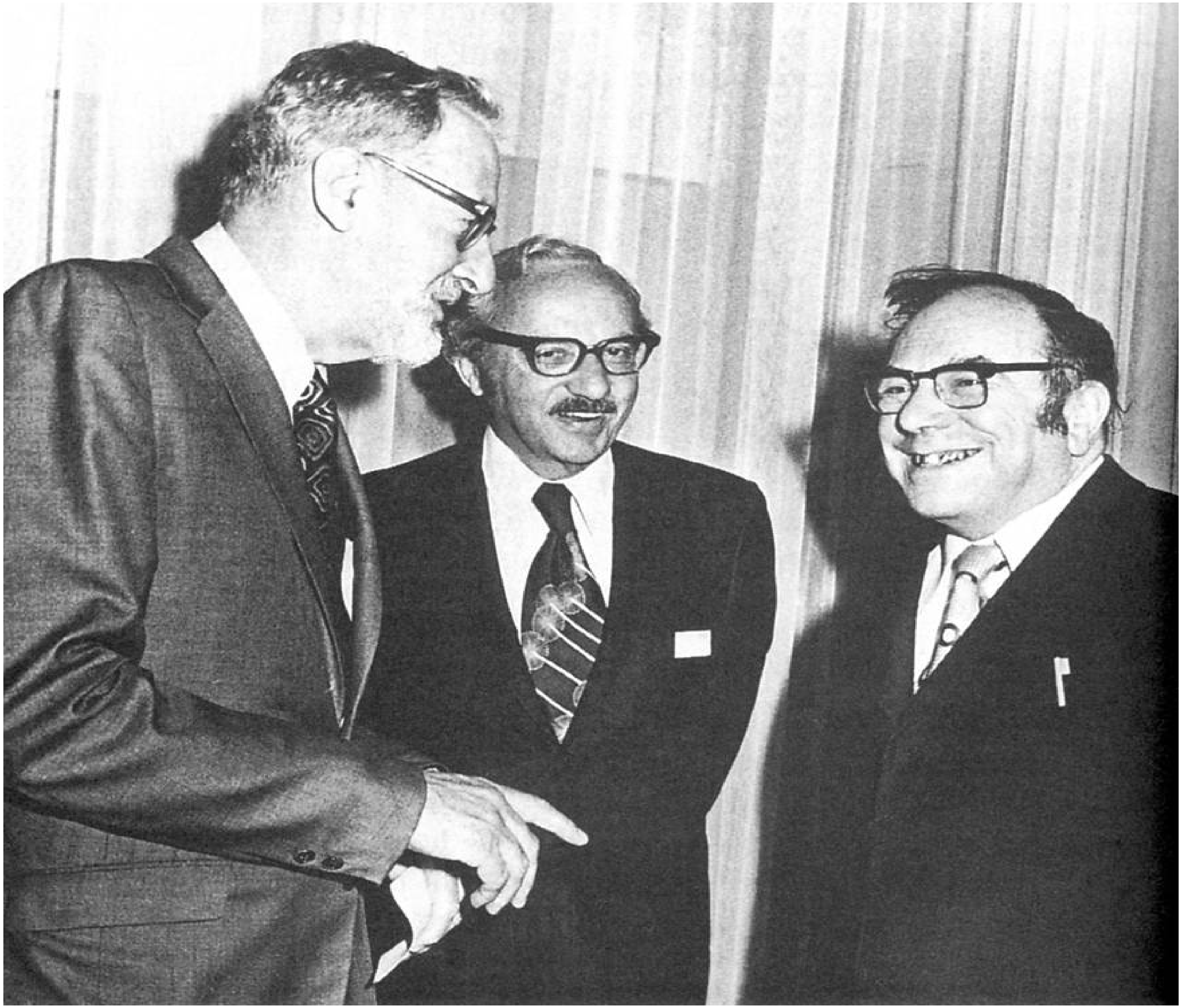
Tseng Lecture
Recent Progresses on Linear Programming and the Simplex Method

Yinyu Ye

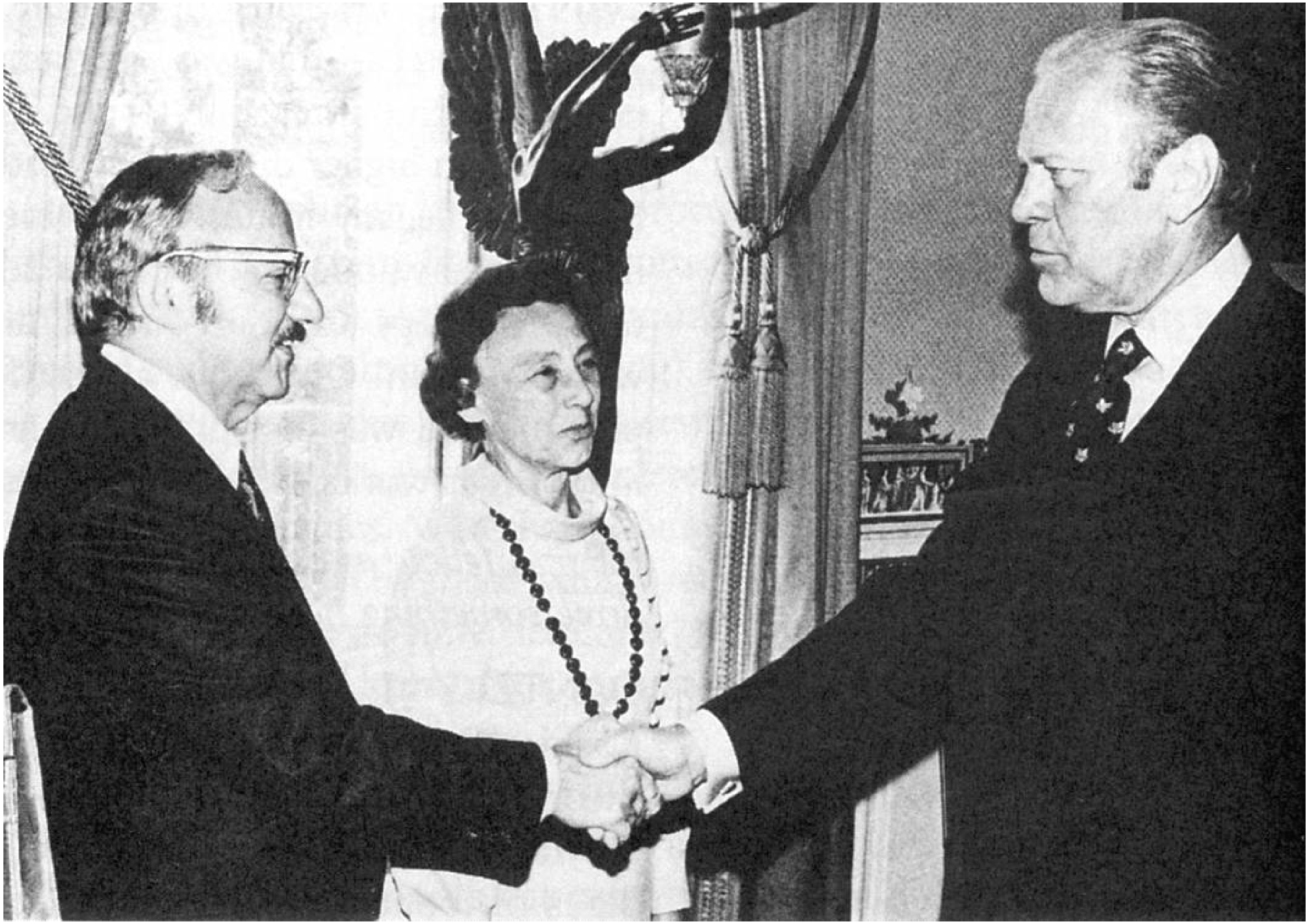
Management Science and Engineering
and
Institute for Computational and Mathematical Engineering
Stanford University

