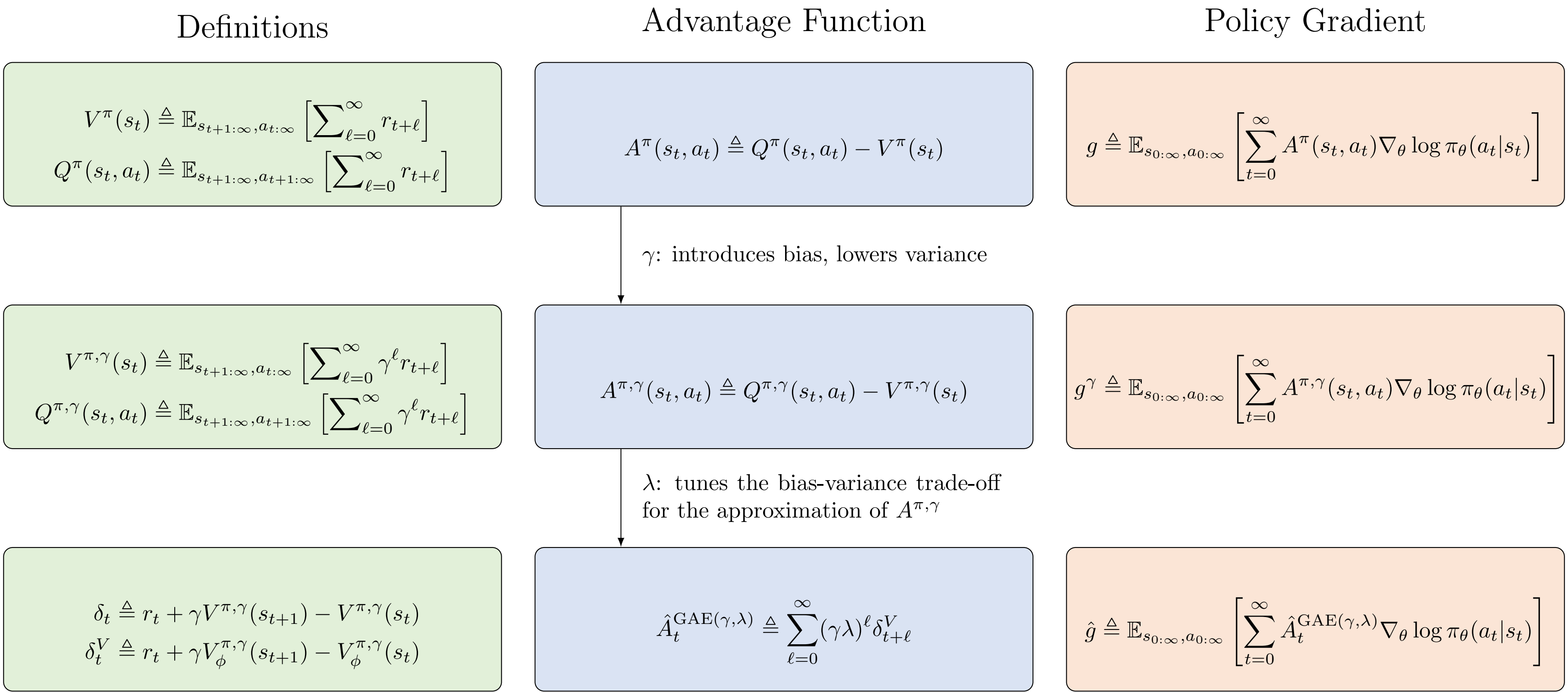

Definitions

$$V^{\pi}(s_t) \triangleq \mathbb{E}_{s_{t+1:\infty},a_{t:\infty}}\left[\sum\nolimits_{\ell=0}^{\infty} r_{t+\ell}\right]$$

$$Q^{\pi}(s_t,a_t) \triangleq \mathbb{E}_{s_{t+1:\infty},a_{t+1:\infty}}\left[\sum\nolimits_{\ell=0}^{\infty} r_{t+\ell}\right]$$

$$V^{\pi,\gamma}(s_t) \triangleq \mathbb{E}_{s_{t+1:\infty},a_{t:\infty}}\left[\sum\nolimits_{\ell=0}^{\infty} \gamma^{\ell} r_{t+\ell}\right]$$

$$Q^{\pi,\gamma}(s_t,a_t) \triangleq \mathbb{E}_{s_{t+1:\infty},a_{t+1:\infty}}\left[\sum\nolimits_{\ell=0}^{\infty} \gamma^{\ell} r_{t+\ell}\right]$$

$$\delta_t \triangleq r_t + \gamma V^{\pi,\gamma}(s_{t+1}) - V^{\pi,\gamma}(s_t)$$

$$\delta_t^V \triangleq r_t + \gamma V_{\phi}^{\pi,\gamma}(s_{t+1}) - V_{\phi}^{\pi,\gamma}(s_t)$$

Advantage Function

$$A^{\pi}(s_t,a_t) \triangleq Q^{\pi}(s_t,a_t) - V^{\pi}(s_t)$$

γ: introduces bias, lowers variance

$$A^{\pi,\gamma}(s_t,a_t) \triangleq Q^{\pi,\gamma}(s_t,a_t) - V^{\pi,\gamma}(s_t)$$

λ: tunes the bias-variance trade-off for the approximation of $A^{\pi,\gamma}$

$$\hat{A}_t^{\mathrm{GAE}(\gamma,\lambda)} \triangleq \sum_{\ell=0}^{\infty} (\gamma\lambda)^{\ell} \delta_{t+\ell}^V$$

Policy Gradient

$$g \triangleq \mathbb{E}_{s_{0:\infty},a_{0:\infty}}\left[\sum_{t=0}^{\infty} A^{\pi}(s_t,a_t) \nabla_{\theta} \log \pi_{\theta}(a_t|s_t)\right]$$

$$g^{\gamma} \triangleq \mathbb{E}_{s_{0:\infty},a_{0:\infty}}\left[\sum_{t=0}^{\infty} A^{\pi,\gamma}(s_t,a_t) \nabla_{\theta} \log \pi_{\theta}(a_t|s_t)\right]$$

$$\hat{g} \triangleq \mathbb{E}_{s_{0:\infty},a_{0:\infty}}\left[\sum_{t=0}^{\infty} \hat{A}_t^{\mathrm{GAE}(\gamma,\lambda)} \nabla_{\theta} \log \pi_{\theta}(a_t|s_t)\right]$$