Lec35

Thursday, April 20, 2023 9:35 AM

- Deterministic SSP: Principle of Optimality:

$$J_k(i) = \min_{j=1,2,\dots,N} \{ a_{ij} + J_{k+1}(j) \}$$

- Finite horizon stochastic SP

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} E_{w_k} \left[g_k(x_k, u_k, w_k) + J_{k+1}\left(f_k(x_k, u_k, w_k)\right) \right]$$

- Infinite-horizon SSP

$$J^*(i) = \min_u \quad g(i,u) + \sum_j P_{ij}(u) J^*(j)$$

- Discounted problems

$$J^*(i) = \min_u \{ g(i,u) + \alpha \sum_j P_{ij}(u) J^*(j) \}$$

future cost is discounted by α

future cost from j

- Average-cost problems

$$\lambda + h(i) = \min_u \{ g(i,u) + \sum_j P_{ij}(u)\, h(j) \}$$

Constrained DP problems: Expected constraints

Saturday, April 25, 2015 3:02 PM

- The linear-program interpretation is even more powerful when we decided with constrained MDP

- We will restrict ourselves to the average cost problem.

- Suppose that in addition to a per-stage cost $g(i,u)$, there is a per-stage penalty of $y(i,u)$.

- We want to minimize the average cost

$$\lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g(x_k, u_k)\right]$$

Subject to a constraint of the average penalty

$$\lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y(x_k, u_k)\right] \le V$$

Linear program

- Let $\lambda_{iu} = \lambda_i q_{iu}$

= The probability of being at state i & use action u

- The average penalty is

$$\sum_{iu} \lambda_{iu} y(i,u)$$

- Thus, the constrained DP can be written as

$$\min \sum_{iu} \lambda_{iu} g(i,u)$$

sub to $\quad \sum_{n} x_{jn} = \sum_{in} x_{in} P_{ij}(n) \quad \forall j$

$$\sum_{in} x_{in}\, y(i,n) \le V \qquad (*)$$

$$\sum_{in} x_{in} = 1$$

- In general, the solution may have multiple non-zero x_{in} for a state i.

  In other words, a probabilistic policy is needed.

Duality

- Since the Linear program is also a convex program, duality holds
- Associate a Lagrange multiplier λ to $(*)$
- The Lagrangian is

$$L(\vec{x}, \lambda) = \sum_{in} x_{in}\, g(i,n) + \lambda \sum_{in} x_{in}\, y(i,n) - \lambda V$$

The dual objective is

$$D(\lambda) = \min L(\vec{x}, \lambda)$$

$$\text{sub to} \quad \sum_{j} x_{jn} = \sum_{in} x_{in} \cdot P_{ij}(n)$$

$$\sum_{in} x_{in} = 1$$

- But this is simply an average-cost problem with per-stage cost

$$g(i,n) + \lambda y(i,n).$$

- In a more symbolic way, we can rewrite the constrained DP as constrained DP as:

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g(x_k, u_k)\right]$$

$$\text{sub to} \quad \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y(x_k, u_k)\right] \le V$$

Associate a Lagrange multiplier λ to the constraint.

$$L = \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g(x_k, u_k) + \lambda y(x_k, u_k)\right] - \lambda V$$

- We should then minimize

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g(x_k, u_k) + \lambda y(x_k, u_k)\right]$$

- This is just an average-cost MDP!

We can then apply all results from duality.

- There exists a policy μ (possibly probablistic) & λ such ~~that~~

KKT
$$\begin{cases} \mu \text{ minimizes } \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g(x_k, u_k) + \lambda y(x_k, u_k)\right] \quad \text{(an MDP)} \\ \lambda \ge 0 \\ \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y(x_k, u_k)\right] \le V \\ \lambda \cdot \left[\lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y(x_k, u_k)\right] - V\right] = 0 \end{cases}$$

- Any pairs of μ, λ that satisfy that KKT condition are also optimal.

- The following iterative algo will converge for λ

At step t:

- Solve the average cost MDP with stage cost

$$g(x_k, u_k) + \lambda y(x_k, u_k)$$

- Update

$$\lambda(t+1) = \left[\lambda(t) + \alpha\left[\lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y(x_k, u_k)\right] - V\right]\right]^+$$

Decomposition

Friday, April 14, 2023 11:41 AM

- As in convex optimization, this type of duality can be very helpful for decomposing a large problem into smaller problems!

- Consider M copies of the MDP

  - The m-th MDP has a per-stage cost
    $$g^m(x_k^m, u_k^m)$$
    and per-stage penalty
    $$y^m(x_k^m, u_k^m)$$

  - If the penalties are not coupled, i.e., each MDP has a separate constraint on
    $$\lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y^m(x_k^m, u_k^m)\right]$$

    Then of course each MDP can be solved independently.

- What if the penalty constraints are coupled

$$\min \quad \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} g^m(x_k^m, u_k^m)\right]$$

$$\text{sub to} \quad \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} y^m(x_k^m, u_k^m)\right] \leq V \quad (*)$$

  - We can no longer solve each MDP seperately!

  - Further, solving this global MDP will likely run into the curse-of-dimensionality!

Further, solving this global LP will likely run into the curse-of-dimensionality!

- Instead, associate a Lagrange multiplier λ to $(*)$.

- The Lagrangian (precise form can be written through the corresponding LP)

$$L(\vec{\pi}, \lambda) = \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} g^m(x_k^m, u_k^m)\right]$$

$$+ \lambda \left[\lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} y^m(x_k^m, u_k^m)\right] - V\right]$$

$$= \sum_{m=1}^{M} \left\{ \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g^m(x_k^m, u_k^m) + \lambda y^m(x_k^m, u_k^m)\right]\right\}$$

$$- \lambda V$$

- Therefore, the m-th MDP can optimize

$$\min_{\pi_m} \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} g^m(x_k^m, u_k^m) + \lambda y^m(x_k^m, u_k^m)\right]$$

- This is a much smaller problem!

- Finally, the Lagrange multiplier can be updated by

$$\lambda(l+1) = \left[\lambda(l) + \alpha\left(\sum_{m=1}^{M} \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} y^m(x_k^m, u_k^m)\right] - V\right)\right]^+$$

using the optimal policy based on $\lambda(l)$.

using the optimal policy based on $\lambda(l)$.

$$\underbrace{\qquad\qquad\qquad\qquad}$$

can be replaced by

$$E\left[\, y^m(x_k^m, u_k^m) \right]$$

assume k is the steady-state