How to choose the state relevance weight in the approximate linear programming approach for dynamic programming?

Yann Le Tallec
and
Theophane Weber

Finite Markov chain framework

- Finite state space X
- For all x in X, finite control space U(x)
- Bounded expected immediate cost $g_u(x)$ of control u in state x
- Transition probability matrix under control u: P_u
- **Proposition:** Any finite Markov chain can be transformed in an equivalent finite Markov chain with $g_u(x)=g(x)$ for all u in U(x).

Linear programming

- Let T be the DP operator for α-discounted problem: $TJ = \min_u g + \alpha P_u J$.
- By monotonicity of T, $J \leq TJ \Rightarrow J \leq TJ \leq T^k J \leq J^*$.
- **Linear programming approach to DP:**

  For all $c > 0$, J^* unique optimal solution of

  (LP): $\max c^T x$ s.t. $J(x) \leq g(x) + \alpha P_u(x,y) J(y), \forall (x,u)$

Approximate linear program

- Curse of dimensionality. Approximate:
  $J^*(x) \approx \Phi(x)r,\ r \in \mathbb{R}^m,\ m \ll |X|$
- **Approximate linear program**, c>0,
  (ALP): $\max_r\ c^T x$ s.t. $\Phi r \leq T\,\Phi r$.
- Unlike (LP), c matters: $r^* = r^*(c)$.
- $\Phi r \leq T\,\Phi r \Rightarrow \Phi r \leq T\,\Phi r \leq J^*$

General performance bound

- **Proposition:**
  For all J in $\mathbb{R}^{|X|}$,

  $$E\left[\| J_{u_J}(x) - J^*(x) |; x \sim \nu\right] = \left\| J_{u_J} - J^* \right\|_{1,\nu} \leq \left\| J - J^* \right\|_{1,\mu_{\nu,u_J}}$$

  where $\mu_{\nu,u} = (1-\alpha)\nu^T (I - \alpha P_u)^{-1}$

- In practice, ν is given by the application.

ALP approximation bound

- **Proposition:**

  Let r* be an optimal solution of (ALP). Then for all v s.t. Φv is a positive Lyapunov function,

$$\left\| J^* - \Phi r^* \right\|_{1,c} \le \frac{2c^T \Phi v}{1-\beta_{\Phi v}} \min_r \left\| J^* - \Phi r \right\|_{\infty,1/\Phi v}$$

- Compare with

$$\left\| J_{u_{\Phi r^*}} - J^* \right\|_{1,\nu} \le \left\| \Phi r^* - J^* \right\|_{1,\mu_{\nu,u_J}}$$

ALP approximation bound

- **Proposition:**

  Let r* be an optimal solution of (ALP). Then for all v s.t. Φv is a positive Lyapunov function,

$$\left\| J^* - \Phi r^* \right\|_{1,c} \leq \frac{2c^T \Phi v}{1 - \beta_{\Phi v}} \min_r \left\| J^* - \Phi r \right\|_{\infty, 1/\Phi v}$$

- Compare with

$$\left\| J_{u_{\Phi r^*}} - J^* \right\|_{1,\nu} \leq \left\| \Phi r^* - J^* \right\|_{1,\mu_{\nu,u_J}}$$

Choose c>0 to relate the 2 bounds in an efficient way

Simple bounds

- We want $\left\| J^* - \Phi r^* \right\|_{1,\mu_{\nu,u_{\Phi r^*}}} \leq K \left\| J^* - \Phi r^* \right\|_{1,c}, K > 0$
  to yield $\left\| J^* - J_{u_{\Phi r^*}} \right\|_{1,\nu} \leq K \frac{2c^T \Phi v}{1 - \beta_{\Phi v}} \min_r \left\| J^* - \Phi r \right\|_{\infty, 1/\Phi v}$
- This relation follows from $\mu_{\nu,u_{\Phi r^*}} \leq Kc$
- But r* depends implicitly on c via (ALP)

1. Trivially, c:=**1**. But poor bound for large state space
2. Algorithm using r*(c)=r*(Kc) for any K>0.
    1. Solve (ALP) for any c>0.
    2. Compute $\mu_{\nu,\Phi r^*}$
    3. If possible, find the smallest K>0 such that $\mu_{\nu,\Phi r^*} \leq Kc$

Find pmf $c=\mu_{\nu,\Phi r^*}$

- If $c=\mu_{\nu,\Phi r^*}>0$, c cannot be big and we have K=1
- Naïve algorithm: $c^k \underset{ALP}{\rightarrow} r^k \underset{greedy}{\rightarrow} u_{\Phi r_k} \rightarrow \mu_{\nu,u_{\Phi rk}} = c^{k+1}$.
- Fixed point? Convergence?
- **Theoretical algorithm**

  Relies on Brower's fixed point theorem of continuous function in convex compact set of $\mathbb{R}^{|X|}$

  - r^k not well defined for multiple optima
  - r^k not continuous in c => randomized c by Gaussian noise N(0,vI), v>0
  - greedy not continuous in r^k => δ-greedy: $P(u) \propto \exp(-\delta^{-1}.(g+P_u\Phi r^k))$

  For all v and δ, there is a fixed point to the naïve algorithm

Reinforced ALP

- Would like to solve (ALP) with the additional constraint

$$c^T = \mu_{\nu,u_{\Phi r^*}}{}^T = (1-\alpha)\nu^T(I-\alpha P_{u_{\Phi r^*}})^{-1}$$

- Recall that $P_{u_{\Phi r^*}}$ is greedy w.r.t Φr^*, i.e.

  $P_{u_{\Phi r}}\ \Phi r^* \le P_u\ \Phi r^*$ for all u.
- Hence,

$$\underbrace{(1-\alpha)\ \nu^T(I-\alpha\ P_{u_{\Phi r}})^{-1}}_{c^T}(I-\alpha\ P_u)\ \Phi r^* \le (1-\alpha)\nu^T\Phi r^*,\ \forall u$$

- Add the necessary linear constraints to (ALP)

  $c^T(I-\alpha\ P_u)\ \Phi r^* \le (1-\alpha)\nu^T\Phi r^*,\ \forall u$

Conclusions

- Some simple bounds on the (ALP) policy but not necessarily tight.
- Theoretical algorithm to find c as a probability distribution.
- Some insight in the role of c in (ALP)
- Need practical algorithms depending on ν and the Markov chain.