla f(y)\|^2\leq L\,\langle\nabla f(x)-\nabla f(y),x-y\rangle$

Theorem Assume ∇f is L-continuous.
Assume $f(x)\geq f(x_*)$, $\forall x\in\mathbb{R}^n$
Assume $f(x)$ is convex.

Then for $x_{k+1}=x_k-\eta\nabla f(x_k)$

where $\eta\in(0,\frac{2}{L})$ is a constant

$$f(x_k)-f(x_*) \le \frac{1}{\frac{1}{f(x_0)-f(x_*)} + k\omega \frac{1}{\|x_0-x_*\|^2}} < \frac{\|x_0-x_*\|^2}{k\omega}$$

$$\omega = \eta\left(1-\frac{L}{2}\eta\right)$$

Proof: Let $r_k = \|x_k - x_*\|$

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \le L\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle.$$

$$\begin{aligned}
r_{k+1}^2 &= \|x_{k+1} - x_*\|^2 \\
&= \|x_k - \eta \nabla f(x_k) - x_*\|^2 \\
&= \|x_k - x_* - \eta \nabla f(x_k)\|^2 \\
&= r_k^2 - 2\eta \langle \nabla f(x_k), x_k - x_*\rangle + \eta^2 \|\nabla f(x_k)\|^2 \\
&= r_k^2 - 2\eta \langle \nabla f(x_k) - \nabla f(x_*), x_k - x_*\rangle + \eta^2 \|\nabla f(x_k)\|^2 \\
&\le r_k^2 - 2\eta \frac{1}{L} \|\nabla f(x_k) - \nabla f(x_*)\|^2 + \eta^2 \|\nabla f(x_k)\|^2 \\
&\le r_k^2 - \eta\left(\frac{2}{L} - \eta\right) \|\nabla f(x_k)\|^2
\end{aligned}$$

Let $R_k = f(x_k) - f(x_*)$

Convexity $\Rightarrow f(x) \ge f(x_k) + \langle \nabla f(x_k), x - x_k\rangle$

$\Rightarrow f(x_*) \ge f(x_k) + \langle \nabla f(x_k), x_* - x_k\rangle$

$\Rightarrow f(x_k) - f(x_*) \le \langle \nabla f(x_k), x_k - x_*\rangle$

$$\leq \|\nabla f(x_k)\| \cdot \|x_k - x_*\|$$

$$\Rightarrow \quad R_k \leq \|\nabla f(x_k)\| \cdot r_k$$

$$\Rightarrow \quad -\|\nabla f(x_k)\| \leq -\frac{R_k}{r_k}$$

Recall we have $\qquad \omega = \eta(1 - \frac{L}{2}\eta)$

$$f(x_{k+1}) \leq f(x_k) - \omega \|\nabla f(x_k)\|^2$$

$$\Rightarrow f(x_{k+1}) - f(x_*) \leq f(x_k) - f(x_*) - \omega \|\nabla f(x_k)\|^2$$

$$0 < R_{k+1} \leq R_k - \frac{\omega}{r_k^2} R_k^2$$

Multiply both sides by $\frac{1}{R_{k+1}} \frac{1}{R_k}$

$$\Rightarrow \quad \frac{1}{R_k} \leq \frac{1}{R_{k+1}} - \frac{\omega}{r_k^2} \frac{R_k}{R_{k+1}}$$

$$\Rightarrow \quad \frac{1}{R_{k+1}} \geq \frac{1}{R_k} + \frac{\omega}{r_k^2} \frac{R_k}{R_{k+1}}$$

$$\geq \frac{1}{R_k} + \frac{\omega}{r_k^2}$$

$r_k \leq r_0$

$\frac{1}{r_k} \geq \frac{1}{r_0}$

Summing up for $k = 0, 1, \cdots, N$

$$\Rightarrow \quad \frac{1}{R_{N+1}} \geq \frac{1}{R_0} + \frac{\omega}{r_0^2}(N+1)$$

$$\Rightarrow \quad R_{N+1} \leq \frac{1}{\frac{1}{R_0} + \frac{\omega}{r_0^2}(N+1)}$$