Efficiency of the Simplex Method

Yinyu Ye

Department of Management Science and Engineering

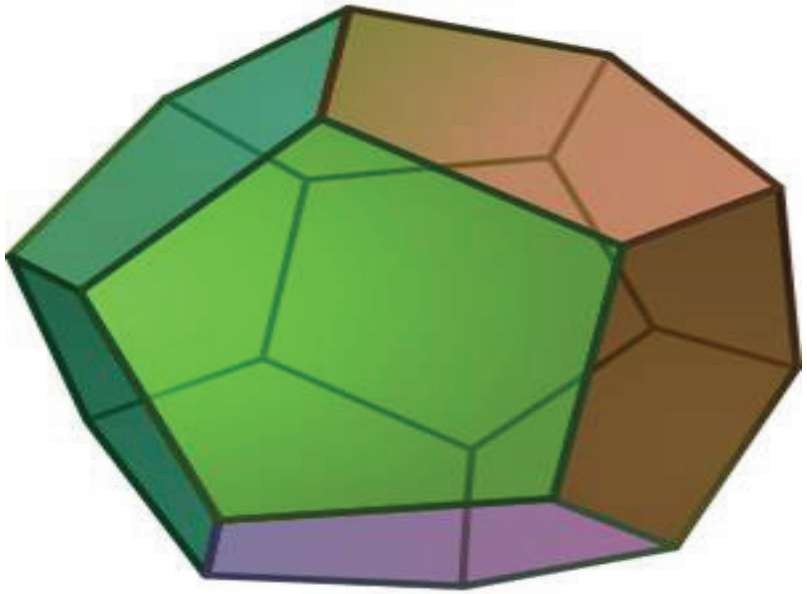
Stanford University

Stanford, CA 94305, U.S.A.

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

Hirsch's Conjecture

Warren Hirsch conjectured in 1957 that the diameter of the graph of a (convex) polyhedron defined by n inequalities in m dimensions is at most $n - m$. The diameter of the graph is the maximum of the shortest paths between every two vertices.



A Counter Example

- Francisco Santos (2010): there is a 43-dimensional polytope with 86 facets and of diameter at least 44.

- There is an infinite family of non-Hirsch polytopes with diameter $(1+\epsilon)n$, even in fixed dimension.

Size of Basic Feasible Solution and Convergence Rate

The simplex method generates a sequence of BFS $\{\mathbf{x}^k\}_{k=0,1,...}$ where the objective value decreases in each step, i.e., $\mathbf{c}^T\mathbf{x}^{k+1} \leq \mathbf{c}^T\mathbf{x}^k$.

Lemma 1 *For every BFS, say $\mathbf{x}_B$, of a LP problem, assume that the sum of its entries is bounded above*

$$\mathbf{e}^T\mathbf{x}_B \leq \Delta,$$

and its smallest entry is bounded below

$$\min\{\mathbf{x}_B\} \geq \delta > 0$$

for some positive constants Δ and δ (non-degenerate case). Then in every pivot step, we have

$$\frac{\mathbf{c}^T\mathbf{x}^{k+1} - z^*}{\mathbf{c}^T\mathbf{x}^k - z^*} \leq 1 - \frac{\delta}{\Delta}$$

where z^* is the minimal objective value of the LP problem.

Proof of the Convergence Rate

Recall at each pivot step,

$$r_e^k = \min_{j \in N}\{r_j^k\} < 0$$

so that

$$\mathbf{c}^T\mathbf{x}^k - z^* \leq -r_e^k \cdot \Delta.$$

On the other hand, we have

$$\mathbf{c}^T\mathbf{x}^{k+1} - \mathbf{c}^T\mathbf{x}^k \leq r_e^k \cdot \delta.$$

Thus

$$(\mathbf{c}^T\mathbf{x}^{k+1} - z^*) - (\mathbf{c}^T\mathbf{x}^k - z^*) \leq r_e \cdot \delta$$

or

$$\frac{\mathbf{c}^T\mathbf{x}^{k+1} - z^*}{\mathbf{c}^T\mathbf{x}^k - z^*} \leq 1 + \frac{r_e \cdot \delta}{\mathbf{c}^T\mathbf{x}^k - z^*} \leq 1 - \frac{\delta}{\Delta}.$$

Implicit Elimination Theorem

Theorem 1 *Let $\mathbf{x}^0$ be any given BFS. Then there is an optimal nonbasic variable $j^0 \in B^0$ and $j^0 \notin B^*$, that would never appear in any of the BFSs generated by the simplex method after $K := \lceil \frac{\Delta}{\delta} \cdot \log\left(\frac{m\Delta}{\delta}\right) \rceil$ steps starting from $\mathbf{x}^0$.*

Then we have

Corollary 1 *For every BFS, say $\mathbf{x}_B$, of a LP problem, let the sum of its entries be bounded above*

$$\mathbf{e}^T \mathbf{x}_B \leq \Delta,$$

and its smallest entry be bounded below

$$\min\{\mathbf{x}_B\} \geq \delta > 0$$

for some positive constants Δ and δ. Then the Simplex method terminates in $\lceil \frac{n\Delta}{\delta} \cdot \log\left(\frac{m\Delta}{\delta}\right) \rceil$ steps.

