Fast Gradient-Descent Methods for Temporal-Difference Learning with Linear Function Approximation

Algorithm derivation and convergence insights

Ahmed Touati

Reinforcement Learning class

Linear value approximation

- Value function: $V(s) = \mathbb{E}\{\sum_{t=0}^{\infty} \gamma^t r_{t+1} | s_0 = s\}$
- Linear approximation: $V_\theta(s) = \theta^T \phi_s$ where $\phi_s \in \mathbb{R}^n$ is a feature vector characterizing state s.
- Conventional linear TD algorithm:
  - We denote by (s_k, s'_k, r_k) the triples of state, next state, and reward with associated feature-vector random variables $\phi_k = \phi_{s_k}$ and $\phi'_k = \phi'_{s_k}$.
  - We define the temporal-difference error:

$$\delta_k = r_k + \gamma \theta^T \phi'_k - \theta^T \phi_k$$

  - parameters update:

$$\theta_{k+1} = \theta_k + \alpha_k \delta_k \phi_k$$

how derive another algorithm with better convergence rate

- Choice of objective function:
- Natural choice: closeness to the true value:

$$\mathrm{MSE}(\theta) = \sum_s d(s)(V_\theta(s) - V(s))^2 = ||V_\theta - V||_D^2$$

- Another option: use an objective function representing how closely the approximate value function satisfies the Bellman equation:

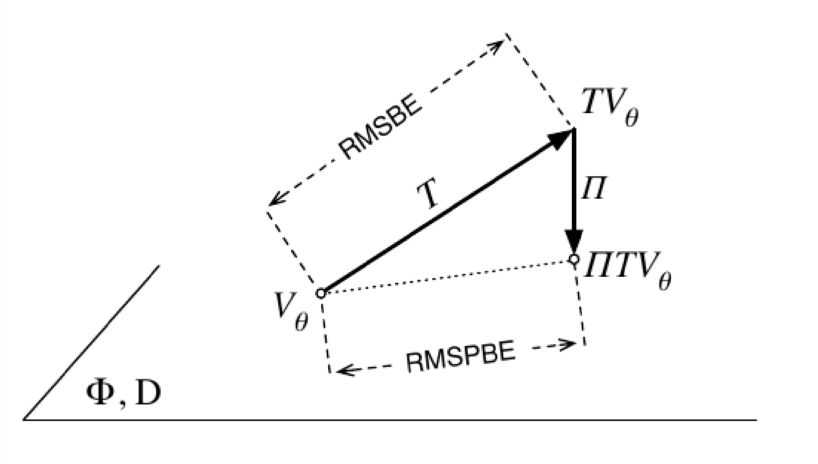
$$\mathrm{MSBE}(\theta) = ||V_\theta - TV_\theta||_D^2$$

  where $TV = R + \gamma PV$ is the Bellman operator

Projected Bellman error

- T takes you out the space. Π projects you back into
-

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



Projection operator

- Π takes any value function v and projects it to the nearest value function representable by the function approximator:
- $\Pi v = V_\theta$ where $\theta = \text{argmin}_\theta ||V_\theta - v||_D$
- If Φ is is the matrix whose rows are the ϕ_s, then,

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

Derivation of the algorithm

- $$\mathrm{MSPBE}(\theta) = ||V_\theta - \Pi T V_\theta||_D^2 = \mathbb{E}[\delta\phi]\mathbb{E}[\phi\phi^T]^{-1}\mathbb{E}[\delta\phi]$$
- $$-\frac{1}{2}\nabla_\theta \mathrm{MSPBE}(\theta) = \mathbb{E}[(\phi - \gamma\phi')\phi^T]\mathbb{E}[\phi\phi^T]^{-1}\mathbb{E}[\delta\phi]$$
- A trick: introduce a second set of weights $w \in \mathbb{R}^n$ to perform a stochastic approximation of the quantity $\mathbb{E}[\phi\phi^T]^{-1}\mathbb{E}[\delta\phi]$:

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

$$\mathbb{E}[\phi\phi^T]w = \mathbb{E}[\delta\phi]$$

$$w_{k+1} = w_k + \beta_k(\delta_k - \phi_k^T w_k)\phi_k$$

- Then:

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

Convergence proof: insights from previous class

TD converges to the TD fixedpoint, $\boldsymbol{\theta}_{TD}$, a biased but interesting answer

TD(0) update:

$$\boldsymbol{\theta}_{t+1} \doteq \boldsymbol{\theta}_t + \alpha\Big(R_{t+1} + \gamma\boldsymbol{\theta}_t^\top\boldsymbol{\phi}_{t+1} - \boldsymbol{\theta}_t^\top\boldsymbol{\phi}_t\Big)\boldsymbol{\phi}_t$$
$$= \boldsymbol{\theta}_t + \alpha\Big(R_{t+1}\boldsymbol{\phi}_t - \boldsymbol{\phi}_t\big(\boldsymbol{\phi}_t - \gamma\boldsymbol{\phi}_{t+1}\big)^\top\boldsymbol{\theta}_t\Big)$$

In expectation:

$$\mathbb{E}[\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t] = \boldsymbol{\theta}_t + \alpha(\mathbf{b} - \mathbf{A}\boldsymbol{\theta}_t),$$

where

$$\mathbf{b} \doteq \mathbb{E}[R_{t+1}\boldsymbol{\phi}_t] \in \mathbb{R}^n \quad \text{and} \quad \mathbf{A} \doteq \mathbb{E}\Big[\boldsymbol{\phi}_t\big(\boldsymbol{\phi}_t - \gamma\boldsymbol{\phi}_{t+1}\big)^\top\Big] \in \mathbb{R}^n \times \mathbb{R}^n$$

Fixedpoint analysis:

$$\begin{aligned} & \mathbf{b} - \mathbf{A}\boldsymbol{\theta}_{TD} = \mathbf{0} \\ \Rightarrow \quad & \mathbf{b} = \mathbf{A}\boldsymbol{\theta}_{TD} \\ \Rightarrow \quad & \boldsymbol{\theta}_{TD} \doteq \mathbf{A}^{-1}\mathbf{b} \end{aligned}$$

Guarantee:

$$\text{MSVE}(\boldsymbol{\theta}_{TD}) \leq \frac{1}{1-\gamma}\min_{\boldsymbol{\theta}} \text{MSVE}(\boldsymbol{\theta})$$

Convergence proof: intuitions

- There are two updates:

$$w_{k+1} = w_k + \beta_k(\delta_k - \phi_k^T w_k)\phi_k$$

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

- Let's set $\alpha_k = \eta\beta_k$ and $\rho_k^T = (d_k^T, \theta_k^T) \in \mathbb{R}^{2n}$ where $d_k = \frac{w_k}{\sqrt{\eta}}$. We obtain this single update:

$$\rho_{k+1} = \rho_k + \alpha_k\sqrt{\eta}(G_{k+1}\rho_k + g_{k+1})$$

where $G_{k+1} = \begin{bmatrix} -\sqrt{\eta}\phi_k\phi_k^T & \phi_k(\gamma\phi_k' - \phi_k)^T \\ (\phi_k - \gamma\phi_k')\phi_k^T & 0 \end{bmatrix}$ and $g_{k+1} = \begin{bmatrix} r_k\phi_k \\ 0 \end{bmatrix}$

Convergence proof

- In expectation:

$$\mathbb{E}[\rho_{k+1}|\rho_k] = \rho_k + \alpha_k(G\rho_k + g)$$

where $G = \mathbb{E}[G_k] = \begin{bmatrix} -\sqrt{\eta}C & -A \\ A^T & 0 \end{bmatrix}$ and $g = \mathbb{E}[g_k] = \begin{bmatrix} b \\ 0 \end{bmatrix}$ with

  - $A = \mathbb{E}[(\phi_k - \gamma\phi_k')\phi_k^T]$
  - $c = \mathbb{E}[\phi_k\phi_k^T]$
  - $b = \mathbb{E}[r_k\phi_k]$

- Fixed Point Analysis: $G\rho + g = 0 \Rightarrow -A\theta + b$: it the TD fixed point !!!

Questions?