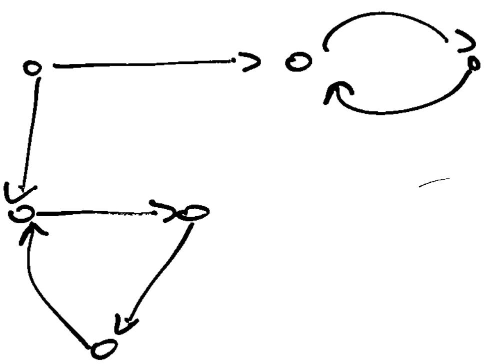
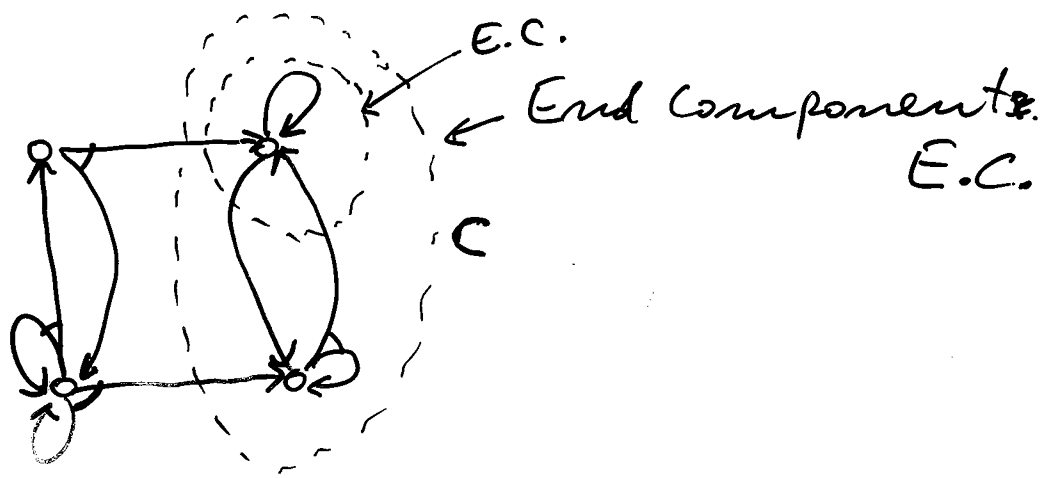
M.C.



MDP



Consider a set $C \subseteq S$.