Proof of the Theorem

If the initial BFS $\mathbf{x}^0$ is not optimal, then we have

$$(\mathbf{s}^*)^T\mathbf{x}^0 = \mathbf{c}^T\mathbf{x}^0 - z^* > 0.$$

Then there must be some index $j^0 \in B^0$ and $j^0 \notin B^*$ such that

$$s^*_{j^0} x^0_{j^0} \geq \frac{\mathbf{c}^T\mathbf{x}^0 - z^*}{m},$$

or

$$s^*_{j^0} \geq \frac{\mathbf{c}^T\mathbf{x}^0 - z^*}{m\Delta}.$$

After $K = \lceil \frac{\Delta}{\delta} \cdot \log\left(\frac{m\Delta}{\delta}\right) \rceil$ steps starting from $\mathbf{x}^0$, from the lemma we must have

$$\mathbf{c}^T\mathbf{x}^K - z^* < \frac{\delta}{m\Delta}(\mathbf{c}^T\mathbf{x}^0 - z^*)$$

and it holds for all subsequent BFSs.

Suppose $j^0 \in B^K$, we have

$$s^*_{j^0} x^K_{j^0} \leq \mathbf{c}^T \mathbf{x}^K - z^* < \frac{\delta}{m\Delta}(\mathbf{c}^T \mathbf{x}^0 - z^*)$$

or

$$s^*_{j^0} < \frac{\mathbf{c}^T \mathbf{x}^0 - z^*}{m\Delta}$$

which gives a contradiction.

The Markov Decision Process

- Markov Decision Processes (MDPs) provide a mathematical framework for modeling sequential decision-making in situations where outcomes are partly random and partly under the control of a decision maker.
- MDPs are useful for studying a wide range of optimization problems solved via Dynamic Programming (DP), where it was known at least as early as the 1950s (cf. Shapley 1953, Bellman 1957).
- Modern applications include dynamic planning, reinforcement learning, social networking, and almost all other dynamic/sequential decision making problems in Mathematical, Physical, Management and Social Sciences.

The Markov Decision Process (continued)

- At each time step, the process is in some state $i \in \{1, ..., m\}$ and the decision maker chooses an action $j \in \mathcal{A}_i$ that is available in state i.
- The process responds at the next time step by randomly moving into a new state i', and giving the decision maker a corresponding cost $c^j(i, i')$.
- The probability that the process changes from i to i' is influenced by the chosen action j in state i. Specifically, it is given by the state transition function $P^j(i, i')$.
- But given i and j, the probability is conditionally independent of all previous states and actions. In other words, the state transitions of an MDP possess the Markov Property.

MDP Stationary Policy

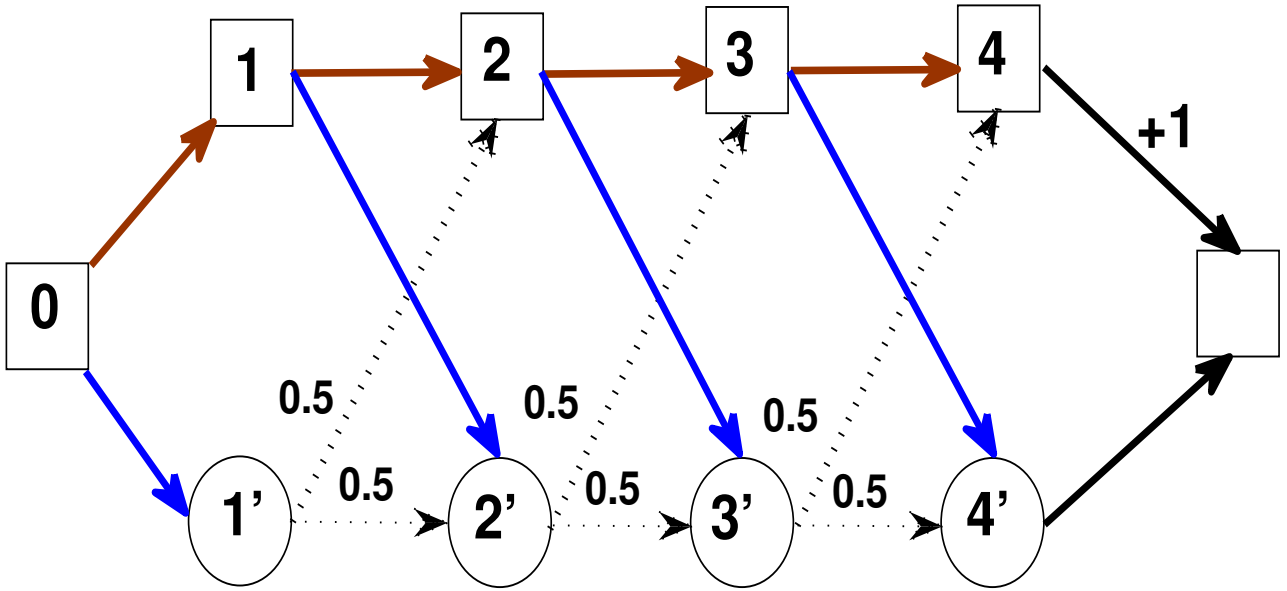
- By a Stationary Policy for the decision maker, we mean a function $\pi = \{\pi_1, \pi_2, \cdots, \pi_m\}$ that specifies an action $\pi_i \in \mathcal{A}_i$ that the decision maker will choose for each state i.

