$\boxed{\text{April 19, 2019}}$ Policy gradient HRL

$\boxed{\text{update April 29.}}$

g_i: Subgoal @ local i, Top agent considers local i subgoal (sub agent) as 'action'.

S_i: State @ local i

θ_i: policy parameter for local i

$\pi_i(g_i|S_i) \doteq \pi_{i\theta_i}(g_i|S_i)$: policy for local i $\qquad$ $\pi_i(V_i, \theta_{iw}) \doteq \pi_{i\theta_{iw}}(V_i, \theta_{iw})$: policy for local i value weight

q_{it} : return, Sample of $q_{\pi_i}(S_i, g_i)$

$\pi_i(V_i, \theta_{iw})$: policy for local i value weight

θ_{iw}: policy parameter for local i value weight

notation ↑

↓

$$\nabla_{\theta_i} J(\theta_i, \theta_{iw}) = \frac{\partial}{\partial \theta_i} J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N} \frac{\partial}{\partial \theta_i} V_{i\pi_{\theta_i}}(S_{i0}) \cdot \pi_i(V_i, \theta_{iw}) = \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \frac{\partial \pi_i(g_i|S_i)}{\partial \theta_i} q_{\pi_i}(S_i, g_i) \cdot \pi_i(V_i, \theta_{iw})$$

$$= \underbrace{\sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i|S_i)} \frac{1}{\pi_i(g_i|S_i)} \frac{\partial \pi_i(g_i|S_i)}{\partial \theta_i} q_{\pi_i}(S_i, g_i) \cdot \pi_i(V_i, \theta_{iw})$$

$$= \mathbb{E}_{\underset{\uparrow \text{global, system}}{\pi}} \underline{\frac{\partial}{\partial \theta_i} \{\ln[\pi_i(g_i|S_i)] q_{\pi_i}(S_i, g_i)\} \cdot \pi_i(V_i, \theta_{iw})}$$

↳ item for updating θ_i

$\because \ \theta_i \leftarrow \theta_i + \alpha \nabla_{\theta_i} J(\theta_i, \theta_{iw})$

$\therefore \ \theta_i \leftarrow \theta_i + \alpha\{\nabla_{\theta_i}[\ln \pi_i(g_i|S_i)] q_{it}\} \pi_i(V_i, \theta_{iw})$

$$\nabla_{\theta_{iw}} J(\theta_i, \theta_{iw}) = \frac{\partial}{\partial \theta_w} J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N} V_{i\pi_{\theta_i}}(S_{i0}) \cdot \frac{\partial}{\partial \theta_{iw}} \pi_i(V_i, \theta_{iw}) = \sum_{i=1}^{N} V_{i\pi_{\theta_i}}(S_{i0}) \cdot \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})$$

$$= \underbrace{\sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i|S_i)} q_{\pi_i}(S_i, g_i) \cdot \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})$$

$$= \mathbb{E}_{\pi} [q_{\pi_i}(S_i, g_i) \cdot \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})]$$

$\because \ \theta_{iw} \leftarrow \theta_{iw} + \beta \nabla_{\theta_{iw}} J(\theta_i, \theta_{iw})$

$\therefore \ \theta_{iw} \leftarrow \theta_{iw} + \beta[q_{it} \cdot \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})]$

Dimitri Bertsekas, John Tsitsiklis. Neuro-Dynamic Programming. 1996

s_t: descent direction

$\nabla f(r_t)' s_t < 0$ 下降方向 (' → Transpose?)

Mar. 25, 2019
update April 28.

P89-94-96
Gradient methods

From the first order Taylor expasion: $f(r_{t+1}) = f(r_t) + \gamma_t \nabla f(r_t)' s_t + O(\gamma_t)$ (纯量)

slope of f at r_t along the direction s_t

A techical condition on the direction s_t: $\forall t$,

$$c_1 \|\nabla f(r_t)\|^2 \le -\nabla f(r_t)^T s_t, \quad \|s_t\| \le c_2 \|\nabla f(r_t)\|, \qquad (3.36)$$

c_1: positive scalar; c_2: positive scalar

guarantees that the vectors s_t and $\nabla f(r_t)$ will not become asymptotically orthogonal.

即保证 $\nabla f(r_t)' s_t$ 不为 0.

