(a) Suppose $\boldsymbol{y} = \boldsymbol{e}_i$, which means that $y_i = 1$ and all other entries are 0. Therefore

$$J^{(t)}(\theta) = -\log \hat{y}_i{}^{(t)}$$

$$\mathrm{PP}^{(t)} = \frac{1}{\hat{y}_i{}^{(t)}} = \exp(J^{(t)}(\theta))$$

The second argument easily follows as exp function is monotone increasing.

$$\mathbb{E}(\mathrm{PP}) = \frac{1}{1/|V|} = |V| = 10000$$

$$\mathbb{E}(J) = \log \mathbb{E}(\mathrm{PP}) = 4\log 10$$

(b) For the convenience of writing, we set

$$\boldsymbol{g}^{(t)} = \boldsymbol{h}^{(t-1)}\boldsymbol{H} + \boldsymbol{e}^{(t)}\boldsymbol{I} + \boldsymbol{b}_1$$

And

$$\frac{\partial J^{(t)}}{\partial \boldsymbol{g}^{(t)}} = \frac{\partial J^{(t)}}{\partial \boldsymbol{h}^{(t)}}\frac{\partial \boldsymbol{h}^{(t)}}{\partial \boldsymbol{g}^{(t)}} = (\hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)})\boldsymbol{U}^t\boldsymbol{h}^{(t)}(1-\boldsymbol{h}^{(t)})$$

where most of the calculation is similar to Homework2 and is not displayed detailedly here.

Then all other partial derivatives can be calculated by the chain rule:

$$\frac{\partial J^{(t)}}{\partial \boldsymbol{b}_2} = \hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)}$$

$$\begin{aligned}
\frac{\partial J^{(t)}}{\partial \boldsymbol{L}_{x^{(t)}}} &= \frac{\partial J^{(t)}}{\partial \boldsymbol{g}^{(t)}}\frac{\partial \boldsymbol{g}^{(t)}}{\partial \boldsymbol{L}_{x^{(t)}}} = (\hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)})\boldsymbol{U}^t\boldsymbol{h}^{(t)}(1-\boldsymbol{h}^{(t)})\boldsymbol{I}^t \\
\frac{\partial J^{(t)}}{\partial \boldsymbol{I}}|_t &= \frac{\partial J^{(t)}}{\partial \boldsymbol{g}^{(t)}}\frac{\partial \boldsymbol{g}^{(t)}}{\partial \boldsymbol{I}} = \boldsymbol{e}^{(t)t}(\hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)})\boldsymbol{U}^t\boldsymbol{h}^{(t)}(1-\boldsymbol{h}^{(t)}) \\
\frac{\partial J^{(t)}}{\partial \boldsymbol{H}}|_t &= \frac{\partial J^{(t)}}{\partial \boldsymbol{g}^{(t)}}\frac{\partial \boldsymbol{g}^{(t)}}{\partial \boldsymbol{H}} = \boldsymbol{h}^{(t-1)t}(\hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)})\boldsymbol{U}^t\boldsymbol{h}^{(t)}(1-\boldsymbol{h}^{(t)}) \\
\frac{\partial J^{(t)}}{\partial \boldsymbol{h}^{(t-1)}} &= \frac{\partial J^{(t)}}{\partial \boldsymbol{g}^{(t)}}\frac{\partial \boldsymbol{g}^{(t)}}{\partial \boldsymbol{h}^{(t-1)}} = (\hat{\boldsymbol{y}}^{(t)} - \boldsymbol{y}^{(t)})\boldsymbol{U}^t\boldsymbol{h}^{(t)}(1-\boldsymbol{h}^{(t)})\boldsymbol{H}^t
\end{aligned}$$