Romain Hollanders, Raphaël Jungers and Jean-Charles Delvenne

present

Policy Iteration has exponential complexity for solving Markov Decision Processes

But there is more to the story...

A Cesame Seminar production

PART 1 : Exponential lower bounds for Policy Iteration

PART 2 : Can we obtain good upper bounds ?

Romain Hollanders, Raphaël Jungers and Jean-Charles Delvenne

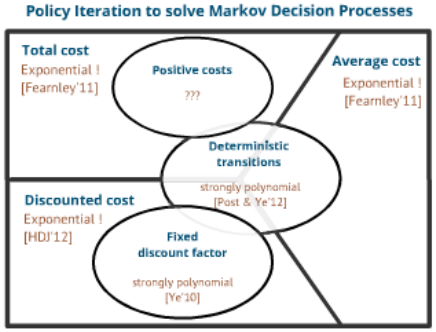
present

Policy Iteration has exponential complexity for solving Markov Decision Processes

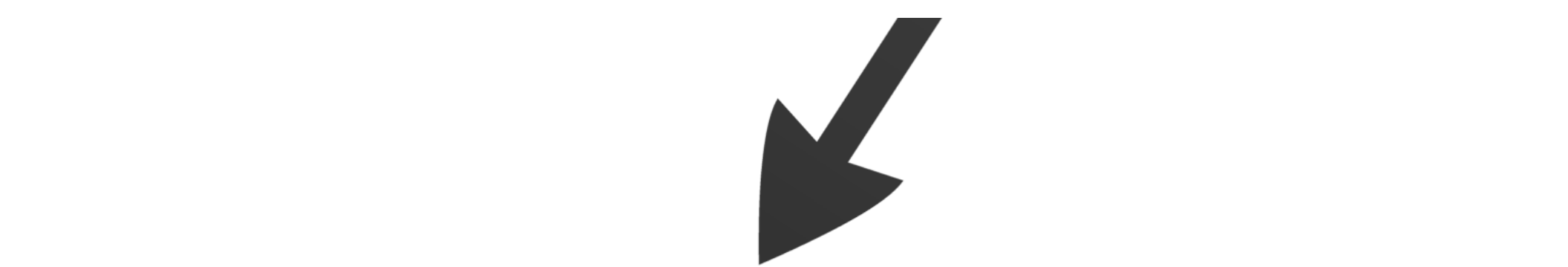
But there is more to the story...

A Cesame Seminar production

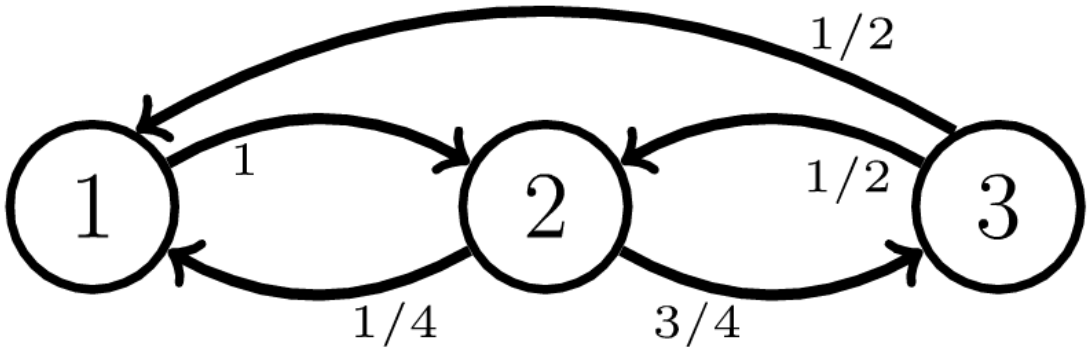
PART 1 : Exponential lower bounds for Policy Iteration



PART 2 : Can we obtain good upper bounds ?



Markov Chains



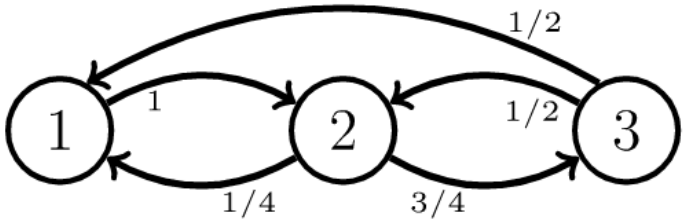
$c_1 = 2$ $c_2 = 5$ $c_3 = 4$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

$$p_k^T = p_0^T P^k$$

How much will I pay

Markov Chains



$c_1 = 2 \qquad c_2 = 5 \qquad c_3 = 4$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

$$p_k^T = p_0^T \, P^k$$

$$\begin{aligned} p_0^T &= 1 \quad 0 \quad 0 \\ p_1^T &= 0 \quad 1 \quad 0 \\ p_2^T &= \tfrac{3}{4} \quad 0 \quad \tfrac{1}{4} \\ p_3^T &= \tfrac{1}{8} \quad \tfrac{7}{8} \quad 0 \\ p_4^T &= \tfrac{21}{32} \quad \tfrac{4}{32} \quad \tfrac{7}{32} \\ &\vdots \end{aligned}$$

How much will I pay if I start from state 1?

