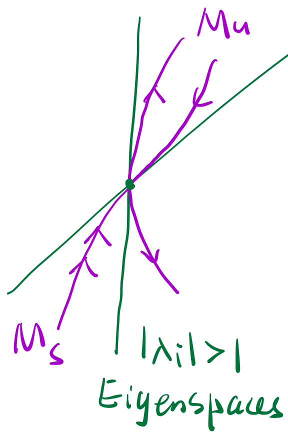
Guest Lecture by Ziang Chen

Reference: Lee et. al. (2019) First order methods almost always avoid strict saddle points



strict saddle points

$\nabla f(x_*)=0$ $\quad \lambda_{min}(\nabla^2 f(x_*))<0$

Gradient Descent $\quad x_{t+1} = x_t - \alpha \nabla f(x_t) = g(x_t)$

$$g = I - \alpha\nabla f$$

Linearization: $A = I - \alpha H$

$$H = \nabla^2 f(x_*)$$

$$\lambda_i(A) = 1 - \alpha\lambda_i(H)$$

If $\lambda_i(H)<0$, then $\lambda_i(A)>1$

$\Rightarrow$ the unstable manifold M_u is $\underbrace{\text{non-trivial}}_{\dim(M_u)\geq 1}$

Assume $\lambda_i(H)\neq 0,\ \forall i \Rightarrow |\lambda_i(A)|\neq 1,\ \forall i$

$\Rightarrow$ no center manifold

$\Rightarrow \dim(M_s) = d - \dim(M_u) \leq d-1 < d$

$\Rightarrow$ Measure $(M_s \cap B) = 0$

saddle point

If $x_t \to x_*$, then $\exists\, t_0$ s.t.

Theorem from last time

$$x_{t_0} \in M_s \cap B$$

$$x_{t_0-1} = g^{-1}(x_{t_0}) \in g^{-1}(M_s \cap B)$$

$$\Rightarrow \quad x_0 \in g^{-t_0}(M_s \cap B)$$

$$\Rightarrow \quad x_0 \in \bigcup_{t=0}^{\infty} g^{-t}(M_s \cap B)$$

$$\Rightarrow \quad \text{measure}\left(\bigcup_{t=0}^{\infty} g^{-t}(M_s \cap B)\right)$$

$$\leq \sum_{t=0}^{\infty} \text{measure}\left(g^{-t}(M_s \cap B)\right)$$

$$= 0$$

<u>Lemma</u>: If $\lambda_i(Dg(x)) \neq 0$, $\forall x \in \mathbb{R}^d$

then $\text{meas}(B) = 0 \Rightarrow \text{meas}(g^{-1}(B)) = 0$

<u>Theorem</u> If x_* is a strict saddle,

then $\{x_0 \mid x_t \to x_*\}$ has measure 0.

Reference: L. Arnold (1998) Random dynamical systems

Prob space $(\Omega, \mathcal{F}, \mathbb{P})$

Ex: $\Omega = \{(i_0, i_1, i_2, \ldots) \mid i_j \in \{1, \ldots, d\}\}$

$\text{Prob}(i_j = k) = \frac{1}{d}$, $\forall k \in \{1, \ldots, d\}$

<u>Def (Metric dynamical system)</u>

$$\theta^t = \theta \circ \theta \circ \cdots \circ \theta$$

$$\theta : \Omega \to \Omega$$

$$\underline{\mathbb{P}(\theta^{-1}(B)) = \mathbb{P}(B), \quad \forall B \in \mathcal{F}}$$

probability preserving

θ is shifting operator

$$\theta(\omega) = \theta(i_0, i_1, \cdots) = (i_1, i_2, \cdots)$$

$$\theta(i_1, i_2, \cdots) = (i_2, i_3, \cdots)$$

<u>Def</u> (Random Dynamical System)

$$\varphi : T \times \Omega \times X \to X$$

$$(t, \omega, x) \longmapsto \varphi(t, \omega, x)$$

Cocycle property:

① $\varphi(0, \omega, x) = x$

② $\varphi(t+s, \omega, x) = \varphi(t, \theta^s \omega, \varphi(s, \omega, x))$

$\forall t, s, \omega, x$

Example: $\omega = (i_0, i_1, i_2, \cdots)$

First iteration, $g(\omega, x) = x - \eta \langle \nabla f(x), e_{i_0} \rangle e_{i_0}$, $i_0 = \pi(\omega)$

$$e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

π is an operator choosing first entry of ω

Second iteration $g(\theta\omega, x) = x - \eta \langle \nabla f(x), e_{\pi(\theta\omega)} \rangle e_{\pi(\theta\omega)}$

$$\pi(\theta\omega) = i_1$$

$$X_0 \xrightarrow{g(\omega,\cdot)} X_1 \xrightarrow{g(\theta\omega,\cdot)} X_2 \longrightarrow \cdots \longrightarrow X_{s-1} \xrightarrow{g(\theta^{s-1}\omega,\cdot)} X_s$$

$$\xrightarrow{g(\theta^{s+1}\omega,\cdot)} X_{s+1} \longrightarrow \cdots \longrightarrow X_{t+s-1} \xrightarrow{g(\theta^{t+s-1}\omega,\cdot)} X_{t+s}$$