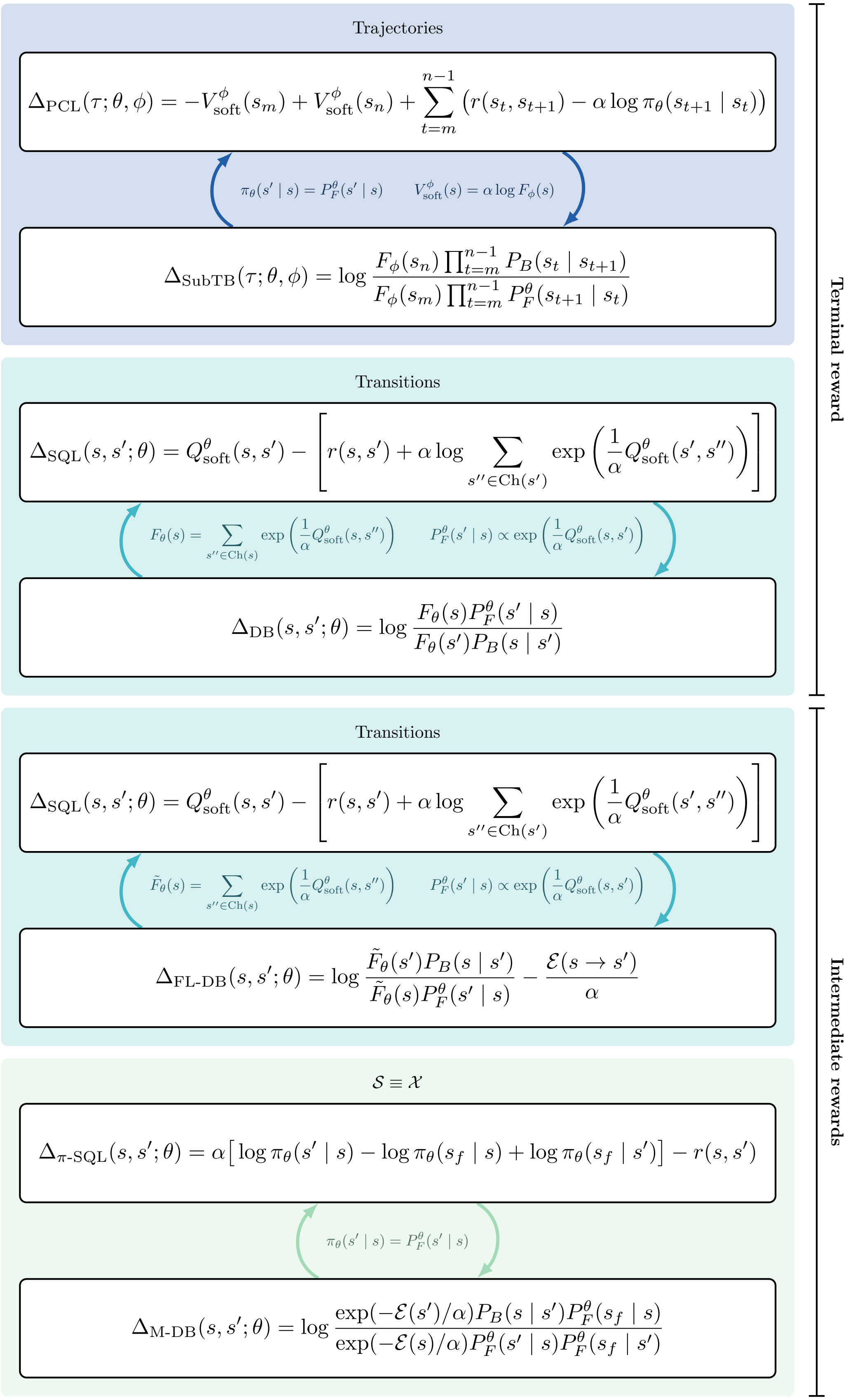

Terminal reward

Trajectories

$$\Delta_{\mathrm{PCL}}(\tau;\theta,\phi) = -V^{\phi}_{\mathrm{soft}}(s_m) + V^{\phi}_{\mathrm{soft}}(s_n) + \sum_{t=m}^{n-1}\big(r(s_t,s_{t+1}) - \alpha\log\pi_\theta(s_{t+1}\mid s_t)\big)$$

$$\pi_\theta(s'\mid s) = P^{\theta}_{F}(s'\mid s) \qquad V^{\phi}_{\mathrm{soft}}(s) = \alpha\log F_\phi(s)$$

$$\Delta_{\mathrm{SubTB}}(\tau;\theta,\phi) = \log\frac{F_\phi(s_n)\prod_{t=m}^{n-1}P_B(s_t\mid s_{t+1})}{F_\phi(s_m)\prod_{t=m}^{n-1}P^{\theta}_{F}(s_{t+1}\mid s_t)}$$

Transitions

$$\Delta_{\mathrm{SQL}}(s,s';\theta) = Q^{\theta}_{\mathrm{soft}}(s,s') - \left[r(s,s') + \alpha\log\sum_{s''\in\mathrm{Ch}(s')}\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s',s'')\right)\right]$$

$$F_\theta(s) = \sum_{s''\in\mathrm{Ch}(s)}\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s,s'')\right) \qquad P^{\theta}_{F}(s'\mid s)\propto\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s,s')\right)$$

$$\Delta_{\mathrm{DB}}(s,s';\theta) = \log\frac{F_\theta(s)P^{\theta}_{F}(s'\mid s)}{F_\theta(s')P_B(s\mid s')}$$

Intermediate rewards

Transitions

$$\Delta_{\mathrm{SQL}}(s,s';\theta) = Q^{\theta}_{\mathrm{soft}}(s,s') - \left[r(s,s') + \alpha\log\sum_{s''\in\mathrm{Ch}(s')}\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s',s'')\right)\right]$$

$$\tilde{F}_\theta(s) = \sum_{s''\in\mathrm{Ch}(s)}\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s,s'')\right) \qquad P^{\theta}_{F}(s'\mid s)\propto\exp\left(\frac{1}{\alpha}Q^{\theta}_{\mathrm{soft}}(s,s')\right)$$

$$\Delta_{\text{FL-DB}}(s,s';\theta) = \log\frac{\tilde{F}_\theta(s')P_B(s\mid s')}{\tilde{F}_\theta(s)P^{\theta}_{F}(s'\mid s)} - \frac{\mathcal{E}(s\to s')}{\alpha}$$

$\mathcal{S}\equiv\mathcal{X}$

$$\Delta_{\pi\text{-SQL}}(s,s';\theta) = \alpha\big[\log\pi_\theta(s'\mid s) - \log\pi_\theta(s_f\mid s) + \log\pi_\theta(s_f\mid s')\big] - r(s,s')$$

$$\pi_\theta(s'\mid s) = P^{\theta}_{F}(s'\mid s)$$

$$\Delta_{\text{M-DB}}(s,s';\theta) = \log\frac{\exp(-\mathcal{E}(s')/\alpha)P_B(s\mid s')P^{\theta}_{F}(s_f\mid s)}{\exp(-\mathcal{E}(s)/\alpha)P^{\theta}_{F}(s'\mid s)P^{\theta}_{F}(s_f\mid s')}$$