① proposition 3.4 $\boxed{\text{Constant Stepsize}}$

proof: $g(\xi) = f(r + \xi z)$ (r, z: vector; ξ: scalar parameter) $\quad \frac{d}{d\xi} g(\xi) = z^T \nabla f(r + \xi z)$ (chain rule)

$$f(r+z) - f(r) = f(r + 1z) - f(r + 0z) = g(1) - g(0) = \int_0^1 \frac{d}{d\xi} g(\xi) d\xi = \int_0^1 z^T \nabla f(r + \xi z) d\xi = \int_0^1 d\xi \left[z^T \nabla f(r) + z^T \nabla f(r + \xi z) - z^T \nabla f(r) \right]$$

$$\le \int_0^1 z^T \nabla f(r) d\xi + \left| \int_0^1 z^T (\nabla f(r + \xi z) - \nabla f(r)) d\xi \right| \le z^T \nabla f(r) \int_0^1 d\xi + \int_0^1 \|z\| \cdot \underline{\|\nabla f(r + \xi z) - \nabla f(r)\|} d\xi$$

$$\underline{\|\nabla f(r + \xi z) - \nabla f(r)\|} \le L \cdot \|(r + \xi z) - r\| \qquad (3.37)$$

$$\le z^T \nabla f(r) + \|z\| \int_0^1 \underline{L \xi \|z\|} d\xi$$

∇f 满足 Lipschitz 连续条件.

$$= z^T \nabla f(r) + \|z\| \cdot \|z\| \cdot L \int_0^1 \xi d\xi = z^T \nabla f(r) + \frac{1}{2} L \|z\|^2 \qquad (3.39)$$

$$\Rightarrow f(r_t + \gamma s_t) - f(r_t) \le \nabla f(r_t)^T \gamma s_t + \frac{1}{2} L \|\gamma s_t\|^2 = \underline{\gamma \nabla f(r_t)^T s_t} + \underline{\frac{1}{2} \gamma^2 L \|s_t\|^2}$$

$\boxed{r_t: \text{ segence generated by gradient method } r_{t+1} = r_t + \gamma_t s_t}$

$$\Rightarrow \underline{f(r_t) - f(r_t + \gamma s_t)} \ge \gamma c_1 \|\nabla f(r_t)\|^2 - [c_2 \|\nabla f(r_t)\|]^2 \frac{1}{2} \gamma^2 L = c_1 \gamma \|\nabla f(r_t)\|^2 - \frac{1}{2} \gamma^2 L c_2^2 \|\nabla f(r_t)\|^2$$

$$= \frac{1}{2} \gamma L c_2^2 \left(\underline{\frac{2c_1}{L c_2^2} - \gamma} \right) \|\nabla f(r_t)\|^2 \qquad (3.40)$$

$$\hookrightarrow 0 < \gamma < \frac{2c_1}{L c_2^2} \qquad (3.38)$$

↓

montorically nonincreasing

↓

$f(r_t) \to -\infty$ or finite value, in ②, $f(r_t) - f(r_{t+1}) \to 0$, so, 3.40 implies $\nabla f(r_t) \to 0$.

March 25, 2019
update April 28

② Proposition 3.5 $\boxed{\text{Diminishing Stepsize}}$

(3.39) $f(r+z) - f(r) \le z^T \nabla f(r) + \frac{1}{2}L\|z\|^2$, $\quad z = \gamma_t s_t$

(3.36) $c_1\|\nabla f(r_t)\|^2 \le -\nabla f(r_t)^T s_t$, $\quad \|s_t\| \le c_2 \|\nabla f(r_t)\|$

$$\Rightarrow f(r_{t+1}) \le f(r_t) + (\gamma_t s_t)^T \nabla f(r) + \frac{1}{2}L\|\gamma_t s_t\|^2$$

$$= f(r_t) + \gamma_t\left(\tfrac{1}{2}\gamma_t L\|s_t\|^2 - |\nabla f(r_t)^T s_t|\right)$$

$$\le f(r_t) - \gamma_t\left(c_1 - \tfrac{1}{2}\gamma_t c_2^2 L\right)\|\nabla f(r_t)\|^2$$

↳ $\to 0$

$\boxed{\nabla f(r_t)^T s_t < 0}$ desent direction

$$\Rightarrow f(r_{t+1}) \le f(r_t) - \gamma_t c\|\nabla f(r_t)\|^2 \quad (3.42)$$

$\forall t \ge \bar t$

$c\sum_{t=\bar t}^{\infty} \gamma_t\|\nabla f(r_t)\|^2 \le f(r_{\bar t}) - \lim_{t\to\infty} f(r_t) < \infty$ 矛盾

$\because \sum_{t=0}^{\infty}\gamma_t = \infty$. ~~∴~~ $\boxed{\|\nabla f(r_t)\|^2 > \epsilon}$ $\forall t > \hat t$, should be incorrect

$\Rightarrow \liminf_{t\to\infty} \|\nabla f(r_t)\| = 0$

if $\bar r$ is a limit point of r_t, $f(r_t)$ 收斂于 converges to the finite value $f(\bar r)$. $\Rightarrow \nabla f(r_t) \to 0$ implying that $\nabla f(\bar r) = 0$

(Q.E.D).

