Stochastic Games

Krishnendu Chatterjee

Institute of Science and Technology

ARiSE Winter School, Vienna, Feb, 2012

Austrian Rigorous Systems Engineering

Quantitative Parity to Quantitative Reachability

- End-components: An end-component generalizes both scc and closed recurrent set. A set U is an end-component if the following properties hold:
  - U is strongly connected.
  - U is closed (no probabilistic edge out).

- Note that player 1 edges may leave the end-component.

- Why is end-component important: it allows us to reason about infinite behaviors.

End Component Property

- End-component property: For an MDP and for all strategies, with probability 1 the set of states visited infinitely often is an end-component.

- Generalizes the scc for graphs and closed recurrent set for Markov chains.

- Proof:
  - Shape of the proof very similar to closed recurrent set.
  - We need to show that if a set U is not an end-component, then cannot be visited infinitely often with positive probability.
  - Assume towards contradiction that there is such a set U.

End Component Property

- End-component property: For an MDP and for all strategies, with probability 1 the set of states visited infinitely often is an end-component.

- Proof:
  - We need to show that if a set U is not an end-component, then cannot be visited infinitely often with positive probability.
  - Assume towards contradiction that there is such a set U.
  - U must be strongly connected.
  - Since U is not end-component, some probabilistic state s with an edge to t going out of U with probability α.
  - Hence the probability that s is visited infinitely often, but the edge to t is taken finitely often is 0.
  - The result follows.

Winning End-component

- An end-component U is winning if the minimum priority of U is even.

- From end-component property for any strategy the probability to satisfy parity is the probability to reach the winning end-components.

- In winning end-components pure memoryless almost-sure winning strategy exists.
  - Proof: Choose successor to shorten distance to the minimum even priority state.

Quantitative Parity to Quantitative Reachability

- The probability to satisfy is the probability to reach winning end-components.
- In winning end-components pure memoryless almost-sure strategy.
- Winning end-components are included in the almost-sure winning set.
- Hence we need quantitative reachability to almost-sure winning set.
- We now need the quantitative reachability to complete the argument.

Quantitative Reachability

- An MDP G, and a target set T.
- $\mathrm{Val}(\mathrm{Reach}(T))(s) = \sup_{\sigma} \Pr_s^{\sigma}(\mathrm{Reach}(T))$.

- $v(s)$ for abbreviation.

- Two properties:
  - Property 1: For $s \in S_P$ we have $v(s) = \sum_{t\in S} v(t)^* \delta(s)(t)$.
  - Property 2: For $s \in S_1$ we have $v(s) = \max \{ v(t) \mid t \in E(s)\}$.

Proof of Property 2

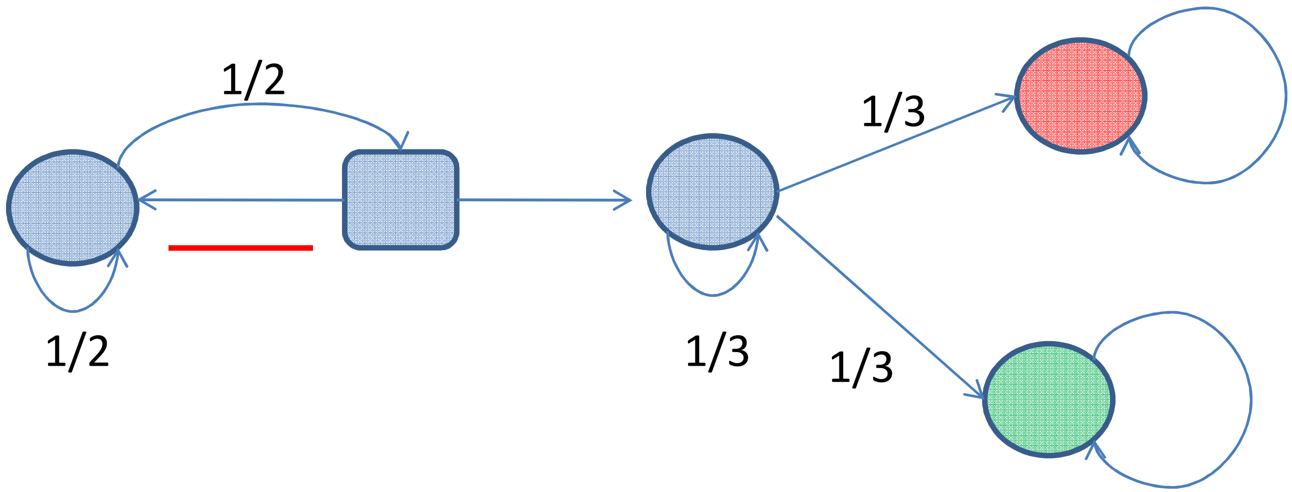
- Inequality 1: $v(s) \geq \max\{ v(t) \mid t \in E(s)\}$
  - Fix ϵ>0.
  - Let t* be the arg max.
  - From s choose t*, and then an ϵ optimal strategy from t* to ensure value at least v(t*)-ϵ.
  - As ϵ>0 is arbitrary, the result follows.

