SUPPLEMENTARY MATERIALS: ITERATION COMPLEXITY OF A SECOND-ORDER AUGMENTED LAGRANGIAN METHOD FOR NONCONVEX UNIT SIMPLEX CONSTRAINED OPTIMIZATION*

HONG ZHU[†] AND MICHAEL K. NG[‡]

SM1. Proof of Lemma 4.3.

Proof. Denote $J_{t1}^+ = \{i \in J_t^+ | 0 \le w_{\cdot t}^i \le \hat{\varepsilon}\}$ and $J_{t2}^+ = \{i \in J_t^+ | 1-\hat{\varepsilon} \le w_{\cdot t}^i \le 1\}$.

If $i \in J_{t1}^+$ and $g_t^i < -\hat{\varepsilon}$, then $\frac{1-w_{\cdot t}^i}{-g_t^i} \ge \frac{1}{2U_g^k}$. For any $\beta \in (0, \frac{1}{2U_g^k}]$, we have

$$(w_{\cdot t}^i - P(w_{\cdot t}^i - \beta g_t^i))g_t^i = \beta (g_t^i)^2 > \beta \hat{\varepsilon}^2. \tag{SM1.1}$$

If $i \in J_{t2}^+$ and $g_t^i > \hat{\varepsilon}$, then $\frac{w_{\cdot t}^i}{g_t^i} \ge \frac{1}{2U_g^k}$. For any $\beta \in (0, \frac{1}{2U_g^k}]$, we have

$$(w_{\cdot t}^i - P(w_{\cdot t}^i - \beta g_t^i))g_t^i = \beta (g_t^i)^2 > \beta \hat{\varepsilon}^2. \tag{SM1.2}$$

If $\|S_t^+ g_t^+\| > \hat{\varepsilon}^2$, then for any $\beta > 0$, we have

$$\begin{aligned}
\hat{\varepsilon}^4 &< \sum_{i \in J_{t1}^+} (w_{\cdot t}^i)^2 (g_t^i)^2 + \sum_{i \in J_{t2}^+} (1-w_{\cdot t}^i)^2 (g_t^i)^2 \\
&= \sum_{i \in J_{t1}^+, w_{\cdot t}^i - 1 \le \beta g_t^i \le w_{\cdot t}^i} (w_{\cdot t}^i)^2 (g_t^i)^2 + \sum_{i \in J_{t1}^+, \beta g_t^i > w_{\cdot t}^i} (w_{\cdot t}^i)^2 (g_t^i)^2 + \sum_{i \in J_{t1}^+, \beta g_t^i < w_{\cdot t}^i - 1} (w_{\cdot t}^i)^2 (g_t^i)^2 \\
&\quad + \sum_{i \in J_{t2}^+, w_{\cdot t}^i - 1 \le \beta g_t^i \le w_{\cdot t}^i} (1-w_{\cdot t}^i)^2 (g_t^i)^2 + \sum_{i \in J_{t2}^+, \beta g_t^i > w_{\cdot t}^i} (1-w_{\cdot t}^i)^2 (g_t^i)^2 \\
&\quad + \sum_{i \in J_{t2}^+, \beta g_t^i < w_{\cdot t}^i - 1} (1-w_{\cdot t}^i)^2 (g_t^i)^2.
\end{aligned}$$

Therefore, at least one of four cases occurs.

i). $\sum_{i \in J_{t1}^+, w_{\cdot t}^i - 1 \le \beta g_t^i \le w_{\cdot t}^i} (w_{\cdot t}^i)^2 (g_t^i)^2 \ge \hat{\varepsilon}^4/6$, combine with $w_{\cdot t}^i \le \hat{\varepsilon}$, $\forall i \in J_{t1}^+$, we get

$$\sum_{i \in J_{t1}^+, -(1-w_{\cdot t}^i) \le \beta g_t^i \le w_{\cdot t}^i} (g_t^i)^2 \ge \hat{\varepsilon}^2/6;$$

