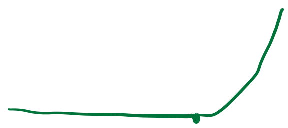
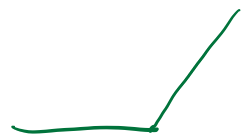
Theorem 2.8. *Assume $\nabla f(\mathbf{x})$ is Lipschitz-continuous with Lipschitz constant L and $f(\mathbf{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex. Then for any $\mathbf{x}, \mathbf{y}$:*

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

An example of $\begin{cases} f \text{ is convex} \\ \nabla f \text{ is } L\text{-continuous} \end{cases}$

$$f(x) = \begin{cases} \frac{x^2}{2}, & x \geqslant 0 \\ 0, & x \leq 0 \end{cases} \qquad f'(x) = \begin{cases} x \\ 0 \end{cases}$$





Proof. Define $\phi(\mathbf{x}) = f(\mathbf{x}) - \langle \nabla f(\mathbf{x}_0), \mathbf{x} \rangle$. Then $\phi(\mathbf{x})$ also has Lipschitz continuous gradient:

$$\|\nabla \phi(\mathbf{x}) - \nabla \phi(\mathbf{y})\| = \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|.$$

Apply Lemma 2.1 to $\phi(\mathbf{x})$:

$$\begin{aligned} \phi(\mathbf{x}) &\leq \phi(\mathbf{y}) + \langle \nabla \phi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2 \\ (|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \|\mathbf{b}\|) \quad &\leq \phi(\mathbf{y}) - \|\nabla \phi(\mathbf{y})\| \|\mathbf{x} - \mathbf{y}\| + \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2 \end{aligned}$$

By Theorem 1.5, $\phi(\mathbf{x})$ is also convex because $-\langle \nabla f(\mathbf{x}_0), \mathbf{x} \rangle$ is convex. Moreover, $\nabla \phi(\mathbf{x}_0) = \mathbf{0}$, thus by Theorem 2.4, $\mathbf{x}_0$ is a global minimizer of $\nabla \phi(\mathbf{x})$. So we get

$$\begin{aligned} \phi(\mathbf{x}_0) = \min_{\mathbf{x}} \phi(\mathbf{x}) &\leq \min_{\mathbf{x}} \left[\phi(\mathbf{y}) - \|\nabla \phi(\mathbf{y})\| \|\mathbf{x} - \mathbf{y}\| + \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2 \right] \\ &\leq \min_{r \geq 0} \left[\phi(\mathbf{y}) - \|\nabla \phi(\mathbf{y})\| r + \frac{L}{2} r^2 \right] \\ &= \phi(\mathbf{y}) - \frac{1}{2L} \|\nabla \phi(\mathbf{y})\|^2. \end{aligned}$$

Thus $\phi(\mathbf{x}_0) \leq \phi(\mathbf{y}) - \frac{1}{2L}\|\nabla\phi(\mathbf{y})\|^2$ implies

$$f(\mathbf{x}_0) - \langle \nabla f(\mathbf{x}_0), \mathbf{x}_0 \rangle \leq f(\mathbf{y}) - \langle \nabla f(\mathbf{x}_0), \mathbf{y} \rangle - \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}_0)\|^2.$$

Since $\mathbf{x}_0, \mathbf{y}$ are arbitrary, we can also write is as

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

which implies

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

Switching $\mathbf{x}$ and $\mathbf{y}$, we get

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

and adding two we get

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

□

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

② $\lim\limits_{x\to\infty} x_k$ may not exist!

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

③ $\lim\limits_{x\to\infty} \|\nabla f(x_k)\| = 0$

④ $\min\limits_{0 \leq k \leq N} \|\nabla f(x_k)\| \leq \frac{1}{\sqrt{N+1}}\sqrt{\frac{1}{\omega}[f(x_0) - f(x_*)]}$

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 contant:

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

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

Remark: $f(x_k) - f(x_*) < \frac{\|x_0 - x_*\|^2}{\omega} \cdot \frac{1}{k}$
gives convergence rate $O(\frac{1}{k})$, under the assumptions of only convexity and L-continuity.

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

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

$$= r_k^2 - 2\eta \langle \nabla f(x_k) - \nabla f(x_*), x_k - x_* \rangle + \eta^2 \|\nabla f(x_k)\|^2$$

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

$$\le r_k^2 - \eta \left(\frac{2}{L} - \eta\right) \|\nabla f(x_k)\|^2 \Rightarrow r_k^2 \downarrow$$

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

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

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

$$\Rightarrow 0 \le f(x_k) - f(x_*) \le \langle \nabla f(x_k), x_k - x_* \rangle$$

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

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

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

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

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

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

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

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

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

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

$$\geq \frac{1}{R_k} + \frac{\omega}{r_k^2} \geq \frac{1}{R_k} + \frac{\omega}{r_0^2}$$

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

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

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

IV. Convergence Rate for Strongly Convex Function

Theorem Strong Convexity & L-conitinuity ∇f

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

Remark: Convexity $(\mu=0)$ & L-conitinuity

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

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

Then for $x_{k+1} = x_k - \eta \nabla f(x_k)$
with any $\eta \in (0, \frac{2}{L+\mu}]$:

$$\|x_k - x^*\|^2 \leq \left(1-\frac{2\eta\mu L}{\mu+L}\right)^k \|x_0 - x^*\|^2$$