反证法证

詳細證明

假设 $\limsup_{t\to\infty} \|\nabla f(r_t)\| > 0$, $\exists\ \epsilon > 0$ let $\|\nabla f(r_t)\| < \frac{1}{2}\epsilon$, $\|\nabla f(r_t)\| > \epsilon$, $i(t) > t$

$\|\nabla f(r_t)\| < \frac{1}{2}\epsilon$, $\|\nabla f(r_{i(t)})\| > \epsilon$,

$\frac{1}{2}\epsilon \le \|\nabla f(r_i)\| \le \epsilon$, $\quad t < i < i(t)$

$\because \|\nabla f(r_{t+1})\| - \|\nabla f(r_t)\| \le \underbrace{\|\nabla f(r_{t+1}) - \nabla f(r_t)\| \le L\|r_{t+1} - r_t\|}_{\text{满足 Lipschitz 連續条件}} = \gamma_t L\|s_t\| \le \underbrace{\gamma_t L c_2}_{<1}\|\nabla f(r_t)\|$,

subset of integers T, $t \in T$

$\because \|s_t\| \le c_2\|\nabla f(r_t)\|$, $\quad \|\nabla f(r) - \nabla f(\bar r)\| \le L\|r - \bar r\|$; Lipschitz condition (3.41)

$\therefore \frac{\epsilon}{2} \le \|\nabla f(r_{i(t)})\| - \|\nabla f(r_t)\| \le \|\nabla f(r_{i(t)}) - \nabla f(r_t)\|$

$\le L\|r_{i(t)} - r_t\| \le L\sum_{i=t}^{i(t)-1}\gamma_i\|s_i\|$

$= Lc_2\sum_{i=t}^{i(t)-1}\gamma_i\|\nabla f(r_i)\| \le Lc_2\epsilon\sum_{i=t}^{i(t)-1}\gamma_i$

$\Leftarrow \boxed{\frac{1}{2Lc_2} \le \sum_{i=t}^{i(t)-1}\gamma_i \quad (3.43)}$ 矛盾

$\forall i = t, t+1, \ldots, i(t)-1$

$\|\nabla f(r_i)\| \ge \frac{\epsilon}{4}$.

$f(r_{i(t)}) \le f(r_t) - c\left(\frac{\epsilon}{4}\right)^2 \underbrace{\sum_{i=t}^{i(t)-1}\gamma_i}_{0}$ 因为①中 $f(r_t)$ 收斂有限值 $\to t\to\infty, t\in T$, $=0$

April 27, 2019

$\hat{q}_{\pi_i}(s_{it}, g_{it})$: local i some unbiased estimator of $q_{\pi_i}(s_{it}, a_{it})$.

q近似 let $f_{x_i}: S \times A \to R$ be our approximation to q_{π_i}, with parameter x_i. It is natural to learn f_{x_i} by following $\pi_i(g_i|s_i)$ and updating x_i by a rule such as

$$\Delta x_{it} \propto \frac{\partial}{\partial x_i}\left[\hat{q}_{\pi_i}(s_{it}, g_{it}) - f_{x_i}(s_{it}, g_{it})\right]^2$$

$$\propto \left[\hat{q}_{\pi_i}(s_{it}, g_{it}) - f_{x_i}(s_{it}, g_{it})\right]\frac{\partial f_{x_i}(s_{it}, g_{it})}{\partial x_i}$$

when such a process has converged to a local optimum, then

$$\sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\pi_i(g_i|s_i)\left[q_{\pi_i}(s_i, g_i) - f_{x_i}(s_i, g_i)\right]\frac{\partial f_{x_i}(s_{it}, g_{it})}{\partial x_i}\cdot \pi_i(V_i, \theta_{iw}) = 0 \qquad (1)$$

f_{x_i} satifies (1) and is compatible with the policy parameterization in the sense that

$$\frac{\partial f_{x_i}(s_i, g_i)}{\partial x_i} = \frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i}\frac{1}{\pi_i(g_i|s_i)} \qquad (2)$$

