A Strongly Polynomial-Time Algorithm for Solving the Markov Decision Problem with Fixed Discount Factor

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Thanks to Kahn Mason, Ben Van Roy and Pete Veinott for many insightful discussions on this subject.

Outline

- Linear programming, complexity, the Markov decision problem;
- Central path and its geometry;
- Combinatorial interior-point algorithm for the MDP;
- Complexity analysis of the algorithm;

Complexity Theory

- a notion of input size,
- a set of basic operations, and
- a cost for each basic operation.

The last two allow one to define the cost of a computation.

The Blum-Shub-Smale model is what we use in this talk, with exact real arithmetic operations (i.e., ignoring round-off errors).

Linear Programming

- $$\begin{aligned} \text{Primal: minimize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

- $$\begin{aligned} \text{Dual: maximize} \quad & \mathbf{b}^T\mathbf{y} \\ \text{subject to} \quad & \mathbf{s} = \mathbf{c} - A^T\mathbf{y} \geq \mathbf{0}, \end{aligned}$$

- $A \in \mathbf{R}^{m \times n}$, $\mathbf{c} \in \mathbf{R}^n$ and $\mathbf{b} \in \mathbf{R}^m$ are given; $\mathbf{x} \in \mathbf{R}^n$ and $(\mathbf{y} \in \mathbf{R}^m, \mathbf{s} \in \mathbf{R}^n)$ are unknown vectors; $\mathbf{s}$ is often called dual slack vector.
- We denote the LP problem as $LP(A, \mathbf{b}, \mathbf{c})$.

- The LP problem is polynomial solvable under the Turing model of computation, proved by Khachiyan and also by Karmarkar and many others. But the problem, whether there is a polynomial-time algorithm for LP under the BSS model of computation, remains open.

- There are some research developments relating complexity of interior-point algorithms with certain “condition or difficulty” measures for linear programming (see Renegar/Peña, Epelman/Freund/Vera, Ho/Tüncel, Todd/Tüncel/Ye, Cucker/Cheung/Cucker, Gonzaga/Hugo, Ye, etc).

- The “layered-step interior point” (LIP) algorithm (Vavasis/Ye, Megiddo/Mizuno/Tsuchiya, Ho/Tüncel, Monteiro/Tsuchiya, etc) interleaves small steps with longer layered least-squares (LLS) steps to follow the central path. The algorithm terminates in $O(n^{3.5}c(A))$ iterations.

$$c(A) = O(\log(\bar{\chi}_A) + \log n). \tag{1}$$

The Markov Decision Problem

$$
\begin{array}{lcccccc}
\text{minimize} & (\mathbf{c}^1)^T\mathbf{x}^1 & \ldots + (\mathbf{c}^i)^T\mathbf{x}^i + \ldots & +(\mathbf{c}^k)^T\mathbf{x}^k & & \\
\text{subject to} & (I - \theta P^1)\mathbf{x}^1 & \ldots + (I - \theta P^i)\mathbf{x}^i + \ldots & +(I - \theta P^k)\mathbf{x}^k & = & \mathbf{e}, \\
 & \mathbf{x}^1 & \ldots \quad \mathbf{x}^i \quad \ldots & \mathbf{x}^k & \geq & \mathbf{0}.
\end{array}
$$

Here, $x^i \in \mathbf{R}^n$ represents the decision variables of all states for action i, I is the $n \times n$ identity matrix, and P^i, $i = 1, \ldots, k$, is an $n \times n$ Markov matrix ($\mathbf{e}P^i = \mathbf{e}$ and $P^i \geq 0$).

$$A = [I - \theta P^1, \ldots, I - \theta P^k] \in \mathbf{R}^{n \times nk}$$

$$\mathbf{b} = \mathbf{e} \in \mathbf{R}^n, \quad \text{and} \quad \mathbf{c} = (\mathbf{c}^1; \ldots; \mathbf{c}^k) \in \mathbf{R}^{nk}.$$

