AVERAGE COST ERROR BOUNDS OF FINITE STATE APPROXIMATIONS OF MARKOV DECISION PROCESSES

Fatih Furkan GURTURK

1. Introduction

Modeling sequential decision making under uncertainty control and optimization problems as Markov Decision Processes (MDPs) is used in a broad range of applications. However, there are computational considerations with large state and action spaces for Markov Decision Processes (MDPs). Using approximate state space instead of large number of states, truncating the state space and quantization of uncountable state space are ways to approach this problem.

One important question is whether this approximation results in an approximation to the optimal value of original model. As number of finite points increase, we may expect the model to have similar results for uncountable state space. The following analysis gives results for which conditions finite state space converges to the true cost functions of uncountable state space, as number of finite state increases. Another question is how much difference is resulted from quantization process, and can we find a bound for performance loss.

Discounting is applied in models when values of the future costs is relatively less important than the current costs. Discount factor represents the ratio of the future value of a gain to its present value. Then the objection function can be written as:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t c_t(X_t, U_t)\right] for\ some\ \beta \in (0,1).$$

When there is no difference of utility in future value and current value, discounted model may not represent the desired model. It happens especially when decision maker makes frequent decisions, decision maker may prefer to calculate average expected cost. In this setting we would like to find the policy u minimizing the average cost in infinite horizon.

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1} c_t(X_t, U_t)\right].$$

There are relations between the discounted cost and average cost in the original problem. For example, in the discounted cost problem, as the β goes to 1, we eliminate the discount. Then, when we find the optimal policy for very large β, we can induce the same policy for average cost problem. Optimizing the discounted cost with large discount factor is a way of solving optimal average cost problems. In the discussion part, such similarities or approaches between discounted cost and the average cost for approximated MDP will be discussed for approximate MDPs. We will try to compare the results and check if there can be a relation such as obtaining the bounds for large discount factor to get a result for average cost criteria.

2. Problem Setting

MDP is specified by following components: (i) the state space $\mathbb{X}$ and action space $\mathbb{U}$, (ii) the transition probability p(· |x, u) on $\mathbb{X}$ defined on $\mathbb{X} \times \mathbb{U}$ which gives the probability of next state given the current state and action are (x, u) (iii) one-stage cost functions defined as $c_t : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$, where in general $c_t = c$. The assumptions we have are:

Assumption 1:

(a) The one-stage cost function c is continuous and bounded.
(b) The transition probabilities $p(\cdot\,|x, u)$ are weakly continuous in (x, u).
(c) State space $\mathbb{X}$ and action space $\mathbb{U}$ are compact.

For the derivation for upper bound on the performance loss due to discretization in terms of the size of the set $\mathbb{X}$, some definitions will be given:

For each $g \in C_b(\mathbb{X})$ the Lipschitz semi-norm of g is:

$$\|g\|_{Lip} := \sup_{z\neq y} \frac{|g(z) - g(y)|}{d_Z(z, y)}$$

If $\|g\|_{Lip}$ is finite, then g is called Lipschitz continuous with Lipschitz constant $\|g\|_{Lip}$. $Lip(\mathbb{X})$ denotes the set of all Lipschitz continuous functions on $\mathbb{X}$, i.e.,

$$Lip(\mathbb{X}) := \{g \in C_b(\mathbb{X}) : \|g\|_{Lip} < \infty\}$$

and $Lip(\mathbb{X}, K)$ denotes the set of all $g \in Lip(\mathbb{X})$ with $\|g\|_{Lip} \leq K$. The Kantorovich distance (The Wasserstein distance of order 1) between two probability measures ζ and ξ over $\mathbb{X}$ is defined as:

$$W_1(\zeta, \xi) := sup\left\{\left|\int_{\mathbb{X}} g\, d\zeta - \int_{\mathbb{X}} g\, d\xi\right| : g \in Lip(\mathbb{X}, 1)\right\}$$

Assumption 2:

(a) The one stage cost function c satisfies $c(\cdot, a) \in Lip(\mathbb{X}, K_1)$ for all $u \in \mathbb{U}$ for some K_1.
(b) The transition probabilities satisfy $W_1\big(p(\cdot\,|\mathrm{z}, \mathrm{a}), p(\cdot\,|\mathrm{y}, \mathrm{a})\big) \leq K_2 d_Z(z, y)$ for all $u \in \mathbb{U}$ for some K_2.
(c) $\mathbb{X}$ is an infinite compact subset of $\mathbb{R}^d$ for some $d \geq 1$, equipped with the Euclidean norm.