(1) (2) $\Rightarrow$ $$\sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\pi_i(g_i|s_i)\left[q_{\pi_i}(s_i, g_i) - f_{x_i}(s_i, g_i)\right]\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i}\frac{1}{\pi_i(g_i|s_i)}\cdot \pi_i(V_i, \theta_{iw}) = 0$$

$$\sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i}\left[q_{\pi_i}(s_i, g_i) - f_{x_i}(s_i, g_i)\right]\cdot \pi_i(V_i, \theta_{iw}) = 0 \qquad (3)$$

which tells us that the error in $f_{x_i}(s_i, g_i)$ is orthogonal to the gradiant of the policy parameterization

$$\frac{\partial}{\partial \theta_i} J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i} q_{\pi_i}(s_i, g_i)\cdot \pi_i(V_i, \theta_{iw}) - \sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i}\left[q_{\pi_i}(s_i, g_i) - f_{x_i}(s_i, g_i)\right]\pi_i(V_i, \theta_{iw})$$

$$= \sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i}\left[q_{\pi_i}(s_i, g_i) - q_{\pi_i}(s_i, g_i) + f_{x_i}(s_i, g_i)\right]\pi_i(V_i, \theta_{iw})$$

$$= \sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i} f_{x_i}(s_i, g_i)\,\pi_i(V_i, \theta_{iw})$$

$$= \underbrace{\sum_{i=1}^{N}\sum_{s} d^{\pi_i}(s_i)\sum_{g}\pi_i(g_i|s_i)}_{\mathbb{E}_\pi}\frac{1}{\pi_i(g_i|s_i)}\frac{\partial \pi_i(g_i|s_i)}{\partial \theta_i} f_{x_i}(s_i, g_i)\,\pi_i(V_i, \theta_{iw})$$

$$= \mathbb{E}_\pi \frac{\partial}{\partial \theta_i}\left\{\ln\left[\pi_i(g_i|s_i)\right] f_{x_i}(s_i, g_i)\right\}\cdot \pi_i(V_i, \theta_{iw})$$

global, system

$\boxed{\text{April 28. 2019}}$. Convergence θ_i, θ_{iw}, $J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N} \sum_{s} d^{\pi_i}(s_i) \sum_{g} \pi_i(g_i|s_i)\, q_{\pi_i}(s_i, g_i) \cdot \pi_i(V_i, \theta_{iw})$

from April 19. 2019, we know: $\nabla_{\theta_i} J(\theta_i, \theta_{iw}) = \underbrace{\sum_{i=1}^{N} \sum_{s} d^{\pi_i}(s_i) \sum_{g} \pi_i(g_i|s_i)}_{\mathbb{E}_\pi} \frac{\partial}{\partial \theta_i} \{\ln[\pi_i(g_i|S_i)]\, \underset{\downarrow\, q_{it}}{q_{\pi_i}(s_i, g_i)}\} \cdot \pi_i(V_i, \theta_{iw})$

$$\nabla_{\theta_{iw}} J(\theta_i, \theta_{iw}) = \underbrace{\sum_{i=1}^{N} \sum_{s} d^{\pi_i}(s_i) \sum_{g} \pi_i(g_i|s_i)}_{\mathbb{E}_\pi} [q_{\pi_i}(s_i, g_i) \cdot \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})]$$

Based on Mar. 25. 2019.

Dimitri Bertsekas, John Tsitsiklis, Neuro-Dynamic Programming, 1996 Pg - 94 - 96 Gradient Methods

From the first order Taylor expansion: $J(\theta_{i_{t+1}}, \theta_{iw}) = J(\theta_{i_t}, \theta_{iw}) + \gamma_t \underbrace{\nabla_{\theta_i} J^T(\theta_{i_t}, \theta_{iw})\, S_t}_{\nabla_{\theta_i} J^T(\theta_{i_t}, \theta_{iw}) S_t < 0} + O(\gamma_t)$. $S_t \to$ deseent direction

and

$$J(\theta_i, \theta_{iw_{t+1}}) = J(\theta_i, \theta_{iw_t}) + \gamma_{wt} \underbrace{\nabla_{\theta_{iw}} J^T(\theta_i, \theta_{iw_t}) S_{wt}}_{<0} + O(\gamma_{wt})$$