- Inequality 2. $v(s) \leq \max\{ v(t) \mid t \in E(s)\}$
  - We have
    - $v(s) \leq \sup_{\mu} \sum_{t \in S} v(t)^* \mu(t) \leq \max \{v(t) \mid t \in E(s)\}$,
      where $\mu \in D(E(s))$.

A Simple Attempt

- For a state s choose a successor that achieves the maximum.
- However this simple construction is not sufficient.

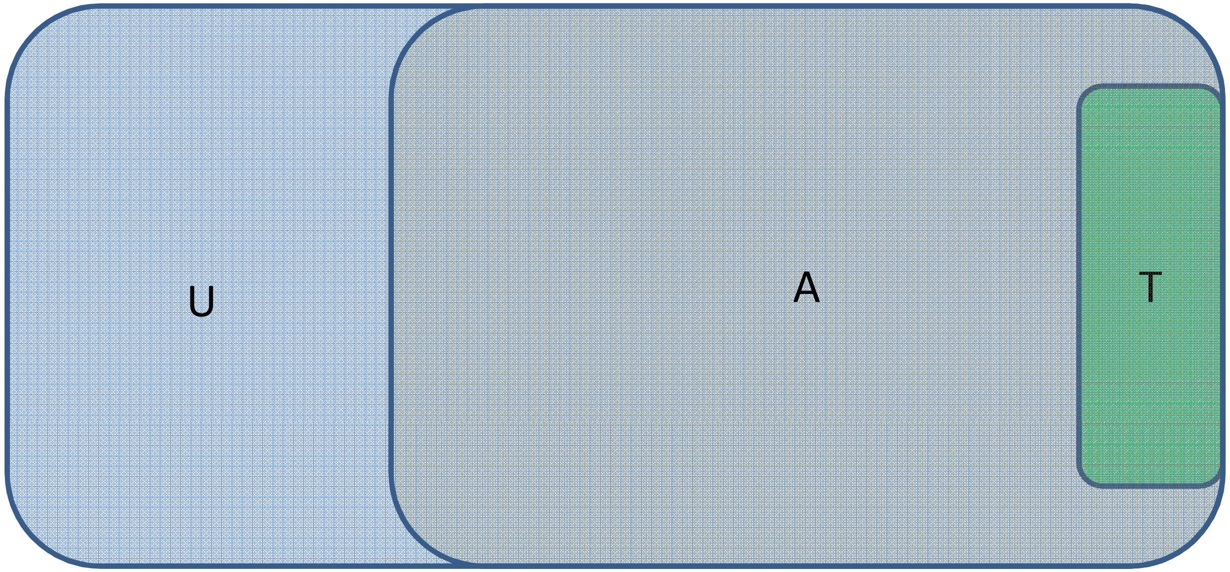
MDP: Simple Fails



In all blue states the value is ½.