3. Performance Loss Due to Discretization

In this section error bounds for discretization will be studied. The analysis will be given for average cost case and will be compared with the discounted cost case. To begin with the value function, we have $K_2\beta < 1$, so that the value function is in $Lip(\mathbb{X})$. This will result in the discounted cost is in $Lip(\mathbb{X}, K)$, where $K = K_1 \frac{1}{1-\beta \mathrm{K}_2}$. The derivation will be for the average cost case and in the discussion part we will compare the results. It starts with the derivation of a rate of convergence in terms of moduli of continuity ω_{c} and ω_{p} of $c(x, u)$ and $p(\cdot\,|x, u)$, and the total variation distance will be used in ω_p. Then the rate of convergence is used under some structural assumptions on ω_c and ω_p to get the error bounds. The modulus of continuity ω_{J^*} of the value function J^* will be used in average cost case, where in the discounted case the Kantorovich distance is used. This insertion will be established if ω_c and ω_p are affine functions (i.e., $\omega_c(r) = K_1 r + L_1$ and $\omega_p(r) = K_2 r + L_2$) using the dual formulation of Kantorovich distance.

$$W_1(\mu, \nu) \coloneqq \sup_{\psi(x)-\varphi(y) \leq d_{\mathbb{X}}(x,y)} \left| \int_{\mathbb{X}} \psi(z)\mu(dz) - \int_{\mathbb{X}} \varphi(z)\,\nu(dz) \right|$$

The modulus of continuity functions in the z variable for $c(x, u)$ and $p(\cdot\,|x, u)$ are:

$$\omega_c(r) = \sup_{u \in \mathbb{U}} \sup_{y,z \in \mathbb{X}} |\, c(z, u) - c(y, u)|$$

$$\omega_p(r) = \sup_{u \in \mathbb{U}} \sup_{y,z \in \mathbb{X}} \|p(\cdot\,|z, u) - p(\cdot\,|y, u)\|$$

These values converge to zero when r goes to zero and when ω_c and ω_p are linear $c(x, u)$ and $p(\cdot\,|x, u)$ are uniformly Lipschitz in z. The rate of convergence will be derived in terms of ω_c and ω_p and then, the convergence rate will be computed for the Lipschitz case. To obtain convergence rates for the average cost, for each $n \geq 1$, let $d_n = 2\alpha(1/\mathrm{n})^{(1/\mathrm{d})}$, where α is a constant larger than zero that satisfies the maximum of minimum distance.

Lemma 1 For all $t \geq 1$, we have

$$\sup_{y\in\mathbb{X}}\left\|p^t(\cdot\,|y,\mathrm{f(y)}) - q_n^t(\cdot\,|y,\mathrm{f(y)})\right\| \le t\omega_p(d_n).$$

The optimal policy f^* is obtained by extending the optimal policy f_n^* to the original MDP where the real state space is used. The optimal average cost ρ_n obtained by approximation is also converges to ρ of the original MDP.