A technical condition on the direction S_t, S_{wt},

θ_i (3.36) $\begin{cases} c_1 \|\nabla_{\theta_i} J(\theta_{i_t}, \theta_{iw})\|^2 \le -\nabla_{\theta_i} J^T(\theta_{i_t}, \theta_{iw}) S_t \\ \|S_t\| \le c_2 \|\nabla_{\theta_i} J(\theta_{i_t}, \theta_{iw})\| \end{cases}$ θ_{iw} (3.36) $\begin{cases} c_{w1} \|\nabla_{\theta_{iw}} (\theta_i, \theta_{iw_t})\|^2 \le -\nabla_{\theta_{iw}} J^T(\theta_i, \theta_{iw_t}) S_{wt} \\ \|S_{wt}\| \le c_{w2} \|\nabla_{\theta_{iw}} J(\theta_i, \theta_{iw_t})\| \end{cases}$

proposition 3.4

① $\boxed{\text{Constant Stepsize}}$ $\boxed{\theta_i}$ $g(\xi) = J(\theta_i + \xi z, \theta_{iw}) \Rightarrow \frac{d}{d\xi} g(\xi) = z^T \nabla J(\theta_i + \xi z, \theta_{iw})$

$$\begin{aligned}
J(\theta_i + z, \theta_{iw}) - J(\theta_i, \theta_{iw}) &= J(\theta_i + 1z, \theta_{iw}) - J(\theta_i + 0z, \theta_{iw}) \\
&= g(1) - g(0) \\
&= \int_0^1 \frac{d}{d\xi} g(\xi)\, d\xi = \int_0^1 z^T \nabla J(\theta_i + \xi z, \theta_{iw})\, d\xi = \int_0^1 d\xi\, [z^T \nabla J(\theta_i, \theta_{iw}) + z^T \nabla J(\theta_i + \xi z, \theta_{iw}) - z^T \nabla J(\theta_i, \theta_{iw})] \\
&\le \int_0^1 z^T \nabla J(\theta_i, \theta_{iw})\, d\xi + \left| \int_0^1 z^T [\nabla J(\theta_i + \xi z, \theta_{iw}) - \nabla J(\theta_i, \theta_{iw})]\, d\xi \right| \\
&\le z^T \nabla J(\theta_i, \theta_{iw}) \int_0^1 d\xi + \int_0^1 \|z\| \cdot \underbrace{\|\nabla J(\theta_i + \xi z, \theta_{iw}) - \nabla J(\theta_i, \theta_{iw})\|}_{\le L \cdot \|(\theta_i + \xi z) - \theta_i\| \quad (3.37)}\, d\xi
\end{aligned}$$

(cont'd)

let's consider $\nabla_{\theta_i} J(\theta_i , \theta_{iw}) = \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i|S_i) \frac{\partial}{\partial \theta_i} \{\ln[\pi_i(g_i|S_i)]\, q_{\pi_i}(S_i, g_i)\} \cdot \pi_i(V_i, \theta_{iw})$

$\nabla_{\theta_i} J(\theta_i, \theta_{iw})$ Satisfy Lipschitz continuity

$\nabla J(\theta_i + \xi z, \theta_{iw})$ Satisfy Lipschitz continuity. too.

Then,

$$J(\theta_i + z, \theta_{iw}) - J(\theta_i, \theta_{iw}) \le z^T \nabla J(\theta_i, \theta_{iw}) \int_0^1 d\xi + \int_0^1 \|z\| \cdot \|\nabla J(\theta_i + \xi z, \theta_{iw}) - \nabla J(\theta_i, \theta_{iw})\| \, d\xi$$

$$\le z^T \nabla J(\theta_i, \theta_{iw}) + \|z\| \int_0^1 L \cdot \|(\theta_i + \xi z) - \theta_i\| \, d\xi$$

$$= z^T \nabla J(\theta_i, \theta_{iw}) + \|z\| \int_0^1 L \xi \|z\| \, d\xi .$$

$$= z^T \nabla J(\theta_i, \theta_{iw}) + \|z\| \cdot \|z\| \cdot L \int_0^1 \xi \, d\xi = z^T \nabla J(\theta_i, \theta_{iw}) + \frac{1}{2} L \|z\|^2 \quad (3.39)$$

$$\Rightarrow J(\theta_i + z, \theta_{iw}) - J(\theta_i, \theta_{iw}) \le z^T \nabla J(\theta_i, \theta_{iw}) + \frac{1}{2} L \|z\|^2 \quad (3.39)$$

