Fast Gradient-Descent Methods for TD Learning with Linear Function Approximation

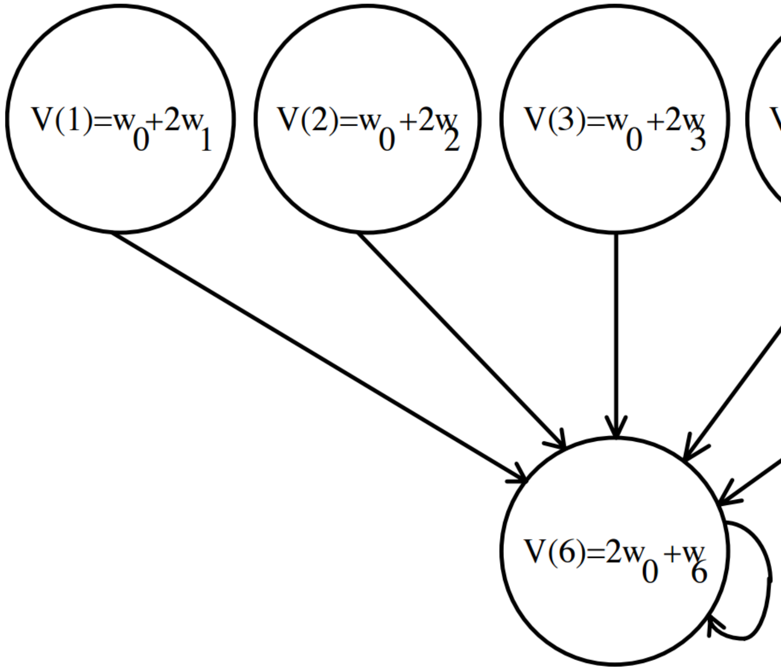
Michael Noseworthy

Sutton, Maei, Precup, Bhatnagar, Silver, Szepesvari, and Wiewiora (2009)

“Gradient” Methods

- TD is **not** a true gradient method
  - Convergence not as robust
  - Not guaranteed for **Off-Policy TD with Linear Function Approx.**
    - Non-gradient approaches (>> O(n))

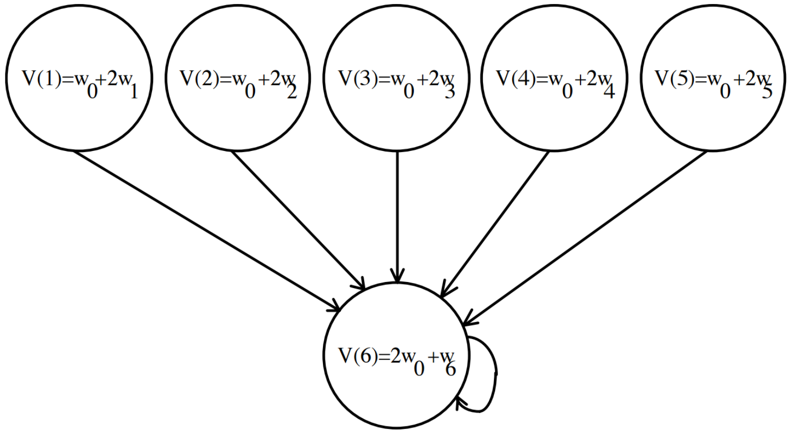
Baird's Counterexample (1995)



- Linear Function Approximation
  - Extra weight for generalization (without this we would converge)

Baird's Counterexample (1995)

- If all weights > 0 and V(6) >> V(1) … V(5)
  - All values diverge!



$$\Delta w = \alpha\left(R + \gamma V(x') - V(x)\right)\frac{\partial V(x)}{\partial w}$$

- W_0 increased 5 times for every time it is decreased

- There are non-pathological examples of divergence as well

Objective Functions (1)

- Mean Squared Error

$$MSE(\theta) = \sum_s d_s \left(V_\theta(s) - V(s)\right)^2 = ||V_\theta - V||_D^2$$

- No convergence guarantees with function approximation

Objective Functions (2)

- Mean Squared Error

$$MSE(\theta) = \sum_s d_s \left(V_\theta(s) - V(s)\right)^2 = ||V_\theta - V||_D^2$$

- Mean Squared Bellman Error

$$MSBE(\theta) = ||V_\theta - TV_\theta||_D^2 = ||V_\theta - R - \gamma P V_\theta||_D^2$$

- But Bellman Operator is unaware of our function approximator
  - What if TV is not representable?