- Total cost

$$x(1) = \sum_{k=0}^{H} p_k^T c$$

- Average cost

$$x(1) = \lim_{H \to \infty} \frac{1}{H} \cdot \sum_{k=0}^{H} p_k^T c$$

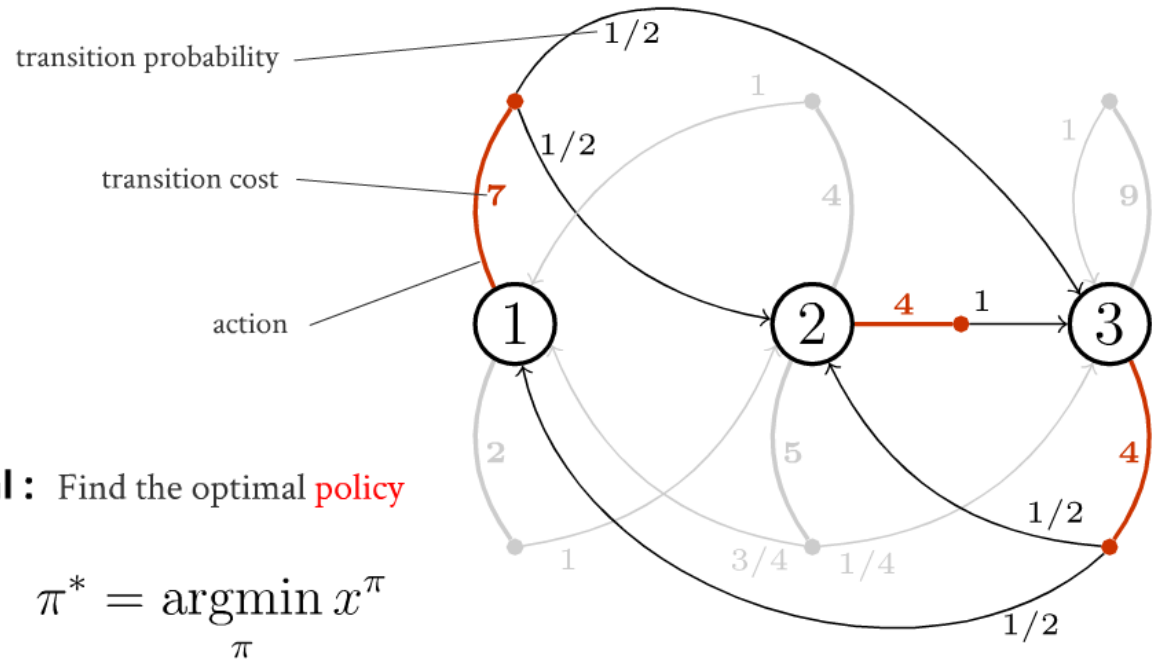
- Discounted cost

$$x(1) = \sum_{k=0}^{\infty} \gamma^k p_k^T c$$

Markov Decision Processes

One action per state

In general



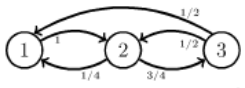
Goal : Find the optimal policy

$$\pi^* = \underset{\pi}{\operatorname{argmin}}\, x^\pi$$

The answer depends on the chosen objective function :

- Total cost
- Average cost
- Discounted cost

Markov Chains



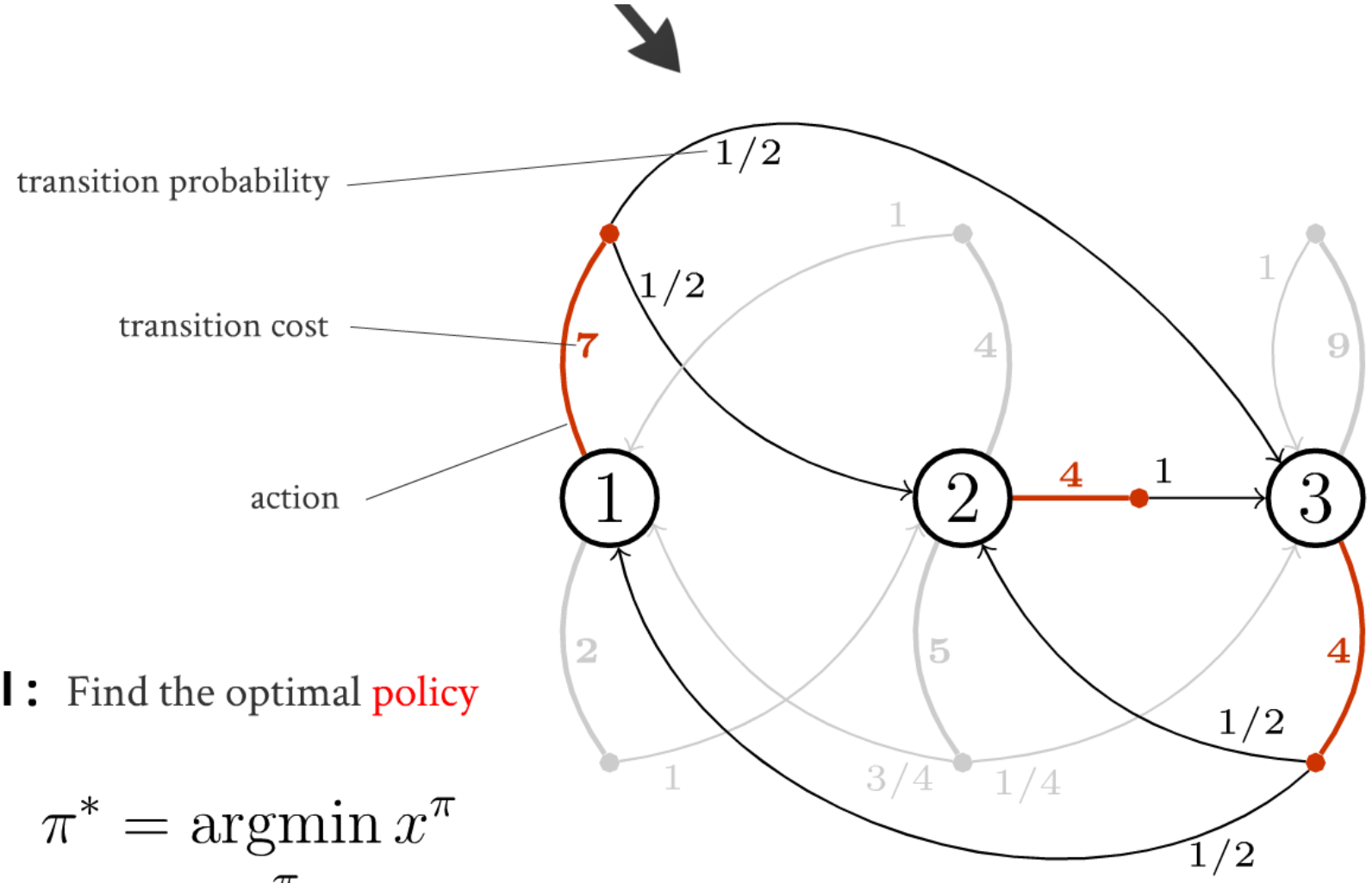
$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

$$p_k^T = p_0^T P^k$$

$$\begin{aligned} p_0^T &= 1 \quad 0 \quad 0 \\ p_1^T &= 0 \quad 1 \quad 0 \\ p_2^T &= \tfrac{3}{4} \quad 0 \quad \tfrac{1}{4} \\ p_3^T &= \tfrac{1}{8} \quad \tfrac{7}{8} \quad 0 \\ p_4^T &= \tfrac{21}{32} \quad \tfrac{4}{32} \quad \tfrac{7}{32} \\ &\vdots \end{aligned}$$

How much will I pay if I start from state 1?

- Total cost
$$x(1) = \sum_{k=0}^{H} p_k^T c$$
- Average cost
$$x(1) = \lim_{H \to \infty} \frac{1}{H} \cdot \sum_{k=0}^{H} p_k^T c$$
- Discounted cost
$$x(1) = \sum_{k=0}^{\infty} \gamma^k p_k^T c$$



Goal : Find the optimal policy

$$\pi^* = \underset{\pi}{\operatorname{argmin}}\, x^\pi$$

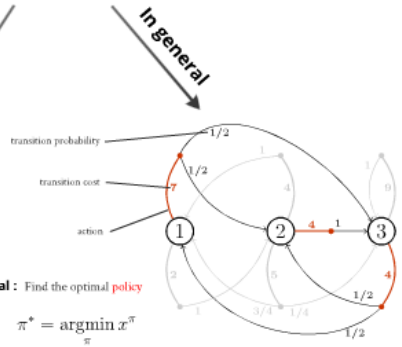
The answer depends on the chosen objective function :

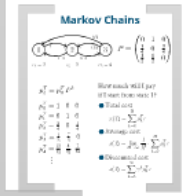
- Total cost
- Average cost
- Discounted cost

Markov Decision Processes

One action per state

In general





Goal : Find the optimal policy

$$\pi^* = \underset{\pi}{\operatorname{argmin}}\, x^\pi$$

The answer depends on the chosen objective function :