<0 descent direction

$$\hookrightarrow J(\theta_{it} + \gamma S_t, \theta_{iw}) - J(\theta_{it}, \theta_{iw}) \le \nabla J^T(\theta_{it}, \theta_{iw}) \gamma S_t + \frac{1}{2} L \|\gamma S_t\|^2 = \gamma \underbrace{\nabla J^T(\theta_{it}, \theta_{iw}) S_t}_{\le\, -C_1 \|\nabla J(\theta_{it}, \theta_{iw})\|^2} + \frac{1}{2} \gamma^2 L \underbrace{\|S_t\|^2}_{\le\, C_2 \|\nabla J(\theta_{it}, \theta_{iw})\|} \quad (3.36)$$

[θ_{it}: Sequence generated by gradient method $\theta_{i\,t+1} = \theta_{it} + \gamma_t S_t$]

$$J(\theta_{it}, \theta_{iw}) - J(\theta_{it} + \gamma S_t, \theta_{iw}) \ge \gamma C_1 \|\nabla J(\theta_{it}, \theta_{iw})\|^2 - [C_2 \|\nabla J(\theta_{it}, \theta_{iw})\|]^2 \frac{1}{2} \gamma^2 L$$

$$= C_1 \gamma \|\nabla J(\theta_{it}, \theta_{iw})\|^2 - \frac{1}{2} \gamma^2 L C_2 \|\nabla J(\theta_{it}, \theta_{iw})\|^2$$

$$= \frac{1}{2} \gamma L_2^2 \Big(\underbrace{\frac{2C_1}{LC_2^2} - \gamma}_{\ge 0}\Big) \|\nabla J(\theta_{it}, \theta_{iw})\|^2 \quad (3.40)$$

$\hookrightarrow 0 < \gamma < \frac{2C_1}{LC_2^2}$ (3.38)

$\Rightarrow$ $J(\theta_{it}, \theta_{iw})$ is montonically nonincreasing

$\Rightarrow$ $J(\theta_{it}, \theta_{iw}) \to -\infty$ or finite value, in ②, $J(\theta_{it}, \theta_{iw}) \to 0$, so 3.40 implies $\nabla J(\theta_{it}, \theta_{iw}) \to 0$

Proposition 3.5

② Diminishing Stepsize θ_i

(3.39) $\begin{cases} J(\theta_i + z, \theta_{iw}) - J(\theta_i, \theta_{iw}) \leq z^T \nabla J(\theta_i) + \frac{1}{2} L \|z\|^2 \\ z = \gamma_t S_t \end{cases}$

<0 descent direction

(3.36) $\begin{cases} C_1 \|\nabla J(\theta_i, \theta_{iw})\|^2 \leq -\boxed{\nabla J(\theta_i, \theta_{iw})^T S_t} \\ \|S_t\| \leq C_2 \|\nabla J(\theta_i, \theta_{iw})\| \end{cases}$

$\Rightarrow J(\theta_{i_{t+1}}, \theta_{iw}) \leq J(\theta_{it}, \theta_{iw}) + (\gamma_t S_t)^T \nabla J(\theta_i) + \frac{1}{2} L \|\gamma_t S_t\|^2 = J(\theta_{it}, \theta_{iw}) + \gamma_t \left(\frac{1}{2}\gamma_t L \|S_t\|^2 - |\nabla J^T(\theta_{it}, \theta_{iw}) S_t|\right)$

$$\leq J(\theta_{it}, \theta_{iw}) - \gamma_t \underbrace{(C_1 - \tfrac{1}{2}\gamma_t C_2^2 L)}_{>0} \|\nabla J(\theta_{it}, \theta_{iw})\|^2$$

when $t > \bar{t}$

$\Rightarrow J(\theta_{i_{t+1}}, \theta_{iw}) \leq J(\theta_{i_t}, \theta_{iw}) - \gamma_t C \|\nabla J(\theta_{it}, \theta_{iw})\|^2$ (3.42)

$\forall t \geq \bar{t}$, By adding 3.42 over all $t \geq \bar{t}$, we obtain

$$C \sum_{t=\bar{t}}^{\infty} \gamma_t \|\nabla J(\theta_{it}, \theta_{iw})\|^2 \leq J(\theta_{i\bar{t}}, \theta_{iw}) - \lim_{t\to\infty} J(\theta_{it}, \theta_{iw}) < \infty$$

$\Updownarrow$ v.s. contradict

$$\sum_{t=0}^{\infty} \gamma_t = \infty, \quad \|\nabla J(\theta_{it}, \theta_{iw})\|^2 > \epsilon \quad \text{for all } t \geq \bar{t}$$

$\Rightarrow \liminf_{t\to\infty} \|\nabla J(\theta_{it}, \theta_{iw})\| = 0$

Then, to show $\lim_{t\to\infty} \nabla J(\theta_{it}, \theta_{iw}) = 0$ 反证法.