Theorem 1 For all $t \ge 1$, there exist positive real numbers R and $K < 1$ such that

$$\left|\rho_{\mathrm{f}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right| \le 4\|c\|RK^t + 2\omega_\mathrm{c}(\mathrm{d_n}) + 2\|c\|t\omega_\mathrm{p}(\mathrm{d_n})$$

Proof Here not all the proof is presented but the idea is summarized. The average cost of optimal policy obtained by extending the optimal policy to all state space and the cost converges to the value function of the original MDP. From other result at [3] it is given:

$$\left|\rho_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right| \le \left|\rho_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^n_{\tilde{\mathrm{f}}_\mathrm{n}^{\,n}{}^*}\right| + \left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*}\right| + \left|\rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right|$$

The first term on the right side is:

$$\left|\rho_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^n_{\tilde{\mathrm{f}}_\mathrm{n}^{\,n}{}^*}\right| \le sup\left|\rho^n_{\mathrm{f}^*} - \rho_{\mathrm{f}_\mathrm{n}^*}\right|$$

$$\left|\rho_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^n_{\tilde{\mathrm{f}}_\mathrm{n}^{\,n}{}^*}\right| \le 2RK^t\|c\| + \|c\| \sup_{y\in\mathbb{X}}\left\|q_n^t(\cdot\,|y,\mathrm{f(y)}) - p^t(\cdot\,|y,\mathrm{f(y)})\right\|$$

$$\left|\rho_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^n_{\tilde{\mathrm{f}}_\mathrm{n}^{\,n}{}^*}\right| \le 2RK^t\|c\| + \|c\|\, t\omega_p(d_n)$$

The second term on the right side is:

$$\left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*}\right| \le \left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \tilde{\rho}^n_{\mathrm{f}^*}\right| + \left|\tilde{\rho}^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*} - \rho^n_{\mathrm{f}^*}\right|$$

$$\left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*}\right| \le 2\sup_{f\in\mathbb{F}}\left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \tilde{\rho}^n_{\mathrm{f}^*}\right|$$

$$\left|\rho^\mathrm{n}_{\tilde{\mathrm{f}}_\mathrm{n}^*} - \rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*}\right| \le 2\omega_c(d_n)$$

The last term on the right side is:

$$\left|\rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right| = \left|\inf_{f\in\mathbb{F}}\rho^\mathrm{n}_\mathrm{f} - \inf_{f\in\mathbb{F}}\rho_\mathrm{f}\right| \le \sup_{f\in\mathbb{F}}|\rho^\mathrm{n}_\mathrm{f} - \rho_\mathrm{f}|$$

$$\left|\rho^\mathrm{n}_{\mathrm{f}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right| \le 2RK^t\|c\| + \|c\|\, t\omega_p(d_n)$$

Combining these three results gives

$$\left|\rho_{\mathrm{f}_\mathrm{n}^*} - \rho_{\mathrm{f}^*}\right| \le 4\|c\|RK^t + 2\omega_\mathrm{c}(\mathrm{d_n}) + 2\|c\|t\omega_\mathrm{p}(\mathrm{d_n})$$

Now with some assumptions on ω_c and ω_p we can calculate a convergence rate. One structural assumption can be linearity. Having linear ω_c and ω_p results in uniform Lipschitz continuous $c(x,u)$ and $p(\cdot\,|x,u)$ in x. Having $\omega_c(r) = K_1 r$ and $\omega_p(r) = K_2 r$ means

$|c(z,u) - c(y,u)| \leq K_1 d_{\mathbb{X}}(z,y)$ for all $z, y \in \mathbb{X}$, $u \in \mathbb{U}$ and $\|p(\cdot|z,u) - p(\cdot|y,u)\| \leq K_2 d_{\mathbb{X}}(z,y)$ for all $z, y \in \mathbb{X}$, $u \in \mathbb{U}$. With this structural assumption we obtain:

$$\left|\rho_{f_n^*} - \rho_{f^*}\right| \leq 4\|c\|RK^t + 4K_1\alpha(1/n)^{1/d} + 4\|c\|K_2\alpha(1/n)^{1/d}t\,.$$

