- Peters & Schaal 2008
  - → Neural Networks
  - → Neuro...

- Generalized Advantage
- Dann, Neumann & Peters ~~[illegible]~~ 2014 JMLR

1) **IDEE** $\pi(a|s) = \cancel{\ldots}\ \rho(a|s,\theta)$ $= \mathcal{N}(a \mid \mu = NN_\theta(s), \sigma^2)$

$$f_W(s,a) = \log \rho(a|s,\theta)^T W$$

$$f_V(s) = \underline{NN}_V(s)$$

"Fitted"

$$\min_{v_t, w_t} \left(r(s,a) + \gamma f_{V_{t-1}}(s') - f_{V_t}(s) + f_{W_t}(s,a) \right)^2$$

2) **IDEE:** Dann-Paper

3) IDEE

$$\pi(a \mid s; \underline{\theta} = \underline{\theta}_0 + \alpha \tilde{\nabla}_\theta J |_{\theta=\theta_0})$$
$$= \mathcal{N}(a \mid \mu_{\theta^1}(s), \theta^2)$$

$$\theta^1 = \theta^1_0 + \alpha \tilde{\nabla}_{\theta^1} J$$
$$\theta^2 = \theta^2_0 + \alpha \tilde{\nabla}_{\theta^2} J = 0$$

$$\Rightarrow \alpha_{max} = \frac{\theta_0^2}{\tilde{\nabla}_{\theta^2} J}$$

$$\alpha = \lambda \alpha_{max} \qquad \lambda \in [0, 1)$$

4) IDEE

- Stochastic
- minibatches

- importance sampling: $\nabla_\theta J = \sum \mu(s) \, \underline{\pi'(a|s)} \, \nabla \log \pi'(a|s) \, Q^{\pi'}(a|s)$

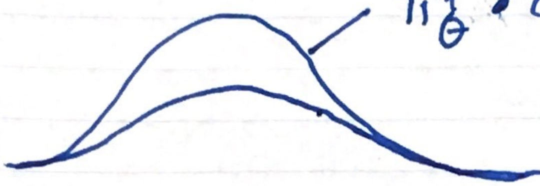
$\pi'(a|s)$

$a \sim \pi(a|s)$

$$\approx \frac{1}{N} \sum \nabla \log \pi'(a|s) \, Q(s,a) = g(\theta)$$

$$\hat{g}(\theta; D_{\pi_\theta}) = \frac{1}{N}\sum_{i=1}^{N} \nabla_\theta \log \pi_\theta(a_i|s_i)\, Q(s_i, a_i)$$

$$\hat{g}(\theta'; D_{\pi_\theta}) = \frac{1}{N}\sum \underbrace{\frac{\pi_{\theta'}(a_i|s_i)}{\pi_\theta(a_i|s_i)}}_{\text{importance weight}} \nabla_{\theta'} \log \pi_{\theta'}(a_i|s_i)\, Q(s_i, a_i)$$



$$\pi_\theta \cdot \underbrace{\alpha}_{\text{Konstante}} \geq \pi_{\theta'} \quad \forall a, s$$

1 Gleichung pro Aktion!

$a_i \sim \pi(a|s_i) \Rightarrow s_i'$

$r(s_i, a_i)$

$$\Rightarrow Q(s_i, a_i) = \underline{r(s_i, a_i)} + \gamma V(s_i')$$

$$V(s_i) + \underbrace{A(s_i, a_i)}$$

Samples

$$\underbrace{f_V(s_i)}_{\phi(s_i)^T v} + \nabla_\theta \log \pi(a_i|s_i)^T w = r(s_i, a_i) + \gamma f_V(s_i')$$

$$\begin{array}{c} \\ i=1 \\ 2 \\ 3 \\ \vdots \\ n \end{array} \left[\begin{array}{cc} \overbrace{\hphantom{\phi^T(s_1)}}^{\Phi} & \overbrace{\hphantom{\nabla_\theta \log \pi(a_1|s_1)}}^{\Psi} \\ \phi^T(s_1) & \nabla_\theta \log \pi(a_1|s_1) \\ \phi^T(s_2) & \nabla_\theta \log \pi(a_2|s_2) \\ \phi^T(s_3) & \nabla_\theta \log \pi(a_3|s_3) \\ \vdots & \vdots \\ \phi^T(s_n) & \nabla_\theta \log \pi(a_n|s_n) \end{array} \right] \begin{bmatrix} \underline{v} \\ \underline{w} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} r(s_1, a_1) \\ r(s_2, a_2) \\ r(s_3, a_3) \\ \vdots \\ r(s_n, a_n) \end{bmatrix}}_{\underline{r}} + \gamma \underbrace{\begin{bmatrix} \phi^T(s_1') \\ \phi^T(s_2') \\ \phi^T(s_3') \\ \vdots \\ \phi^T(s_n') \end{bmatrix}}_{P\Phi} \underline{v}$$