ii). $\sum_{i \in J_{t1}^+, \beta g_t^i > w_{\cdot t}^i} (w_{\cdot t}^i)^2 (g_t^i)^2 \ge \hat{\varepsilon}^4/6$, which leads

$$\sum_{i \in J_{t1}^+, \beta g_t^i > w_{\cdot t}^i} w_{\cdot t}^i g_t^i \ge \hat{\varepsilon}^2/\sqrt{6};$$

iii). $\sum_{i \in J_{t1}^+, \beta g_t^i < w_{\cdot t}^i - 1} (w_{\cdot t}^i)^2 (g_t^i)^2 \ge \hat{\varepsilon}^4/6$, combine with $w_{\cdot t}^i < 1 - w_{\cdot t}^i$, we have

$$\sum_{i \in J_{t1}^+, \beta g_t^i < w_{\cdot t}^i - 1} -(1-w_{\cdot t}^i) g_t^i \ge \hat{\varepsilon}^2/\sqrt{6};$$

*Submitted to the editors DATE.

Funding: The first author's research was partially supported by NSF of China grant NSFC11971149, 12271217. M.K. Ng's research is partially supported by the HKRGC GRF 17201020, 17300021, C1013-21GF, C7004-21GF and Joint NSFC-RGC N-HKU76921.

[†]School of Mathematical Sciences, Jiangsu University, Xuefu Road, Zhenjiang, 2120123, Jiangsu, China (zhuhongmath@126.com).

[‡]Department of Mathematics, HongKong Baptist University, Kowloon City, Hong Kong, 999077, China (michael-ng@hkbu.edu.hk).

iv). $\sum_{i\in J_{t2}^+, w_{\cdot t}^i-1\le \beta g_t^i\le w_{\cdot t}^i}(1-w_{\cdot t}^i)^2(g_t^i)^2\ge \hat{\varepsilon}^4/6$, combine with $1-w_{\cdot t}^i\le\hat{\varepsilon}$, we get

$$\sum_{i\in J_{t2}^+,\beta g_t^i\le w_{\cdot t}^i}(g_t^i)^2\ge\hat{\varepsilon}^2/6;$$

v). $\sum_{i\in J_{t2}^+,\beta g_t^i> w_{\cdot t}^i}(1-w_{\cdot t}^i)^2(g_t^i)^2\ge \hat{\varepsilon}^4/6$, combine with $1-w_{\cdot t}^i< w_{\cdot t}^i$, we have

$$\sum_{i\in J_{t2}^+,\beta g_t^i> w_{\cdot t}^i} w_{\cdot t}^i g_t^i\ge\hat{\varepsilon}^2/\sqrt{6}.$$

vi). $\sum_{i\in J_{t2}^+,\beta g_t^i< w_{\cdot t}^i-1}(1-w_{\cdot t}^i)^2(g_t^i)^2\ge \hat{\varepsilon}^4/6$, we get

$$\sum_{i\in J_{t2}^+,\beta g_t^i< w_{\cdot t}^i-1}(w_{\cdot t}^i-1)g_t^i\ge\hat{\varepsilon}^2/\sqrt{6}.$$

Therefore, for any $\beta>0$, we have

$$\begin{aligned}
&\sum_{i\in J_t^+}(w_{\cdot t}^i-P(w_{\cdot t}^i-\beta g_t^i))g_t^i\\
=&\sum_{i\in J_{t1}^+,w_{\cdot t}^i-1\le\beta(g_t^i)^2\le w_{\cdot t}^i}\beta g_t^i+\sum_{i\in J_{t1}^+,\beta g_t^i>w_{\cdot t}^i}w_{\cdot t}^i g_t^i+\sum_{i\in J_{t1}^+,\beta g_t^i<w_{\cdot t}^i-1}(w_{\cdot t}^i-1)g_t^i\\
&+\sum_{i\in J_{t2}^+,w_{\cdot t}^i-1\le\beta(g_t^i)^2\le w_{\cdot t}^i}\beta g_t^i+\sum_{i\in J_{t2}^+,\beta g_t^i>w_{\cdot t}^i}w_{\cdot t}^i g_t^i+\sum_{i\in J_{t2}^+,\beta g_t^i<w_{\cdot t}^i-1}(w_{\cdot t}^i-1)g_t^i\\
\ge&\min\{\beta/6,1/\sqrt{6}\}\hat{\varepsilon}^2.
\end{aligned}\tag{SM1.3}$$

