$$
\begin{aligned}
||\omega^f_{mT}-\omega^*_{mT}|| \leq& ||\omega^f_{mT}-\omega^c_{mT}+\omega^c_{mT}-\omega^*_{mT}|| \\
\leq& ||\omega^f_{mT}-\omega^c_{mT}||+||\omega^c_{mT}-\omega^*_{mT}||
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
||\omega^f_{mT}-\omega^c_{mT}|| \leq& \frac{1}{K}\sum_{i=1}^{K}[(1+\eta\lambda)^T||\omega^k_{(m-1)T}-\omega^c_{(m-1)T}|| \\
&+\eta||p^k_l-p_o||_1(\sum_{j=2}^{T}\mathbf{g}_1\omega^c_{mT-j}(1+\eta\lambda)^{j-1})]
\end{aligned} \tag{2}
$$

and

$$
\begin{aligned}
||\omega^c_{mT}-\omega^*_{mT}|| \leq& (1+\eta\lambda)^T||\omega^c_{(m-1)T}-\omega^*_{(m-1)T}|| \\
&+\eta||p_o-p_u||_1(\sum_{j=1}^{T}\mathbf{g}_2\omega^*_{mT-j}(1+\eta\lambda)^{j-1})
\end{aligned} \tag{3}
$$

we will prove the inequality in Equation 2 in Proof 1 and the inequality in Equation 3 in Proof 2

1 PROOF 1

$$
\begin{aligned}
\|\omega^f_{mT}-\omega^c_{mT}\| =& \|\frac{1}{K}\sum_{k=1}^{K}\omega^k_{mT}-\omega^c_{mT}\| \\
=& \|\frac{1}{K}\sum_{k=1}^{K}(\omega^k_{mT-1}-\eta\sum_{i=1}^{C}p^k(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^k_{mT-1})]) \\
&-(\omega^c_{mT-1}-\eta\sum_{i=1}^{C}p(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})])\| \\
\leq& \|\frac{1}{K}\sum_{k=1}^{K}\omega^k_{mT-1}-\omega^c_{mT-1}\|+\eta\|\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{C}p^k(y=i)(\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^k_{mT-1})] \\
&-\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})])\| \\
\overset{(1)}{\leq}& \frac{1}{K}\sum_{k=1}^{K}\|\omega^k_{mT-1}-\omega^c_{mT-1}\|+\eta\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{C}p^k(y=i)\lambda\|\omega^k_{mT-1}-\omega^c_{mT-1}\| \\
=& \frac{1}{K}\sum_{k=1}^{K}(1+\eta\lambda)\|\omega^k_{mT-1}-\omega^c_{mT-1}\|
\end{aligned} \tag{4}
$$

The inequality (1) holds because we assume $\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega)]$ is λ-Lipschitz for $x,y\sim p$.

Notice that:

$$
\begin{aligned}
\|\omega^k_{mT-1} - \omega^c_{mT-1}\| =& \|\omega^k_{mT-2} - \eta \sum_{i=1}^{C} p^k(y=i) \nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega^k_{mT-2})] \\
& - \omega^c_{mT-2} + \eta \sum_{i=1}^{C} p(y=i) \nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega^c_{mT-2})]\| \\
\leq & \|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta \| \sum_{i=1}^{C} p^k(y=i) (\nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega^k_{mT-2})] \\
& - \nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega^c_{mT-2})])\| \\
& + \eta \| \sum_{i=1}^{C} (p^k(y=i) - p(y=i)) \nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega^c_{mT-2})]\| \\
\leq & \|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta \sum_{i=1}^{C} p^k(y=i) \lambda \|\omega^k_{mT-2} - \omega^c_{mT-2}\| \\
& + \eta g_{max}(\omega^c_{mT-2}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)| \\
=& [1 + \eta\lambda \sum_{i=1}^{C} p^k(y=i)] \|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta g_{max}(\omega^c_{mT-2}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)| \\
=& [1 + \eta\lambda] \|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta g_{max}(\omega^c_{mT-2}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)|
\end{aligned}
\tag{5}
$$

(Note. $g_{max}(\omega) = \max_{i=1}^{C} \|\nabla_\omega \mathbb{E}_{x|y=i}[-\log f_i(x, \omega)]\|$)

So, by induction, we have:

$$
\begin{aligned}
\|\omega^k_{mT-1} - \omega^c_{mT-1}\| \leq & (1 + \eta\lambda) \|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta g_{max}(\omega^c_{mT-2}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)| \\
\leq & (1 + \eta\lambda)^2 \|\omega^k_{mT-3} - \omega^c_{mT-3}\| + (1 + \eta\lambda) \eta g_{max}(\omega^c_{mT-3}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)| \\
& + \eta g_{max}(\omega^c_{mT-2}) \sum_{i=1}^{C} |p^k(y=i) - p(y=i)| \\
\leq & (1 + \eta\lambda)^{T-1} \|\omega^k_{(m-1)T} - \omega^c_{(m-1)T}\| \\
& + \sum_{j=2}^{T} \eta g_{max}(\omega^c_{mT-j}) (1 + \eta\lambda)^{j-2} \sum_{i=1}^{C} |p^k(y=i) - p(y=i)|
\end{aligned}
\tag{6}
$$

Therefore:

$$\begin{aligned}
\|\omega^f_{mT} - \omega^c_{mT}\| \leq& \frac{1}{K}\sum_{i=1}^{K}(1+\eta\lambda)\|\omega^k_{mT-1} - \omega^c_{mT-1}\| \\
\leq& \frac{1}{K}\sum_{i=1}^{K}[(1+\eta\lambda)^T\|\omega^k_{(m-1)T} - \omega^c_{(m-1)T}\| \\
&+ (\sum_{i=1}^{C}|p^k(y=i) - p(y=i)|)(\sum_{j=2}^{T}\eta g_{max}(\omega^c_{mT-j})(1+\eta\lambda)^{j-1})]
\end{aligned} \tag{7}$$

2 PROOF 2

$$\begin{aligned}
\|\omega^c_{mT} - \omega^*_{mT}\| =& \|\omega^c_{mT-1} - \eta\sum_{i=1}^{C}p(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] \\
&- \omega^*_{mT-1} + \eta\sum_{i=1}^{C}p^u(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^*_{mT-1})]\| \\
\leq& \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\|\sum_{i=1}^{C}p(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] \\
&- \sum_{i=1}^{C}p^u(y=i)\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^*_{mT-1})]\| \\
\leq& \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\|\sum_{i=1}^{C}p(y=i)(\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] \\
&- \nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^*_{mT-1})])\| \\
&+ \eta\|\sum_{i=1}^{C}(p(y=i) - p^u(y=i))\nabla_\omega\mathbb{E}_{x|y=i}[-\log f_i(x,\omega^*_{mT-1})]\| \\
\leq& \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\sum_{i=1}^{C}p(y=i)\lambda\|\omega^c_{mT-1} - \omega^*_{mT-1}\| \\
&+ \eta g_{max}(\omega^*_{mT-1})\sum_{i=1}^{C}|p(y=i) - p^u(y=i)| \\
=& (1+\eta\lambda)\|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta g_{max}(\omega^*_{mT-1})\sum_{i=1}^{C}|p(y=i) - p^u(y=i)| \\
\leq& (1+\eta\lambda)^T\|\omega^c_{(m-1)T} - \omega^*_{(m-1)T}\| \\
&+ (\eta\sum_{i=1}^{C}|p(y=i) - p^u(y=i)|)(\sum_{j=1}^{T}(1+\eta\lambda)^{j-1}g_{max}(\omega^*_{mT-j}))
\end{aligned} \tag{8}$$