- Total cost
- Average cost
- Discounted cost

Resolution

Linear Programming

Value Iteration

Policy Iteration

0. Choose an initial policy π_0

while $\pi_k \neq \pi_{k-1}$

1. Evaluate π_k (Bellman)

$$x^{\pi_k} = c^{\pi_k} + \gamma\, P^{\pi_k} x^{\pi_k}$$

2. Improve π_k

$$\pi_{k+1} = \underset{\pi}{\operatorname{argmin}}\; c^\pi + \gamma\, P^\pi x^{\pi_k}$$

$$k \leftarrow k+1$$

end

Linear Programming

Value Iteration

Policy Iteration

0. Choose an initial policy π_0

while $\pi_k \neq \pi_{k-1}$

1. Evaluate π_k (Bellman)

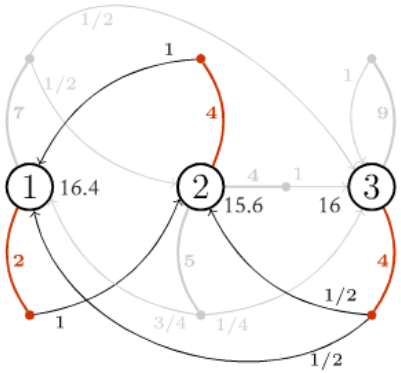
$$x^{\pi_k} = c^{\pi_k} + \gamma\, P^{\pi_k} x^{\pi_k}$$

2. Improve π_k

$$\pi_{k+1} = \underset{\pi}{\operatorname{argmin}}\; c^{\pi} + \gamma\, P^{\pi} x^{\pi_k}$$

$$k \leftarrow k+1$$

end



$\gamma = 0.8$

Policy Iteration

0. Choose an initial policy π_0

while $\pi_k \neq \pi_{k-1}$

1. Evaluate π_k (Bellman)

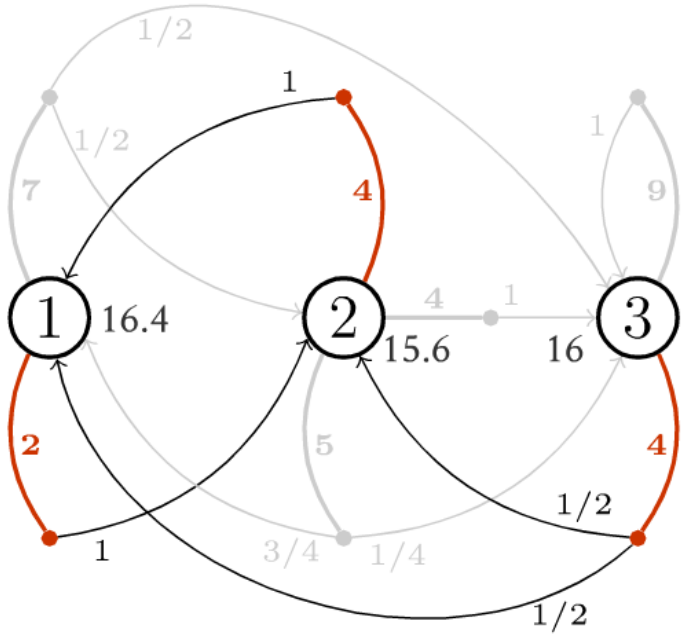
$$x^{\pi_k} = c^{\pi_k} + \gamma\, P^{\pi_k} x^{\pi_k}$$

2. Improve π_k

$$\pi_{k+1} = \operatorname*{argmin}_{\pi}\; c^{\pi} + \gamma\, P^{\pi} x^{\pi_k}$$