Notice that $g_t^i(w_{\cdot t}^i-P(w_{\cdot t}^i-\beta g_t^i))\ge 0$. For any $\beta\in(0,\frac{1}{2U_g^k}]$, we have

$$\begin{aligned}
g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\beta g_t))=\sum_{i=1}^n g_t^i(w_{\cdot t}^i-P(w_{\cdot t}^i-\beta g_t^i))&\ge\sum_{i\in J_t^+}g_t^i(w_{\cdot t}^i-P(w_{\cdot t}^i-\beta g_t^i))\\
&\overset{(??)-(??)}{\ge}\min\{\beta/6,1/\sqrt{6}\}\hat{\varepsilon}^2.
\end{aligned}$$

By (4.2a), for any $\beta\in(0,\frac{1}{l_g^k}]$, we have

$$\begin{aligned}
l(P(w_{\cdot t}-\beta g_t))\le&l(w_{\cdot t})+g_t^\top(P(w_{\cdot t}-\beta g_t)-w_{\cdot t})+\frac{l_g^k}{2}\|P(w_{\cdot t}-\beta g_t)-w_{\cdot t}\|^2\\
\le&l(w_{\cdot t})-g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\beta g_t))+\frac{l_g^k}{2}\beta g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\beta g_t))\\
\le&l(w_{\cdot t})-\frac{1}{2}g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\beta g_t)).
\end{aligned}$$

Therefore, if $\hat{\varepsilon}\le\min\{\frac{1}{l_g^k},\frac{2}{U_g^k},1\}$, we can set $\beta\ge\hat{\varepsilon}$ in **(??)** -(**??**), which yields

$$g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\beta g_t))\ge\frac{1}{6}\hat{\varepsilon}^3.$$

Combine with the line search rule in Algorithm 4.1, we have $\theta^{\tilde{m}_t}\ge\frac{\theta}{6}$ and

$$l(w_{\cdot t})-l(P(w_{\cdot t}-\theta^{\tilde{m}_t}g_t))\ge\frac{1}{2}g_t^\top(w_{\cdot t}-P(w_{\cdot t}-\theta^{\tilde{m}_t}g_t))\ge\frac{\theta}{12}\hat{\varepsilon}^3.$$

SM2. Proof of Lemma 4.4.

Proof. Notice that for $J_t^- \neq \emptyset$ and $\|g_t^-\| > \hat{\varepsilon}$, Algorithm 4.2 will be invoked and $d_t^+ = 0$.

If $\alpha\|d_t\| \leq \hat{\varepsilon}$ for some $\alpha > 0$, then $P(w_{\cdot t} + \alpha d_t) = w_{\cdot t} + \alpha d_t$ and

$$l(P(w_{\cdot t} + \alpha d_t)) = l(w_{\cdot t} + \alpha d_t) \overset{(4.2c)}{\leq} l(w_{\cdot t}) + \alpha g_t^\top d_t + \frac{\alpha^2}{2} d_t^\top H_t d_t + \frac{l_h^k \alpha^3}{6}\|d_t\|^3$$

$$= l(w_{\cdot t}) + \alpha (g_t^-)^\top d_t^- + \frac{\alpha^2}{2}(d_t^-)^\top H_t^- d_t^- + \frac{l_h^k \alpha^3}{6}\|d_t^-\|^3. \tag{SM2.1}$$