The Dual of MDP

And its dual (by adding slack variables) is

$$
\begin{array}{lrcl}
\text{maximize} & \mathbf{e}^T\mathbf{y} & & \\
\text{subject to} & (I-\theta P^1)^T\mathbf{y}+\mathbf{s}^1 & = & \mathbf{c}^1, \\
 & \dots & \dots & \dots \\
 & (I-\theta P^i)^T\mathbf{y}+\mathbf{s}^i & = & \mathbf{c}^i, \\
 & \dots & \dots & \dots \\
 & (I-\theta P^k)^T\mathbf{y}+\mathbf{s}^k & = & \mathbf{c}^k, \\
 & \mathbf{s}^1,\dots,\mathbf{s}^i,\dots,\mathbf{s}^k & \geq & \mathbf{0}.
\end{array}
$$

Discount factor: $\theta < 1$.

For simplicity, consider $k = 2$ throughout this talk.

Complexity Results on MDP

Value-Iter.	Policy-Iter.	LP-Alg.	Combinatorial IPA
$n^2 \cdot \frac{L(P^i,c^i,\theta)}{1-\theta}$	$n^3 \cdot \frac{2^n}{n}$	$n^{2.5} \cdot n^{0.5} L(P^i, c^i, \theta)$	

New Result on MDP

Value-Iter.	Policy-Iter.	LP-Alg.	Combinatorial IPA
$n^2 \cdot \frac{L(P^i,c^i,\theta)}{1-\theta}$	$n^3 \cdot \frac{2^n}{n}$	$n^{2.5} \cdot n^{0.5} L(P^i, c^i, \theta)$	$n^{1.5} \cdot n^{2.5} \ln \frac{1}{1-\theta}$

Termination: Why $L(\cdot)$?

All polynomial algorithms are continuous algorithms and it denotes how small the error should be in order to round an exact optimal solution (policy)?

$$\mathbf{c}^T\mathbf{x} - z^* \leq 2^{-L(A,\mathbf{b},\mathbf{c})}$$

or

$$\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} \leq 2^{-L(A,\mathbf{b},\mathbf{c})}$$

This talk presents a combinatorial interior-point algorithm for MDP.

We remark that the condition measure $\bar{\chi}_A$ mentioned earlier cannot be bounded by $1/(1-\theta)$:

$$A = \begin{bmatrix} 1-\theta & 0 & 1-\theta(1-\epsilon) & 0 \\ 0 & 1-\theta & -\theta\cdot\epsilon & 1-\theta \end{bmatrix}$$

Here, for any given $\theta > 0$, $\|(A_B)^{-1}A\|$ can be arbitrarily large as $\epsilon \longrightarrow 0^+$ when

$$A_B = \begin{pmatrix} 1-\theta & 1-\theta(1-\epsilon) \\ 0 & -\theta\cdot\epsilon \end{pmatrix}.$$

In fact, all other condition measures used in complexity analyses for general LP can be arbitrarily bad for the MDP.

The Two-Action MDP

Comparing to the LP standard form,

$$A = [I - \theta P^1,\ I - \theta P^2] \in \mathbf{R}^{n \times 2n},$$

$$\mathbf{b} = \mathbf{e} \in \mathbf{R}^n, \quad \text{and} \quad \mathbf{c} = (\mathbf{c}^1; \mathbf{c}^2) \in \mathbf{R}^{2n}.$$

Any feasible basis

$$A_B = I - \theta P$$

$$(A_B)^{-1} = (I - \theta P)^{-1} = I + \theta P + \theta^2 P^2 + \ldots.$$

MDP Properties

- Both the primal and dual MDPs have interior feasible points if $0 \leq \theta < 1$.
- The feasible set of the primal MDP is bounded. More precisely,

$$\mathbf{e}^T \mathbf{x} = \frac{n}{1-\theta},$$

where $\mathbf{x} = (\mathbf{x}^1; \mathbf{x}^2)$.