- The min-present cost MDP is to find a stationary policy to minimize the expected discounted sum over an infinite horizon:

$$\sum_{t=0}^{\infty} \gamma^t E[c^{\pi_{i^t}}(i^t, i^{t+1})],$$

  where $0 \leq \gamma < 1$ is a discount rate.

- Typically, we use $\gamma = \frac{1}{1+\rho}$ where ρ is the interest rate.



An MDP Example: Actions are colored in red, blue and black; and all actions have zero cost except the one sending the state 4 to the absorbing state.

Algorithmic Events of the MDP Methods I

- Shapley (1953) and Bellman (1957) developed a method called the Value-Iteration (VI) method to approximate the optimal state values.
- Another best known method is due to Howard (1960) and is known as the Policy-Iteration (PI) method, which generate an optimal policy in finite number of iterations in a distributed and decentralized way.
- de Ghellinck (1960), D'Epenoux (1960) and Manne (1960) showed that the MDP has an LP representation, so that it can be solved by the Simplex method of Dantzig (1947) in finite number of steps, and the Ellipsoid method of Kachiyan (1979) in polynomial time.

The Fixed Point Model of the MDP

$$\begin{cases} y_1 = \min_{j\in\mathcal{A}_1}\{c_j + \gamma \mathbf{p}_j^T \mathbf{y}\} \\ \vdots \\ y_i = \min_{j\in\mathcal{A}_i}\{c_j + \gamma \mathbf{p}_j^T \mathbf{y}\} \\ \vdots \\ y_m = \min_{j\in\mathcal{A}_m}\{c_j + \gamma \mathbf{p}_j^T \mathbf{y}\}, \end{cases}$$

where $\mathcal{A}_i$ represents all actions available in state i, and $\mathbf{p}_j$ is the state transition probabilities from state i to all states when action jth in state i is taken.

The (Dual) LP Form of the MDP

$$\begin{aligned}
\text{maximize}_{\mathbf{y}} \quad & \textstyle\sum_{i=1}^{m} y_i \\
\text{subject to} \quad & y_1 - \gamma \mathbf{p}_j^T \mathbf{y} \leq c_j,\ j \in \mathcal{A}_1 \\
& \vdots \\
& y_i - \gamma \mathbf{p}_j^T \mathbf{y} \leq c_j,\ j \in \mathcal{A}_i \\
& \vdots \\
& y_m - \gamma \mathbf{p}_j^T \mathbf{y} \leq c_j,\ j \in \mathcal{A}_m.
\end{aligned}$$

The Interpretations of the LP Dual Formulation

The LP variables $\mathbf{y} \in \mathbf{R}^m$ represent the expected present cost-to-go values of the m states, respectively, for a given policy.

The LP problem entails choosing variables in $\mathbf{y}$, one for each state i, that maximize $\mathbf{e}^T\mathbf{y}$ so that it is the fixed point

$$y_i^* = \min_{j \in \mathcal{A}_i} \{\mathbf{c}_{j_i} + \gamma \mathbf{p}_{j_i}^T \mathbf{y}\}, \ \forall i,$$

with an optimal policy

$$\pi_i^* = \arg\min \{\mathbf{c}_j + \gamma \mathbf{p}_j^T \mathbf{y}, \ j \in \mathcal{A}_i\}, \ \forall i.$$

It is well known that there exist a unique optimal stationary policy $(\mathbf{y}^*, \pi^*)$ where, for each state i, y_i^* is the minimum expected present cost that an individual in state i and its progeny can incur.

The MDP-LP Primal Formulation

$$\min_{\mathbf{x}} \quad \sum_{j\in\mathcal{A}_1} c_j x_j + \quad \ldots \quad + \sum_{j\in\mathcal{A}_m} c_j x_j$$

$$\begin{array}{lccccccc} \text{s.t.} & \sum_{j\in\mathcal{A}_1}(\mathbf{e}_1 - \gamma\mathbf{p}_j)x_j + & \ldots & + \sum_{j\in\mathcal{A}_m}(\mathbf{e}_m - \gamma\mathbf{p}_j)x_j & = & \mathbf{e}, \\ & \ldots & x_j & \ldots & \geq & 0, \ \forall j, \end{array}$$