Assume the contrary. that is. $\limsup_{t\to\infty} \|\nabla J(\theta_{it}, \theta_{iw})\| > 0$.

Then, $\exists\, \epsilon > 0$, let $\begin{cases} \|\nabla J(\theta_{it}, \theta_{iw})\| < \frac{\epsilon}{2}, \text{ for inf many } t \\ \|\nabla J(\theta_{it}, \theta_{iw})\| > \epsilon, \text{ for inf many } t \\ \frac{\epsilon}{2} \leq \|\nabla J(\theta_{it}, \theta_{iw})\| < \epsilon, \text{ if } t < i < i(t) \end{cases}$

Since $\|\nabla J(\theta_{i_{t+1}}, \theta_{iw})\| - \|\nabla J(\theta_{i_t}, \theta_{iw})\| \leq \underbrace{\|\nabla J(\theta_{i_{t+1}}, \theta_{iw}) - \nabla J(\theta_{i_t}, \theta_{iw})\| \leq L \|\theta_{i_{t+1}} - \theta_{i_t}\|}_{\text{Lipschitz continuity}}$

$$= \gamma_t L \|S_t\| \leq \underbrace{\gamma_t}_{\text{all } t \in \mathcal{T} \text{ so that } <1} L C_2 \|\nabla J(\theta_{it}, \theta_{iw})\|$$

当 $t \in \mathcal{T}$ 时, 可使:

$\Rightarrow$ 也就是说 $\|\nabla J(\theta_{i_{t+1}}, \theta_{iw})\| \leq 2 \|\nabla J(\theta_{it}, \theta_{iw})\|$

$\because$ $\left\{ \begin{array}{l} \|S_t\| \le C_2 \|\nabla J(\theta_{i_t}, \theta_{iw})\| \\ \|\nabla J(\theta_i, \theta_{iw}) - \nabla J(\bar{\theta}_i, \theta_{iw})\| \le L \|\theta_i - \bar{\theta}_i\| \end{array} \right.$; Lipschitz condition (3.41) $\}$

$$\therefore \frac{\epsilon}{2} \le \|\nabla J(\theta_{i(t)}, \theta_{iw})\| - \|\nabla J(\theta_i, \theta_{iw})\| \le \|\nabla J(\theta_{i(t)}, \theta_{iw}) - \nabla J(\theta_{i_t}, \theta_{iw})\|$$
$$\le L \|\theta_{i(t)} - \theta_{i_t}\| \le L \sum_{i=t}^{i(t)-1} \gamma_i \|S_i\|$$
$$= L C_2 \sum_{i=t}^{i(t)-1} \gamma_i \|\nabla J(\theta_{i(t)}, \theta_{iw})\| \le L C_2 \epsilon \sum_{i=t}^{i(t)-1} \gamma_i$$

$\Rightarrow \frac{1}{2LC_2} \le \sum_{i=t}^{i(t)-1} \gamma_i$ (3.43) 即 $\sum_{i=1}^{i(t)-1} \gamma_i \ge \frac{1}{2LC_2} > 0$

$\left. \begin{array}{l} \Rightarrow \|\nabla J(\theta_{i_{t+1}}, \theta_{iw})\| \le 2 \|\nabla J(\theta_{i_t}, \theta_{iw})\| \\ \Rightarrow \|\nabla J(\theta_{i_{t+1}}, \theta_{iw})\| \ge \frac{\epsilon}{2} \end{array} \right\} \Rightarrow \|\nabla J(\theta_{i_t}, \theta_{iw})\| \ge \frac{\epsilon}{4}$ for $i = t, t+1, \cdots, i(t)-1$.

contradicting. 矛盾

(3.42) $J(\theta_{i_{t+1}}, \theta_{iw}) \le J(\theta_{i_t}, \theta_{iw}) - \gamma_t C \|\nabla J(\theta_{i_t}, \theta_{iw})\|^2$

$\left. \begin{array}{l} J(\theta_{i(t)}, \theta_{iw}) \le J(\theta_{i_t}, \theta_{iw}) - C\left(\frac{\epsilon}{4}\right)^2 \sum_{i=t}^{i(t)-1} \gamma_i \\ \because \text{①}. J(\theta_{i_t}, \theta_{iw}) \text{ converge to a finite value,} \end{array} \right\} \Rightarrow \lim_{\substack{t\to\infty, \\ t\in\tau}} \sum_{i=t}^{i(t)-1} \gamma_i = 0$

So, $\lim_{t\to\infty} \nabla J(\theta_{i_t}, \theta_{iw}) = 0$.