- Let $\hat{\mathbf{x}}$ be a basic feasible solution of the MDP. Then, any basic variable, say $\hat{\mathbf{x}}_i$, has its value

$$\hat{\mathbf{x}}_i \geq 1.$$

- Let B^* and N^* be the optimal partition for the MDP. Then, B^* contains at least one feasible basis, i.e., $|B^*| \geq n$ and $|N^*| \leq n$; and for any $j \in B^*$

there is an optimal solution $\mathbf{x}^*$ such that

$$\mathbf{x}_j^* \geq 1.$$

- Let A_B be any feasible basis and A_N be any submatrix of the rest columns of the MDP constraint matrix, then

$$\|(A_B)^{-1}A_N\| \leq \frac{2n\sqrt{n}}{1-\theta}.$$

The partition of LP variables

- If $LP(A, \mathbf{b}, \mathbf{c})$ has an optimal solution pair, then there exists a unique index set $B^* \subset \{1, ..., n\}$ and $N^* = \{1, ..., n\} \setminus B^*$, such that the optimal faces are

$$A_{B^*}\mathbf{x}_{B^*} = \mathbf{b}, \quad \mathbf{x}_{B^*} \geq \mathbf{0}, \quad \mathbf{x}_{N^*} = \mathbf{0}$$

$$\mathbf{s}_{B^*} = \mathbf{c}_{B^*} - A_{B^*}^T\mathbf{y} = \mathbf{0}, \quad \mathbf{s}_{N^*} = \mathbf{c}_{N^*} - A_{N^*}^T\mathbf{y} \geq \mathbf{0}.$$

- This partition is called the strict complementarity partition:

$$A_{B^*}\mathbf{x}_{B^*} = \mathbf{b}, \quad \mathbf{x}_{B^*} > \mathbf{0}, \quad \mathbf{x}_{N^*} = \mathbf{0}$$

$$\mathbf{s}_{B^*} = \mathbf{c}_{B^*} - A_{B^*}^T\mathbf{y} = \mathbf{0}, \quad \mathbf{s}_{N^*} = \mathbf{c}_{N^*} - A_{N^*}^T\mathbf{y} > \mathbf{0}.$$

The Central Path of LP

$$
\begin{aligned}
A\mathbf{x} &= \mathbf{b}, \\
A^T\mathbf{y} + \mathbf{s} &= \mathbf{c}, \\
SX\mathbf{e} &= \mu\mathbf{e}, \\
\mathbf{x} > \mathbf{0}, & \quad \mathbf{s} > \mathbf{0}.
\end{aligned}
$$

The solution to these equations, written $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$, is called the central path point for μ, and the aggregate of all points, as μ ranges from 0 to ∞, is the central path of the LP problem.

The following is a geometric property of the central path:

Lemma 1 *Let* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *and* $(\mathbf{x}(\mu'), \mathbf{y}(\mu'), \mathbf{s}(\mu'))$ *be two central path points such that* $0 \le \mu' < \mu$. *Then for any* i,

$$s(\mu')_i \le ns(\mu)_i \quad \text{and} \quad x(\mu')_i \le nx(\mu)_i.$$

In particular, if $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ *is optimal, then, for any* $\mu > 0$ *and any* i,

$$\mathbf{s}_i^* \le n\mathbf{s}(\mu)_i \quad \text{and} \quad \mathbf{x}_i^* \le n\mathbf{x}(\mu)_i.$$