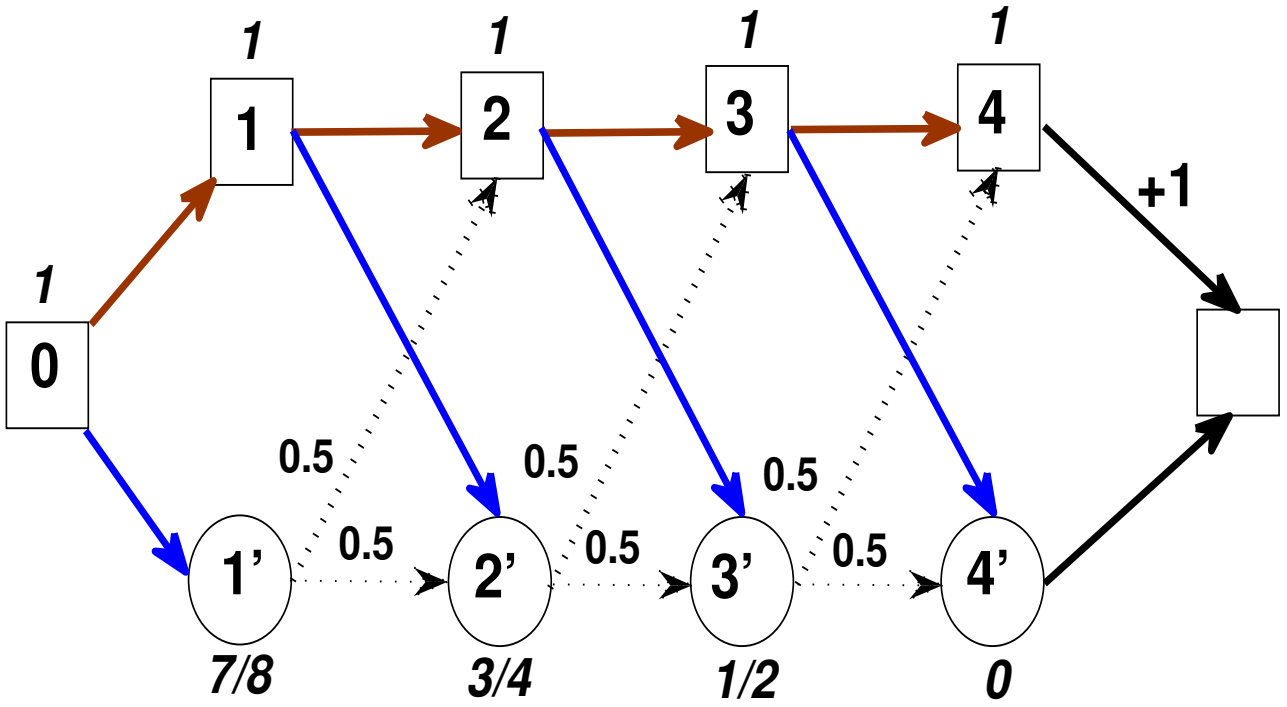
where $\mathbf{e}$ is the vector of ones, and $\mathbf{e}_i$ is the unit vector with 1 at the i-th position.

Variable x_j, $j \in \mathcal{A}_i$, is the state-action frequency, or the expected present value of the number of times in which an individual is in state i and takes state-action j.

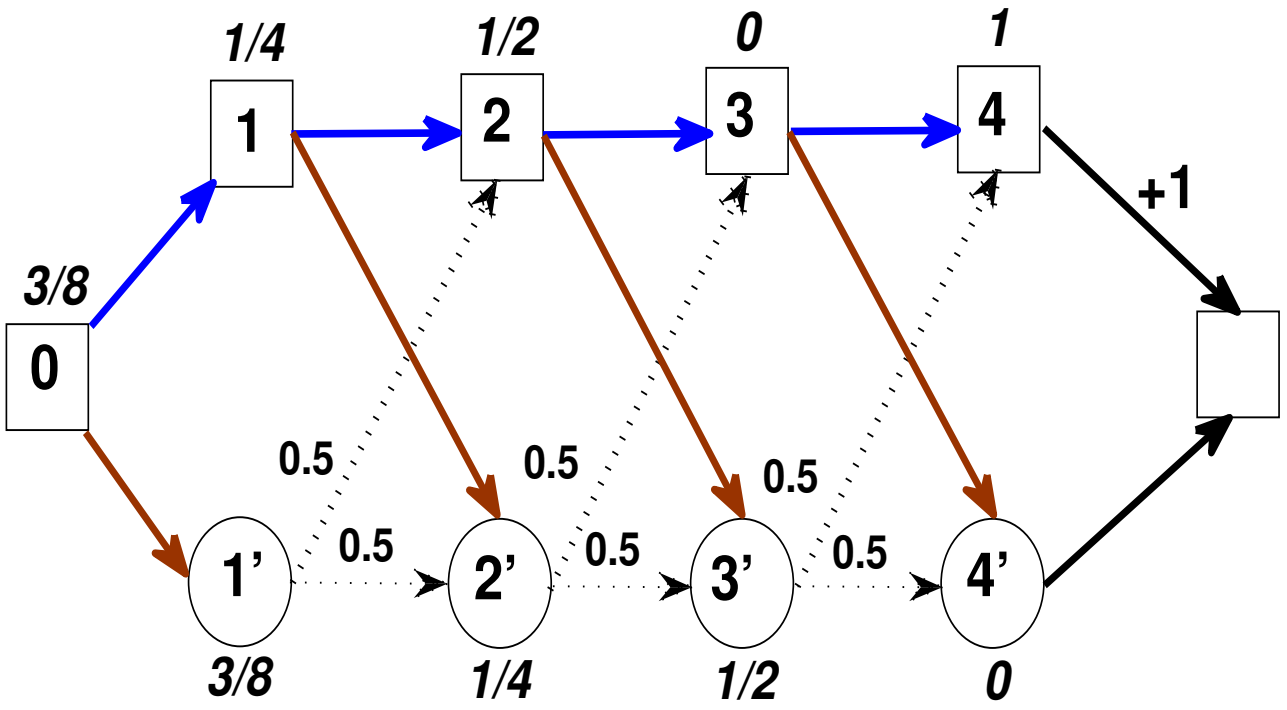
Thus, solving the problem entails choosing a state-action frequencies that minimize the expected present value sum of total costs.

