CS294-40 Learning for Robotics and Control Lecture 5 - 11/13/2008

Linear Programming Approach

Lecturer: Pieter Abbeel *Scribe: Nimbus Goehausen*

1 Overview

LP: essentially alternative to VI PI (very hard to get guarantees with function approximation)

LP: easy to get guarantees with function approximation

2 review of Value iteration

2.1 basic VI

$$initialize\ V = 0 \tag{1}$$
$$iterate: \tag{2}$$
$$\forall s: \bar{V}(s) = \max_a \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s')) \tag{3}$$
$$V = \bar{V} \tag{4}$$
$$\tag{5}$$

the last two items are simply the bellman backup: $V = TV$ This algorithm will converge to $V^* : V^* = TV^*$

2.2 VI with function approximation

The algorithm with function approximation

$$initalize\ V = 0 \tag{6}$$
$$iterate: \tag{7}$$
$$\forall s: \bar{V}(s) = \max_a \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s')) \tag{8}$$
$$V = \phi r\ for\ r\ = \arg\min r\|\bar{V}\phi \cdot r\| \tag{9}$$

Though this is just a theoretical algorithm, in practice one would do this:

$$r = 0 \tag{10}$$
$$iterate: \tag{11}$$
$$\bar{S} \in S:\ \forall s \in \bar{S}: \bar{V}(s) = \max_a \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s')) \tag{12}$$
$$\text{V} = \phi r\ \text{ for }\ \text{r} = \arg\min_r \textstyle\sum_{s\in\bar{S}}(\bar{V}(s) - (\phi r)(s)) \tag{13}$$

Where we're only looking at some subset of the state space.

2.3 Tetris Example

cost: height of the current wall, discount = .9
22 features:
ϕ_0 *to* ϕ_9: height of columns 0 through 9
ϕ_{10} *to*ϕ_{18}: height differential of consecutive columns
ϕ_{19}: holes
ϕ_{20}:max height
ϕ_{21}:1

3 LP approach

3.1 relaxing Value Iteration

In Value iteration we solve for V in the following way
V(s) = $\max_a \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s'))$

This is nonlinear, so lets investigate an alternative linear approach. We "relax" the problem and allow for a larger set of solutions

Instead we find V s.t.:
$\forall s: \ V(s) \geq \max_a \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s'))$

this is equivalent to:
$\forall s,a: \ V(s) \geq \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s'))$

We find a solution by solving
min $d^T V$ s.t. $\forall s,a: \ V(s) \geq \sum_{s'} P(s'|s,a)(R(s,a,s') + \gamma V(s'))$

where d is an arbitrary vector
In shorthand this is: $\min d^T V$ s.t. V $\geq TV$
With this approach we have linear constraints and we can find a solution efficiently as it merely requires solving a linear optimization problem

Our choice of d will affect which solution we find. Can we ensure that we find V*?
For a feasible solution to LP we must have:

$$V \geq TV \tag{14}$$

Recall

$$V_1 \geq V_2 \Rightarrow TV_1 \geq TV_2 \tag{15}$$

this imples

$$TV \geq T^2 V TV^2 \geq T^3 V ... \Rightarrow V \geq TV \geq T^2 V \geq ... \geq TV = V^* \tag{16}$$

Every feasible solution of V satisfies V $\geq V^*$
We can find V* by solving the following liner program:

$$\min_V \sum_s V(s) \text{ s.t. } V \geq TV \tag{17}$$

more generally this is:

$$\min_V d^T V V \geq TV \tag{18}$$

This gives V^* as a solution as long as $\forall s : d(s) > 0$

3.2 LP function approximation

$$\min_{V,r} d^T V \tag{19}$$
$$\text{s.t. } V \geq TV \tag{20}$$
$$V = \phi r \tag{21}$$

$\downarrow$

$$\min_{V,r} d^T \phi r \tag{22}$$
$$\text{s.t. } \phi r \geq T\phi r \tag{23}$$
$$V = \phi r \tag{24}$$

$\Updownarrow$

$$\min_r \sum_{s \in \bar{s}} d(s) \sum_i \phi_i(s) \cdot r_i \tag{25}$$
$$\text{s.t.} \forall s \in \bar{s}, a : \tag{26}$$
$$\sum_i \phi_i(s) \cdot r_i \geq \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma \sum_j \phi_j(s') \cdot r_j] \tag{27}$$
$$\tag{28}$$

lets call this solution r* :
is r$^* \in \arg\min_r \|V^* - \phi r^*\|$? No
$\|V^* - \phi r^*\|$ is comparable to $\min_r \|V^* - \phi r\|$

3.3 relevant referances

Ben Van Roy
de Farias Van Roy LP approach
Farias Van Roy Tetris case study