@zblu Reinforcement Learning Theory

INF8250AE Course Notes

RL 7: Value Function Approximation

Function Approximation, AVI, BRM, PBE, LSTD

Key Topics: Function approximation in RL, Approximate Value Iteration, Bellman residual minimization, Projected Bellman Error, LSTD
Reference: `https://amfarahmand.github.io/IntroRL/`

Notes on Reinforcement Learning – Oct 2025

POLYTECHNIQUE MONTRÉAL

Mila

Reinforcement Learning 5: Value Function Approximation

Zhuobie

2026 年 4 月 13 日

Lecture Goal and Roadmap

This lecture studies *value function learning under function approximation*. We introduce several loss functions (population and sample-based) and corresponding algorithms for

- approximating V^π, Q^π, V^*, or Q^* in a restricted function space $\mathcal{F}$,
- using either batch data (offline RL) or online interaction.

High-level structure of the lecture.

1. Motivation and basic examples of function approximation (FA).
2. Population-level value computation in a function space (AVI, BRM, PBE, LSTD).
3. Batch RL: turning those population objectives into empirical ones (FQI/FVI, empirical BRM, LSTD/LSPI).
4. Online RL with FA: TD-like stochastic approximation and semi-gradient viewpoint.

All later algorithms (DQN, LSPI, TD with FA, etc.) are instantiations of these basic ideas.

1. Motivation and Examples

1.1 Why Function Approximation?

In many real-world problems, the state-action space $\mathcal{X} \times \mathcal{A}$ is huge:

- $\mathcal{X} \subset \mathbb{R}^d$ continuous ($d \geq 1$): we cannot tabulate an arbitrary function on $\mathbb{R}^d$.
- $\mathcal{X}$ finite but enormous (millions or billions of states).

So we approximate key objects (value functions, policies, models) with a parametric or structured representation that fits in memory and can generalize.

Three major approximation targets in RL.

- **Value function approximation:** $\hat{V}^\pi \approx V^\pi$ or $\hat{V}^* \approx V^*$.
- **Policy approximation:** $\hat{\pi}_\theta \approx \pi^*$ (next lecture).
- **Model approximation:** $\hat{P} \approx P$ (model-based RL).

This lecture concentrates on value function approximation.

1.2 Linear Function Approximation

We fix basis functions $\phi_1, \ldots, \phi_p$ and approximate

$$\hat{V}(x) = \phi(x)^\top w = \sum_{i=1}^{p} \phi_i(x) w_i,$$

with $w \in \mathbb{R}^p$ and feature map $\phi : \mathcal{X} \to \mathbb{R}^p$. The *value function space* is

$$\mathcal{F} = \{x \mapsto \phi(x)^\top w : w \in \mathbb{R}^p\}. \tag{1}$$

It is a linear subspace in terms of the parameters w (not necessarily linear in the state x).

1.3 Approximation Error vs. Estimation Error

When learning a function V from data within a function class $\mathcal{F}$, the total error can be decomposed into two fundamental components: *approximation error* and *estimation error*.

Definition 1.1 (Approximation Error (Bias)) *The approximation error measures how well the function class $\mathcal{F}$ can represent the target function V:*

$$\inf_{\hat{V} \in \mathcal{F}} |\hat{V} - V|_\infty.$$

This error depends on:

- the structure of the function class $\mathcal{F}$,
- the regularity of V.

If V is L-Lipschitz and $\mathcal{F}_\varepsilon$ is a piecewise constant class with resolution ε, then

$$\inf_{\hat{V}\in\mathcal{F}_\varepsilon} |\hat{V} - V|_\infty \leq L\varepsilon.$$

Thus, smoother functions admit smaller approximation error.

Definition 1.2 (Estimation Error (Variance)) *Given n samples, we estimate $\hat{V}$ from data. The statistical error behaves as*

$$O\left(\sqrt{\frac{\log|\mathcal{F}|}{n}}\right),$$

where $|\mathcal{F}|$ quantifies the complexity of the function class.

In contrast, for mean estimation (a single parameter),

$$\mathbb{E}[|m_n - \mu|] = O\left(\frac{1}{\sqrt{n}}\right),$$

which corresponds to the case $|\mathcal{F}| \approx 1$.

1.3.1 Trade-off

The two errors exhibit opposite behavior:

- Increasing the richness of $\mathcal{F}$ reduces approximation error,
- but increases estimation error through larger complexity.

Balancing these two terms is the key principle in statistical function estimation.

1.4 Other Function Approximators

Beyond linear FA:

- Deep neural networks (DNNs),
- RKHS / kernel methods,
- decision trees and random forests,
- local methods (kNN, kernel smoothing, etc.).