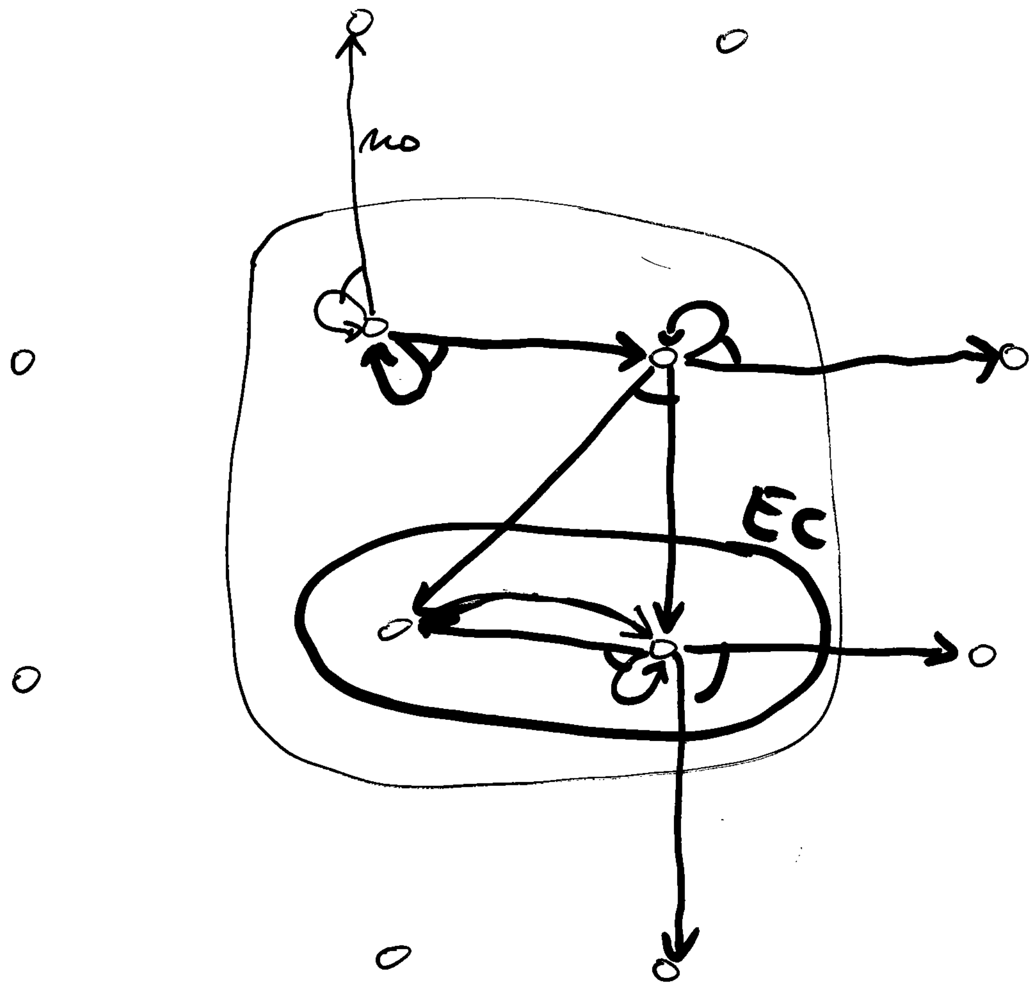
For $s \in C$, let $\Gamma^*_C(s) = \{a \in \Gamma(s) \mid \delta(s,a) \subseteq C\}$.

Let $E_C \subseteq C \times C$ consist of all $(s,t) \in C \times C$

such that $\exists a \in \Gamma_C(s).\ t \in \delta(s,a)$.

"you can go $s \longrightarrow t$ with pos prob. while staying in C with prob. 1".

C is an E.C. iff the graph (C, E_C) is strongly connected.



Algo:

1. Remove all actions leaving C.
2. Find ~~str~~ maximal strongly connected components in (C, E_c). Let them ~~Pa~~ be $C_1, C_2, C_3 \ldots C_m$
3. Iterate on $C_1, C_2, C_3 \ldots C_m$.

Algo for finding the maximal E.C. in a set C.

let maxec (C : set) : set of sets =

1. Build (C, E_C)

   (* That is, remove all actions that can lead out of C, and consider the resulting graph on C. *)

2. Find the maximal strongly connected components in (C, E_C): let them be $B_1, B_2, \ldots, B_m$.

~~3. Return~~

3. if $m = 1$ and $B_1 = C$, return $\{C\}$.