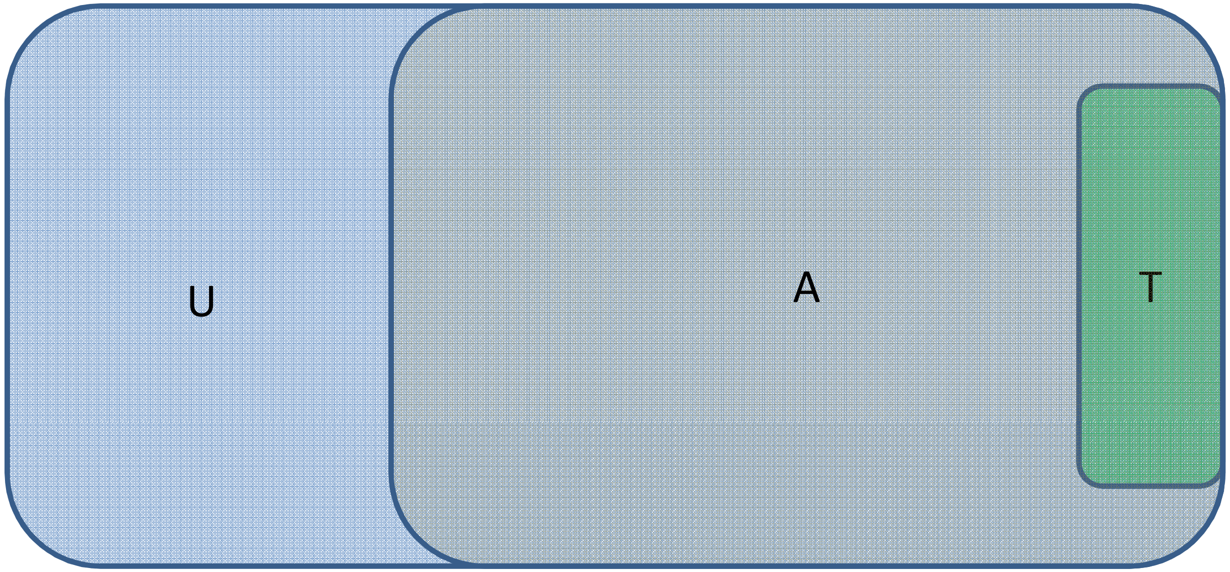
However the choice of red edge is bad.

Quantitative Reachability



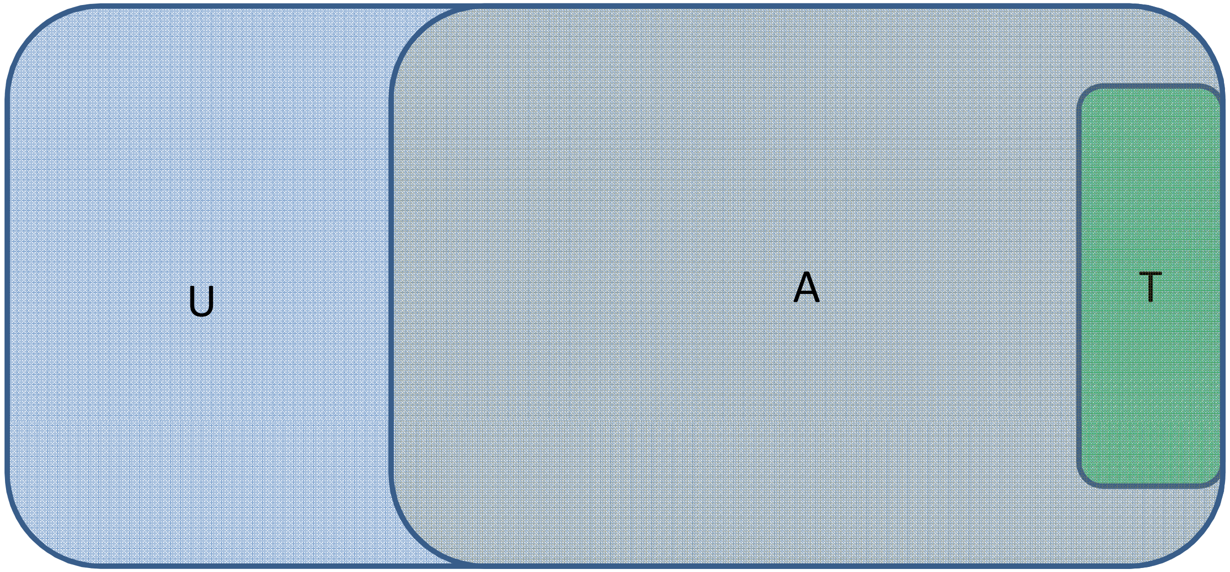
- Original MDP is connected.
- Compute simple reachability to T.
- From U, there is no path so value is 0.
- From A, the value is positive everywhere as there is a path.