(a) Since $d_t^- = -\|g_t^-\|^{-1/2} g_t^-$, we have $\|d_t^-\| = \|g_t^-\|^{\frac{1}{2}} \geq \hat{\varepsilon}$, $\|d_t^-\| = \|g_t^-\|^{\frac{1}{2}} \leq u_g^{\frac{1}{2}}$, $(d_t^-)^\top g_t^- = -\|g_t^-\|^{3/2} = -\|d_t^-\|^3 \leq 0$ and

$$|(d_t^-)^\top H_t d_t^-| = \frac{1}{\|g_t^-\|_2}|(g_t^-)^\top H_t g_t^-| \leq \hat{\varepsilon}\|g_t^-\| = \hat{\varepsilon}\|d_t^-\|^2.$$

Combine with (**??**), if $0 < \alpha \leq \min\{\frac{\hat{\varepsilon}}{\|d_t\|}, \frac{-(\frac{1}{2}+\eta)+\sqrt{(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k\|d_t^-\|^2/\hat{\varepsilon}^2}}{\frac{1}{3}l_h^k}\frac{\hat{\varepsilon}}{\|d_t^-\|}\}$, we have

$$l(P(w_{\cdot t} + \alpha d_t)) \leq l(w_{\cdot t}) - \alpha\|d_t^-\|^3 + \frac{\alpha^2}{2}\hat{\varepsilon}\|d_t^-\|^2 + \frac{l_h^k\alpha^3}{6}\|d_t^-\|^3 \leq l(w_{\cdot t}) - \eta\alpha^2\hat{\varepsilon}\|d_t^-\|^2.$$

Since backtracking is used to get α_t, we have

$$\begin{aligned}
\alpha_t \geq& \min\{\theta\frac{\hat{\varepsilon}}{\|d_t^-\|}\min\{1, \frac{-(\frac{1}{2}+\eta)+\sqrt{(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k\|d_t^-\|^2/\hat{\varepsilon}^2}}{\frac{1}{3}l_h^k}\}, 1\} \\
\geq& \min\{\theta\frac{\hat{\varepsilon}}{u_g^{1/2}}\min\{1, \frac{-(\frac{1}{2}+\eta)+\sqrt{(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k}}{\frac{1}{3}l_h^k}\}, 1\} \\
=& \min\{\theta\frac{\hat{\varepsilon}}{u_g^{1/2}}\min\{1, \frac{2}{(\frac{1}{2}+\eta)+\sqrt{(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k}}\}, 1\}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
l(P(w_{\cdot t} + \alpha d_t)) \leq& l(w_{\cdot t}) - \eta\alpha^2\hat{\varepsilon}\|d_t\|^2 \\
\leq& l(w_{\cdot t}) - \eta\theta^2\min\{1, \frac{4}{((\frac{1}{2}+\eta)+\sqrt{(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k})^2}\}\hat{\varepsilon}^3
\end{aligned}$$

and $m_t \leq \max\{\log_\theta(\theta\hat{\varepsilon}/\sqrt{u_g}), \log_\theta(2\theta\varepsilon_h/(\sqrt{u_g}(\frac{1}{2}+\eta)+\sqrt{u_g(\frac{1}{2}+\eta)^2+\frac{2}{3}l_h^k u_g}))\}$ by noting that $m_t = \log_\theta \alpha_t$. This completes the proof of statement (a).

(b) Define

$$\begin{aligned}
s_t \triangleq& \min\{s \in \mathbb{N} | \theta^s\|d_t\| \leq \hat{\varepsilon}\} \\
j_t \triangleq& \min\{j \geq s_t,\ j \in \mathbb{N} | \theta^j g_t^\top d_t + \frac{\theta^{2j}}{2}d_t^\top H d_t + \frac{l_h^k\theta^{3j}}{6}\|d_t\|^3 \leq -\eta\theta^{2j}\hat{\varepsilon}\|d_t\|^2\}.
\end{aligned}$$