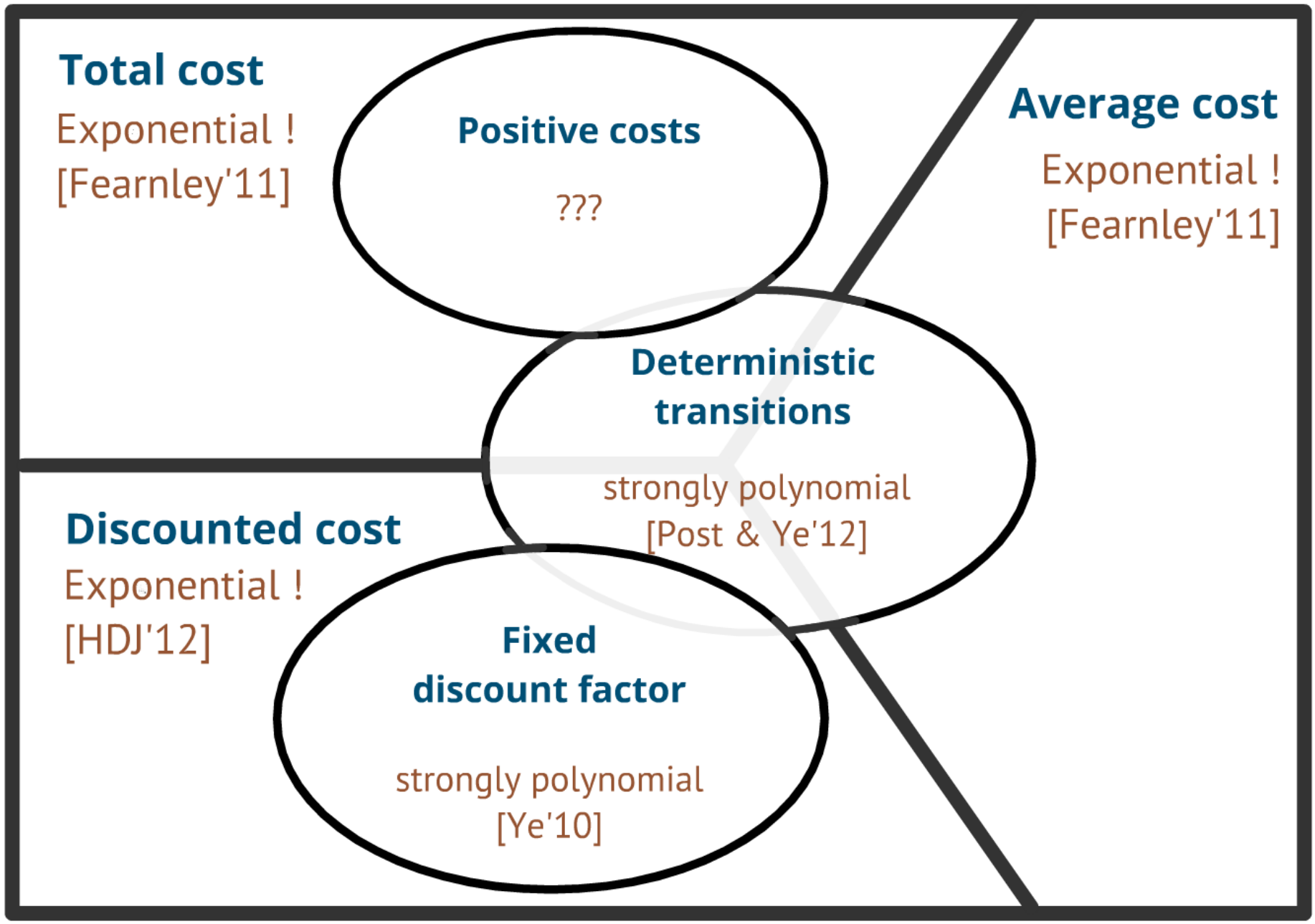
$$k \leftarrow k+1$$

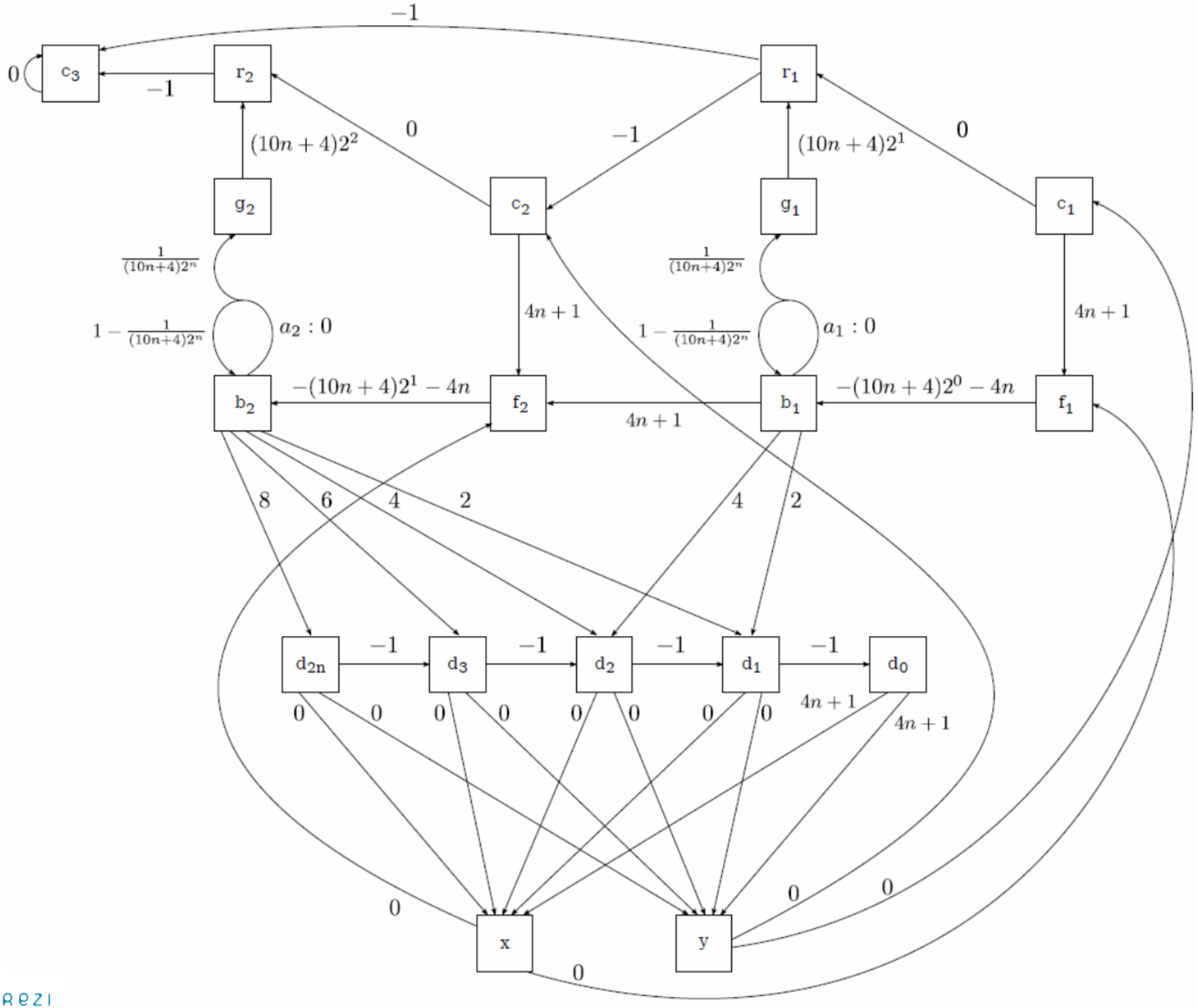
end



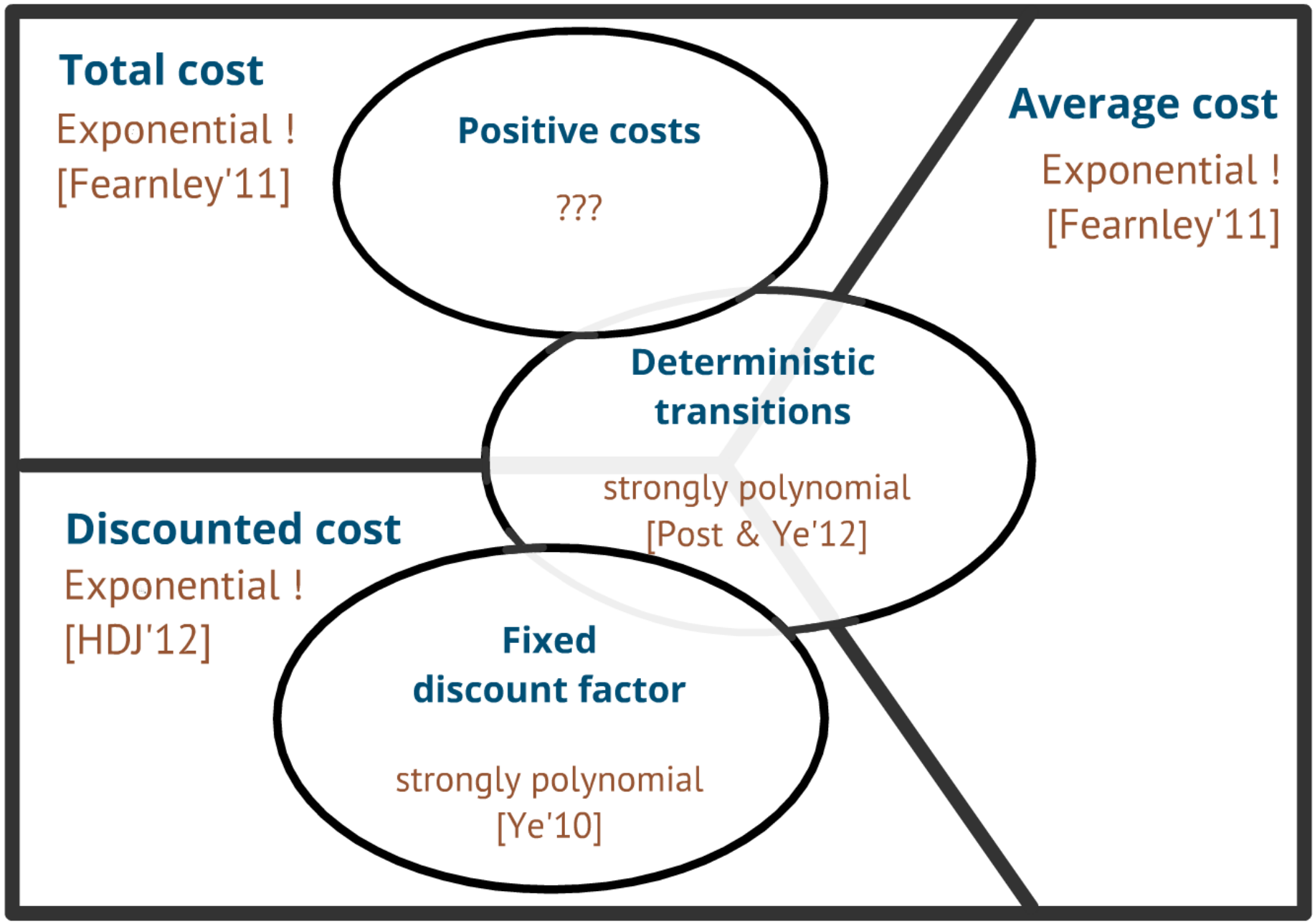
$\gamma = 0.8$

Policy Iteration to solve Markov Decision Processes

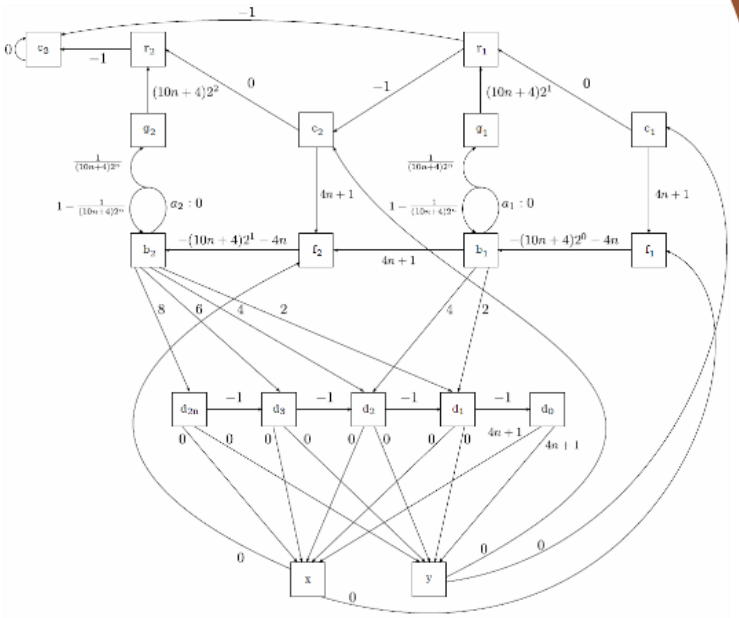



c_3
r_2
r_1
g_2
c_2
g_1
c_1
b_2
f_2
b_1
f_1
d_{2n}
d_3
d_2
d_1
