Linear Programming Approach to Dynamic Programming

Jonatan Schroeder

University of British Columbia

Mar 5, 2008

Outline

Based on the lecture notes by Daniela P. de Farias

DP considerations

$$(TJ)(x) = \min_a \left\{ g_a(x) + \alpha \sum_y P_a(x,y) J(y) \right\}$$

- Stationary policies
- $T^k J$ approaches J^* when k is large enough, $TJ^* = J^*$
- Infinite horizon, discounted-cost

Optimization Problem

- Consider the optimization problem:

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & TJ \geq J \end{array}$$

- Vector c is strictly positive
- J^* is the unique solution to this problem

Linear Programming Problem

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & TJ \geq J \end{array}$$

- Strictly, this is not a linear program:

$$(TJ)(x) \geq J(x)$$

$$\min_a \left\{ g_a(x) + \alpha \sum_y P_a(x,y) J(y) \right\} \geq J(x)$$

Linear Programming Problem

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & TJ \geq J \end{array}$$

- Strictly, this is not a linear program:

$$(TJ)(x) \geq J(x)$$

$$\min_a \left\{ g_a(x) + \alpha \sum_y P_a(x,y)J(y) \right\} \geq J(x)$$

- But can be converted to linear program:

$$g_a(x) + \alpha \sum_y P_a(x,y)J(y) \geq J(x), \forall a \in \mathcal{A}_x$$

Exact LP

Exact LP

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & g_a(x) + \alpha \sum_y P_a(x,y) J(y) \geq J(x), \forall a \in \mathcal{A}_x \end{array}$$

Exact LP

Exact LP

$$
\begin{array}{ll}
\max_J & c^T J \\
\text{subject to} & g_a(x) + \alpha \sum_y P_a(x,y) J(y) \geq J(x), \forall a \in \mathcal{A}_x
\end{array}
$$

- Variables: states in the system

Exact LP

Exact LP

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & g_a(x) + \alpha \sum_y P_a(x,y) J(y) \geq J(x), \forall a \in \mathcal{A}_x \end{array}$$

- Variables: states in the system
- Constraints: state-action pairs

Dual Linear Programming

- Consider a dual LP:

$$
\begin{aligned}
\min_{\mu} \quad & \sum_{x,a} \mu(x,a) g_a(x) \\
\text{subject to} \quad & \sum_{y} \sum_{a} \mu(y,a) P_a(y,x) = \sum_{a} \mu(x,a), \forall x \\
& \sum_{x,a} \mu(x,a) = 1 \\
& \mu(x,a) \geq 0, \forall x, a
\end{aligned}
$$

- $\mu(x,a)$: probability over the state-action space that action a is taken when current state is x

Randomized Policies

- Usually a policy is a mapping from states to actions
- A *randomized policy* is a function u which prescribes a probability $u(x, a)$ for taking action a when current state is x

Randomized Policies

- Usually a policy is a mapping from states to actions
- A *randomized policy* is a function u which prescribes a probability $u(x, a)$ for taking action a when current state is x

$$g_u(x) = \sum_a u(x, a) g_a(x)$$

Randomized Policies

- Usually a policy is a mapping from states to actions
- A *randomized policy* is a function u which prescribes a probability $u(x, a)$ for taking action a when current state is x

$$g_u(x) = \sum_a u(x, a) g_a(x)$$

$$\pi(x) = \sum_a \mu(x, a)$$

Randomized Policies

- Usually a policy is a mapping from states to actions
- A *randomized policy* is a function u which prescribes a probability $u(x, a)$ for taking action a when current state is x

$$g_u(x) = \sum_a u(x, a) g_a(x)$$

$$\pi(x) = \sum_a \mu(x, a)$$

$$u(x, a) = \frac{\mu(x, a)}{\pi(x)}$$

Randomized Policies (cont.)

- Transition matrix P_u: $P_u(x, y) = \sum_a u(x, a) P_a(x, y)$

Randomized Policies (cont.)

- Transition matrix P_u: $P_u(x, y) = \sum_a u(x, a) P_a(x, y)$
- Stationary distribution π_u:

$$\pi_u^T P_u = \pi_u^T$$

$$\sum_x \pi_u(x) = 1$$

$$\pi_u(x) \geq 0$$

Randomized Policies (cont.)

- Transition matrix P_u: $P_u(x,y) = \sum_a u(x,a) P_a(x,y)$
- Stationary distribution π_u:

$$\pi_u^T P_u = \pi_u^T$$

$$\sum_x \pi_u(x) = 1$$

$$\pi_u(x) \geq 0$$

- Proposal: The dual LP solution finds a stationary distribution π_u

Dual LP - Proof

Constraints (1)

$$\sum_{y} \sum_{a} \mu(y, a) P_a(y, x) = \sum_{a} \mu(x, a)$$

Dual LP - Proof

Constraints (1)

$$
\begin{aligned}
\sum_y \sum_a \mu(y,a) P_a(y,x) &= \sum_a \mu(x,a) \\
\sum_y \sum_a \pi(y) u(y,a) P_a(y,x) &= \pi(x)
\end{aligned}
$$