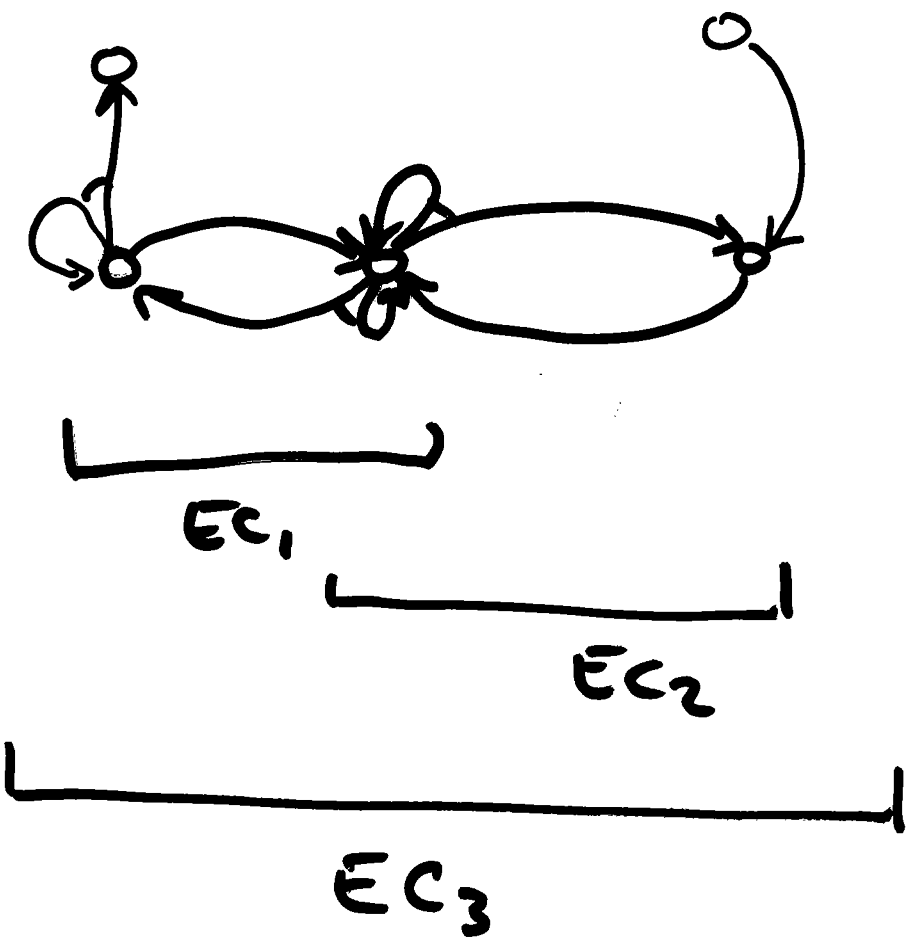
4. Otherwise, return

$$\bigcup_{i=1}^{m} \text{maxec}(B_i).$$

Theorem

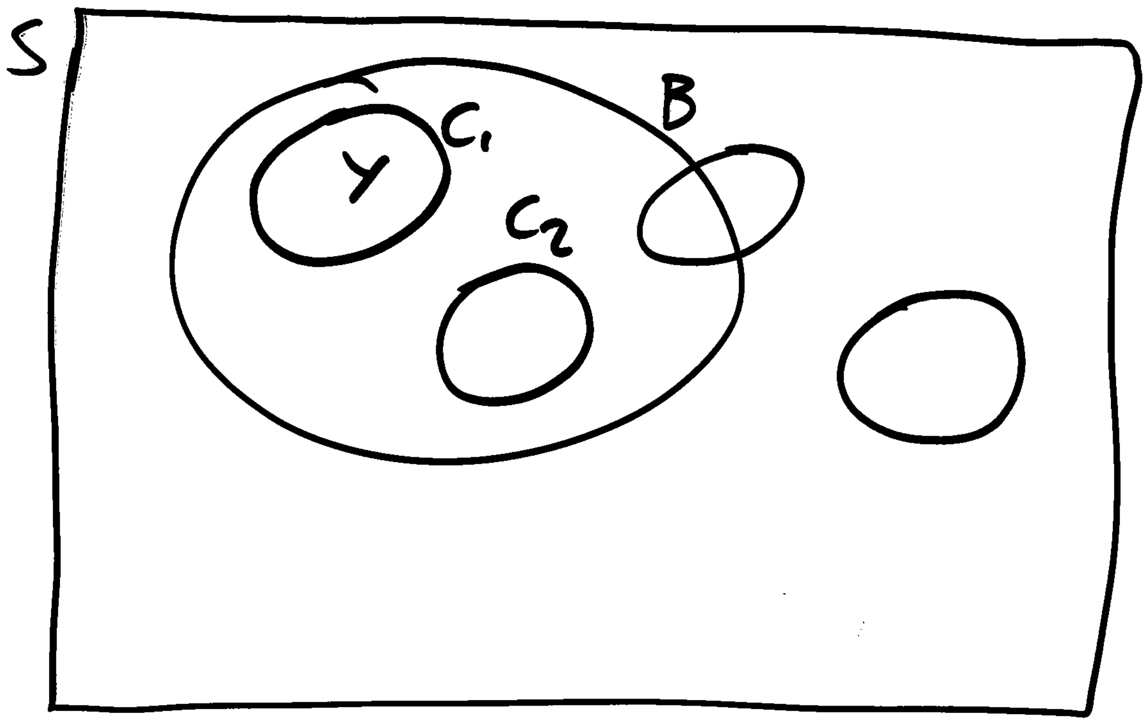
For all strategies π_1, and all $s \in S$,