d_0
x
y
$a_2 : 0$
$a_1 : 0$
$(10n+4)2^2$
$(10n+4)2^1$
$\frac{1}{(10n+4)2^n}$
$1 - \frac{1}{(10n+4)2^n}$
$4n+1$
$-(10n+4)2^1 - 4n$
$-(10n+4)2^0 - 4n$
-1
0
8
6
4
2

Policy Iteration to solve Markov Decision Processes



We add discount $\gamma = 1 - \varepsilon$

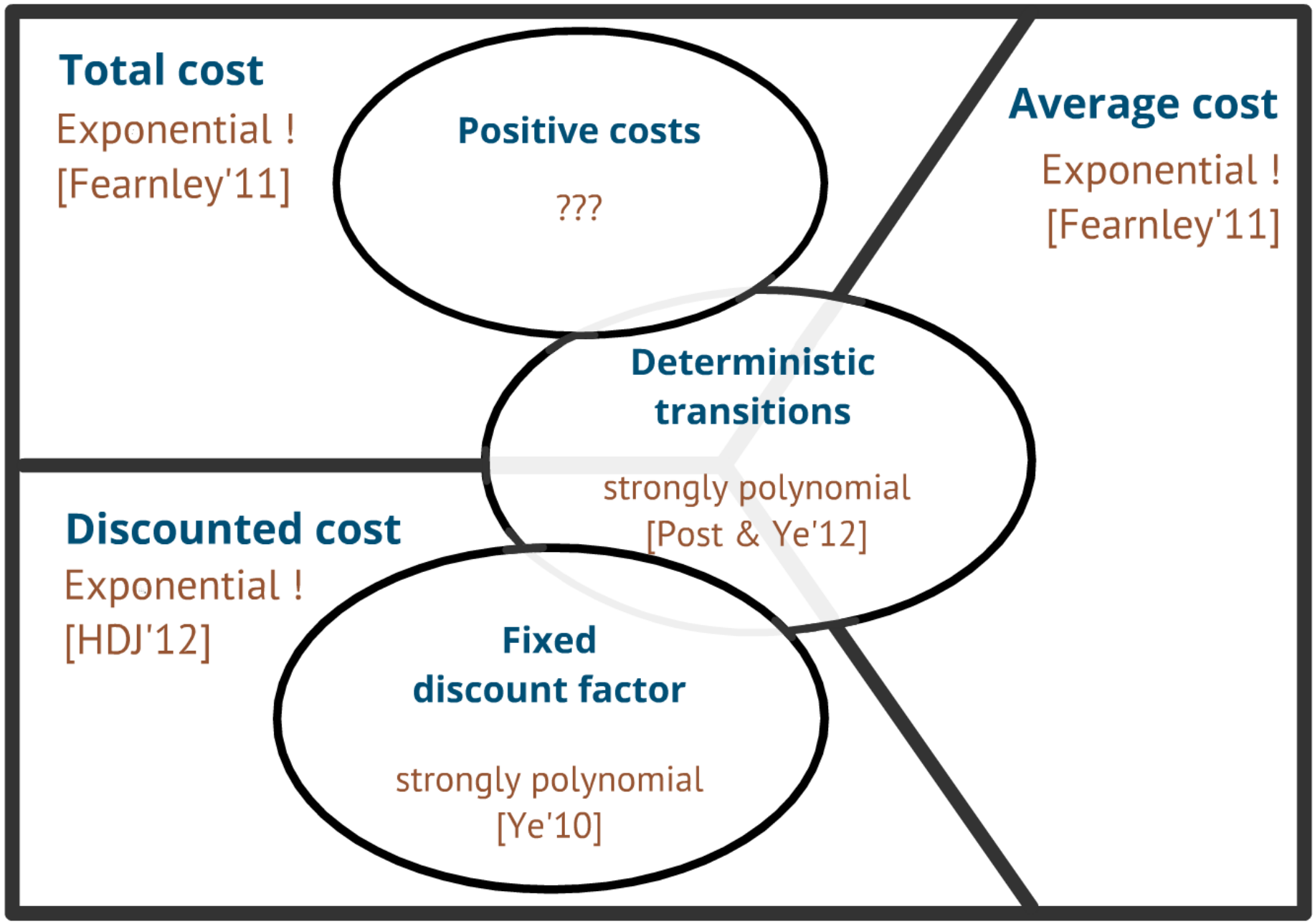


→ How much perturbation?