Quantitative Reachability



- From U, there is no path so value is 0.
- From A, the value is positive everywhere as there is a path.
- Retain only the edges that attains the max in A (remove all the other). Make U and T absorbing.
- Easy to show that there is still path to T from A.
- Choose the edge that shortens distance to T.

Quantitative Reachability



- Retain only the edges that attains the max in A (remove all the other). Make U and T absorbing.
- Easy to show that there is still path to T from A.
- Choose the edge that shortens distance to T.
- Markov chain where all closed recurrent states are U or T.
- The values v(s) satisfies the Markov chain equality. Hence the memoryless strategy achieves v(s).

Quantitative Reachability

- An MDP G, and a target set T.
- Val(Reach(T))(s) = $\sup_\sigma \Pr_s^\sigma$(Reach(T)).

- Existence of pure memoryless optimal strategies.

- Algorithm: Linear programming. Variable x_s for all states s.
  - $x_s = 0$ $\quad s \in U$
  - $x_s = 1$ $\quad s \in T$
  - $x_s = \sum_{t \in S} x_t * \delta(s)(t)$ $\quad s \in S_P$
  - $x_s = \max_{t \in E(s)} x_t$ $\quad s \in S_1$.

Quantitative Reachability

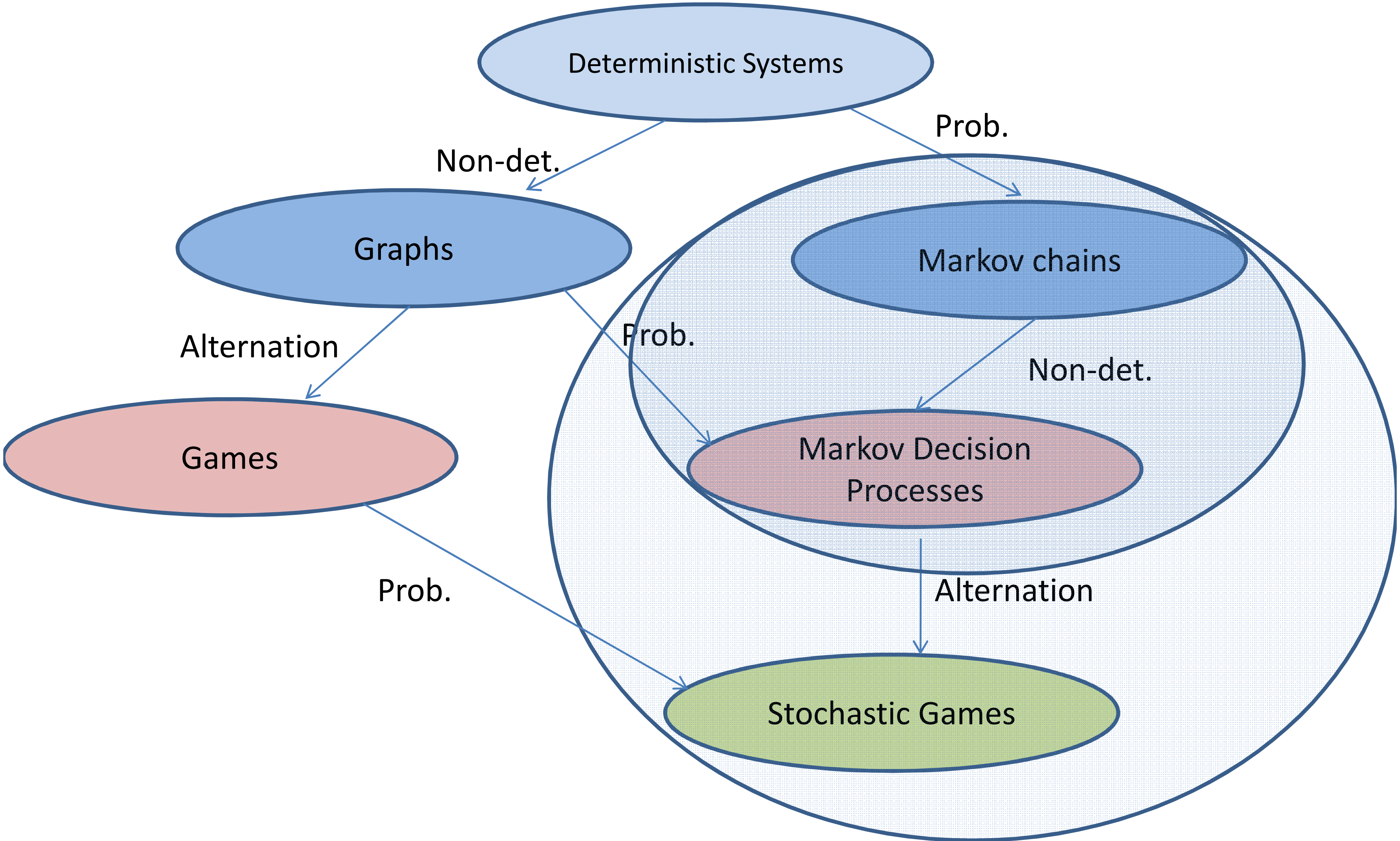
- Algorithm: Linear programming. Variable x_s for all states s.
  - $x_s = 0$ $\quad s \in U$
  - $x_s = 1$ $\quad s \in T$
  - $x_s = \sum_{t \in S} x_t * \delta(s)(t)$ $\quad s \in S_P$
  - $x_s = \max_{t \in E(s)} x_t$ $\quad s \in S_1$.