We will eliminate the dependency on t so that the convergence rate as a function of only n. The upper bound in the above equation can be minimized with respect to t for each n. By defining the constants $I_1 = 4\|c\|R$, $I_2 = 4K_1\alpha$, and $I_3 = 4\|c\|K_2\alpha$ the upper bound will be:

$$I_1K^t + I_2(1/n)^{1/d} + I_3(1/n)^{1/d}t$$

The derivative of the above has zero values at:

$$t'(n) = ln\left(\frac{n^{1/d}}{I_4}\right)\frac{1}{ln\left(\frac{1}{K}\right)}$$

where $I_4 = \frac{I_3}{I_1 ln\left(\frac{1}{K}\right)}$. So, by putting the zero valued for each n for t, we will obtain:

$$\left|\rho_{f_n^*} - \rho_{f^*}\right| \leq (I_1I_4 + I_2)(1/n)^{1/d} + \frac{I_3}{ln(1/K)}(1/n)^{1/d}ln\left(\frac{n^{1/d}}{I_4}\right).$$

This result gives a time independent error bound value for approximation.

4. Discussion

The average cost of optimal policy obtained by extending the optimal policy to all state space and the cost converges to the value function of the original MDP. So, it is sufficient to compute the optimal policy of MDP with n discrete state spaces and extend the optimal policy obtained for sufficiently large n to the original MDP. Then the resulting cost value will be in the bounds of the result given in Section 3.

The error bounds resulted from average cost case analysis have very different characteristics than the discounted cost case. Even if we can obtain an optimal discounted cost policy and apply to average cost policy, we may not able to conclude the bounds.

Error bound for state approximation on discounted cost policy is given as:

$$\left|\rho_{f_n^*} - \rho_{f^*}\right| \leq \left(\left(\beta K_2 + \frac{\beta+3}{(1-\beta)^2}\right)\|\rho_{f^*}\|_{Lip} + \frac{K_1}{1-\beta}\right)2\alpha(1/n)^{1/d}.$$

The result found for average cost model is very much different. When we think about the approach where we deduce the optimal policy for average cost using the optimal policy for β values very close to 1, we observe it is not applicable to this result.

Getting a β value close to 1 will result in βK_2 to disappear next to $\frac{\beta+3}{(1-\beta)^2}$. So the first term can be approximated as $\frac{\beta+3}{(1-\beta)^2}\|\rho_{f^*}\|_{Lip} + \frac{K_1}{1-\beta}$. Also, $1-\beta$ will be dominated by $(1-\beta)^2$ term, so we can go on the discussion using these terms.

$$\left|\rho_{f_n^*} - \rho_{f^*}\right| \leq \left(\frac{4}{(1-\beta)^2}\right)\|\rho_{f^*}\|_{Lip} 2\alpha(1/n)^{1/d}.$$

The terms appearing here are still does not resemble to the values we obtained for average cost, which is:

$$\left|\rho_{f_n^*} - \rho_{f^*}\right| \leq (I_1 I_4 + I_2)(1/n)^{1/d} + \frac{I_3}{ln(1/K)}(1/n)^{1/d} ln\left(\frac{n^{1/d}}{I_4}\right).$$

The difference here can be attributed to the additional assumptions we have for the analysis of average cost. To show the convergence we had ω_c and ω_p to be linear, so this will affect the terms. Moreover, the approach of eliminating t in the analysis by inserting the value that gives the zero for the first derivative is another reason that we obtained different analytical formula. This result is applicable since it shows the maximum that can occur, to eliminate the time dependency occurring in the average cost formulation. Analytical difference is expected considering the discounted cost policy and errors does not depend on time, whereas average cost policy and errors depend on time.

References

1. M.L. Puterman, *Markov Decision Processes* (Wiley, Hoboken, NJ, 2005)
2. D.P. Bertsekas, Dynamic Programming and Optimal Control: Volume II (Athena Scientific, Belmont, 1995)
3. N. Saldi, T. Linder, S. Yüksel, Finite state approximations of Markov decision processes with general state and action spaces, in American Control Conference, Chicago (2015)
4. N. Saldi, S. Yüksel, T. Linder, Finite-state approximation of Markov decision processes with unbounded costs and Borel spaces, in IEEE Conference Decision Control, Japan (2015)
5. N. Saldi, S. Yüksel, T. Linder, Asymptotic optimality of finite approximations to Markov decision processes with Borel spaces. Math. Oper. Res. 42(4), 945–978 (2017)