Review

① Convex Function: $\forall x, y \in \mathbb{R}^n$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in (0,1)$$

Example: $A \in \mathbb{R}^{n \times n}$

$f(A) = \|A\|$ is convex w.r.t. A.

Convex function satisfies

1) $f(\underbrace{\sum_{i=1}^{n} a_i x_i}) \leq \sum_{i=1}^{n} a_i f(x_i), \quad \forall x_i \in \mathbb{R}^n$

Convex combination $\boxed{a_i \geq 0, \ \sum_{i=1}^{m} a_i = 1}$

2) If $g(\cdot)$ is an integrable function,

$$f\left(\frac{1}{b-a} \int_a^b g(t)\, dt\right) \leq \frac{1}{b-a} \int_a^b f[g(t)]\, dt$$

Example: $\left\| \int_0^1 \nabla^2 f(x + t(y-x))\, dt \right\| \leq \int_0^1 \|\nabla^2 f(\ )\|\, dt$

3) $\forall x, y, \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

if ∇f exists.

4) $\nabla^2 f(x) \succeq 0$ if $\nabla^2 f$ exists

② <u>Def</u> (Lipschitz Continuity)

$f(x)$ is Lipschitz continuous if

$$\exists L > 0, \ \forall x, y \in \mathbb{R}^n, \ |f(x) - f(y)| \leq L \|x - y\|$$

Theorem: $\|\nabla^2 f(x)\| \leq L$, $\forall x$

$\Rightarrow \nabla f(x)$ is Lipschitz continuous

③

Descent Lemma Assume $\nabla f(x)$ is L-Lipschitz

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

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

Taylor Theorem $f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x)$

$z = x + \theta(y-x)$

Courant-Fischer-Weyl Minmax Principle:

$A \in \mathbb{R}^{n\times n}$ is real symmetric with eigenvalues

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then

$$\lambda_n \leq \frac{x^T A x}{x^T x} \leq \lambda_1, \quad \forall x \in \mathbb{R}^n$$

Example: Assume $\|\nabla^2 f(x)\| \leq L$, $\forall x$

$\Rightarrow |\lambda_i(\nabla^2 f)| = \sigma_i(\nabla^2 f) \leq \sigma_1 \leq L$, $\forall x$

$$\Rightarrow \left|\frac{h^T \nabla^2 f(z) h}{h^T h}\right| \leq \max\{|\lambda_1|, |\lambda_n|\}$$

$= \sigma_1 \leq L$

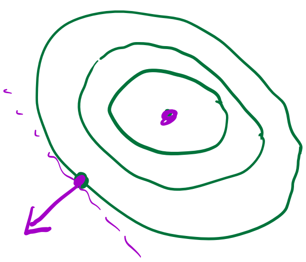
$$-\frac{1}{2}L\|y-x\|^2 \leq \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \leq \frac{1}{2}L\|y-x\|^2$$

$\boxed{\text{Sufficient Decrease Lemma}}$

Assume $\nabla f(x)$ is L-Lipschitz. Then $\forall x \in \mathbb{R}^n$, $\forall \eta > 0$

$$f(x) - f(x - \eta\nabla f(x)) \geq \eta\left(1 - \frac{L}{2}\eta\right)\|\nabla f(x)\|^2$$

Gradient Descent $x_{k+1} = x_k - \eta\nabla f(x_k)$, $\eta > 0$



Contour lines of $f(x,y) = (x-1)^2 + (y-2)^2$

Proof: Descent Lemma $f(y) \leq f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2$

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

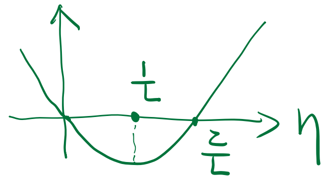
$$= f(x) - \eta\left(1 - \frac{L}{2}\eta\right)\|\nabla f(x)\|^2.$$

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

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

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

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



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

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.

