1 Proof 1

$$\|\omega_{mT}^{f} - \omega_{mT}^{c}\| = \|\frac{1}{K}\sum_{k=1}^{K}\omega_{mT}^{k} - \omega_{mT}^{c}\|$$

$$=\|\frac{1}{K}\sum_{k=1}^{K}(\omega_{mT-1}^{k} - \eta\sum_{i=1}^{C}p^{k}(y=i)\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-1}^{k})])$$

$$-(\omega_{mT-1}^{c} - \eta\sum_{i=1}^{C}p(y=i)\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-1}^{c})])\|$$

$$\leq\|\frac{1}{K}\sum_{k=1}^{K}\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\| + \eta\|\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{C}p^{k}(y=i)(\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-1}^{k})]$$

$$-\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-1}^{c})])\|$$

$$\leq^{(1)}\frac{1}{K}\sum_{k=1}^{K}\|\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\| + \eta\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{C}p^{k}(y=i)\lambda\|\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\|$$

$$=\frac{1}{K}\sum_{k=1}^{K}(1+\eta\lambda)\|\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\| \tag{1}$$

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

$$=\frac{1}{K}\sum_{k=1}^{K}(1+\eta\lambda)\|\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\|$$

Notice that

$$\|\omega_{mT-1}^{k} - \omega_{mT-1}^{c}\| = \|\omega_{mT-2}^{k} - \eta\sum_{i=1}^{C}p^{k}(y=i)\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-2}^{k})] - \omega_{mT-2}^{c} + \eta\sum_{i=1}^{C}p(y=i)\nabla$$

$$\leq \|\omega_{mT-2}^{k} - \omega_{mT-2}^{c}\| + \eta\|\sum_{i=1}^{C}p^{k}(y=i)(\nabla_{\omega}E_{x|y=i}[-\log f_i(x,\omega_{mT-2}^{k})] - \nabla_{\omega}E_{x|y=i}[$$

$$\leq \|\omega_{mT-2}^{k} - \omega_{mT-2}^{c}\| + \eta\sum_{i=1}^{C}p^{k}(y=i)\lambda\|\omega_{mT-2}^{k} - \omega_{mT-2}^{c}\| + \eta g_{max}(\omega_{mT-2}^{c})\sum_{i=1}^{C}|p^{k}$$

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

$$= [1+\eta\lambda\sum_{i=1}^{C}p^{k}(y=i)]\|\omega_{mT-2}^{k} - \omega_{mT-2}^{c}\| + \eta g_{max}(\omega_{mT-2}^{c})\sum_{i=1}^{C}|p^{k}(y=i) - p(y=i)|$$

$$= [1+\eta\lambda]\|\omega_{mT-2}^{k} - \omega_{mT-2}^{c}\| + \eta g_{max}(\omega_{mT-2}^{c})\sum_{i=1}^{C}|p^{k}(y=i) - p(y=i)|$$

So, by induction, we have

$$\|\omega_{mT-1}kf - \omega^c_{mT-1}\| \leq (1+\eta\lambda)\|\omega^k_{mT-2} - \omega^c_{mT-2}\| + \eta g_{max}(\omega^c_{mT-2})\sum_{i=1}^{C}|p^k(y=i) - p(y=i)|$$

$$\leq (1+\eta\lambda)^2\|\omega^k_{mT-3} - \omega^c_{mT-3}\| + (1+\eta\lambda)\eta g_{max}(\omega^c_{mT-3})\sum_{i=1}^{C}|p^k(y=i) - p(y=i)$$

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

Therefore:

$$\|\omega^f_{mT} - \omega^c_{mT}\| \leq \frac{1}{K}\sum_{i=1}^{K}(1+\eta\lambda)\|\omega^k_{mT-1} - \omega$$

$$\leq \frac{1}{K}\sum_{i=1}^{K}\Bigg[(1+\eta\lambda)^T\|\omega^k_{(m-1)T} - \omega^c_{(m-1)T}\| + \left(\sum_{i=1}^{C}|p^k(y=i) - p(y=i)|\right)\left(\sum_{j=2}^{T}\eta g_{max}(\omega^c_{mT-j})(1+\eta\lambda)\right.$$

2 Proof 2

$$\|\omega^c_{mT} - \omega^*_{mT}\| = \|\omega^c_{mT-1} - \eta\sum_{i=1}^{C}p(y=i)\nabla_\omega E_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] - \omega^*_{mT-1} + \eta\sum_{i=1}^{C}p^u(y=i)\nabla_\omega E_{x|y=}$$

$$\leq \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\|\sum_{i=1}^{C}p(y=i)\nabla_\omega E_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] - \sum_{i=1}^{C}p^u(y=i)\nabla_\omega E_{x|y=}$$

$$\leq \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\|\sum_{i=1}^{C}p(y=i)\left(\nabla_\omega E_{x|y=i}[-\log f_i(x,\omega^c_{mT-1})] - \nabla_\omega E_{x|y=i}[-\log f_i(x\right.$$

$$\leq \|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta\sum_{i=1}^{C}p(y=i)\lambda\|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta g_{max}(\omega^*_{mT-1})\sum_{i=1}^{C}|p(y=i) - p$$

$$= (1+\eta\lambda)\|\omega^c_{mT-1} - \omega^*_{mT-1}\| + \eta g_{max}(\omega^*_{mT-1})\sum_{i=1}^{C}|p(y=i) - p^u(y=i)|$$

$$\leq (1+\eta\lambda)^T\|\omega^c_{(m-1)T} - \omega^*_{(m-1)T}\| + \left(\eta\sum_{i=1}^{C}|p(y=i) - p^u(y=i)|\right)\left(\sum_{j=1}^{T}(1+\eta\lambda)^{j-1}g_{max}(\omega\right.$$