A Non-asymptotic Analysis of Non-parametric Temporal-Difference Learning

E. Berthier, Z. Kobeissi and F. Bach

Inria & Ecole Normale Supérieure, PSL Research University,
Institut Louis Bachelier,
Paris, France

NeurIPS@Paris 2022
TD(0) with linear function approximation

Linear approximation of the value function:

\[ V^*(x) \simeq \xi^\top \varphi(x), \text{ for some } \xi \in \mathbb{R}^p. \]

TD(0): sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[ \xi_n = \xi_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] \varphi(x_n), \]
TD(0) with linear function approximation

Linear approximation of the value function:

\[ V^*(x) \simeq \xi^\top \varphi(x), \text{ for some } \xi \in \mathbb{R}^p. \]

**TD(0):** sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[ \xi_n = \xi_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] \varphi(x_n), \]

Converges under classical assumptions for stochastic approximation, \(\triangle\) to something different from \(V^*\) if \(V^* \notin \text{span}(\varphi_1, ..., \varphi_p)\).

[Tsitsiklis and Van Roy, 1997], [Bhandari et al., 2018]
TD(0) with linear function approximation

Linear approximation of the value function:

\[ V^*(x) \simeq \xi^\top \varphi(x), \text{ for some } \xi \in \mathbb{R}^p. \]

TD(0): sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[ \xi_n = \xi_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] \varphi(x_n), \]

Converges under classical assumptions for stochastic approximation, \( \Delta \) to something different from \( V^* \) if \( V^* \notin \text{span}(\varphi_1, \ldots, \varphi_p) \).

[Tsitsiklis and Van Roy, 1997], [Bhandari et al., 2018]

*Can we fix this with a universal approximator?*
**TD(0) with linear function approximation**

Linear approximation of the value function:

\[ V^*(x) \simeq \xi^\top \varphi(x), \text{ for some } \xi \in \mathbb{R}^p. \]

**TD(0):** sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[ \xi_n = \xi_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] \varphi(x_n), \]

Converges under classical assumptions for stochastic approximation, ! to something different from \(V^*\) if \(V^* \not\in \text{span}(\varphi_1, \ldots, \varphi_p, \ldots)\).

[Tsitsiklis and Van Roy, 1997], [Bhandari et al., 2018]

*Can we fix this with a universal approximator?*
Non-parametric TD(0)

Sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[
V_n = V_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] K(x_n, \cdot),
\]

where \(K\) is the reproducing kernel of an RKHS \(\mathcal{H} \subset L^2\).
Non-parametric TD(0)

Sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[
V_n = V_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] K(x_n, \cdot),
\]

where \(K\) is the reproducing kernel of an RKHS \(\mathcal{H} \subset L^2\).

▶ the iterates are in \(\mathcal{H}\) (functional space)
Non-parametric TD(0)

Sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[
V_n = V_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] K(x_n, \cdot),
\]

where \(K\) is the reproducing kernel of an RKHS \(\mathcal{H} \subset L^2\).

- the iterates are in \(\mathcal{H}\) (functional space)
- recovers linear approximation with \(K(x, y) = \varphi(x)^\top \varphi(y)\)
Non-parametric TD(0)

Sample a transition \((x_n, r(x_n), x'_n)\) and update:

\[
V_n = V_{n-1} + \rho_n \left[ r(x_n) + \gamma V_{n-1}(x'_n) - V_{n-1}(x_n) \right] K(x_n, \cdot),
\]

where \(K\) is the reproducing kernel of an RKHS \(\mathcal{H} \subset L^2\).

- the iterates are in \(\mathcal{H}\) (functional space)
- recovers linear approximation with \(K(x, y) = \varphi(x)\top \varphi(y)\)
- universal kernel such that \(\overline{\mathcal{H}} = L^2\) (Sobolev kernel).
  \[\rightarrow\text{convergence to } V^* \text{ in } L^2\text{-norm, even if } V^* \notin \mathcal{H}.\]
Main convergence result

Theorem

Assume that for some \( \theta \in (-1, 1] \):

\[
\| \Sigma^{-\theta/2} V^* \|_{\mathcal{H}} < +\infty .
\]  

(source condition)

Then with suitable regularization, step size and averaging scheme:

\[
\mathbb{E} \left[ \| \overline{V}_n - V^* \|_{L^2}^2 \right] = O \left( (\log n)^2 n^{-\frac{1+\theta}{2+\theta}} \right).
\]
Main convergence result

Theorem

Assume that for some $\theta \in (-1, 1]$:

$$\|\Sigma^{-\theta/2} V^*\|_{\mathcal{H}} < +\infty.$$  

(source condition)

Then with suitable regularization, step size and averaging scheme:

$$E \left[ \|\overline{V}_n - V^*\|_{L^2}^2 \right] = O \left( (\log n)^2 n^{-\frac{1+\theta}{2+\theta}} \right).$$

▶ $\theta = 0$: $V^* \in \mathcal{H}$ recovers known $1/\sqrt{n}$ parametric rate.
▶ $\theta \in (0, 1]$: stronger assumption, faster rate.
▶ $\theta = -1$: $V^* \in L^2$, only asymptotic convergence.
▶ $\theta \in (-1, 0)$: $V^* \notin \mathcal{H}$, weaker assumption, slower rate.
Main convergence result

**Theorem**

Assume that for some $\theta \in (-1, 1]$:

$$\| \Sigma^{-\theta/2} V^* \|_H < +\infty.$$  \hspace{1cm} \text{(source condition)}

Then with suitable regularization, step size and averaging scheme:

$$\mathbb{E} \left[ \| \bar{V}_n - V^* \|_{L^2}^2 \right] = O \left( (\log n)^2 n^{-\frac{1+\theta}{2+\theta}} \right).$$

Theorem proved in the *i.i.d.* sampling setting.

Extends to sampling from a Markov chain with exponential mixing, with an additional boundedness assumption.
Numerical experiment

Sobolev kernel of regularity $s$ on the 1d torus.
Source condition $\theta$: decrease of Fourier coefficients of $V^*$.

- Predicted slope: $-0.43$
- Observed slope: $-0.58$

$\rightarrow$ Influence of mixing in the constants.
See you at the poster session!