Finally, if $\bar{\theta}_i$ is a limit point of θ_{i_t}, then $J(\theta_{i_t}, \theta_{iw})$ converges to the finite value $J(\bar{\theta}_i, \theta_{iw})$.

Thus we have $\nabla J(\theta_{i_t}, \theta_{iw}) \to 0$, implying that $\nabla J(\bar{\theta}_i, \theta_{iw}) = 0$. Q.E.D.

(3.36.) θ_i
$$\begin{cases} C_1 \| \nabla_{\theta_i} J(\theta_{i_t}, \theta_{iw}) \|^2 \le -\nabla_{\theta_i} J^T(\theta_{i_t}, \theta_{iw}) S_t \\ \| S_t \| \le C_2 \| \nabla_{\theta_i} J(\theta_{i_t}, \theta_{iw}) \| \end{cases}$$

$$\nabla_{\theta_i} J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) \frac{\partial}{\partial \theta_i} \{ \ln[\pi_i(g_i | S_i)] \, q_{\pi_i}(S_i, g_i) \} \cdot \pi_i(V_i, \theta_{iw})$$

technical condition :
$$\begin{cases} C_1 \| \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) \frac{\partial}{\partial \theta_{it}} \{ \ln[\pi_i(g_i | S_i)] \, q_{\pi_i}(S_i, g_i) \} \cdot \pi_i(V_i, \theta_{iw}) \|^2 \le \\ - \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) \frac{\partial}{\partial \theta_{it}} \{ \ln[\pi_i(g_i | S_i)] \, q_{\pi_i}(S_i, g_i) \} \cdot \pi_i(V_i, \theta_{iw}) \cdot S_t \\ \| S_t \| \le C_2 \| \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) \frac{\partial}{\partial \theta_{it}} \{ \ln[\pi_i(g_i | S_i)] \, q_{\pi_i}(S_i, g_i) \} \cdot \pi_i(V_i, \theta_{iw}) \| \end{cases}$$

(3.36) θ_{iw}
$$\begin{cases} C_{w1} \| \nabla_{\theta_{iw}} (\theta_i, \theta_{iw_t}) \|^2 \le -\nabla_{\theta_{iw}} J^T(\theta_i, \theta_{iw_t}) S_{wt} \\ \| S_{wt} \| \le C_{w2} \| \nabla_{\theta_{iw}} J(\theta_i, \theta_{iw_t}) \| \end{cases}$$

$$\nabla_{\theta_{iw}} J(\theta_i, \theta_{iw}) = \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) [q_{\pi_i}(S_i, g_i) \nabla_{\theta_{iw}} \pi_i(V_i, \theta_{iw})]$$

technical condition :
$$\begin{cases} C_{w1} \| \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) [q_{\pi_i}(S_i, g_i) \nabla_{\theta_{iwt}} \pi_i(V_i, \theta_{iwt})] \|^2 \le \\ - \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) [q_{\pi}(S_i, g_i) \nabla_{\theta_{iwt}} \pi_i(V_i, \theta_{iwt})] S_{wt} \\ \| S_{wt} \| \le C_{w2} \| \sum_{i=1}^{N} \sum_{S} d^{\pi_i}(S_i) \sum_{g} \pi_i(g_i | S_i) [q_{\pi_i}(S_i, g_i) \nabla_{\theta_{iwt}} \pi_i(V_i, \theta_{iw})] \| \end{cases}$$

For $\boxed{\theta_{iw}}$, $\lim_{t \to \infty} \nabla J(\theta_i, \theta_{iwt}) = 0$ proof is similar to $\boxed{\theta_i}$. $\lim_{t \to \infty} \nabla J(\theta_{it}, \theta_{iw}) = 0$ when $\begin{cases} \theta_i \text{ or } \theta_{iw} \text{ satifies } (3.36) \\ \nabla J(\theta_i, \theta_{iw}) \text{ satifies Lipschitz continuity} \end{cases}$