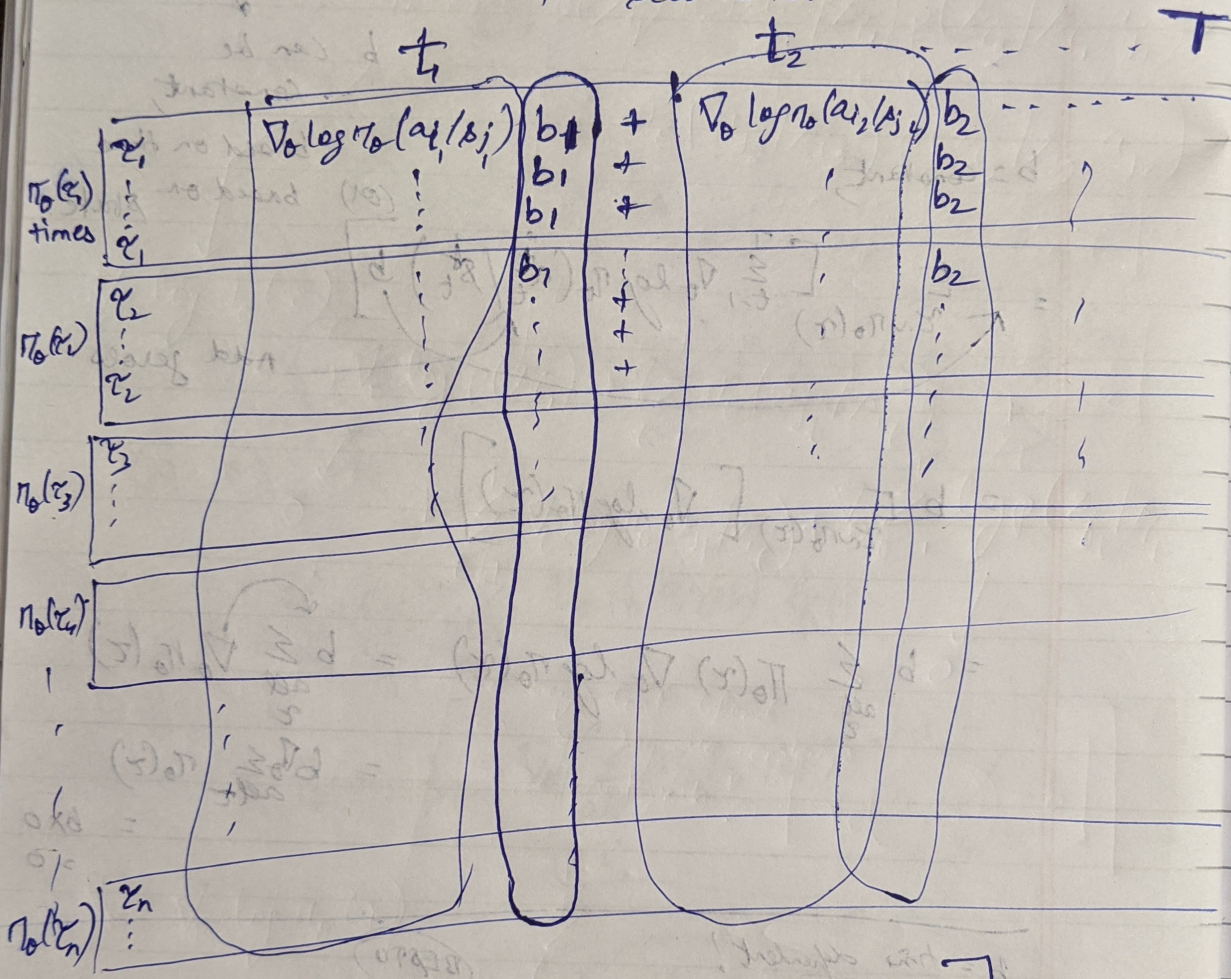
Baseline as function of time

$$\nabla_\theta J(\theta) = \underset{\tau \sim \pi_\theta(\tau)}{E}\left[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_i/s_j)\left[\tilde{Q}(a_i^t, s_j^t) - b_t\right]\right]$$