Research supported in part by AFOSR

Linear Programming started…



… with the simplex method



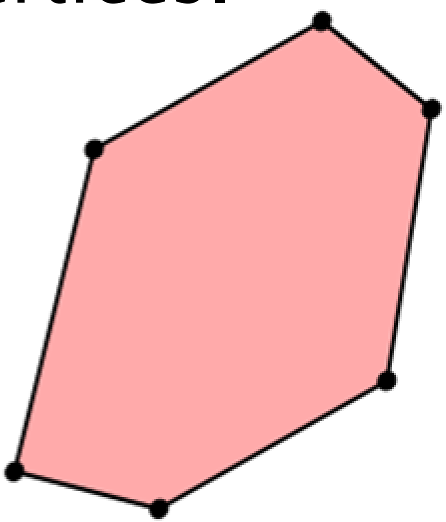
Outlines

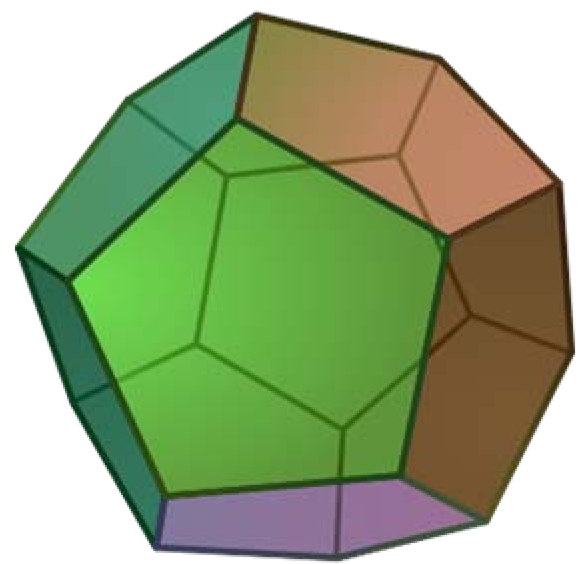
- Counterexamples to the Hirsch conjecture
- More pivoting rules and their behavior
- Simplex and policy-iteration methods for stochastic Markov Decision Process (MDP) and Zero-Sum Game with fixed discounts
- Simplex method for deterministic MDP with variable discounts
- Other efficient methods and results for linear programming

De Loera, "New Insights into the Complexity and Geometry of Linear Optimization," OPTIMA, 2011.

Hirsch's Conjecture

- Warren Hirsch conjectured in 1957 that the diameter of the graph of a polyhedron defined by *n* inequalities in *d* dimensions is at most *n*−*d*.
- The diameter of the graph is the maximum of the shortest paths between every two vertices.





Counter examples to Hirsch's conjecture

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 $(1 + \varepsilon)n$, even in fixed dimension.
- Santos' construction is an extension of a result of Klee and Walkup (1967), where they proved that the Hirsch conjecture could be proved true from just the case $n = 2d$.

More pivoting rules …

- The simplex method is governed by a pivot rule, i.e. a method of choosing adjacent vertices with a better objective function value.
- Klee and Minty (1972) showed that Dantzig's original greedy pivot rule may require exponentially many steps.
- The random edge pivot rule chooses, from among all improving pivoting steps (or edges) from the current basic feasible solution (or vertex), one uniformly at random.
- The Zadeh pivot rule chooses the decreasing edge or the entering variable that has been entered least often in the previous pivot steps.

… and they fall as well

- No non-polynomial lower bounds were known until now for these two pivot rules.
- Friedmann, Hansen and Zwick (2011) gave an example that the random edge pivot rule needs sub-exponentially many steps.
- Friedman (2011) developed an example that the Zadeh pivot rule needs exponentially many steps.
- These examples explore the connection of linear programming and Markov Decision Process (MDP), and the close relation between the simplex method for solving linear programs and the policy iteration method for MDP.

(The diameter of MDP polytopes is bounded by *d*.)

Markov Decision Process

- Markov decision process provides 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, 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, Economics, and Social Sciences.

States and Actions

- At each time step, the process is in some state $i = 1, \ldots, m$, and the decision maker chooses an action $j \in A_i$ that is available for state *i,* say of total *n* actions.
- The process responds at the next time step by randomly moving into a new state *i'*, and giving the decision maker an immediate corresponding cost c_j.
- The probability that the process enters *i'* as its new state is influenced by the chosen action *j* . Specifically, it is given by the state transition probability distribution P_j.
- But given action *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.