Objective Functions (3)

- Mean Squared Error

$$MSE(\theta) = \sum_s d_s \left(V_\theta(s) - V(s)\right)^2 = ||V_\theta - V||_D^2$$

- Mean Squared Bellman Error

$$MSBE(\theta) = ||V_\theta - TV_\theta||_D^2 = ||V_\theta - R - \gamma P V_\theta||_D^2$$

- Mean Squared Projected Bellman Error

$$MSPBE(\theta) = ||V_\theta - \Pi T V_\theta||_D^2$$

- Projection Operator:

$$\Pi v = V_\theta \;\; where \; \theta = \operatorname*{argmin}_\theta ||V_\theta - v||_D^2$$

Projection Operator

- Weighted Least Squares Problem

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} ||\Phi\theta - v||_D^2$$

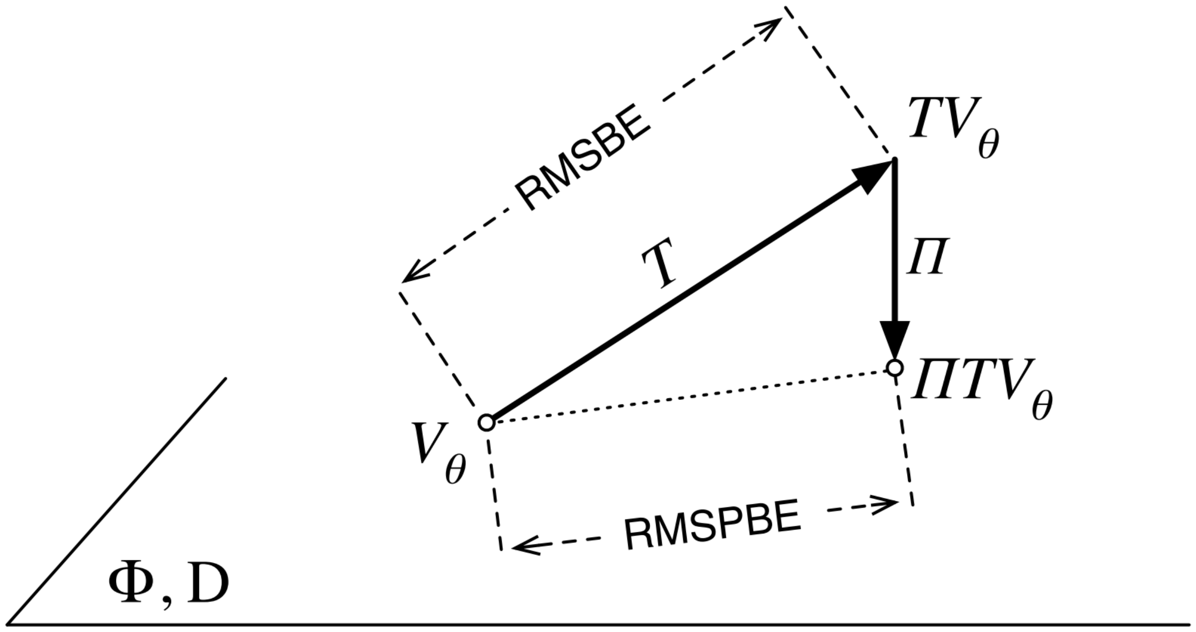
$$\hat{\theta} = (\Phi^T D \Phi)^{-1} \Phi^T D v$$

- Projection does not depend on v

$$\Pi v = V_\theta$$

$$\Pi v = \Phi \hat{\theta}$$

$$\Pi = (\Phi^T D \Phi)^{-1} \Phi^T D$$



GTD2

- **Goal**: Stochastic Gradient Algorithm
- **Step 1**: Rewrite MSPBE

$$MSPBE(\theta) = \mathbb{E}\left[\delta\phi\right]^T \mathbb{E}\left[\phi\phi^T\right] \mathbb{E}\left[\delta\phi\right]$$

- **Step 2**: Take the gradient

$$-\frac{1}{2}\nabla MSBPE(\theta) = \mathbb{E}\left[(\phi - \gamma\phi)\phi^T\right] \mathbb{E}\left[\phi\phi^T\right] \mathbb{E}\left[\delta\phi\right]$$

- **Step 3**: Sample / quasi-stationary estimate

$$\theta_{k+1} = \theta_k + \alpha_k(\phi_k - \gamma\phi_k')(\phi_k^T w)$$
$$w_{k+1} = w_k + \beta_k(\delta_k - \phi_k^T w_k)\phi_k$$

TDC (TD with Gradient Correction)

$$-\frac{1}{2}\nabla MSBPE(\theta) = \mathbb{E}\left[\delta\phi\right] - \gamma\mathbb{E}\left[\phi'\phi^T\right]w$$

$$\theta_{k+1} = \theta_k + \boxed{\alpha_k\delta_k\phi_k} - \boxed{\alpha_k\gamma\phi'_k(\phi_k^T w_k)}$$

TD Update

Gradient Correction to Follow MSPBE