$$Pr_s^{\pi_1}(\mathrm{inf}(\sigma) \text{ is an E.C.}) = 1.$$



Safety is in Poly Time:

Algo:



$C_1 \cup C_2$ is the goal: try to reach it.

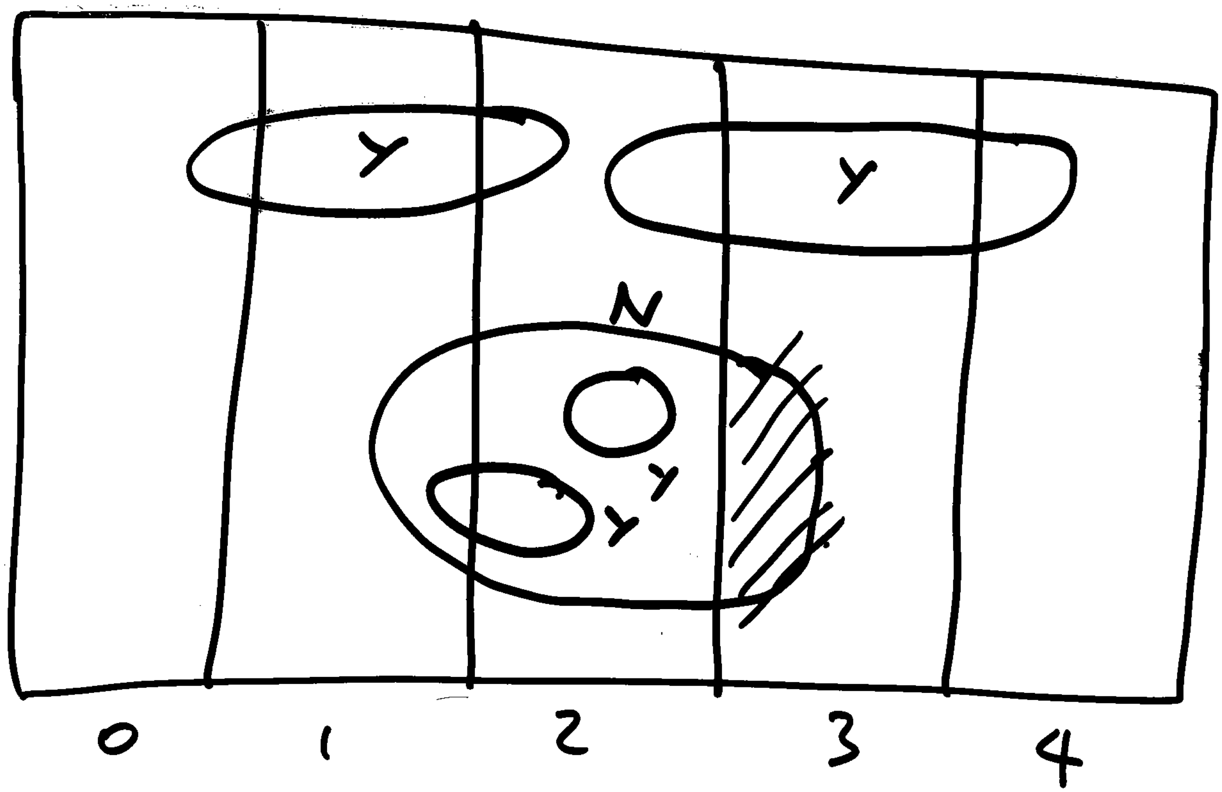
Goal: $\Box B$.

