$$\nabla_\omega J(\omega) = E_\pi [\gamma^t R_t \nabla_\omega \log \pi(a, s_t, \omega)]$$

$$E_\pi [\gamma^t V(s_t) \nabla_\omega \log \pi(a, s_t, \omega)] =$$

$$= E_\pi \left[\gamma^t \sum_a \pi(a, s_t, \omega) V(s_t) \frac{\nabla_\omega \pi(a, s_t, \omega)}{\pi(a, s_t, \omega)}\right]$$

$$= E_\pi [\gamma^t V(s_t) \sum_a \nabla_\omega \pi(a, s_t, \omega)]$$

$$= E_\pi [\gamma^t V(s_t) \underbrace{\nabla_\omega \underbrace{\sum_a \pi(a, s_t, \omega)}_{1}}_{0}] = 0$$

$$\nabla_\omega J(\omega) = E_\pi [\gamma^t [R_t - V(s_t)] \nabla_\omega \log \pi(a, s_t, \omega)]$$

$$VAR[X-Y] = VAR[X] - 2COV[X,Y] + VAR[Y] < VAR[X]$$

if X and Y are strongly correlated

$$\nabla_\omega J(\omega) = E_\pi [\gamma^t [\underbrace{Q_\pi(s_t, a) - V_\pi(s_t)}_{A_\pi(s_t, a)}] \nabla_\omega \log \pi(a, s_t, \omega)]$$

$$\nabla_\omega J(\omega) = E_\pi [\gamma^t A(s_t, a) \nabla_\omega \log \pi(a, s_t, \omega)]$$

$$\gamma^t A(s_t, a, \theta) = \nabla_\omega \log \pi(a, s_t, \omega)^T \theta$$

$$\nabla_\omega J(\omega) = E_\pi [\nabla_\omega \log \pi(a, s_t, \omega)^T \theta \nabla_\omega \log \pi(a, s_t, \omega)]$$

$$\nabla_\omega J(\omega) = E_\pi [\nabla_\omega \log \pi(a, s_t, \omega) \nabla \log \pi(a, s_t, \omega)^T] \theta$$

doesn't depend on π

$$\theta = G_\omega^{-1} \nabla_\omega J(\omega)$$

$$G_\omega = E_\pi [\nabla_\omega \log \pi(a, s_t, \omega) \nabla_\omega \log \pi(a, s_t, \omega)^T]$$

$$\omega = \omega + \alpha \theta$$

$$\left| \sum_{t=0}^{T-1} \gamma^t A(s_t, a_t, \theta) - \sum_{t=0}^{T-1} \gamma^t A(s_t, a_t) \right|^2$$

$$\sum_{t=0}^{T-1} \gamma^t A(s_t, a_t) = A(s_0, a_0) + \gamma A(s_1, a_1) + \dots + \gamma^{T-1} A(s_{T-1}, a_{T-1}) =$$

$$= Q(s_0, a_0) - V(s_0) + \gamma Q(s_1, a_1) - \gamma V(s_1) + \dots + \gamma^{T-1} Q(s_{T-1}, a_{T-1}) + \gamma^{T-1} V(s_{T-1})$$

$$\cancel{\approx r_1 + \gamma V(s_1) - V(s_0) + \gamma r_2 - \gamma V(s_1) + \gamma r_3 - \gamma V(s_2) + \dots + \gamma^{T-1} r_T + \gamma^T \underbrace{V(s_T)}_{=0} - \gamma^{T-1} V(s_{T-1})}$$

$$\approx r_1 + \gamma \cancel{V(s_1)} - V(s_0) + r_2 + \gamma^2 \cancel{V(s_2)} - \cancel{\gamma V(s_1)} + \dots + \gamma^{T-1} r_T + \gamma^T \underbrace{V(s_T)}_{=0} - \gamma^{T-1} \cancel{V(s_{T-1})}$$

$$\approx R_0 - J$$

$$\left| \sum_t \theta^T \phi(s_t, a_t) + J - R \right|^2 = \left| [\theta^T \; J] \begin{bmatrix} \overset{\nabla_w \log \pi(a_t, s_t, w)}{\overset{\parallel}{\phi}} \\ 1 \end{bmatrix} - R \right|^2$$