Then by the definition of s_t and j_t, we have

$$\begin{aligned}
l(P(w_{\cdot t} + \theta^{j_t}d_t)) =& l(w_{\cdot t} + \theta^{j_t}d_t) \leq l(w_{\cdot t}) + \theta^{j_t}g_t^\top d_t + \frac{\theta^{2j_t}}{2}d_t^\top H d_t + \frac{l_h^k\theta^{3j}}{6}\|d_t\|^3 \\
<& l(w_{\cdot t}) - \eta\theta^{2j}\hat{\varepsilon}\|d_t\|^2.
\end{aligned}$$

Recall the line search condition for this case (see Algorithm (4.2)), we known that $m_t \le j_t$. Notice that

$$\|d_t\| = \|d_t^-\| \overset{(4.5b)}{\le} \gamma\hat{\varepsilon}^{-1}\|g_t^-\| \le \gamma\hat{\varepsilon}^{-1}\|g_t\| \le \gamma\hat{\varepsilon}^{-1}U_g^k, \tag{SM2.2}$$

thus

$$s_t \le \left[\log_\theta\left(\frac{\hat{\varepsilon}}{\|d_t\|}\right)\right]_+ + 1 \le \left[\log_\theta\left(\frac{\hat{\varepsilon}^2}{\gamma U_g^k}\right)\right]_+ + 1.$$

Notice that for any $j \in \mathbb{N}$, by the definition of $\hat{r}_t^-$, we have

$$\begin{aligned}
&\theta^j(g_t^-)^\top d_t^- + \frac{\theta^{2j}}{2}(d_t^-)^\top H^- d_t^- + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3\\
=&\theta^j(\hat{r}_t^- - (H^- + \tau_t\|g_t^-\|^\delta I)d_t^-)^\top d_t^- + \frac{\theta^{2j}}{2}(d_t^-)^\top H^- d_t^- + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3\\
\le&-\theta^j(1-\frac{\theta^j}{2})(d_t^-)^\top(H^- + \tau_t\|g_t^-\|^\delta I)d_t^- - \hat{\varepsilon}\theta^{2j}\|d_t^-\|^2 + \theta^j(\hat{r}_t^-)^\top d_t^- + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3\\
\overset{(4.5a)}{\le}&-\theta^j(1-\frac{\theta^j}{2})\hat{\varepsilon}\|d_t^-\|^2 + \theta^j\|\hat{r}_t^-\|\|d_t^-\| + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3\\
\overset{(4.5c)}{\le}&-\frac{\theta^j}{2}\hat{\varepsilon}\|d_t^-\|^2 + \frac{\theta^j}{2}\hat{\varepsilon}\zeta\|d_t^-\|^2 + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3\\
=&-\frac{\theta^j}{2}(1-\zeta)\hat{\varepsilon}\|d_t^-\|^2 + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3,
\end{aligned} \tag{SM2.3}$$

where the first inequality follows from $\tau_t \ge \frac{2\hat{\varepsilon}}{\|g_t^-\|^\delta}$, which leads to $\frac{\tau_t}{2}\theta^{2j}\|g_t^-\|^\delta\|d_t^-\|^\delta \ge \hat{\varepsilon}\theta^{2j}\|d_t^-\|^2$.

It is easy to check that when $j > \log_\theta\left(\frac{1-\zeta}{\eta+\sqrt{\eta^2+\frac{2-\zeta}{6}\frac{\gamma U_g^k l_h^k}{\hat{\varepsilon}^2}}}\right)$, we have

$$\theta^j \le \frac{1-\zeta}{\eta+\sqrt{\eta^2+\frac{1-\zeta}{3}\frac{\gamma U_g^k l_h^k}{\hat{\varepsilon}^2}}} \overset{(C.1)}{\le} \frac{1-\zeta}{\eta+\sqrt{\eta^2+\frac{1-\zeta}{3}\frac{\|d_t^-\| l_h^k}{\hat{\varepsilon}}}}$$

and

$$-\frac{\theta^j}{2}(1-\zeta)\hat{\varepsilon}\|d_t^-\|^2 + \frac{l_h^k\theta^{3j}}{6}\|d_t^-\|^3 \le -\eta\theta^{2j}\hat{\varepsilon}\|d_t^-\|^2.$$