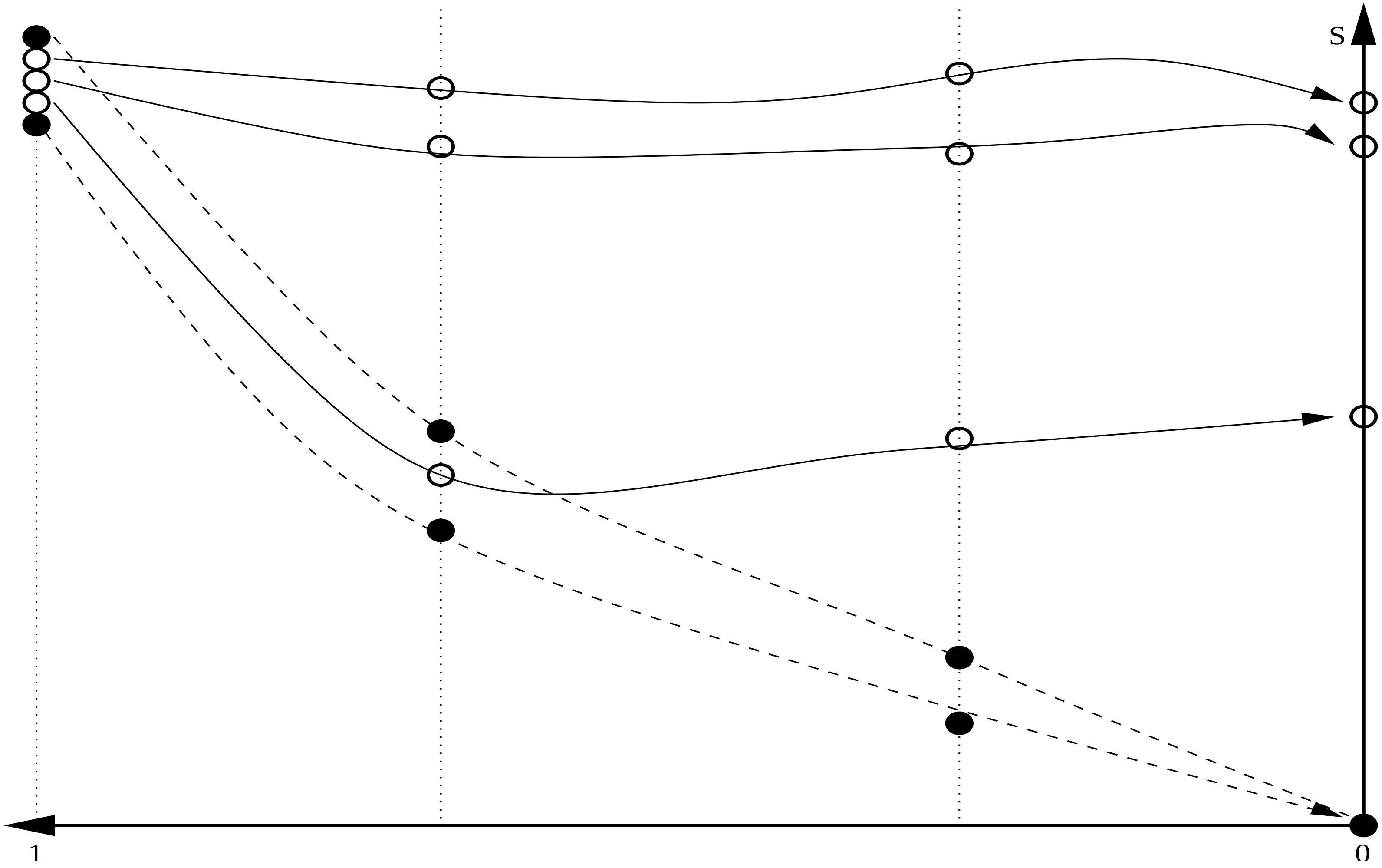

Figure 1: Individual variables on the central path

Corollary 1 *For any $\mu \in (0, \mu^0]$, the central path pair of the MDP satisfies*

$$\mathbf{x}(\mu)_j \leq \frac{n}{1-\theta} \quad \textit{and} \quad \mathbf{s}(\mu)_j \geq \frac{1-\theta}{n}\mu \quad \textit{for every } j = 1, \ldots, 2n;$$

and

$$\mathbf{x}(\mu)_j \geq \frac{1}{2n} \quad \textit{and} \quad \mathbf{s}(\mu)_j \leq 2n\mu \quad \textit{for evey } j \in B^*.$$