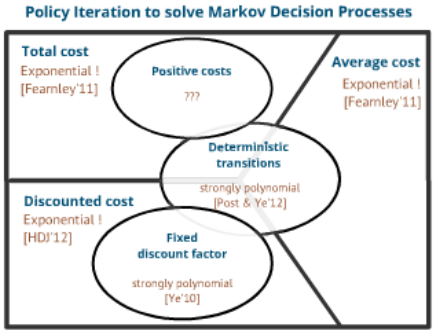
$$x^{\pi'}(s)$$
$$\leq F(n,\delta,\kappa)\,\varepsilon$$
$$\tilde{x}^{\pi'}(s)$$
$$\tilde{x}^{\pi_k}(s)$$
$$\leq F(n,\delta,\kappa)\,\varepsilon$$
$$x^{\pi_k}(s)$$
$$\geq G(n,\delta)$$

→ OK for some $\varepsilon \sim \dfrac{1}{2^{q(n,\delta,\kappa)}}$

Policy Iteration to solve Markov Decision Processes



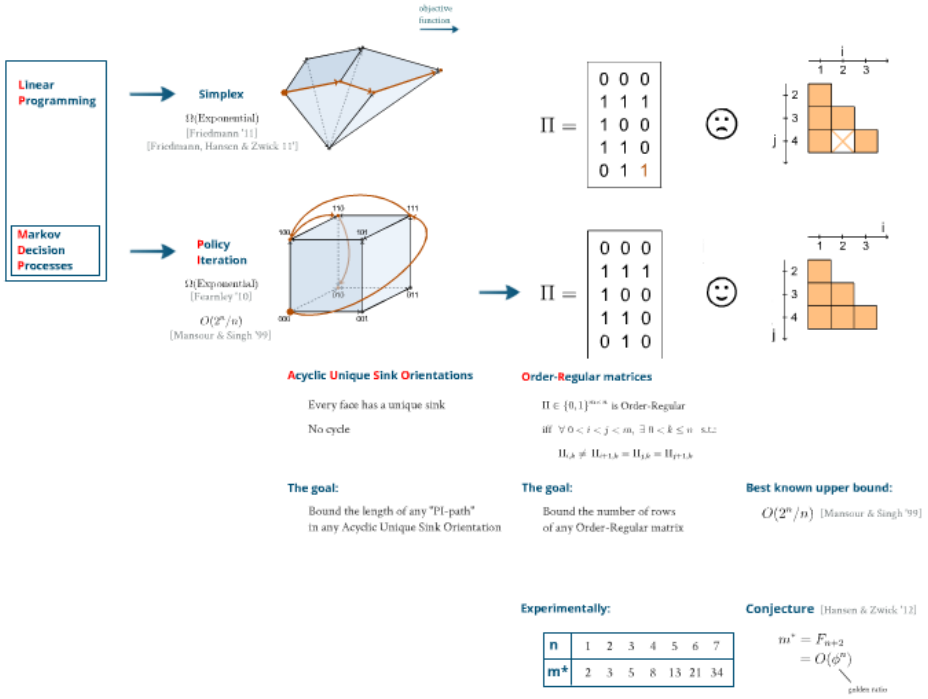
PART 1 : Exponential lower bounds for Policy Iteration



PART 2 : Can we obtain good upper bounds ?

PART 1 : Exponential lower bounds for Policy Iteration

PART 2 : Can we obtain good upper bounds?



Summary on the complexity of Policy Iteration

Lower bounds

In general: exponential in the worst case

With some particular structures: strongly polynomial

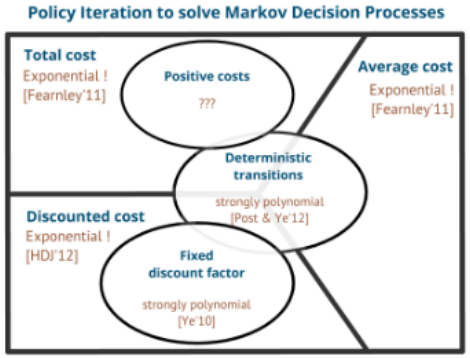
What about the average complexity?

Upper bounds

Best known bound: $O(2^n/n)$

Our hope: $O(1.618^n)$

PART 1 : Exponential lower bounds for Policy Iteration



PART 2 : Can we obtain good upper bounds?

The state of the art and some inspiring ideas

Upper bounds on m* $m^* \leq O(2^n/n) \leq O(2^n)$

Conjecture on m* $m^* \sim \Theta(\phi^n) \quad = O(1.618^n)$

Lower bounds on m* $m^* \geq \Omega(\sqrt{2}^n) = \Omega(1.4142^n)$

objective function

Linear Programming

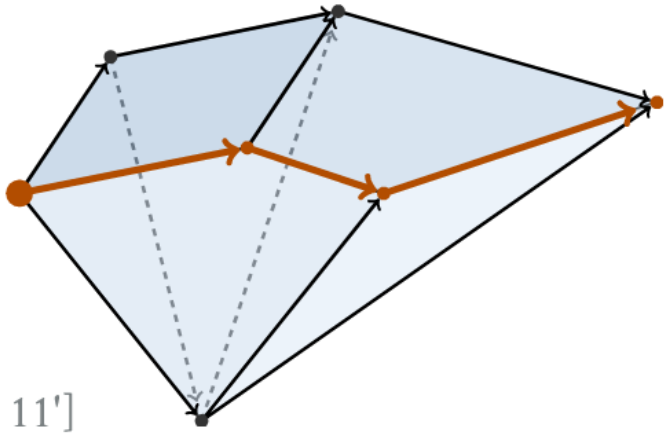
Markov Decision Processes

Simplex

$\Omega(\text{Exponential})$

[Friedmann '11]

[Friedmann, Hansen & Zwick 11']



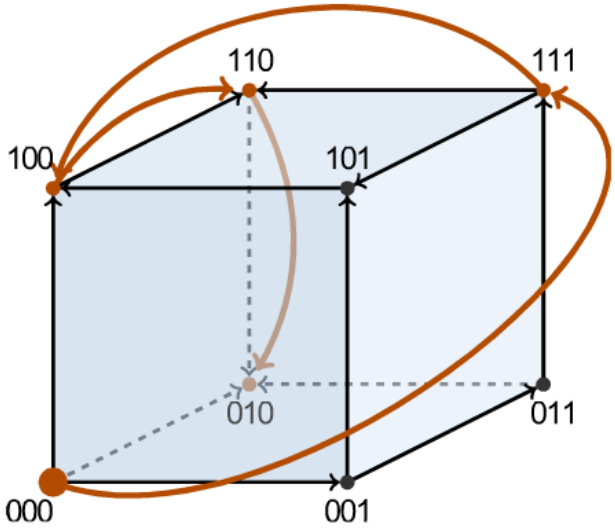
Policy Iteration

$\Omega(\text{Exponential})$

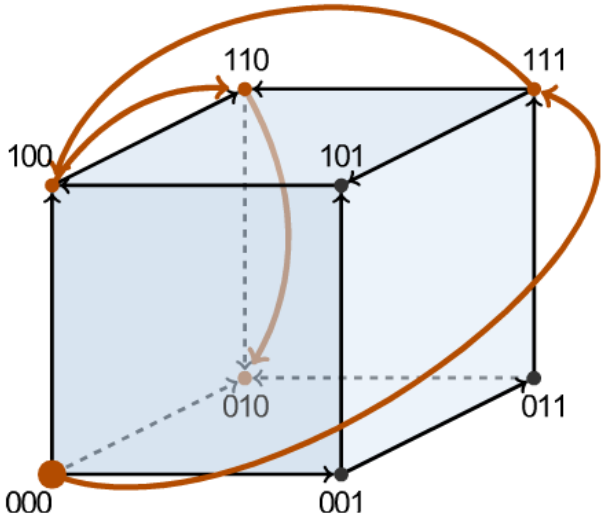
[Fearnley '10]