Combine with (C.2), recall the definition of s_t and j_t, we have

$$j_t \le \min\{\left[\log_\theta\left(\frac{\hat{\varepsilon}^2}{\gamma U_g^k}\right)\right]_+, \left[\log_\theta\left(\frac{1-\zeta}{\eta+\sqrt{\eta^2+\frac{2-\zeta}{6}\frac{\gamma U_g^k l_h^k}{\hat{\varepsilon}^2}}}\right)\right]_+\} + 1,$$

which is also an upper bound for m_t.

To get the lower bound for $\theta^{2m_t}\hat{\varepsilon}\|d_t\|^2$, we consider the following four cases.

Case 1. $j_t = s_t = 0$. In this case we have $m_k = 0$, $\alpha_t = 1$ and $\|d_t^-\| = \|d_t\| \le \hat{\varepsilon}$.

Therefore, $P(w_{\cdot t} + \alpha_t d_t) = w_{\cdot t} + \alpha_t d_t$. Then we have

$$\begin{aligned}
\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\| =& \|\nabla l(w_{\cdot t} + d_t)\big|_{J_t^-}\| \\
=& \|\nabla l(w_{\cdot t} + d_t)\big|_{J_t^-} - g_t^- + r_t^- - (H^- + \tau_t\|g_t^-\|^\delta I)d_t^-\| \\
\overset{(4.2b)}{\leq}& \frac{l_h^k}{2}\|d_t^-\|^2 + \tau_t\|g_t^-\|^\delta\|d_t^-\| + \|r_t^-\| \\
\overset{(4.5c)}{\leq}& \frac{l_h^k}{2}\|d_t^-\|^2 + \frac{4\hat{\tau}+\zeta}{2}\hat{\varepsilon}\|d_t^-\|,
\end{aligned}$$

where the second equality follows from the definition of $\hat{r}_t$. By solving the quadratic equation w.r.t. $\|d_k^-\|$, we have

$$\begin{aligned}
\|d_k^-\| \geq& \left(-(\frac{1}{2}\zeta + 2\hat{\tau}) + \sqrt{(\frac{1}{2}\zeta + 2\hat{\tau})^2 + 2l_h^k\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|/\hat{\varepsilon}^2}\right)\frac{\hat{\varepsilon}}{l_h^k} \\
\geq& \left(-(\frac{1}{2}\zeta + 2\hat{\tau}) + \sqrt{(\frac{1}{2}\zeta + 2\hat{\tau})^2 + 2l_h^k\min\{\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|/\hat{\varepsilon}^2, 1\}}\right)\frac{\hat{\varepsilon}}{l_h^k} \\
=& \frac{2\min\{\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|/\hat{\varepsilon}^{-1}, \hat{\varepsilon}\}}{\sqrt{(\frac{1}{2}\zeta + 2\hat{\tau})^2 + 2l_h^k\min\{\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|/\hat{\varepsilon}^2, 1\}} + (\frac{1}{2}\zeta + 2\hat{\tau})} \\
\geq& \frac{4}{\sqrt{(\zeta + 4\hat{\tau})^2 + 8l_h^k} + (\zeta + 4\hat{\tau})}\min\{\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|/\hat{\varepsilon}^{-1}, \hat{\varepsilon}\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&l(P(w_{\cdot t} + \alpha_t d_t)) \leq l(w_{\cdot t}) - \eta\alpha_t^2\hat{\varepsilon}\|d_t\|^2 \\
\leq& l(w_{\cdot t}) - \eta\frac{16}{(\sqrt{(\zeta + 4\hat{\tau})^2 + 8l_h^k} + (\zeta + 4))^2}\min\{\|\nabla l(P(w_{\cdot t} + \alpha_t d_t))\big|_{J_t^-}\|^2/\hat{\varepsilon}^{-1}, \hat{\varepsilon}^3\}.
\end{aligned}$$