Primal initial interior point

$$(\mathbf{x}^i)^0 = (I - \theta P^i)^{-1} e, \quad i = 1, 2$$

and

$$\mathbf{x}^0 = \begin{pmatrix} \frac{1}{2}(\mathbf{x}^1)^0 \\ \frac{1}{2}(\mathbf{x}^2)^0 \end{pmatrix}.$$

Thus, $\mathbf{x}^0$ is an interior feasible point for the MDP and

$$\mathbf{x}^0 \geq \frac{1}{2}\mathbf{e} \in \mathbf{R}^{2n}.$$

Dual initial interior point

$$\mathbf{y}^0 = -\gamma \mathbf{e} \quad \text{and} \quad \mathbf{s}^0 = \begin{pmatrix} (\mathbf{s}^1)^0 \\ (\mathbf{s}^2)^0 \end{pmatrix} = \begin{pmatrix} \mathbf{c}^1 + \gamma(1-\theta)\mathbf{e} \\ \mathbf{c}^2 + \gamma(1-\theta)\mathbf{e}. \end{pmatrix}$$

where γ is chosen sufficiently large such that

$$\mathbf{s}^0 > 0 \quad \text{and} \quad \gamma \geq \frac{\mathbf{c}^T \mathbf{x}^0}{n}.$$

Potential of the initial point pair

Denote $\mu^0 = (\mathbf{x}^0)^T\mathbf{s}^0/2n$ and consider the TTY potential function

$$\phi(\mathbf{x},\mathbf{s}) = 2n\log(\mathbf{s}^T\mathbf{x}) - \sum_{j=1}^{2n}\log(\mathbf{s}_j\mathbf{x}_j) \geq 2n\log(2n).$$

$$\begin{aligned}
\phi(\mathbf{x}^0,\mathbf{s}^0) &= 2n\log(\mathbf{c}^T\mathbf{x}^0 + \gamma(1-\theta)\frac{n}{1-\theta}) - \sum_{j=1}^{2n}\log(\mathbf{s}_j^0\mathbf{x}_j^0) \\
&\leq 2n\log(\mathbf{c}^T\mathbf{x}^0 + \gamma\cdot n) - \sum_{j=1}^{2n}\log(\mathbf{s}_j^0/2) \quad (\text{since } \mathbf{x}_j^0 \geq 1/2) \\
&= 2n\log(2n) - \sum_{j=1}^{2n}\log\frac{2n(\mathbf{c}_j/2 + \gamma(1-\theta)/2)}{\mathbf{c}^T\mathbf{x}^0 + \gamma\cdot n}
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad 2n\log(2n) - \sum_{j=1}^{2n} \log \frac{n\gamma(1-\theta)}{\mathbf{c}^T\mathbf{x}^0 + \gamma \cdot n} \\
&\leq \quad 2n\log(2n) - \sum_{j=1}^{2n} \log \frac{n\gamma(1-\theta)}{2\gamma \cdot n} \\
&= \quad 2n\log(2n) + 2n\log(\frac{2}{1-\theta}).
\end{aligned}
$$

Approximately centered pair

"Approximately centered" point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ such that

$$\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) := \|SX\mathbf{e}/\mu - \mathbf{e}\| \leq \eta_0,$$

where, say, $\eta_0 = 0.2$ throughout this talk.

