A Doubly Stochastic Gradient

To derive doubly stochastic gradient for equation (5), we first denote (5) as $J(\Theta)$ with $\Theta = \{P, Q\}$ and resolve the expectation form as:

$$\begin{aligned} J(\theta) &= \mathbb{E}_{z_{ik}\sim\pi(\cdot|\bar{C}_t)}[\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik})] \\ &= \sum_k \pi(z_{ik} \mid \bar{C}_t) \log \bar{\mathcal{L}}(z_{jl}|z_{ik}). \end{aligned}$$

Denote $\Theta = \{P, Q\}$ as the parameter set for policy π. The gradient with respect to Θ should be:

$$\begin{aligned} &\frac{\partial J(\theta)}{\partial \Theta} = \frac{\partial}{\partial \Theta} \sum_k \pi(z_{ik} \mid \bar{C}_t) \log \bar{\mathcal{L}}(z_{jl}|z_{ik}) \\ &= \sum_k \log \bar{\mathcal{L}}(z_{jl}|z_{ik}) \frac{\partial \pi(z_{ik} \mid \bar{C}_t)}{\partial \Theta} \\ &= \sum_k \log \bar{\mathcal{L}}(z_{jl}|z_{ik}) (\frac{\partial \log \pi(z_{ik} \mid \bar{C}_t)}{\partial \Theta})(\pi(z_{ik} \mid \bar{C}_t)) \\ &= \mathbb{E}_{z_{ik}\sim\pi(\cdot|\bar{C}_t)}[\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) \frac{\partial \log \pi(z_{ik} \mid \bar{C}_t)}{\partial \Theta}] \end{aligned}$$

Accordingly, if we conduct typical stochastic gradient ascent training on $J(\Theta)$ with respect to Θ from samples z_{ik} with a learning rate η, the update formula will be:

$$\Theta = \Theta + \eta \log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) \frac{\partial \log \pi(z_{ik} \mid \bar{C}_t)}{\partial \Theta}.$$

However, the collocation log likelihood should always be non-positive: $\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) \leq 0$. Therefore, as long as the collocation log likelihood $\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik})$ is negative, the update formula is to minimize the likelihood of choosing z_{ik}, despite the fact that z_{ik} may be good choices. On the other hand, if the log likelihood reaches 0, according to (4), it indicates:

$$\begin{aligned} &\log \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) = 0 \Rightarrow \quad \bar{\mathcal{L}}(z_{jl} \mid z_{ik}) = 1 \\ \Rightarrow \quad &U_{z_{ik}}^T V_{z_{jl}} \to \infty, \;\; U_{z_{ik}}^T V_{z_{uv}} \to \infty, \;\; \forall z_{uv}, \end{aligned}$$

which leads to computational overflow from an infinity value.