The MDP Example in LP Form

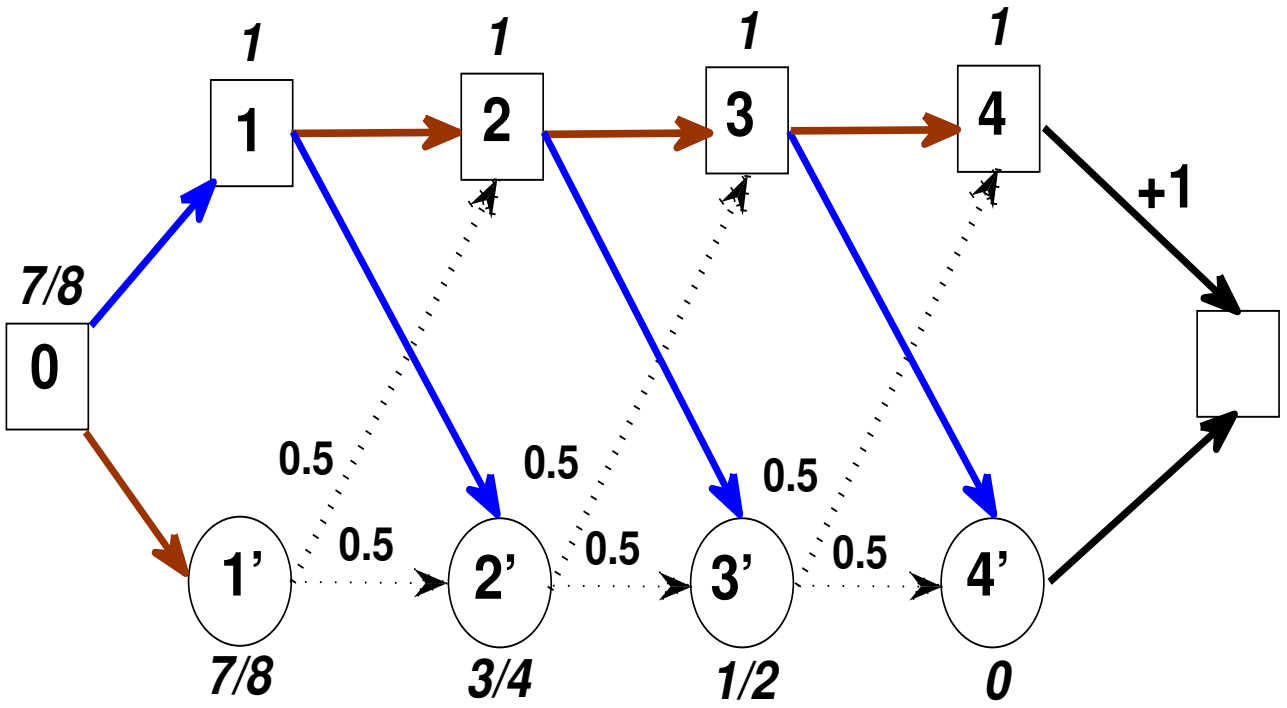
a:	(0_1)	(0_2)	(1_1)	(1_2)	(2_1)	(2_2)	(3_1)	(3_2)	(4_1)	$(4_1')$
c:	0	0	0	0	0	0	0	0	1	0
(0)	1	1	0	0	0	0	0	0	0	0
(1)	$-\gamma$	0	1	1	0	0	0	0	0	0
(2)	0	$-\gamma/2$	$-\gamma$	0	1	1	0	0	0	0
(3)	0	$-\gamma/4$	0	$-\gamma/2$	$-\gamma$	0	1	1	0	0
(4)	0	$-\gamma/8$	0	$-\gamma/4$	0	$-\gamma/2$	$-\gamma$	0	$1-\gamma$	0
$(4')$	0	$-\gamma/8$	0	$-\gamma/4$	0	$-\gamma/2$	0	$-\gamma$	0	$1-\gamma$