Policy and Discount Factor

- A policy of MDP is a set function $\pi = \{j_1, j_2, \cdots, j_m\}$ that specifies one action $j_i \in A_i$ that the decision maker will choose for each state i.
- The MDP is to find an optimal (stationary) policy to minimize the expected discounted sum over an infinite horizon with a discount factor $0 \leq \gamma < 1$.
- One can obtain an LP that models the MDP problem in such a way that there is a one-to-one correspondence between policies of the MDP and basic feasible solutions of the (dual) LP, and between improving switches and improving pivots.

de Ghellinck (1960), D'Epenoux (1960) and Manne (1960)

Cost-to-Go values and LP formulation

- Let $y \in R^m$ represent the expected present cost-to-go values of the m states, respectively, for a given policy. Then, the cost-to-go vector of the optimal policy is a Fixed Point of

$$y_i = \min\{c_j + \gamma p_j^T y, j \in A_i\}, \forall i,$$

$$j_i = \arg\min\{c_j + \gamma p_j^T y, j \in A_i\}, \forall i.$$

- Such a fixed point computation can be formulated as an LP

$$\max \quad \sum_{i=1}^{m} y_i$$

$$\text{s.t.} \quad y_i \le c_j + \gamma p_j^T y, \forall j \in A_i; \forall i.$$

The dual of the MDP-LP

$$\min \quad \sum_{i=1}^{n} c_j x_j$$

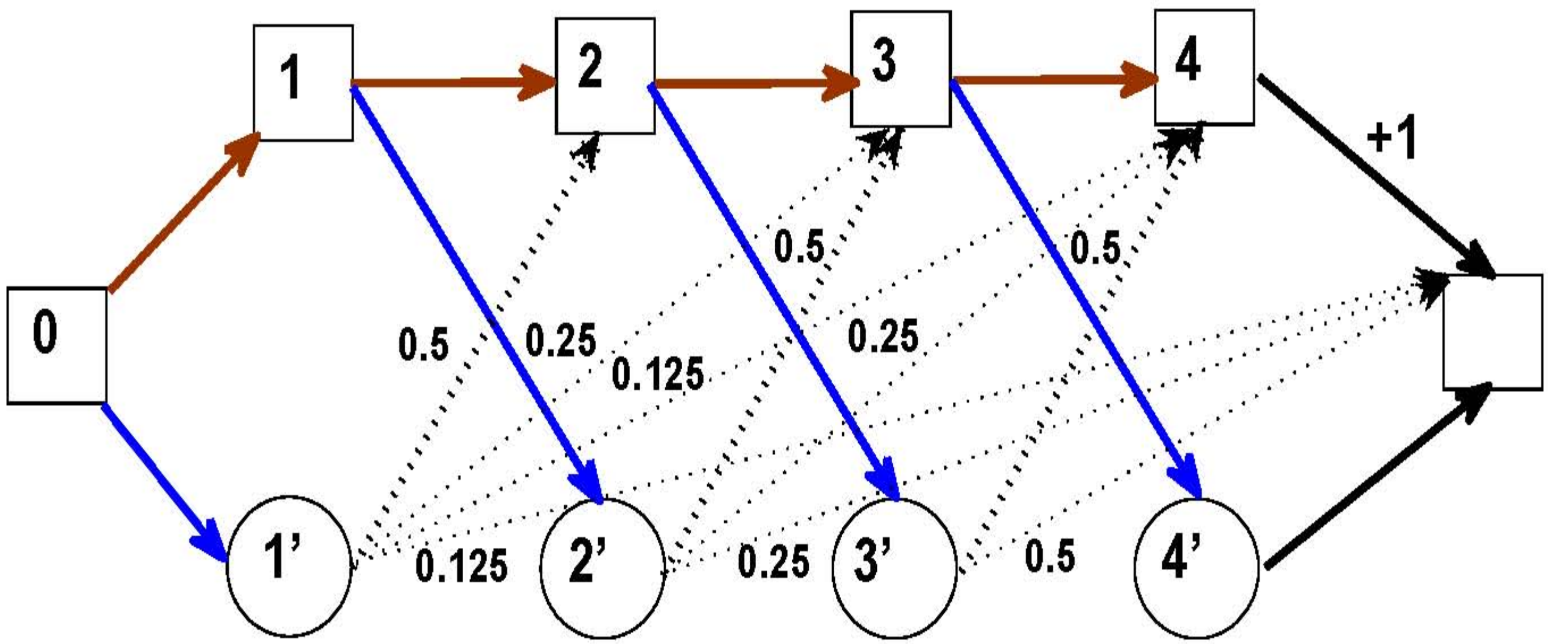
$$\text{s.t.} \quad \sum_{j=1}^{n} (e_{ij} - \gamma p_{ij}) x_j = 1, \forall i,$$

$$x_j \geq 0, \forall j.$$

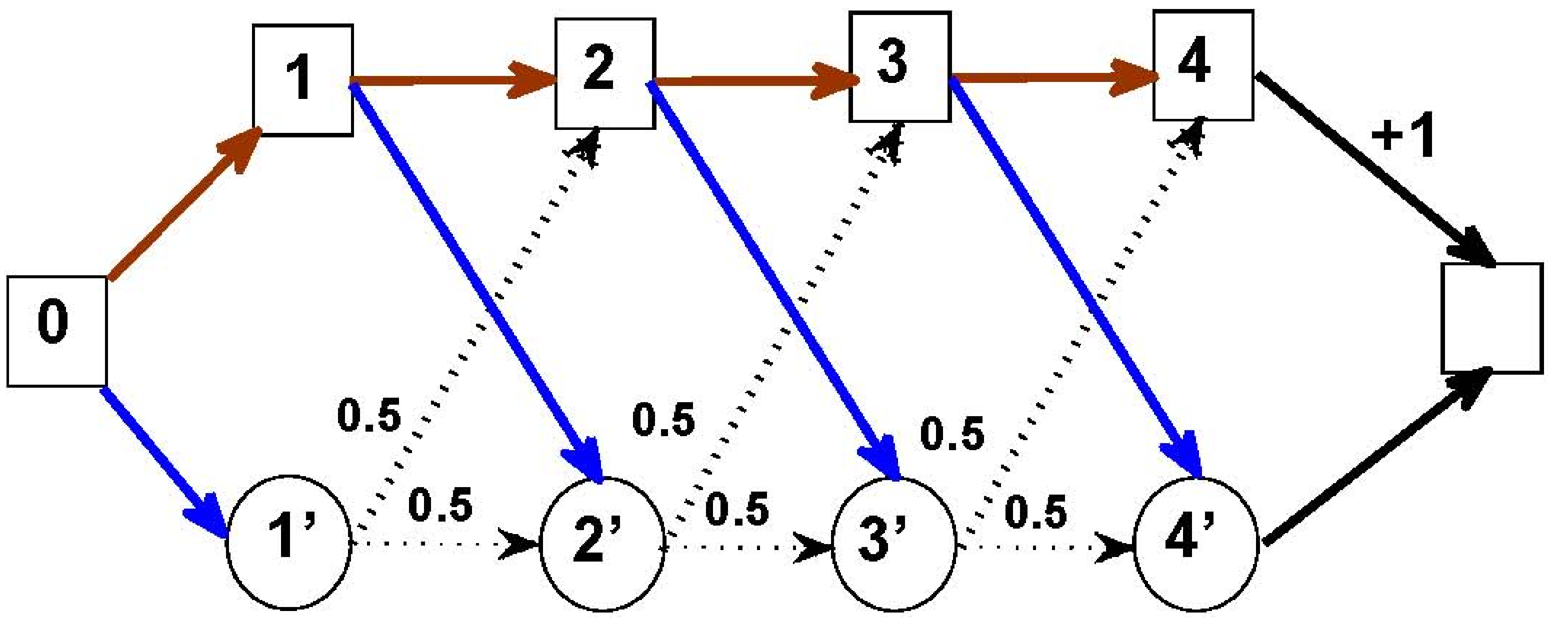
where $e_{ij} = 1$ if $j \in A_i$ and 0 otherwise.