2. Value Function Computation with a Function Approximator

We now assume a function space $\mathcal{F}$ is fixed and ask: how to compute an approximate value function using it? Most approaches define a **population loss function** and then try to minimize it.

2.1 Lp Norms and Distances Between Functions

For a probability distribution $\nu \in \mathcal{M}(\mathcal{X})$ and a function $V : \mathcal{X} \to \mathbb{R}$, the $L_p(\nu)$ norm ($1 \leq p < \infty$) is

$$\|V\|_{p,\nu}^p \triangleq \int_{\mathcal{X}} |V(x)|^p \, d\nu(x). \tag{2}$$

The sup norm is

$$\|V\|_\infty \triangleq \sup_{x\in\mathcal{X}} |V(x)|.$$

We will often define distances as

$$d(V_1, V_2) = \|V_1 - V_2\|_{p,\mu},$$

for some distribution μ on $\mathcal{X}$ or $\mathcal{X} \times \mathcal{A}$.

2.2 Approximate Value Iteration (Population)

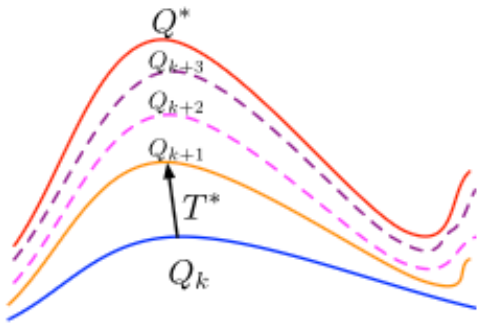
Recall exact value iteration (VI) with Bellman operator T (either T^π or T^*):

$$V_{k+1} \leftarrow TV_k.$$

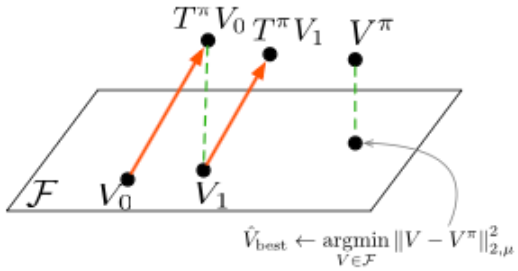
T is a γ-contraction in $\|\cdot\|_\infty$, so VI converges to the unique fixed point (V^π or V^*).

Idea of Approximate Value Iteration (AVI): at each iteration, instead of applying T exactly, we *project back* to $\mathcal{F}$:

$$V_{k+1} \leftarrow \arg\min_{V\in\mathcal{F}} \|V - TV_k\|_{p,\mu}. \tag{4}$$



(a) Contraction



(b) Approximate Value Iteration (Population)

2.3 Bellman Residual Minimization (Population)

Geometric View.

We aim to solve the Bellman fixed-point equation

$$V = T^\pi V,$$

whose unique solution is V^π due to the contraction property of T^π. Equivalently, V^π is the unique function satisfying

$$\|V - T^\pi V\| = 0.$$

In practice, we restrict the search to a function class $\mathcal{F}$ (e.g., a linear subspace). Since $V^\pi \notin \mathcal{F}$ in general, we cannot achieve zero residual. Thus, Bellman Residual Minimization (BRM)

$$\min_{V\in\mathcal{F}} \|V - T^\pi V\|$$

seeks the function in $\mathcal{F}$ whose Bellman residual is smallest.

The key question is: *why does minimizing* $\|V - T^\pi V\|$ *lead to a good approximation of* V^π*?* This is justified by the following theoretical bound.

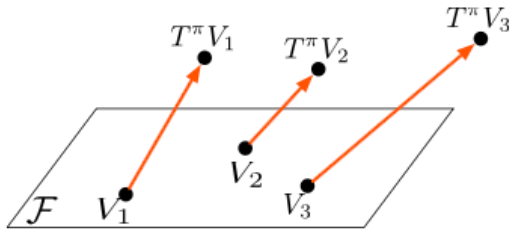


Figure 2: Bellman Residual Minimization: geometric intuition.

Theoretical View.

The Bellman error controls the value error through the contraction property of T^π.

Supremum norm bound. For any V, we have

$$\|V - V^\pi\|_\infty \leq \frac{\|V - T^\pi V\|_\infty}{1-\gamma}.$$

Thus, the Bellman residual serves as a surrogate loss for the true error.

Stationary distribution and L_2 bound. Let ρ^π be the stationary distribution of the Markov chain induced by policy π, i.e.,

$$\rho^\pi = \rho^\pi P^\pi.$$

Under ρ^π, the Bellman operator T^π is a γ-contraction w.r.t. $\|\cdot\|_{2,\rho^\pi}$.

