Control With Gradient TD Methods + The Nonlinear Case

COMP 767

Matthew Smith

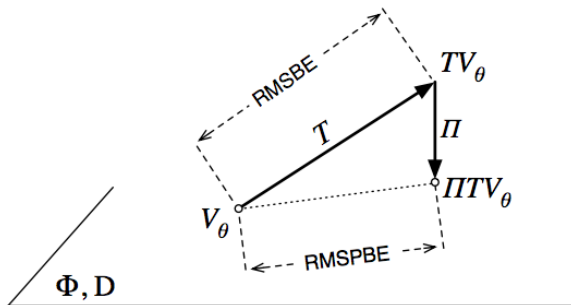# Overview

Gradient TD

Control With Gradient Methods

This slide is just to add slides.

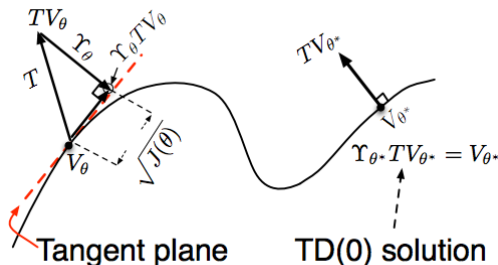# The part we've covered:

▶ This Picture:

▶

$$MSPBE = ||(\Pi_\theta TV_\theta - V_\theta)||_D^2$$

$$-1/2\nabla MSPBE = -\mathbb{E}[\delta\phi] - \gamma\mathbb{E}[\phi'\phi^\top]w$$

$$w = \mathbb{E}[\phi\phi^\top]^{-1}\mathbb{E}[\delta\phi]$$

The part we haven't:

- This Picture:

- MSPBE now projects onto the tangent space of the nonlinear function which we assume to be smooth enough to be locally linear.

$$MSPBE = ||\Pi_\theta(TV_\theta - V_\theta)||_D^2$$
$$-1/2\nabla MSPBE = -\mathbb{E}[\delta\phi] - \gamma\mathbb{E}[\phi'\phi^\top]w + h(\theta, w)$$
$$w = \mathbb{E}[\phi\phi^\top]^{-1}\mathbb{E}[\delta\phi]$$
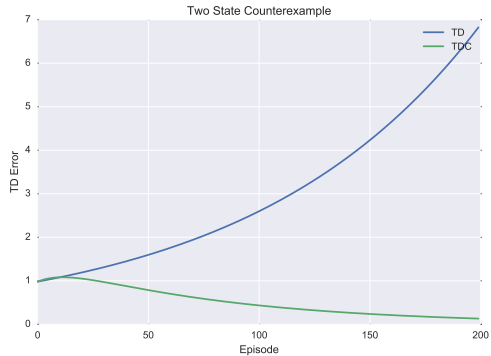
The part we haven't:

- Update rules are much the same but now there are second order terms.

$$\theta_{k+1} = \Gamma \left[\theta_k + \alpha_k \{ \delta_k \phi_k - \gamma \phi'_k (\phi_k^\top w_k) - h_k \} \right]$$
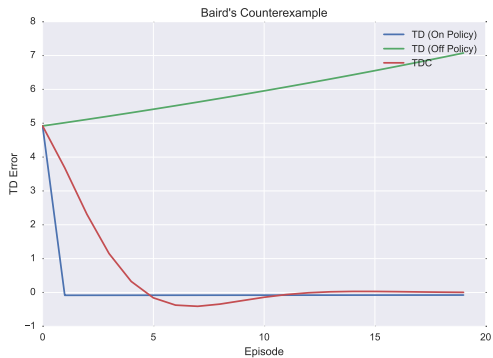
$$h_k = \delta_k - (\phi_k^\top w_k) \delta^2 V_\theta(s_k) w_k$$

- note that now $\phi_k = \nabla V_\theta(s_k)$
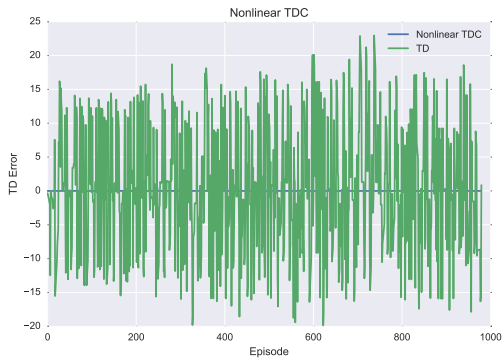- Also, Lee and Anderson, 2014 do this + control with small neural nets
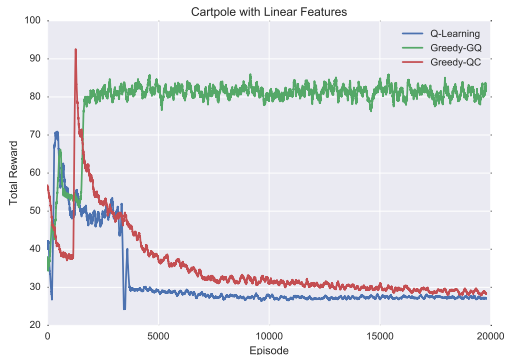
Quick Results

Quick Results

Baird's Counterexample

Quick Results

Nonlinear TDC — TD Error vs Episode

Simply Plug TDC or GTD2 into SARSA or Q-learning!

This is in the original feature space!

Simply Plug TDC or GTD2 into SARSA or Q-learning!

This is in the original feature space!

Cartpole with Linear Features