Dual variable x_j represents the expected action flow or visit-frequency, that is, the expected present value of the number of times action j is used.

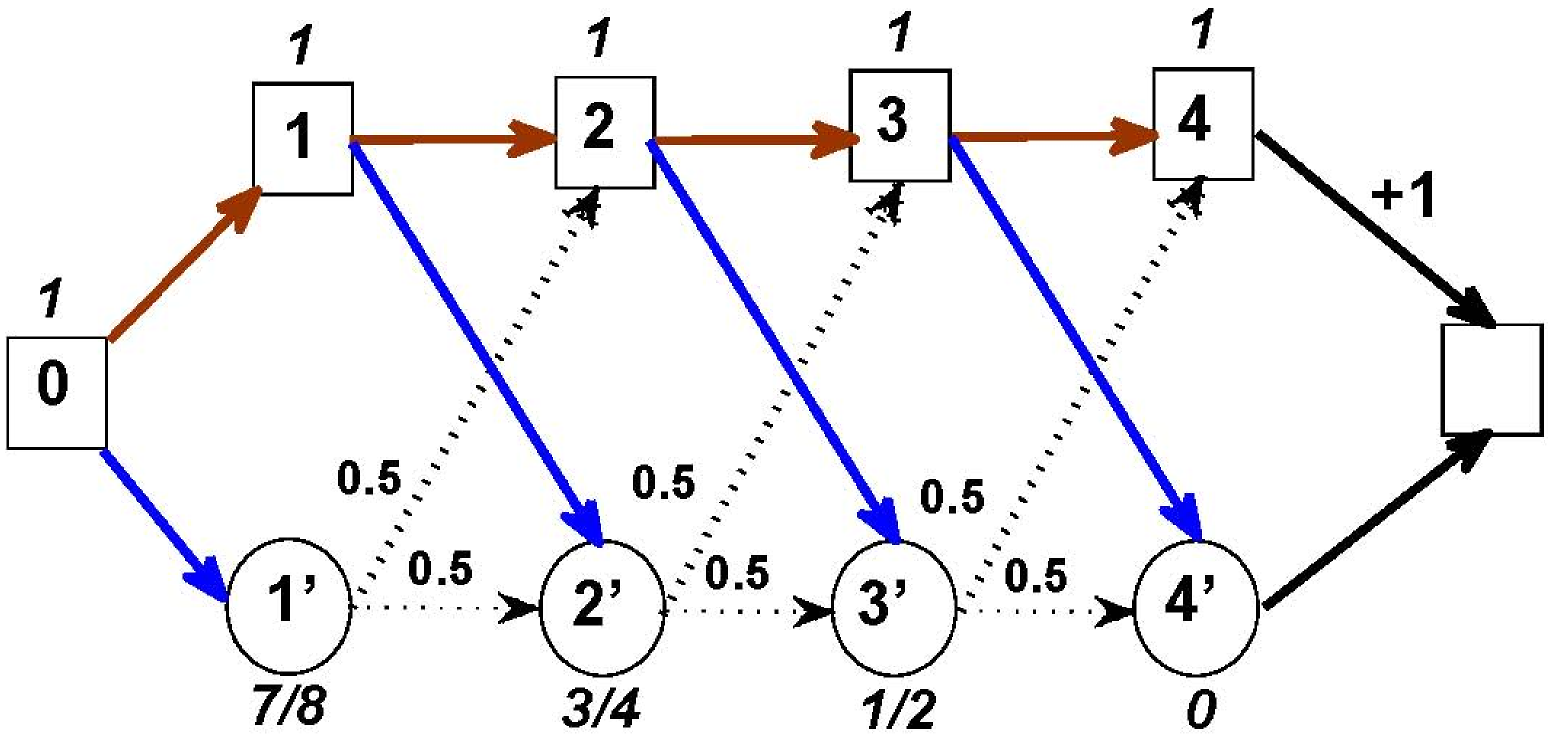
A Simple MDP Problem I



A Simple MDP Problem II

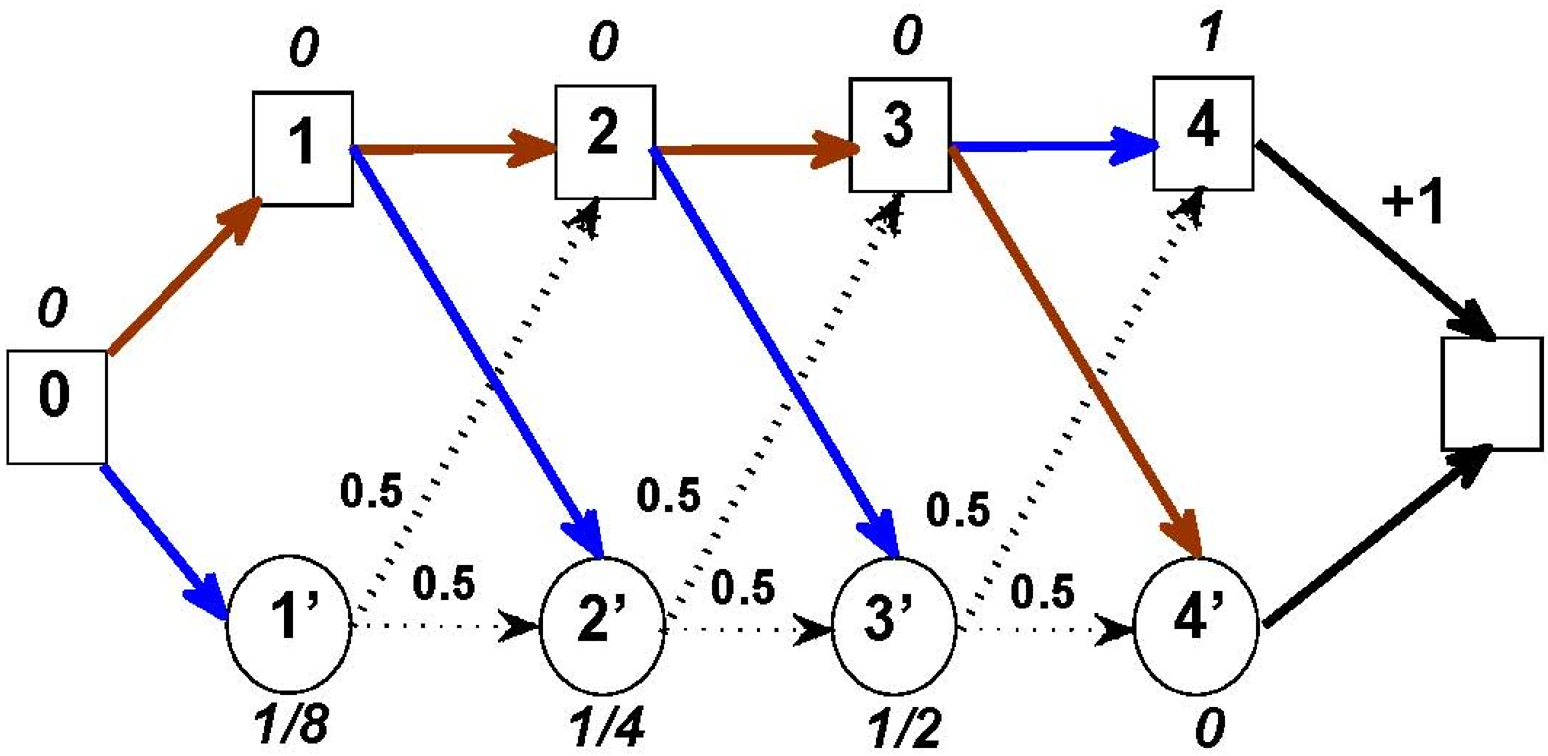


Cost-to-Go values



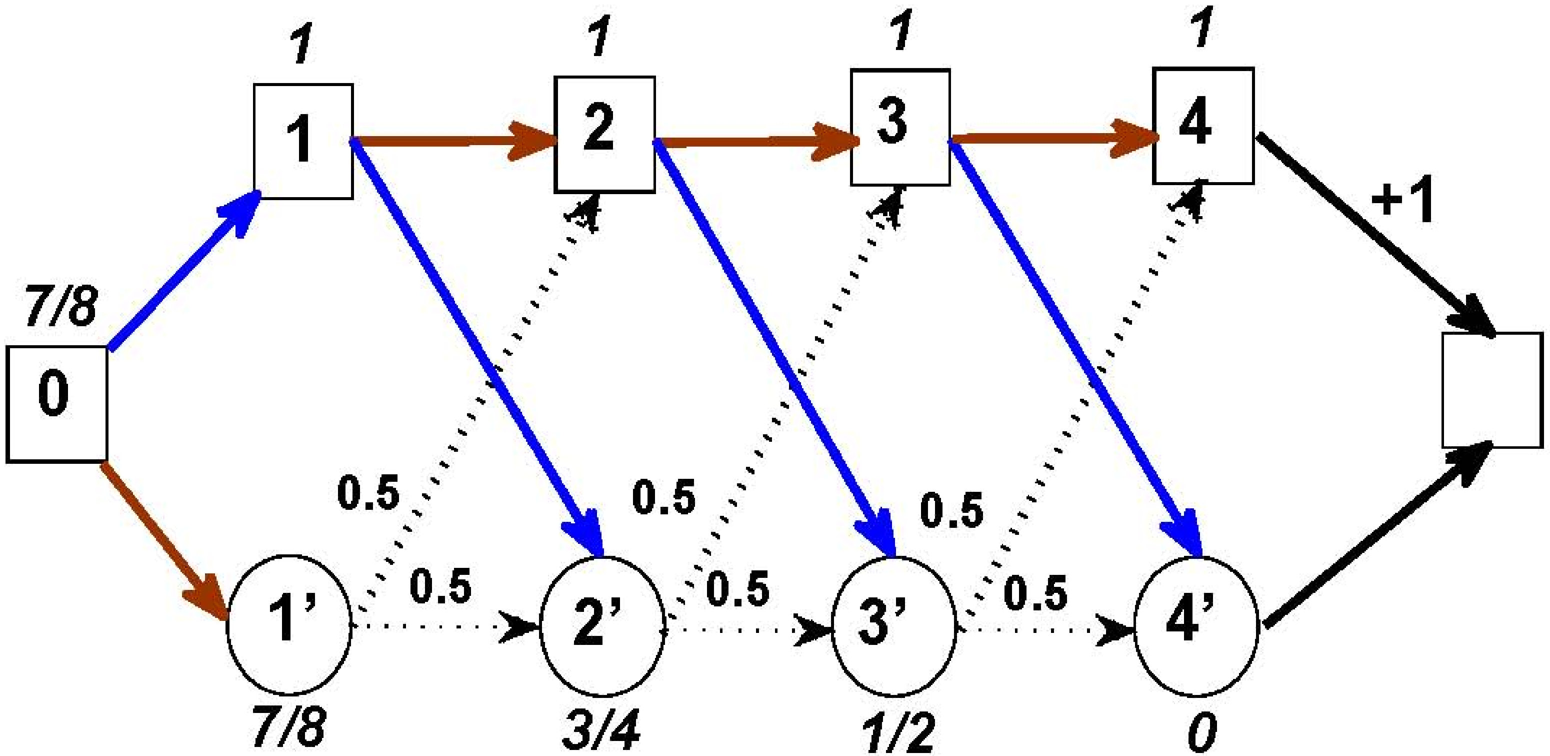
Chosen actions in Red

Greedy Simplex Rule



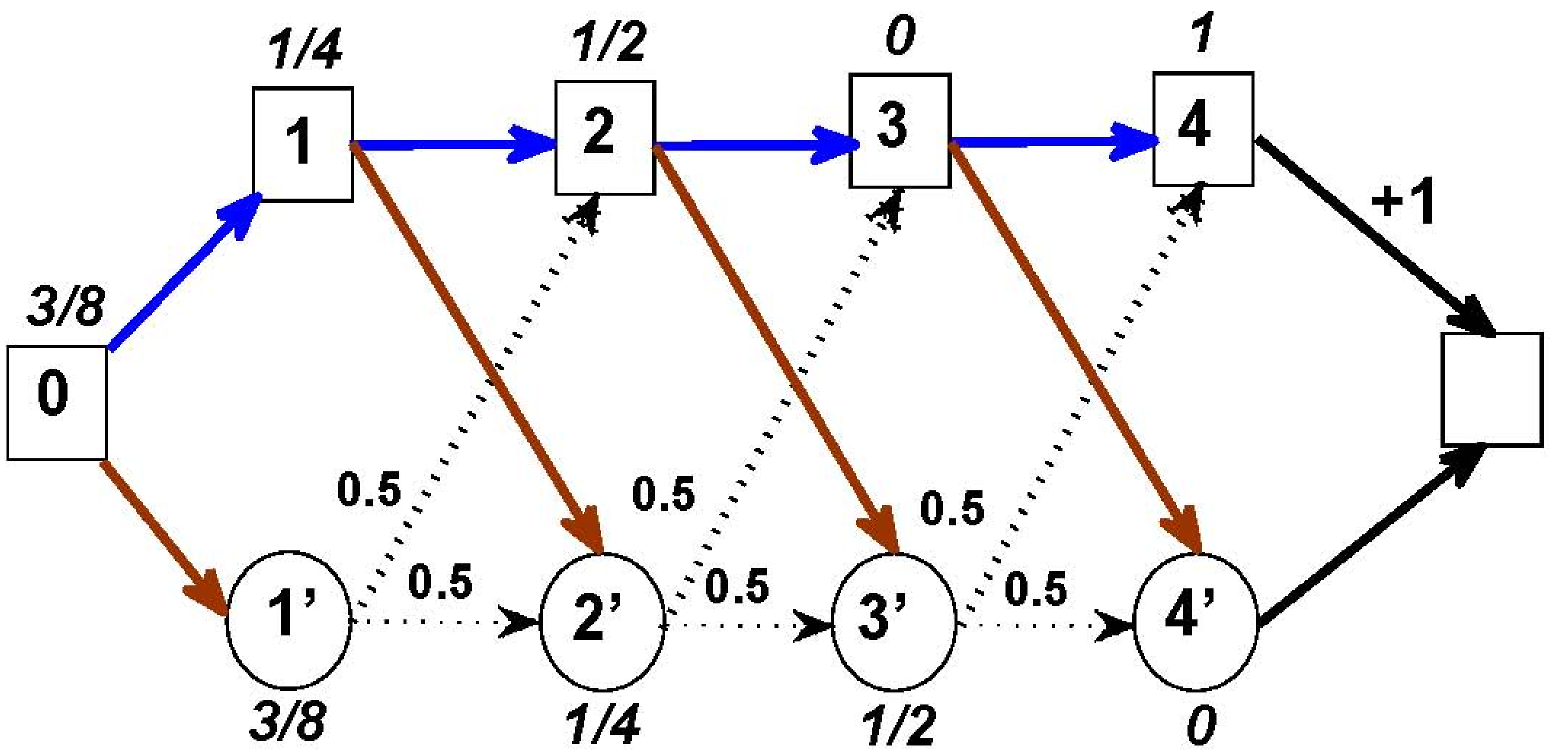
Chosen actions in Red

Lowest-Index Simplex Rule



Chosen actions in Red

Policy Iteration Rule (Howard 1960)



Chosen actions in Red

Efficiency of simplex/policy methods

- Early work included that of Paul Tseng (1990)
- Melekopoglou and Condon (1990) showed that the simplex method with the smallest index pivot rule needs an exponential number of iterations to compute an optimal policy for a specific MDP problem regardless of discount factors.