$O(2^n/n)$

[Mansour & Singh '99]



0 1 1



$$\Pi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Acyclic Unique Sink Orientations

Every face has a unique sink

No cycle

Order-Regular matrices

$\Pi \in \{0,1\}^{m \times n}$ is Order-Regular

iff $\forall\ 0 < i < j < m,\ \exists\ 0 < k \leq n$ s.t.:

$$\Pi_{i,k} \neq \Pi_{i+1,k} = \Pi_{j,k} = \Pi_{j+1,k}$$

The goal:

Bound the length of any "PI-path"
in any Acyclic Unique Sink Orientation

The goal:

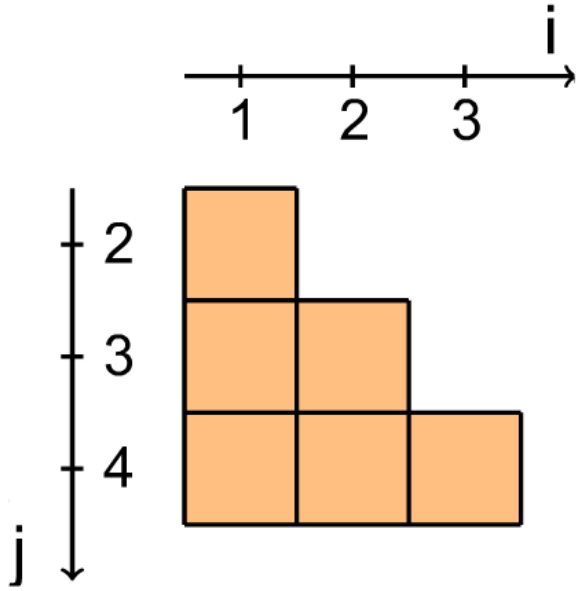
Bound the number of rows
of any Order-Regular matrix

Best

O

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$$

$$\Pi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$



Order-Regular matrices

$\Pi \in \{0,1\}^{m \times n}$ is Order-Regular

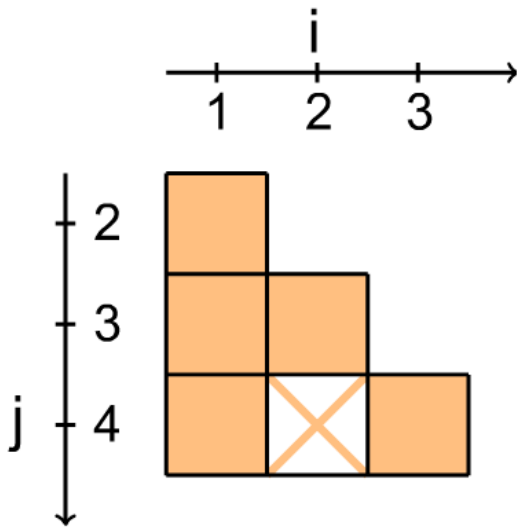
iff $\forall\ 0 < i < j < m,\ \exists\ 0 < k \leq n$ s.t.:

$$\Pi_{i,k} \neq \Pi_{i+1,k} = \Pi_{j,k} = \Pi_{j+1,k}$$

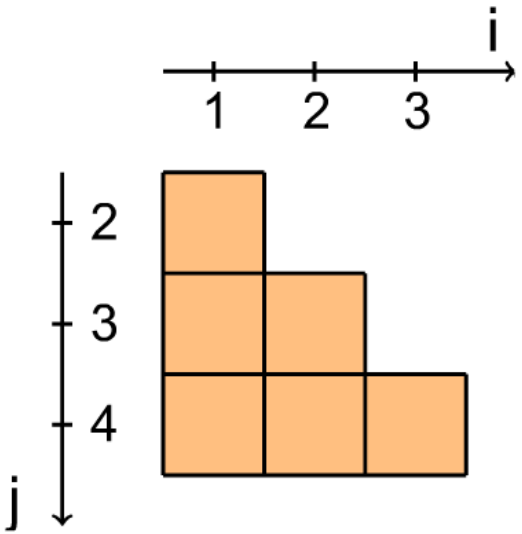
The goal:

Best known upper bound:

$$\Pi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$



$$\Pi = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$



$$\Pi_{i,k} \neq \Pi_{i+1,k} = \Pi_{j,k} = \Pi_{j+1,k}$$

The goal:

Bound the number of rows
of any Order-Regular matrix

Best known upper bound:

$O(2^n/n)$ [Mansour & Singh '99]

Experimentally:

n	1	2	3	4	5	6	7
m*	2	3	5	8	13	21	34

Conjecture [Hansen & Zwick '12]