Complexity to compute an initial central-path point

Therefore, using the primal-dual potential reduction algorithm, we can generate an (approximate) central path point $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0)$ such that

$$\eta(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0, \mu^0) \le \eta_0.$$

in at most $O(n(\log \frac{2}{1-\theta}))$ interior-point algorithm iterations where each iteration uses $O(n^3)$ arithmetic operations,

Combinatorial Algorithm: Separation of variables

$$g = \frac{10n^2(1+\eta_0)}{(1-\theta)\sqrt{1-\eta_0}}.$$

For any given approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ such that

$$\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0,$$

define

$$J_1(\mu) = \{j : \mathbf{s}_j \leq \frac{8n\mu}{3}\},$$

$$J_3(\mu) = \{j : \mathbf{s}_j \geq \frac{8n\mu \cdot g}{3}\}$$

and $J_2(\mu)$ be the rest of indices. Thus, for any $j_1 \in J_1(\mu)$ and $j_3 \in J_3(\mu)$, we have

$$\frac{\mathbf{s}_{j_1}}{\mathbf{s}_{j_3}} \leq \frac{1}{g}.$$

For any $j \in B^*$, we observe

$$\mathbf{s}_j = \frac{\mathbf{s}_j}{\mathbf{s}(\mu)_j}\mathbf{s}(\mu)_j \leq \frac{\mathbf{s}_j}{\mathbf{s}(\mu)_j}2n\mu \leq \frac{4}{3}2n\mu = \frac{8n\mu}{3}.$$

Therefore,

Lemma 2 *Let* $J_1(\mu)$ *be defined above at any* $0 < \mu \leq \mu^0$*. Then, every variable of* B^* *is in* $J_1(\mu)$ *or* $B^* \subset J_1(\mu)$ *for any* $0 < \mu \leq \mu^0$*, and, thereby,* $J_1(\mu)$ *always contains an optimal basis.*

Combinatorial Algorithm: Elimination of variables

If $J_3(\mu)$ is not empty, we can now eliminate all its primal variables and dual constraints from further consideration, since they must be all in N^* and take zero value at any optimal solution.

To restore the primal feasibility after elimination, we solve the least squares problem:

$$\min_{\delta \mathbf{x}_1} \quad \|D_1^{1/2} \delta \mathbf{x}_1\| \text{ subject to } \quad A_1 \delta \mathbf{x}_1 = A_3 \mathbf{x}_3.$$

Then, we have

$$A_1(\mathbf{x}_1 + \delta \mathbf{x}_1) + A_2 \mathbf{x}_2 = A_1 \mathbf{x}_1 + A_2 \mathbf{x}_2 + A_3 \mathbf{x}_3 = \mathbf{b}.$$

Combinatorial Algorithm: Restoration of the central path

Lemma 3 *Not only* $A_1(\mathbf{x}_1 + \delta\mathbf{x}_1) + A_2\mathbf{x}_2 = \mathbf{b}$ *and* $(\mathbf{x}_1 + \delta\mathbf{x}_1; \mathbf{x}_2) > 0$*, but also*

$$\eta((\mathbf{x}_1 + \delta\mathbf{x}_1; \mathbf{x}_2), \mathbf{y}, (\mathbf{s}_1; \mathbf{s}_2), \mu) \leq 2\eta_0.$$

That is, they are a near central-path point pair for the same μ *of the MDP after eliminating every primal variables and dual constraints in* $J_3(\mu)$*.*

How to Make $J_3(\mu) \neq \emptyset$

We apply a predictor-corrector method of Mizuno-Todd-Ye.

$$\epsilon^0 := \frac{1}{\sqrt{\mu^0}} \|D^{-1/2}(\delta\bar{\mathbf{s}} + \mathbf{s})\| = \frac{1}{\sqrt{\mu^0}} \|D^{1/2}\delta\bar{\mathbf{x}}\| \tag{2}$$

is strictly greater than 0. Let

$$\bar{\alpha} = \max\left\{0, 1 - \frac{\sqrt{n}\epsilon^0}{\eta_0}\right\}. \tag{3}$$

$$\bar{\mathbf{x}} = \mathbf{x} + \bar{\alpha}\delta\bar{\mathbf{x}},$$

$$\bar{\mathbf{y}} = \mathbf{y} + \bar{\alpha}\delta\bar{\mathbf{y}},$$

and

$$\bar{\mathbf{s}} = \mathbf{s} + \bar{\alpha}\delta\bar{\mathbf{s}}.$$

If $\bar{\alpha} < 1$, we have the new iterate $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ nearly centered and strictly feasible.