- Fearnley (2010) showed that the policy-iteration method needs an exponential number of iterations for a undiscounted finite-horizon MDP.
- In practice, the policy-iteration method, including the simplex method with greedy pivot rule, has been remarkably successful and shown to be most effective and widely used.
- Are the policy-iteration method and the simplex method efficient for MDP with discounts, or are they strongly polynomial-time algorithms? – A vindication?

Bound on the simplex/policy methods

- Y (2011): The classic simplex and policy iteration methods, with the greedy pivoting rule, terminate in no more than

$$\frac{mn}{1-\gamma}\log\left(\frac{m^2}{1-\gamma}\right)$$

  pivot steps, where n is the total number of actions in an m–state MDP with discount factor γ.
- This is a strongly polynomial-time upper bound when γ is bounded above by a constant less than one.

Roadmap of proof

- Define a combinatorial event that cannot repeats more than *n* times. More precisely, at any step of the pivot process, there exists a non-optimal action *j* that will never re-enter future policies or bases after
$$\frac{m}{1-\gamma}\log\left(\frac{m^2}{1-\gamma}\right)$$
pivot steps
- There are at most *(n – m)* such non-optimal action to eliminate from appearance in any future policies generated by the simplex or policy-iteration method.
- The proof relies on the duality, the reduced-cost vector at the current policy and the optimal reduced-cost vector to provide a lower and upper bound for a non-optimal action when the greedy rule is used.