1. Compute $\mathrm{maxec}(B) = L$.

2. $\langle\langle 1 \rangle\rangle \Box B = \langle\langle 1 \rangle\rangle \Diamond \bigcup L$.

$\Rightarrow$ If I reach $\bigcup L$, I win. Easy.

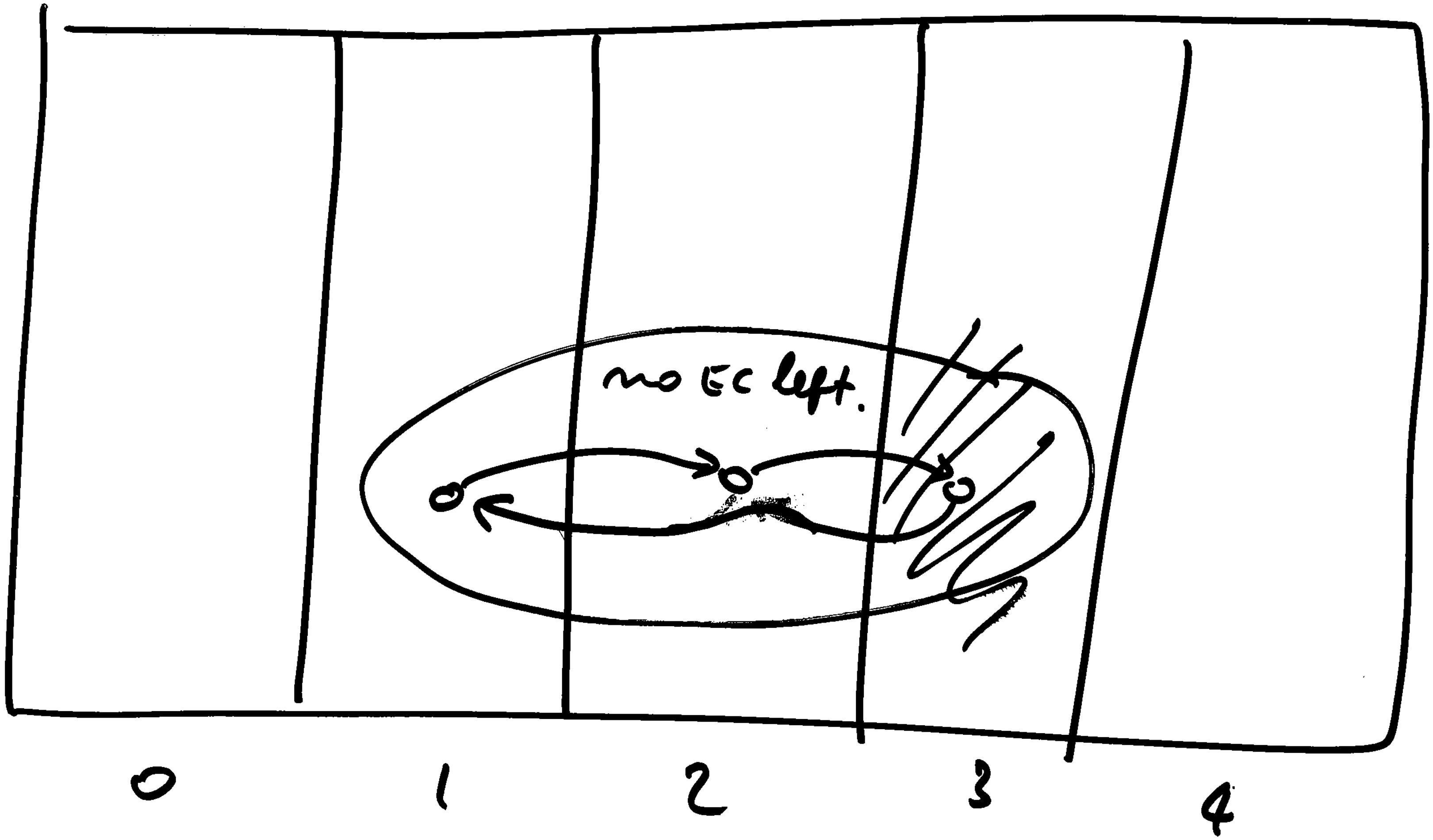
$\Leftarrow$ If I don't reach $\bigcup L$, I win with prob. 0.