Lemma 4 *If $\epsilon^0 > 0$, then there must be a variable indexed $\bar{j}$ such that $\bar{j} \in N^*$, and the central-path value*

$$\mathbf{s}(\mu)_{\bar{j}} \geq \frac{\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^{2.5}} \cdot \epsilon^0,$$

for all $\mu \in (0, \mu^0]$.

Now consider two cases:

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} \geq 1. \tag{4}$$

and

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} < 1. \tag{5}$$

Complexity: Case 1

$\bar{\alpha} = 0$, and

$$\epsilon^0 \geq \frac{\eta_0}{\sqrt{n}} \quad \text{and} \quad \mathbf{s}(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^3},$$

where index $\bar{j} \in N^*$ is the one singled out in Lemma 4. In this case, we continue apply the predictor-corrector path-following algorithm reducing μ from μ^0. Thus, as soon as

$$\frac{\mu}{\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)}{8\sqrt{2}n^4 g},$$

we have

$$\mathbf{s}(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^3} \geq 4n\mu \cdot g.$$

That is, $\bar{j} \in J_3(\mu)$.

Complexity: Case 2

$$1 - \bar{\alpha} = \frac{\sqrt{n}\epsilon^0}{\eta_0} \quad \text{and} \quad \mathbf{s}(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)(1-\bar{\alpha})\mu^0}{2\sqrt{2}n^3}$$

where again index $\bar{j}$ is the one singled out in Lemma 4. Note that the first predictor step has reduced μ^0 to $(1-\bar{\alpha})\mu^0$. Then, we continue apply the predictor-corrector algorithm reducing μ from $(1-\bar{\alpha})\mu^0$. As soon as

$$\frac{\mu}{(1-\bar{\alpha})\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)}{8\sqrt{2}n^4 g},$$

we have again

$$\mathbf{s}(\mu)_{\bar{j}} \geq 4n\mu \cdot g.$$

That is, $\bar{j} \in J_3(\mu)$.

Complexity to Make $J_3(\mu) \neq \emptyset$

In at most $O(n^{0.5}(\log \frac{1}{1-\theta} + \log n))$ predictor-corrector interior-point algorithm iterations, we have $J_3(\mu) \neq \emptyset$.