$$\left[\Phi - \gamma P\Phi \quad \Psi \right] \begin{bmatrix} \underline{v} \\ \underline{w} \end{bmatrix} = \underline{r}$$

$$V^\pi(s) = \phi(s)^T v$$

$$E = \underline{r} - \left[\Phi - \gamma P \Phi \quad \Psi\right] \begin{matrix} v \\ w \end{matrix}$$

$$= [\Phi, \Psi][\Phi, \Psi]^+$$

$$\begin{bmatrix} v \\ w \end{bmatrix} = \left([\Phi, \Psi] \begin{bmatrix} \Phi - \gamma P \Phi \\ \Psi \end{bmatrix} \right)^{-1} \begin{bmatrix} \Phi \\ \Psi \end{bmatrix}^T \underline{r}$$

⇒ LSPI Lagoudakis & Parr

LSTD Boyan

LSTD → GTD

$$J(v, w) = \frac{1}{2} \sum_i \Big(\underbrace{\overbrace{r_i + \gamma \phi(s_i')^T \hat{v}}^{t} - \phi(s_i) v - \nabla_\theta \log \pi(a_i | s_i)^T w}_{E_i(v, w)} \Big)^2$$

$$\nabla_v J = -\sum \phi(s_i) E_i(v, \hat{v}, w)$$

$$\nabla_w J = -\sum \nabla_\theta \log \pi(a_i | s_i) E_i(v, \hat{v}, w)$$

$$v' = v - \alpha \nabla_v J = v + \alpha \sum_i \phi(s_i) E_i(v, \hat{v}, w)$$

w' analog!

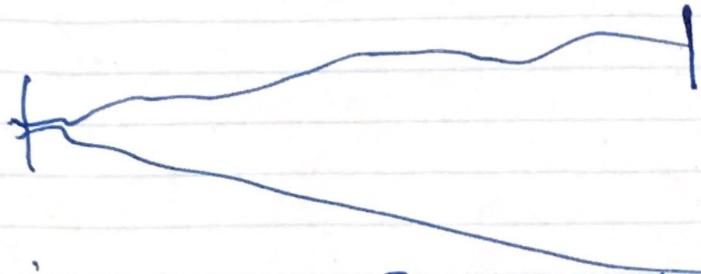
(Bhatnagar Ingantel NAC)

Λ Gleichy pro Trajektorii

$$
\begin{aligned}
& A(s,a) = r(s,a) + \cancel{\gamma V(s')} - V(s) \\
+ \; & \gamma A(s',a') = \gamma r(s',a') + \gamma^2 V(s'') - \cancel{\gamma V(s')} \\
\hline
& A(s,a) + \gamma A(s',a') = r(s,a) + \gamma r(s',a') + \gamma^2 V(s'') - V(s)
\end{aligned}
$$

$$\sum_{i=0}^{N} \gamma^i A(s_i, a_i) = \sum_{i=0}^{N} \gamma^i r(s_i, a_i) + \gamma^N V(s_{N+1}) - \underbrace{V(s_0)}_{J}$$

$$\gamma^N \to 0$$



$$\underbrace{\sum \gamma^i \nabla_\theta \log \pi(a_i|s_i)^T w + 1 \cdot J} = \underbrace{\sum \gamma^i r(s_i, a_i)}_{R_e}$$

$$\Phi_e = \left[\sum \gamma^i \underline{\nabla_\theta \log \pi(a_i|s_i)}, \; 1 \right]$$

$$\underline{\Phi} \begin{bmatrix} w \\ J \end{bmatrix} = \underline{R}$$

$$\begin{bmatrix} w \\ J \end{bmatrix} = (\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{R}$$