③ Consider solving Ordinary Differential Equation

$$\frac{d}{dt} u = F(u) \qquad u^n \approx u(n\Delta t)$$

The forward Euler scheme is

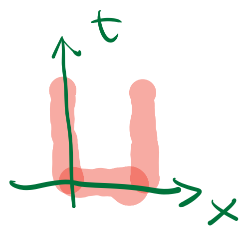
$$\frac{u^{n+1} - u^n}{\Delta t} = F(u^n)$$

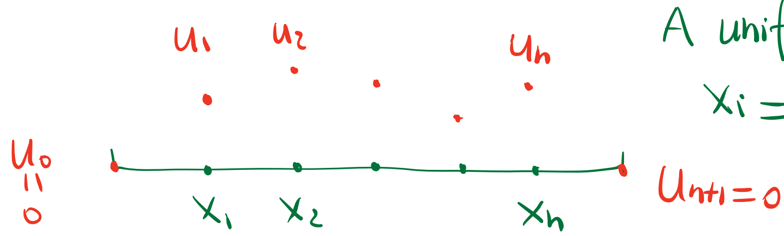
$$u^{n+1} = u^n + \Delta t F(u^n)$$

same as GD if $F = -\nabla f$

Example: Initial-boundary value problem of 1D heat equation

$$\begin{cases} \frac{\partial}{\partial t} u(x,t) = \frac{\partial^2}{\partial x^2} u(x,t) \\ u(x,0) = u_0(x) \\ u(0,t) = u(1,t) = 0 \end{cases} \qquad x \in [0,1]$$





A uniform grid

$x_i = i\Delta x$

Second order difference $\quad u''(x) \approx \dfrac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{\Delta x^2}$

$$K = \frac{1}{\Delta x^2}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{pmatrix}$$

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad -KU = \frac{1}{\Delta x^2}\begin{pmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & -2 & 1 \\ & & 1 & -2 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$$

A semi-discrete scheme for heat equation:

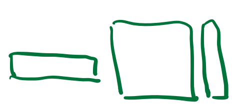
$$\frac{d}{dt} U(t) = -K\, U(t)$$

Forward Euler Scheme

$$U^{n+1} = U^n - \Delta t\, K U^n$$

GD for $\quad \nabla f(U) = KU$

$$f(U) = \frac{1}{2} U^T K U$$



See my MA 615 Notes for eigenvalues of K:

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

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

Stability for forward Euler $\left(\|U^{n+1}\| \leq \|U^n\|\right)$

requires $\Delta t \leq \frac{1}{2}\Delta x^2$

Let's regard it as GD minimizing $f(U)$

$$\lambda_i = \frac{2 - 2\cos(\pi i \Delta x)}{\Delta x^2} = 4\frac{1}{\Delta x^2}\sin^2\left(\frac{\pi \Delta x}{2} i\right)$$

$$\nabla^2 f = K \Rightarrow \|\nabla^2 f\| \leq \max_i \sigma_i = \max_i \lambda_i \leq 4\frac{1}{\Delta x^2}$$

$$\Rightarrow \nabla f \text{ is } L\text{-continuous with } L = 4\frac{1}{\Delta x^2}$$

$$\boxed{\eta < \frac{2}{L} \Rightarrow \Delta t < \frac{1}{2}\Delta x^2}$$

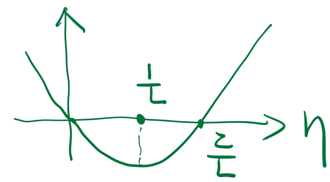
Remark: Forward Euler $u^{n+1} = u^n + \Delta t F(u^n)$

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

if $F = -\nabla f$. But FE is usually used for approximating time dynamics $u(x,t)$. GD is used for approximating minimizer (steady state ODE sol).

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

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



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

$\min_{\eta} f(x_k - \eta \nabla f(x_k))$

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

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

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

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

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

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

① Sum it for $k = 0, 1, 2, \ldots$

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

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

Completeness Theorem of Real numbers

A 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

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

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

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

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

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

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

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

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

③ $\min_{0 \leq k \leq N} \|\nabla f(x_k)\| \leq \frac{1}{\sqrt{N+1}} \sqrt{\frac{1}{\omega}[f(x_0) - f(x_*)]}$

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

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

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

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

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

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

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

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

Descent Lemma on $\phi(x)$:

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

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

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

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

$\parallel$

$\phi(x_0)$

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

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

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

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

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

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

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

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

Add two inequalities $\Rightarrow$

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

<u>Theorem</u> 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 $r_k = \|x_k - x_*\|$

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

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

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

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

$$\Rightarrow f(x_k) - f(x_*) \leq \langle \nabla f(x_k), x_k - x_* \rangle$$
$$\leq \|\nabla f(x_k)\| \cdot \|x_k - x_*\|$$

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

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

$$f(x_{k+1}) \leq f(x_k) - \omega \|\nabla f(x_k)\|^2$$
$$\Rightarrow f(x_{k+1}) - f(x_*) \leq f(x_k) - f(x_*) - \omega \|\nabla f(x_k)\|^2$$
$$0 < R_{k+1} \leq R_k - \frac{\omega}{r_k^2} R_k^2$$

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

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

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

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

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