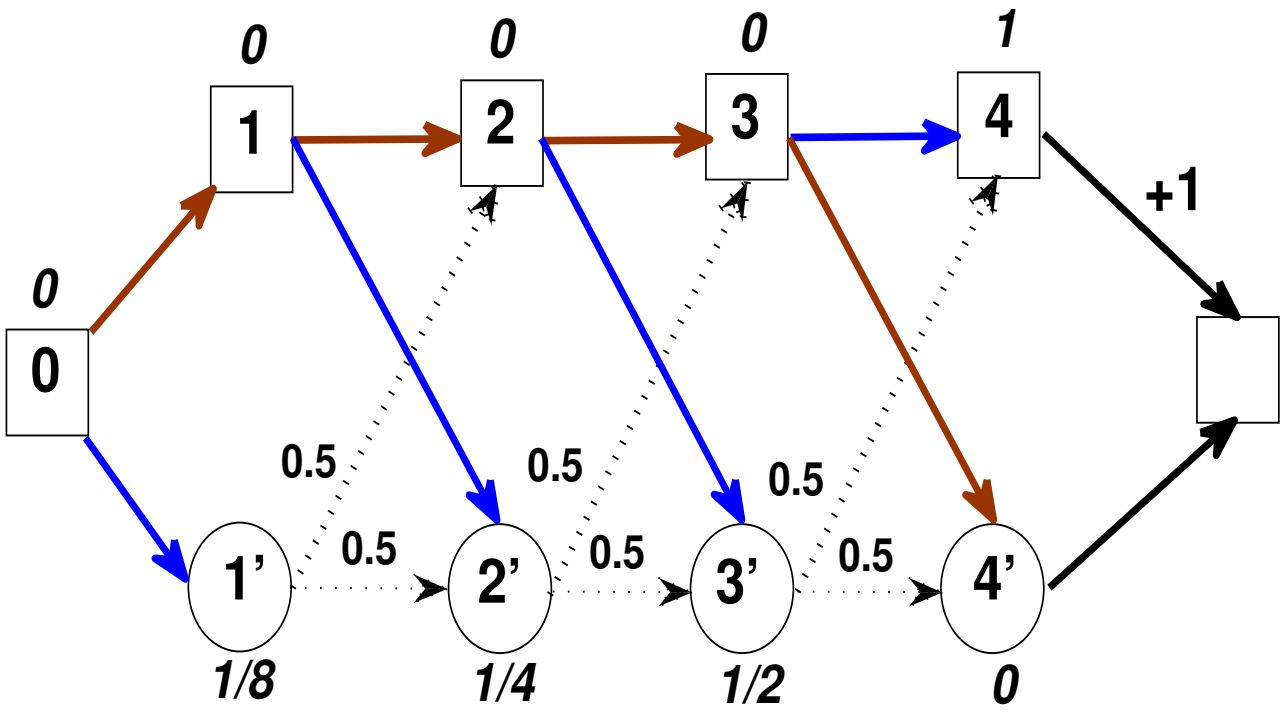
Pricing: the Cost-to-Go Values of the States: cost-to-go values on each state when actions colored in red are taken; the policy is not optimal.



The Policy Iteration (PI): New values on each state when actions in red are taken.



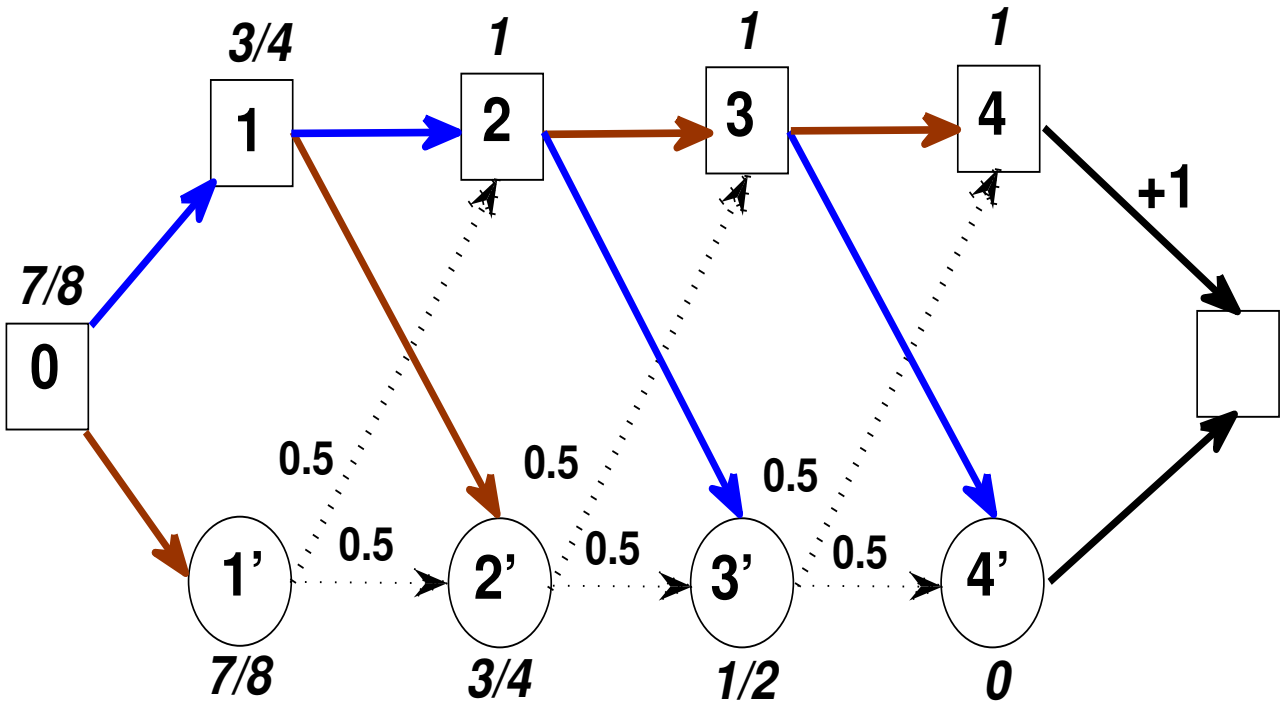
The Simplex or Simple Policy Index-Rule Iteration: New values on each state when actions in red are taken.