Improvement and extension

Hansen, Miltersen and Zwick (2011):

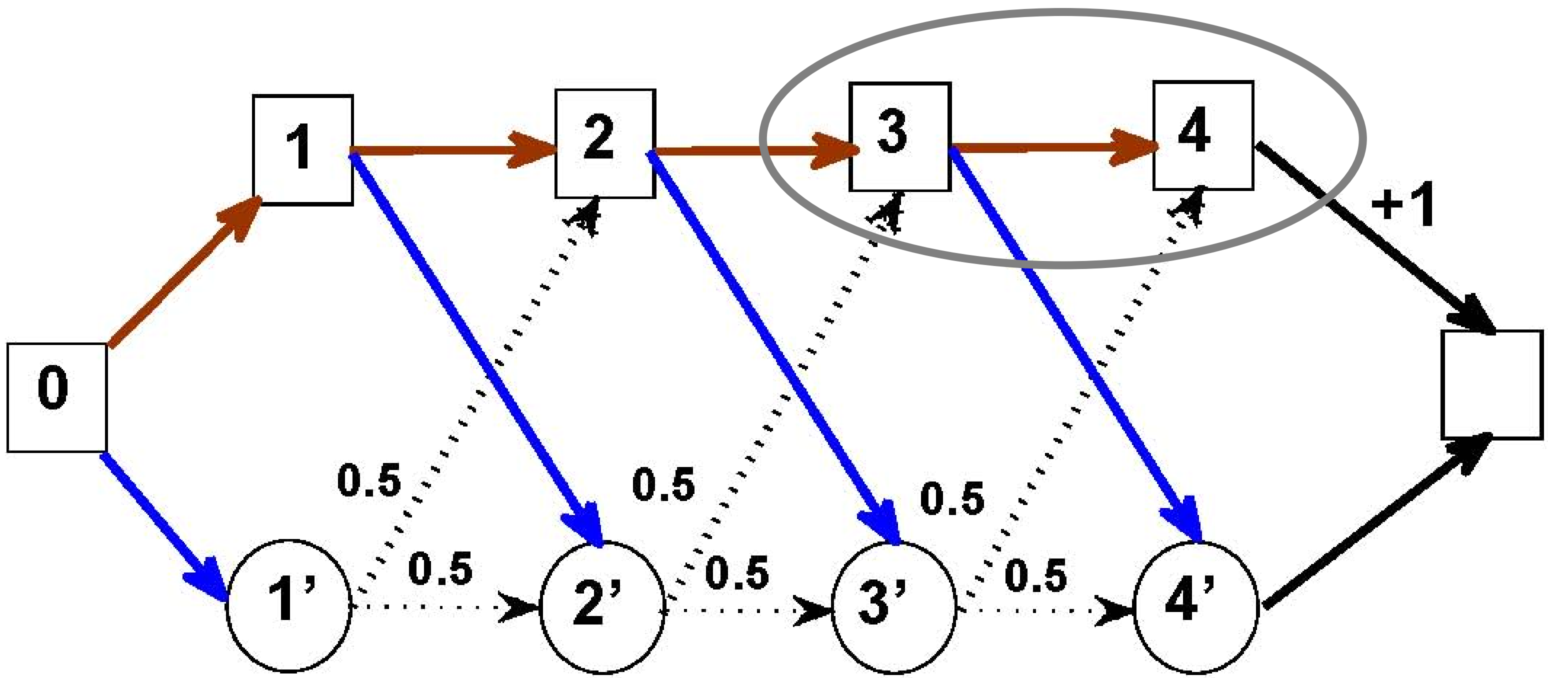
- For the policy iteration method, there exists a non-optimal action *j* that will never re-enter policies after

$$\frac{1}{1-\gamma}\log\left(\frac{m^2}{1-\gamma}\right)$$

  pivot steps.

- The simplex and policy iteration methods, with the greedy pivoting rule, are strongly polynomial-time algorithms for Turn-Based Two-Person Zero-Sum Stochastic Game with any fixed discount factor, which problem cannot even be formulated as an LP.

A Turn-Based Zero-Sum Game



Improvement and extension

- Kitahara and Mizuno (2011) extended the bound to solving general non-degenerate LPs:

$$\min \quad \sum_{i=1}^{n} c_j x_j$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i, \forall i; \ x_j \geq 0, \forall j.$$

- The simplex method terminates in at most

$$\frac{mn}{\sigma} \log\left(\frac{m^2}{\sigma}\right)$$

pivot steps, when the ratio of the minimum value over the maximum value, in all basic feasible solution entries, is bounded below by σ.

Deterministic MDP with discounts

Distribution vector $p_j \in R^m$ contains exactly one *1* and *0* everywhere else

$$y_i = \min\{c_j + \gamma_j p_j^T y,\ j \in A_i\}, \forall i,$$

$$j_i = \arg\min\{c_j + \gamma_j p_j^T y,\ j \in A_i\}, \forall i.$$

$$\begin{array}{ll} \max & \sum_{i=1}^{m} y_i \\ \text{s.t.} & y_i \le c_j + \gamma_j p_j^T y, \forall j \in A_i; \forall i. \end{array}$$

It has uniform discounts if all γ_j are identical.

The dual resembles generalized flow

$$\min \quad \sum_{i=1}^{n} c_j x_j$$

$$\text{s.t.} \quad \sum_{j=1}^{n} (e_{ij} - \gamma_j p_{ij}) x_j = 1, \forall i,$$

$$x_j \geq 0, \forall j.$$

where $e_{ij} = 1$ if $j \in A_i$ and 0 otherwise.

Dual variable x_j represents the expected action flow or frequency, that is, the expected present value of the number of times action j is chosen.

Efficiency of simplex/policy methods

- They are not known to be polynomial-time algorithms for deterministic MDP even with uniform discounts.
- There are quadratic lower bounds on these methods for solving MDP with uniform discounts.
- Ian Post and Y (2012): The Simplex method with the greedy pivot rule terminates in at most

$$0(m^3 n^2 \log^2 m)$$

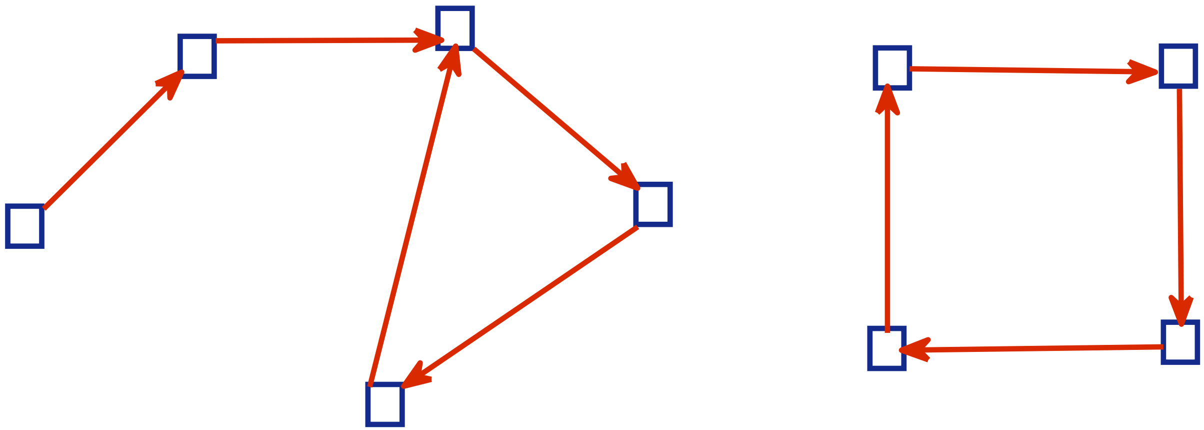
pivot steps when discount factors are uniform, or in at most

$$0(m^5 n^3 \log^2 m)$$

pivot steps with non-uniform discounts.