Recursively:

1) Find maximal EC.
2) If their rightmost end is odd, shave it off, and find ~~EC~~ maximal ECs in the remainder.

Collect all surviving ECs, and try to get there.


no EC left.
0
1
2
3
4

Strategy iteration for MDPs.

Problem: reachability.

Consider $\mu X. ([R] \cup QPre(X))$

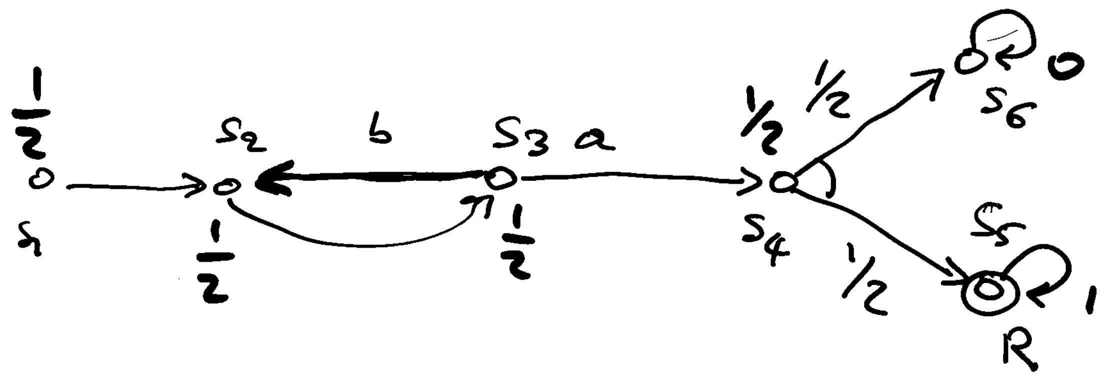
least solution of

$$X(s) = \begin{cases} 1 & \text{if } s \in R \\ \max_{a \in \Gamma(s)} \sum_{t \in S} \tau(s,a)(t) \cdot X(t) \end{cases}$$

At every $s \in S$, there is a set $\Gamma_R(s)$ of "optimal actions", that realize the max.

Thm 1 If at every s you play all moves in $\Gamma_R(s)$ unif. at random, you achieve the max. prob of $\Diamond R$.

<u>Fact</u>: if you deterministically play a move in $T_R(s)$, you may not play optimally.



<u>Both</u> a, and b, are optimal at s_3.

~~Basi~~
<u>Thm 2</u>: There <u>is</u> a determ π_1 st

$$Pr_s^{\pi_1}(\Diamond R) = \langle\langle 1 \rangle\rangle \Diamond R\ (s).$$

Policy } Iteration algo for MDPs.
Strategy }

Let $\eta : S \mapsto Moves$ indicate a det, memoryless strategy.

Algo:

Fix η_0 arbitrarily.

The MDP, under η_0, is a $MC(\eta_0)$

$$P_{st} = \tau(s, \eta_0(s))(t).$$

Compute $v_0(s)$ = prob. of $\Diamond R$ from s in $MC(\eta_0)$.

$$\forall s, \text{ Let } \eta_1(s) = \underset{a \in \Gamma(s)}{\arg\max} \sum_t \tau(s,a)(t) \cdot v_0(t).$$

Go on.

η_2

η_3

$\vdots$

fixpoint.