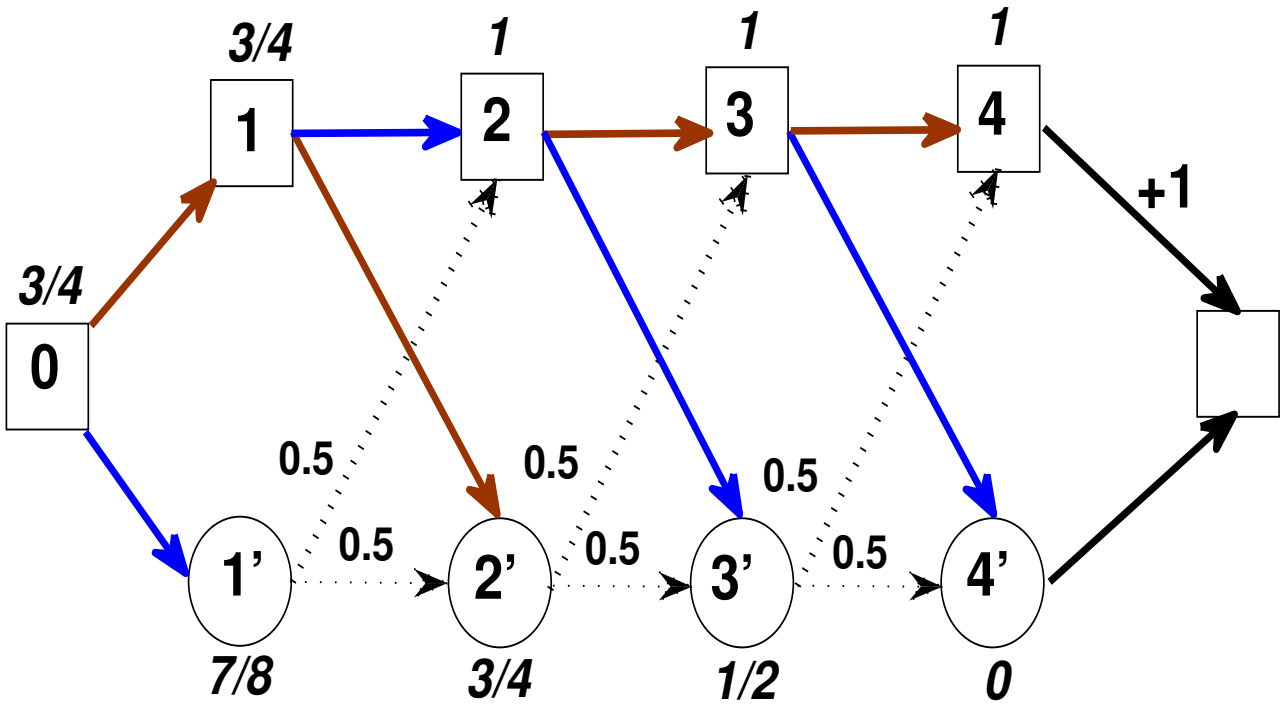
The Simplex or Simple Policy Greedy-Rule Iteration: New values on each state when actions in red are taken.

Complexity of the Policy Iteration and Simplex Methods

- In practice, the Policy Iteration (PI) method, including the simple policy iteration or Simplex method, has been remarkably successful and shown to be most effective and widely used.
- Mansour and Singh in 1994 gave an upper bound on the number of iterations, $2^m/m$, for the policy-iteration method when each state has 2 actions.
- A negative result, similar to Klee and Minty (1972), of Melekopoglou and Condon (1990) showed that a simple Policy Iteration method, where in each iteration only the action for the state with the smallest index is updated, needs an exponential number of iterations to compute an optimal policy for a specific MDP problem regardless of the discount rates.
- In the past 50 years, many efforts have been made to resolve the worst-case complexity issue of the Policy Iteration method or the Simplex method, and to answer the question: are they (strongly) polynomial-time algorithms?



The Simplex or Simple Policy Index-Rule Iteration II: New values on each state when actions in red are taken.



The Simplex or Simple Policy Index-Rule Iteration III: New values on each state when actions in red are taken

The Discounted MDP Properties

Lemma 2 *The discounted MDP primal LP formulation has the following properties:*

1. *The feasible set is bounded. More precisely, for every feasible* $\mathbf{x} \geq \mathbf{0}$, $\mathbf{e}^T\mathbf{x} = \frac{m}{1-\gamma}$
2. *There is a one-to-one correspondence between a stationary policy of the original discounted MDP and a basic feasible solution (BFS) of the primal.*
3. *Every policy or BFS basis has the Leontief substitution form* $A_\pi = I - \gamma P_\pi$.
4. *Let* $\mathbf{x}^\pi$ *be a basic feasible solution. Then any basic variable, say* $\mathbf{x}^\pi_i$, *has its value* $1 \leq \mathbf{x}^\pi_i \leq \frac{m}{1-\gamma}$.