Dual LP - Proof

Constraints (1)

$$\begin{aligned}
\sum_y \sum_a \mu(y,a) P_a(y,x) &= \sum_a \mu(x,a) \\
\sum_y \sum_a \pi(y) u(y,a) P_a(y,x) &= \pi(x) \\
\sum_y \pi(y) P_u(y,x) &= \pi(x)
\end{aligned}$$

Dual LP - Proof
Constraints (1)

$$\begin{aligned}
\sum_y \sum_a \mu(y,a) P_a(y,x) &= \sum_a \mu(x,a) \\
\sum_y \sum_a \pi(y) u(y,a) P_a(y,x) &= \pi(x) \\
\sum_y \pi(y) P_u(y,x) &= \pi(x) \\
\pi^T P_u &= \pi^T
\end{aligned}$$

Dual LP - Proof

Constraints (2)

$$\sum_x \sum_a \mu(x,a) = 1$$

Dual LP - Proof

Constraints (2)

$$\begin{aligned} \sum_x \sum_a \mu(x, a) &= 1 \\ \sum_x \pi(x) &= 1 \end{aligned}$$

Dual LP - Proof
Constraints (2)

$$\begin{aligned} \sum_x \sum_a \mu(x, a) &= 1 \\ \sum_x \pi(x) &= 1 \end{aligned}$$

- π is a stationary distribution associated with policy u ($\pi = \pi_u$)

Dual LP - Proof

Goal function

$$\sum_{x,a} \mu(x,a) g_a(x)$$

Dual LP - Proof

Goal function

$$\sum_{x,a} \mu(x,a) g_a(x) \quad = \quad \sum_x \sum_a \mu(x) u(x,a) g_a(x)$$

Dual LP - Proof

Goal function

$$\begin{aligned}\sum_{x,a} \mu(x,a) g_a(x) &= \sum_x \sum_a \mu(x) u(x,a) g_a(x) \\ &= \sum_x \pi_u(x) g_u(x)\end{aligned}$$

Dual LP - Proof

Goal function

$$
\begin{aligned}
\sum_{x,a} \mu(x,a) g_a(x) &= \sum_x \sum_a \mu(x) u(x,a) g_a(x) \\
&= \sum_x \pi_u(x) g_u(x) \\
&= \lambda_u
\end{aligned}
$$

Dual LP - Proof
Goal function

$$
\begin{aligned}
\sum_{x,a} \mu(x,a) g_a(x) &= \sum_x \sum_a \mu(x) u(x,a) g_a(x) \\
&= \sum_x \pi_u(x) g_u(x) \\
&= \lambda_u
\end{aligned}
$$

- The dual LP goal corresponds to the average cost λ_u of policy u.

Recap – Approximate DP

- $\tilde{J}(\cdot, r) \approx J^*(\cdot)$
  - $\tilde{J}(x, r)$ is an approximation to $J^*(x)$
- Consider Φ a matrix such that $\tilde{J} = \Phi\tilde{r}$
- $\tilde{r}$ is “simpler” than J

New Optimization Problem

- The original exact DP was:

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & TJ \geq J \end{array}$$

New Optimization Problem

- The original exact DP was:

$$\begin{array}{ll} \max_J & c^T J \\ \text{subject to} & TJ \geq J \end{array}$$

- A close approximation $\tilde{r}$ can be computed by:

$$\begin{array}{ll} \max_r & c^T \Phi r \\ \text{subject to} & T\Phi r \geq \Phi r \end{array}$$

Approximate LP

$$\begin{array}{ll} \max_r & c^T \Phi r \\ \text{subject to} & g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x,y)(\Phi r)(y) \geq (\Phi r)(x) \end{array}$$

Approximate LP

$$\begin{array}{ll} \max_r & c^T \Phi r \\ \text{subject to} & g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x,y)(\Phi r)(y) \geq (\Phi r)(x) \end{array}$$

- Smaller number of variables
- Same number of constraints

Reduced Linear Program (RLP)

- To reduce the number of constraints, we may use Reduced Linear Program
- Based on:
  - A constraint sample size m
  - A probability measure Ψ over the set of state-action pairs
  - A bounding set $\mathcal{N} \in \Re^K$
- A set $\mathcal{X}$ is constructed with m state-action pairs sampled according to Ψ

Reduced Linear Program (RLP)

- The RLP is defined by:

$$
\begin{array}{ll}
\max_r & c^T \Phi r \\
\text{subject to} & g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x, y)(\Phi r)(y) \geq (\Phi r)(x), \\
& \qquad\qquad\qquad \forall (x, a) \in \mathcal{X} \\
& r \in \mathcal{N}
\end{array}
$$

Reduced Linear Program (RLP)

- The RLP is defined by:

$$\begin{array}{ll} \max_r & c^T \Phi r \\ \text{subject to} & g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x,y)(\Phi r)(y) \geq (\Phi r)(x), \\ & \qquad\qquad \forall (x,a) \in \mathcal{X} \\ & r \in \mathcal{N} \end{array}$$

- m, Ψ and $\mathcal{X}$ should be carefully chosen

Linear Programming Approach to Dynamic Programming

Jonatan Schroeder

University of British Columbia

Mar 5, 2008