As a result, for any V and $p \geq 1$, we have

$$\|V - V^\pi\|_{1,\rho^\pi} \leq \frac{\|V - T^\pi V\|_{p,\rho^\pi}}{1-\gamma}.$$

This provides a more practical bound compared to the supremum norm, as it weights states according to their long-run visitation frequencies.

Proof

For any V, we write

$$V - V^\pi = (V - T^\pi V) + (T^\pi V - T^\pi V^\pi).$$

Taking norms and using the contraction property,

$$\|V - V^\pi\| \leq \|V - T^\pi V\| + \gamma \|V - V^\pi\|.$$

Rearranging yields

$$(1-\gamma)\|V - V^\pi\| \leq \|V - T^\pi V\|.$$

Under the stationary distribution ρ^π, the expectation of the second term is preserved due to

$$\rho^\pi = \rho^\pi P^\pi,$$

which ensures the same contraction argument holds in $\|\cdot\|_{2,\rho^\pi}$ (and more generally $\|\cdot\|_{p,\rho^\pi}$).

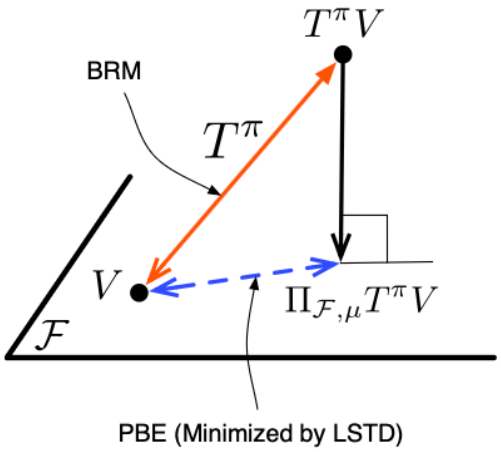
2.4 Projected Bellman Error（Population）

2.4.1 Problem Definition

Given a function class $\mathcal{F}$, the Bellman update $T^\pi V$ generally lies outside $\mathcal{F}$. To obtain a well-defined approximation within $\mathcal{F}$, we project $T^\pi V$ back onto $\mathcal{F}$ and seek a fixed point:

$$V = \Pi_{\mathcal{F},\mu} T^\pi V.$$

This defines the *projected Bellman equation*, where the update is constrained to remain in $\mathcal{F}$.



The projection operator is defined as

$$\Pi_{\mathcal{F},\mu} V := \arg\min_{V' \in \mathcal{F}} \|V' - V\|_{2,\mu}, \quad \|f\|_{2,\mu}^2 = \sum_x \mu(x) f(x)^2,$$

where μ is a probability distribution over states.

$$\min_{V \in \mathcal{F}} \|V - \Pi_{\mathcal{F},\mu} T^\pi V\|_{2,\mu}.$$

Remark

Problem structure: Nested (coupled) optimization. We can equivalently see PBE as solving two coupled problems:

$$\begin{cases} \tilde{V}(V') \leftarrow \arg\min_{V'' \in \mathcal{F}} \|V'' - T^\pi V'\|_{2,\mu}^2, \\ V \leftarrow \arg\min_{V' \in \mathcal{F}} \|V' - \tilde{V}(V')\|_{2,\mu}^2. \end{cases}$$

Here $\tilde{V}(V')$ is the projection of $T^\pi V'$ back onto $\mathcal{F}$.

Regularized variants: add $\lambda J(V')$ and $\lambda J(V'')$ to control overfitting when $\mathcal{F}$ is large.

2.4.2 Solving approaches

Assumption: Linear Function Approximator

Assume a linear function approximator:

$$\mathcal{F} = x \mapsto \phi(x)^\top w : w \in \mathbb{R}^p, \qquad V = \Phi w, \tag{1}$$

where $\Phi \in \mathbb{R}^{N \times p}$ is the feature matrix.

The projected Bellman equation becomes:

$$\Phi w = \Pi_{\mathcal{F},\mu} T^\pi(\Phi w). \tag{2}$$

Thus, solving PBE reduces to solving a linear system in w.

Remark

[Solving Linear Systems] Consider an overdetermined linear system $Ax \approx b$.

Optimization	**Direct Solution**	**Fixed-Point Iteration**
$\nabla = A^\top(Ax - b)$	$x^* = (A^\top A)^{-1} A^\top b$	$x_{k+1} = (1-\alpha)x_k + \alpha[(I - A)x_k + b]$
$x^* = \arg\min_x \|Ax - b\|_2^2$	$A^\top A x = A^\top b$	$x_{k+1} = x_k - \alpha A^\top(Ax_k - b)$

In what follows, we focus on the direct and fixed-point approaches.