Precise Complexity Results

- The classic simplex and policy iteration methods, with the greedy pivoting rule, are a strongly polynomial-time algorithm for MDP with fixed discount rate. The method terminates in a number of steps bounded by $\frac{mn}{1-\gamma} \cdot \log\left(\frac{m^2}{1-\gamma}\right)$, and each step uses at most $O(mn)$ arithmetic operations, where n is the total number of actions.

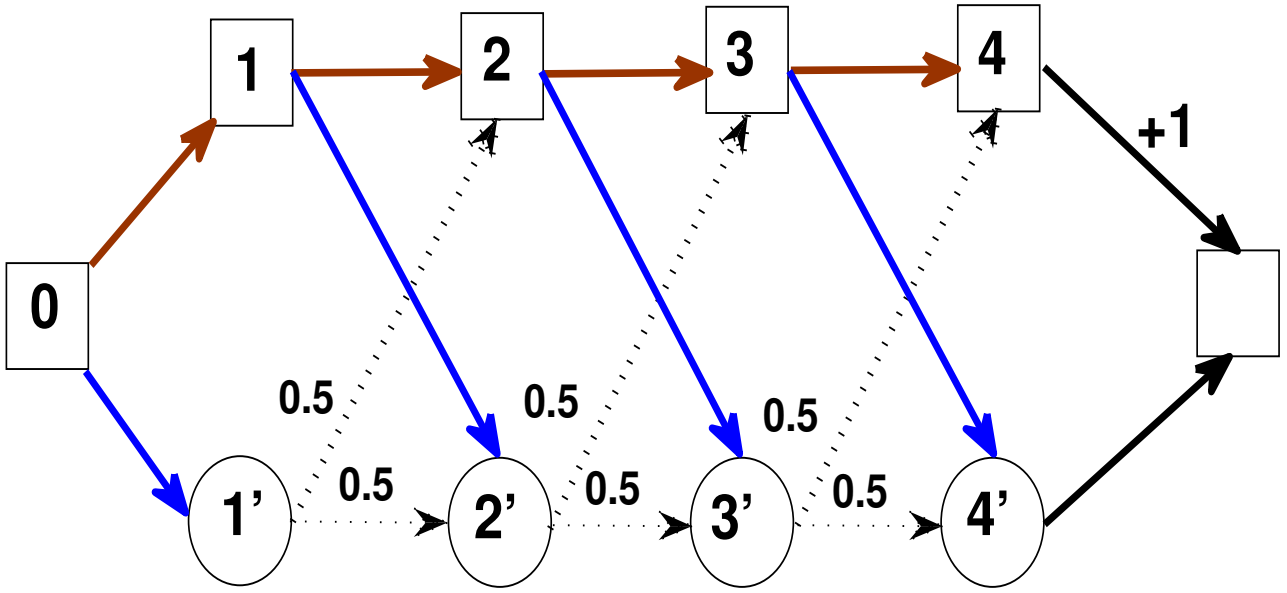
- The policy-iteration method terminates in no more

$$\frac{n}{1-\gamma} \cdot \log\left(\frac{m}{1-\gamma}\right),$$

steps and each step uses at most $m^2 n$ arithmetic operations (Hansen, Miltersen, and Zwick, September 2010).

The Shapley Two-Person Zero-Sum Stochastic Game

- Similar to the Markov decision process, but the states is partitioned to two sets where one is to maximize and the other is to minimize.
- It has no linear programming formulation, and it is unknown if it can be solved in polynomial time in general.
- For a fixed discount rate, it can be solved in polynomial time (Littman 1996) using the value iteration method.
- Hansen, Miltersen and Zwick (2010) very recently proved that the strategy iteration method solves it in strongly polynomial time when discount rate is fixed. This is the first strongly polynomial time algorithm for solving the discounted game.



A Markov Game Process Example: $\{3, 4\}$ want to maximize while $\{0, 1, 2\}$.

Remarks and Open Questions I

- The performance of the simplex method is very sensitive to the pivoting rule.
- Tatonnement and decentralized process works under the Markov property.
- Greedy or Steepest Descent works when there is a discount!
- Multi-updates or pivots work better than a single-update does; policy iteration vs. simplex.
- The proof techniques are generalized to solving general linear programs by Kitahara and Mizuno (2010).

Remarks and Open Questions II

- Can the iteration bound for the simplex method be reduced to linear in the number of actions?
- Is the simplex or policy iteration method polynomial for the MDP regardless of discount rate γ or input data?
- Is there an MDP algorithm whose running time is strongly polynomial regardless of discount rate γ?
- Is there a Stochastic Game algorithm whose running time is polynomial regardless of discount rate γ?
- Is there a strongly polynomial-time algorithm for LP?