$$m^* = F_{n+2}$$
$$= O(\phi^n)$$

golden ratio

The state of the art and some inspiring ideas

Upper bounds on m*

$$m^* \leq O(2^n/n) \leq O(2^n)$$

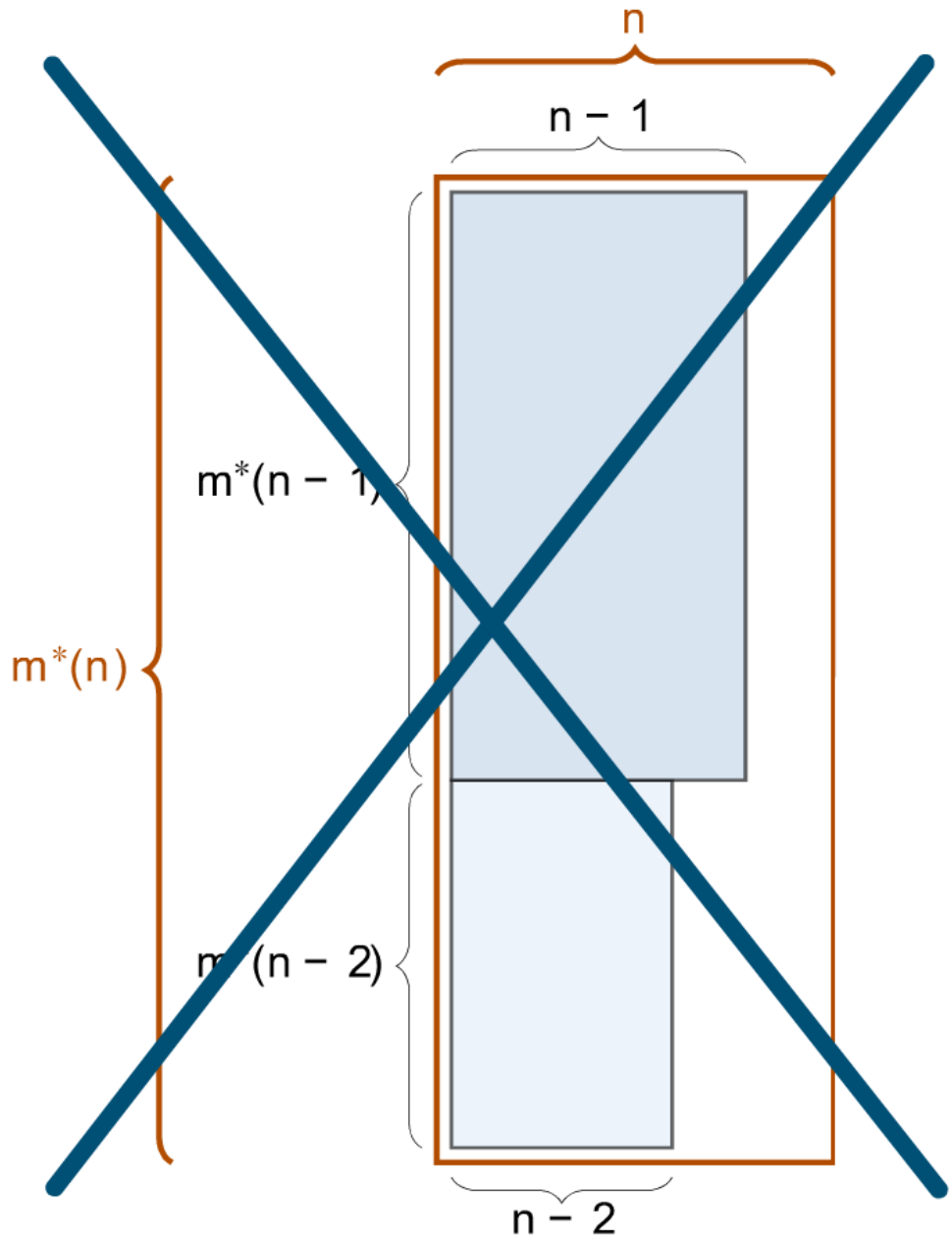
Conjecture on m*

$$m^* \sim O(\phi^n) = O(1.618^n)$$

Lower bounds on m*

$$m^* \geq \Omega\left(\sqrt{2}^n\right) = \Omega(1.4142^n)$$

A natural idea



fails...

The state of the art and some inspiring ideas

Upper bounds on m*

$$m^* \leq O(2^n/n) \leq O(2^n)$$

Conjecture on m*

$$m^* \sim \Theta(\phi^n) = O(1.618^n)$$

Lower bounds on m*

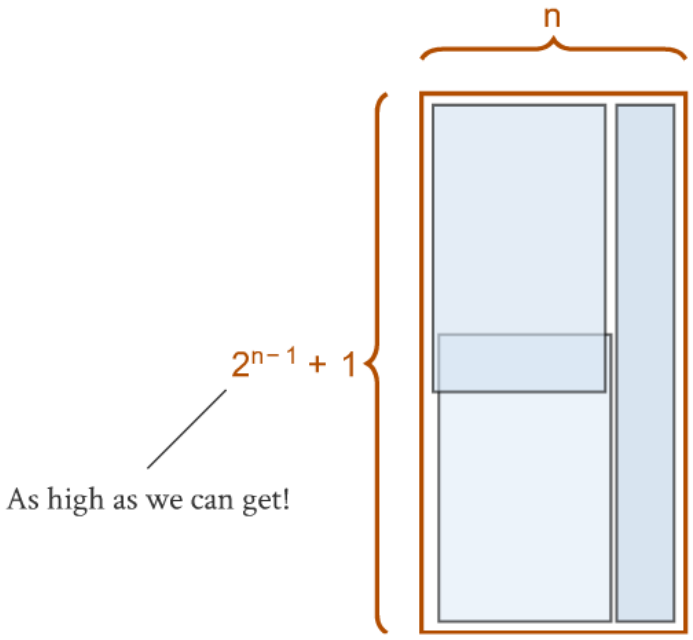
$$m^* \geq \Omega\left(\sqrt{2}^n\right) = \Omega(1.4142^n)$$

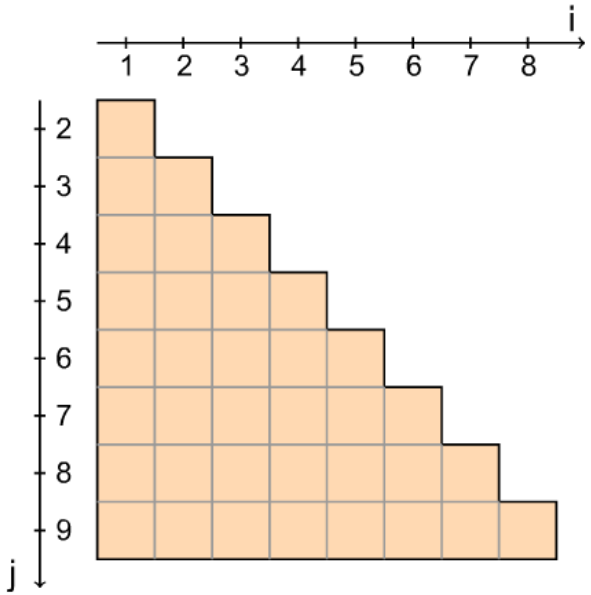
A relaxation

$\Pi \in \{0,1\}^{m \times n}$ is quasi-Order-Regular

iff $\forall\ 0 < i < j < m,\ \exists\ 0 < k \leq n$ s.t.:

$$\Pi_{i,k} \neq \Pi_{i+1,k} = \Pi_{j,k} \not\equiv \Pi_{j+1,k}$$





We can build quasi-Order-Regular matrices with $2^{n-1} + 1$ rows!

The state of the art and some inspiring ideas

Upper bounds on m*

$$m^* \leq O(2^n/n) \leq O(2^n)$$

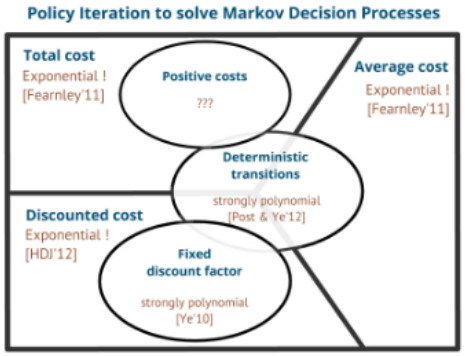
Conjecture on m*

$$m^* \sim O(\phi^n) = O(1.618^n)$$

Lower bounds on m*

$$m^* \geq \Omega\left(\sqrt{2}^n\right) = \Omega(1.4142^n)$$

PART 1 : Exponential lower bounds for Policy Iteration



PART 2 : Can we obtain good upper bounds?

The state of the art and some inspiring ideas

Upper bounds on m* $m^* \leq O(2^n/n) \leq O(2^n)$

Conjecture on m* $m^* \sim \Theta(\phi^n) \quad = O(1.618^n)$

Lower bounds on m* $m^* \geq \Omega(\sqrt{2}^n) = \Omega(1.4142^n)$

Summary on the complexity of Policy Iteration

Lower bounds

In general: exponential in the worst case

With some particular structures: strongly polynomial

What about the average complexity?

Upper bounds

Best known bound: $O(2^n/n)$

Our hope: $O(1.618^n)$