Projection Operator and linear reformulation of PBE Define the μ-weighted inner product:

$$\langle V_1, V_2 \rangle_\mu = V_1^\top M V_2, \qquad M = \text{diag}(\mu). \tag{3}$$

The projection onto $\mathcal{F}$ is:

$$\Pi_{\mathcal{F},\mu} V = \arg\min_{w \in \mathbb{R}^p} \|\Phi w - V\|_{2,\mu}^2, \tag{4}$$

which yields

$$\Pi_{\mathcal{F},\mu} V = \Phi(\Phi^\top M \Phi)^{-1} \Phi^\top M V. \tag{5}$$

The Bellman operator satisfies:

$$T^\pi(\Phi w) = r^\pi + \gamma P^\pi \Phi w. \tag{6}$$

Substituting into the projected Bellman equation:

$$\Phi w = \Phi(\Phi^\top M \Phi)^{-1} \Phi^\top M \big(r^\pi + \gamma P^\pi \Phi w\big).$$

Approach 1: Direct Solution (LSTD: Least Squares Temporal Difference) Multiplying both sides by $\Phi^\top M$:

$$\Phi^\top M \Phi w = \Phi^\top M \big(r^\pi + \gamma P^\pi \Phi w\big).$$

Rearranging:

$$\Phi^\top M \big(r^\pi + \gamma P^\pi \Phi w - \Phi w\big) = 0,$$

which leads to

$$\underbrace{\left(\Phi^\top M(\Phi - \gamma P^\pi \Phi)\right)}_{A} w = \underbrace{\Phi^\top M r^\pi}_{b}.$$

Thus,

$$w = \left[\Phi^\top M(\Phi - \gamma P^\pi \Phi)\right]^{-1} \Phi^\top M r^\pi.$$

This is the **Least Squares Temporal Difference (LSTD)** solution.

Remark

Geometric intuition: the LSTD solution makes the Bellman residual $T^\pi V(w) - V(w)$ orthogonal (in $\langle \cdot, \cdot \rangle_\mu$) to all basis functions ϕ_i:

$$\langle \phi_i, T^\pi V(w) - V(w) \rangle_\mu = 0, \quad i = 1, \ldots, p.$$

Approach 2: Fixed-Point Iteration We can also solve the projected Bellman equation via iteration:

$$V_{k+1} = (1 - \alpha) V_k + \alpha \Pi_{\mathcal{F}, \rho^\pi} T^\pi V_k. \tag{7}$$

Under the stationary distribution ρ^π, the operator is a contraction, and the iteration converges.
In parameter form, this becomes:

$$w_{k+1} = (1 - \alpha) w_k + \alpha (\Phi^\top D_\pi \Phi)^{-1} \Phi^\top D_\pi \left(r^\pi + \gamma P^\pi \Phi w_k\right).$$

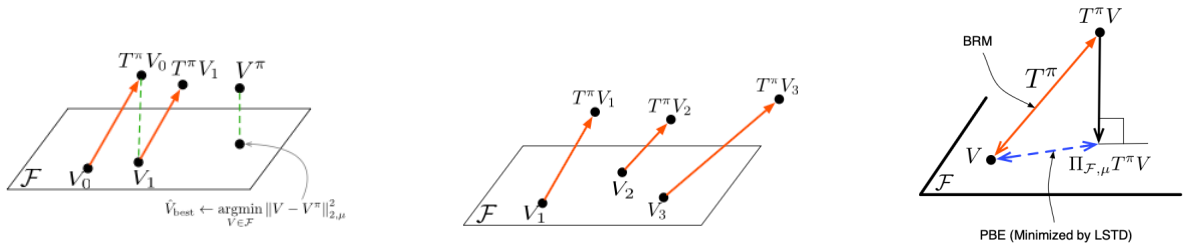
where $D_\pi = \text{diag}(\rho^\pi)$.

Error Bound for the LSTD Solution If ρ^π is the stationary distribution of π, then

$$|V_{TD} - V^\pi| * 2, \rho^\pi \leq \frac{|\Pi * \mathcal{F}, \rho^\pi V^\pi - V^\pi|_{2, \rho^\pi}}{\sqrt{1 - \gamma^2}}. \tag{8}$$

Implication. This bound shows that the TD solution is within a constant factor of the best possible approximation in $\mathcal{F}$. In particular, if $\mathcal{F}$ represents V^π well, then V_{TD} is also close to V^π.

3. Summary



(a) AVI geometry (b) BRM geometry (c) PBE geometry

Figure 3: Summary of core geometric ideas in value function approximation.