The definition of $\hat{\eta}_\pi(\tilde{\pi})$ is

$$\hat{\eta}_\pi(\tilde{\pi}) = \eta(\pi) + \sum_s \rho_\pi(s) \sum_a \tilde{\pi}(a|s) A_\pi(s,a), \tag{1}$$

we can derive Eq.1 and Eq.2 by replacing $\tilde{\pi}$ with parameterized π_{θ_0} due to $\sum_s \rho_\pi(s) \sum_a \pi_{\theta_0}(a|s) A_\pi(s,a) = 0$.

We know that $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$, let $p(\theta) = \sigma(\tau(r_t(\theta) - 1))$, then we have

$$\nabla_\theta p(\theta) = \sigma(\tau(r_t(\theta) - 1))(1 - \sigma(\tau(r_t(\theta) - 1)))\tau\nabla_\theta r_t(\theta) \tag{2}$$

$$= p(\theta)(1 - p(\theta))\tau\nabla_\theta r_t(\theta) \tag{3}$$

Then we use $\nabla_\theta p(\theta)$ to simplify $\nabla_\theta L^{sc}$,

$$\nabla_\theta L^{sc} = \nabla_\theta E_{a \sim \pi_{\theta_{old}}}[p(\theta)\frac{4}{\tau}\hat{A}] \tag{4}$$

$$= \nabla_\theta \int_{\pi_{\theta_{old}}} p(\theta)\frac{4}{\tau}\hat{A}\mathrm{d}a \tag{5}$$

$$= \int_{\pi_{\theta_{old}}} \nabla_\theta p(\theta)\frac{4}{\tau}\hat{A}\mathrm{d}a \tag{6}$$

$$= \int_{\pi_{\theta_{old}}} 4p(\theta)(1 - p(\theta)\nabla_\theta r_t(\theta)\hat{A}\mathrm{d}a \tag{7}$$

$$= E_{a \sim \pi_{\theta_{old}}}[4p(\theta)(1 - p(\theta)\nabla_\theta r_t(\theta)\hat{A}]. \tag{8}$$

In the line (Eq.4), we rewrite the $\nabla_\theta L^{sc}$ using $p(\theta)$. In (Eq.5), we expand the definition of the expectation as an integral. In (Eq.6), we exchange the order of the integral and the derivative, the validity of this operation has been discussed in the paper "Monte Carlo Gradient Estimation in Machine Learning"(Section 4.3.1). In (Eq.8), we rewrite the expression as an expectation with respect to the distribution$\pi_{\theta_{old}}(a|s)$. In our paper, E_t represents the expectation of empirical trajectory return.