If $\eta = \frac{2}{L+\mu}$, then

$$\|x_k - x^*\| \le \left(\frac{L/\mu - 1}{L/\mu + 1}\right)^k \|x_0 - x^*\|$$

$$f(x_k) - f(x_*) \le \frac{L}{2}\left(\frac{L/\mu - 1}{L/\mu + 1}\right)^{2k} \|x_0 - x^*\|^2$$

Remark: For $f(x) = \frac{1}{2} x^T K x - x^T b$

$\nabla^2 f(x) = K > 0$ with $\lambda_i(K) = \frac{1}{\Delta x^2}\left(2 - 2\cos(\pi i \Delta x)\right)$

$i = 1, 2, \cdots, n$

$$L = \|\nabla^2 f\| = \max_i \sigma_i(K)$$

$$\mu = \min_i \lambda_i(K) = \min_i \sigma_i(K)$$

because $K > 0$

So $\frac{L}{\mu} = \frac{\sigma_1}{\sigma_n}$ is the condition number of $\nabla^2 f = K$.

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

$$r_{k+1}^2 = \|x_{k+1} - \eta \nabla f(x_k)\|^2$$

$$= \|x_k - x^* - \eta \nabla f(x_k)\|^2$$

$$= r_k^2 + 2\langle -\eta \nabla f(x_k), x_k - x^* \rangle + \eta^2 \|\nabla f(x_k)\|^2$$

$$= r_k^2 - 2\eta \langle \nabla f(x_k) - \nabla f(x_*), x_k - x^* \rangle + \eta^2 \|\nabla f(x_k)\|^2$$

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

$$\langle \nabla f(x_k) - \nabla f(x_*), x_k - x_* \rangle \geq \frac{\mu L}{\mu + L}\|x_k - x_*\|^2 + \frac{1}{\mu + L}\|\nabla f(x_k)\|^2$$

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

$$= \left(1 - 2\eta \frac{\mu L}{\mu + L}\right) r_k^2 + \eta\left(\eta - \frac{2}{\mu + L}\right)\|\nabla f(x_k)\|^2$$

So $\eta \in \left(0, \frac{2}{\mu + L}\right] \Rightarrow 0 \leq 1 - 2\eta \frac{\mu L}{\mu + L} < 1$

$$r_{k+1}^2 \leq \left(1 - 2\eta \frac{\mu L}{\mu + L}\right) r_k^2$$

$\Leftrightarrow 2\eta \frac{\mu L}{\mu + L} < 1$

$\frac{2}{\mu + L} \cdot \frac{2\mu L}{\mu + L} \leq 1$

$$\Rightarrow r_k^2 \leq \left(1 - \frac{2\eta \mu L}{\mu + L}\right)^k r_0^2$$

$\Leftrightarrow 4\mu L \leq (\mu + L)^2$

If $\eta = \frac{2}{\mu + L} \Rightarrow r_k^2 \leq \left[1 - \frac{2\mu L}{(\mu + L)^2}\right]^k r_0^2$

$$\Rightarrow \quad r_k \le \left[\frac{L-\mu}{L+\mu}\right]^k r_0$$

$$\boxed{\begin{array}{l}\text{Descent Lemma} \\ f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|x-y\|^2\end{array}}$$

$$\Rightarrow \quad f(x_k) \le f(x_*) + \nabla f(x_*)^T (x_k - x_*) + \frac{L}{2}\|x_k - x_*\|^2$$

$$\Rightarrow \quad f(x_k) - f(x_*) \le \frac{L}{2}\|x_k - x_*\|^2$$

$$\le \frac{L}{2}\left[\frac{L-\mu}{L+\mu}\right]^{2k} \|x_0 - x_*\|^2$$

Convergence Rates of GD

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

$\eta \in \left(0, \frac{2}{L}\right)$

① ∇f is L-continuous : $O\left(\frac{1}{\sqrt{k}}\right)$

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

② ∇f is L-continuous & $f(x)$ is convex : $O\left(\frac{1}{k}\right)$

$\eta \in \left(0, \frac{2}{L}\right)$

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

③ ∇f is L-continuous & $f(x)$ is strongly convex with μ

$O(c^k)$ $\quad 0 < c < 1$ $\quad \eta \in \left(0, \frac{2}{L+\mu}\right]$

$$\|x_k - x^*\|^2 \le \left(1 - \frac{2\eta\mu L}{\mu + L}\right)^k \|x_0 - x^*\|^2$$

$$\|x_k - x^*\| \le \left(\sqrt{1 - \frac{2\eta\mu L}{\mu + L}}\right)^k \|x_0 - x^*\|$$

$c = \sqrt{1 - \frac{2\eta\mu L}{\mu + L}}$

$$f(x_k) - f(x_*) \le \frac{L}{2}\|x_k - x_*\|^2$$

$$\le \left(\sqrt{1 - \frac{2\eta\mu L}{\mu + L}}\right)^{2k} \|x_0 - x^*\|^2$$

$$0 < 1 - \frac{2\eta\mu L}{\mu + L} < 1 \iff 0 < \eta < \frac{\mu + L}{2\mu L}$$

$$0 < \eta \leq \frac{2}{L+\mu} \Rightarrow 0 < \eta < \frac{\mu+L}{2\mu L}$$

Remark: Larger $\mu \Rightarrow$ smaller $c = \sqrt{1-2\eta\frac{L}{1+L/\mu}}$

$\Rightarrow$ faster convergence

Example: $f(x) = ax^2 \quad x \in \mathbb{R}, \ a > 0$

$\mu = 2a$