Case 2. $j_t = s_t \geq 1$. By the definition of s_t, we have $\theta^{s_t}\|d_t\| > \theta\hat{\varepsilon}$. Since $\alpha_t = \theta^{m_t}$ with $m_t \leq j_t = s_t$, we have

$$\alpha_t^2\hat{\varepsilon}\|d_t\|^2 = (\theta^{m_t}\|d_t\|)^2\hat{\varepsilon} \geq (\theta^{s_t}\|d_t\|)^2\hat{\varepsilon} \geq \theta^2\hat{\varepsilon}^3.$$

Case 3. $j_t > l_t = 1$. By the definition of j_t, we know that $j = 0$ and $j = j_t - 1$ does not satisfies line search condition. Combine with (C.2) and $\|d_t\| = \|d_t^-\|$, we have

$$-\eta\hat{\varepsilon}\|d_t\|^2 \leq -\frac{1}{2}(1-\zeta)\hat{\varepsilon}\|d_t\|^2 + \frac{l_h^k}{6}\|d_t\|^3 \tag{SM2.4}$$

and

$$-\eta\theta^{2(j_t-1)}\hat{\varepsilon}\|d_t\|^2 \leq -\frac{\theta^{j_t-1}}{2}(1-\zeta)\hat{\varepsilon}\|d_t\|^2 + \frac{l_h^k\theta^{3(j_t-1)}}{6}\|d_t\|^3. \tag{SM2.5}$$

By (C.3a), we have

$$\|d_t\| \geq \frac{3\hat{\varepsilon}}{l_h^k}(1-\zeta-2\eta).$$

By (C.3b), we have

$$\frac{1-\zeta}{2}\hat{\varepsilon} \le \frac{l_h^k \theta^{2(j_t-1)}}{6}\|d_t\| + \eta\hat{\varepsilon}\theta^{j_t-1} \le (\frac{l_h^k}{6} + \frac{\eta\hat{\varepsilon}}{\|d_t^-\|})\theta^{j_t-1}\|d_t\|,$$

which yields

$$\theta^{j_t-1}\|d_t\| \ge \frac{(1-\zeta)\hat{\varepsilon}}{l_h^k/3 + 2\eta\hat{\varepsilon}/\|d_t\|}. \tag{SM2.6}$$

Hence, $\theta^{j_t-1}\|d_t\| \ge \frac{3(1-\zeta-2\eta)\hat{\varepsilon}}{l_h^k}$. Therefore,

$$l(P(w_{\cdot t} + \alpha_t d_t)) \le l(w_{\cdot t}) - \eta\alpha_t^2\hat{\varepsilon}\|d_t\|^2 \le l(w_{\cdot t}) - \eta\theta^2\frac{9(1-\zeta-2\eta)^2}{(l_h^k)^2}\hat{\varepsilon}^3.$$

Case 4. $j_t > s_t \ge 1$. By the definition of s_t, we have $\|d_t\| > \hat{\varepsilon}$. Combine with (C.4), we have

$$\theta^{j_t-1}\|d_t\| \ge \frac{(1-\zeta)\hat{\varepsilon}}{l_h^k/3 + 2\eta}.$$

Therefore,

$$l(P(w_{\cdot t} + \alpha_t d_t)) \le l(w_{\cdot t}) - \eta\alpha_t^2\hat{\varepsilon}\|d_t\|^2 \le l(w_{\cdot t}) - \eta\theta^2\frac{9(1-\zeta)^2}{(l_h^k + 6\eta)^2}\hat{\varepsilon}^3.$$

(c) If d_t^- is an approximate solution of the regularized Newton equation, then the proof of statement (c) is exactly the same as the proof of statement (b) of this Lemma. If d_t^- is obtained from the negative curvature of H^-, then d_t^- is a negative curvature of H^- and . the proof of statement (c) is exactly the same as the proof of statement (b) of this Lemma. ∎