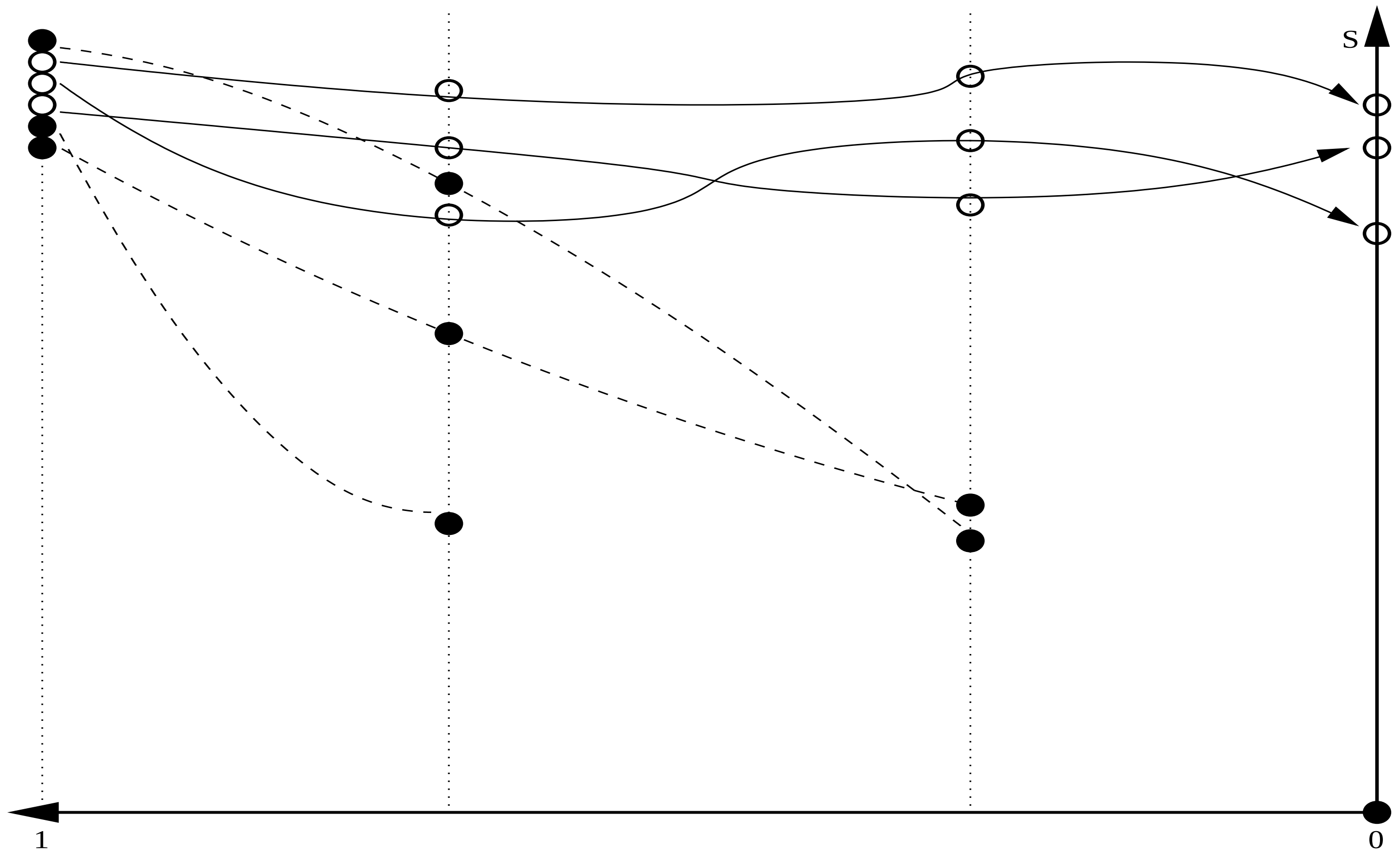

Figure 2: Markov Decision Problem

Complexity Theorem

Theorem 1 *The combinatorial interior-point algorithm generates an optimal solution of the MDP in at most n major eliminating steps, and each step uses $O(n^{0.5}(\log\frac{1}{1-\theta}+\log n))$ predictor-corrector interior-point algorithm iterations.*

Using the Karmakar rank-one updating scheme, the average number of arithmetic operations of each predictor-corrector interior-point iteration is $O(n^{2.5})$. Thus,

Theorem 2 *The combinatorial interior-point algorithm generates an optimal solution of the MDP in at most $O(n^4(\log\frac{1}{1-\theta}+\log n))$ arithmetic operations.*

Extensions to general MDP

Corollary 2 *The combinatorial interior-point algorithm generates an optimal solution of the MDP in at most* $(k-1)n$ *major eliminating steps, and each step uses* $O((nk)^{0.5}(\log \frac{1}{1-\theta} + \log n + \log k))$ *predictor-corrector interior-point algorithm iterations, where* n *is the number of states and* k *is the number of actions for each state. The total arithmetic operations to solve the MDP is bounded by* $O(n^4 k^2 (\log \frac{1}{1-\theta} + \log n + \log k))$*.*

What's next

- Get rid of θ?
- Does θ have to play a role in the complexity of the MDP?