↓ visualize the double sum for only the second term



$$\sum_{t=1}^{T} b_t \left[\sum_{\substack{\text{all} \\ s_i, a_j}} P_\theta(s_i, a_j)\, \nabla_\theta \log \pi_\theta(a_j/s_i)\right]$$

$$\sum_{t=1}^{T} b_t \left[\sum_{\substack{\text{all} \\ s_i, a_j}} P_\theta(s_i)\, \pi_\theta(a_j/s_i)\, \nabla_\theta \log \pi_\theta(a_j/s_i)\right]$$

$$\sum_{t=1}^{T} b_t \sum_{\text{all } s_i, a_j} \left[P_\theta(s_i)\, \nabla_\theta \pi_\theta(a_j/s_i)\right]$$

Taking s_i common!

$$\sum_{t=1}^{T} b_t \sum_{\text{all } s_i} P_\theta(s_i) \left[\sum_{\substack{\text{all} \\ a_j}} \nabla_\theta \pi_\theta(a_j/s_i)\right]$$

$\nabla_\theta(1) = 0$

↳ Softmax & NN

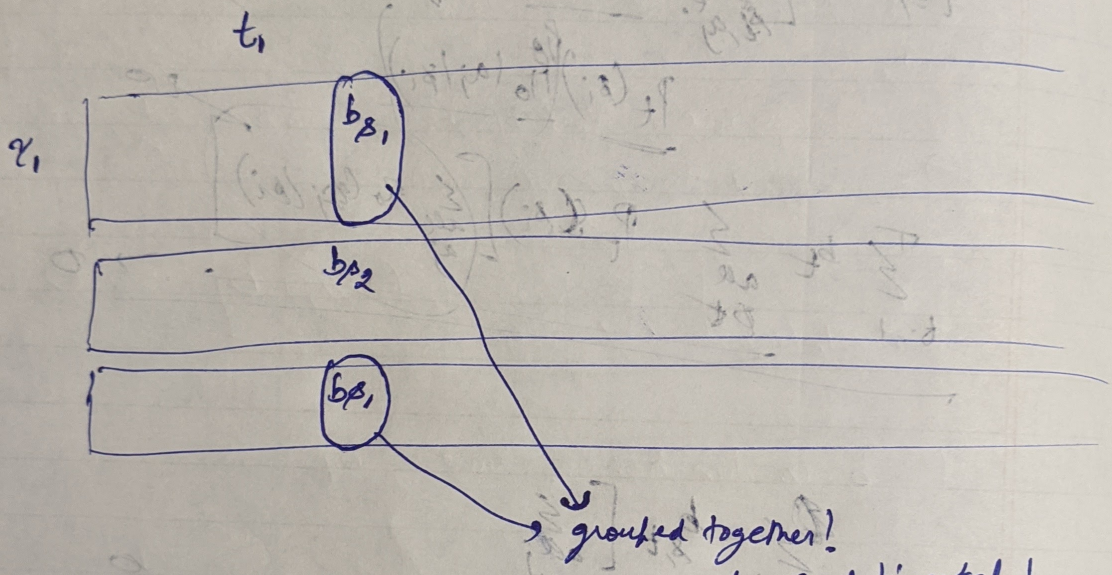
$= \boxed{0}$!

Baseline as function of both state & time

$$\mathbb{E}_{\tau \sim \pi_\theta(\tau)}\left[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_i/s_j)\left[Q(a_i,s_j) - b_{s_t}\right]\right]$$

visualizing double sum

for only the 2nd term



$$\sum_{t=1}^{T} \sum_{\substack{\text{all} \\ s_j}} b_{s_j,t} \sum_{\substack{\text{all} \\ a_i}} P_\theta^t(a_i,s_j) \times \nabla_\theta \log \pi_\theta(a_i/s_j)$$

$$= \sum_{t=1}^{T} \sum_{\substack{\text{all} \\ s_j}} b_{s_j,t} \sum_{\substack{\text{all} \\ a_i}} P_\theta^t(s_j)\, \pi_\theta(a_i/s_j)\, \nabla_\theta \log \pi_\theta(a_i/s_j)$$

$$= \sum_{t=1}^{T} \sum_{\substack{\text{all} \\ s_j}} b_{s_j,t}\, P_\theta^t(s_j) \underbrace{\sum_{\substack{\text{all} \\ a_i}} \nabla_\theta \pi_\theta(a_i/s_j)}_{\to 0}$$

sum of softmax of a NN = 1

$\nabla_\theta(1) = 0$

$$= \boxed{0} \quad !$$