We are not yet able to prove such results hold for the policy iteration method.

Policy structures with uniform factors



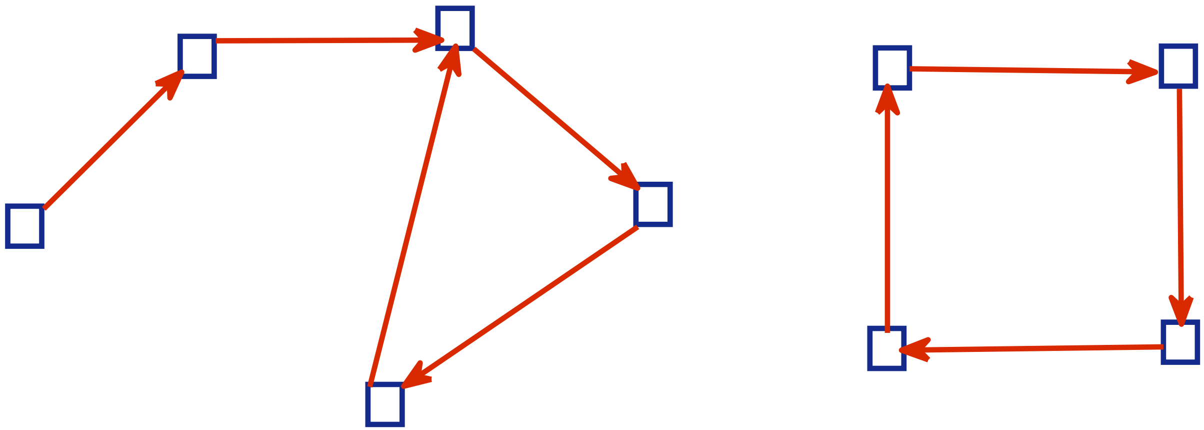
Each chosen action can be either a path–edge or cycle–edge.

x_j in [1, m] if it is a path–action,
x_j in [1/(1-γ), m/(1-γ)] if it is a cycle–action, so that they form two possible polynomial layers.

Roadmap of proof

- There two types of pivots: the newly chosen action is either on a path or on a cycle of the new policy.
- In every $m^2 n \log(m)$ consecutive pivot steps, there must be at least one step that is a cycle pivot.
- After every $m \log(m)$ cycle pivot steps, there is an action that would never re-enter as a cycle or path action.
- There are at most n action for such a down-grade.
- Item 2 result remains true when discounts are not uniform, but others do not hold.

Policy structures of general factors



The flow value of x_j depends on the smallest discount factor (dominating factor γ_a) on a same cycle.

There are n different discount factors, so that there are n possible different polynomial layers of x_js.

Decomposed "s-dual" of MDP-LP

$$\min \quad \sum_{i=1}^{n} c_j x_j$$

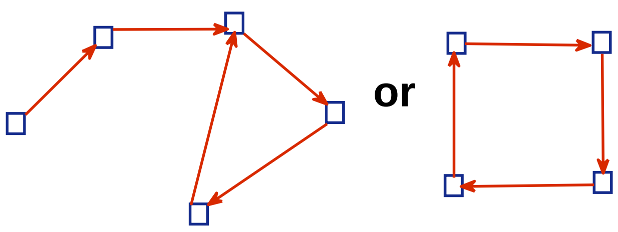
$$\text{s.t.} \quad \sum_{j=1}^{n} (e_{ij} - \gamma_j p_{ij}) x_j = 1, i = s,$$

$$\sum_{j=1}^{n} (e_{ij} - \gamma_j p_{ij}) x_j = 0, i \neq s,$$

$$x_j \geq 0, \forall j.$$

There are *m* such "dual" LPs, and the optimal policy is also optimal for each of them.

x_j of a given policy on each "s-dual" form a single path+cycle or a single cycle.



Roadmap of Proof

- Let (s, γ_a) denote a policy where the cycle for the s-dual is dominated by γ_a.
- In every $m^2 n \log(m)$ consecutive pivot steps, there must be at least one step that is a cycle pivot.
- After every $m^2 \log(m)$ cycle pivot steps, there is an action that would never re-enter to form a (s, γ_a) policy.
- There are at most nm such combinations, and at most n actions for such a down-grade.
- This gives the overall pivot step bound.

Other efficient methods and results

- Chubanov (2011) announced a new polynomial time algorithm to determine the feasibility of a system given in certain form:

  There exists a strongly polynomial algorithm which either finds a solution of a linear system $Ax = b$, $0 \leq x \leq 1$, or correctly decides that the system has no $\{0, 1\}$ solutions.
- Bertsimas and Vempala (2004) and Dunagan and Vempala (2008) present random–walk type methods of which they can prove run in polynomial time.
- Spielman and Teng (2004) and later with significant improvements by Vershynin (2009) have provided new probabilistic insights, called smoothed analysis, into why we observe a good practical performance of the simplex algorithm.
- Dedieu, Malajovich, and Shub (2005), Deza, Terlaky and Zinchenko (2009), Loera, Sturmfels, and Vinzant (2010) provided new insights on total curvature of the central path.

Remarks and Open Problems

- Is the policy iteration method a strongly polynomial time algorithm for deterministic MDP?

- Is there strongly polynomial time algorithm for MDP with variable discounts, generalized network flow, or even LP?

- New LP applications?

- Solve LPs with a huge size (billion-dimension) in practice?

Linear Programming and the Simplex Method Story Continues …