- The above optimization to linear program
  - Objective function: min $\sum_{t \in S} x_t$
  - $x_s \geq x_t$ $\quad s \in S_1,\ t \in E(s)$.

MDP Summary

	Reachability	**Liveness**	**Parity**
Qualitative	O(n m)	O(n m)	O(n m d)
Quantitative	Linear programming	Linear programming	Linear programming

Stochastic Games

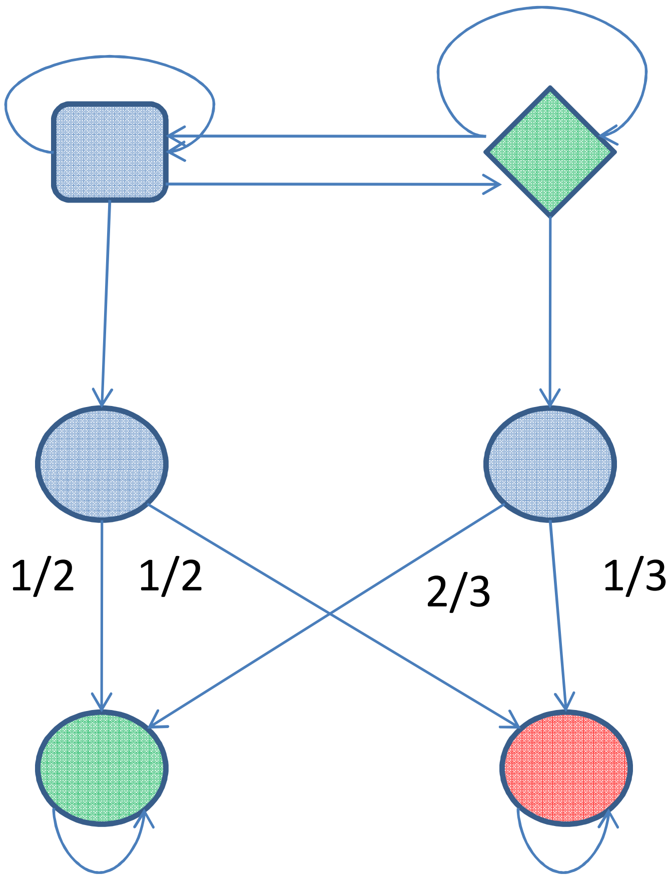


Stochastic Games

- Stochastic games
  - Non-determinism: angelic vs. demonic non-determinism (alternation).
  - Probability.
  - Generalizes non-deterministic systems and Markov chains, alternating games, MDPs.

- An MDP $G= ((S,E), (S_1, S_2,S_P), \delta)$
  - $\delta : S_P \rightarrow D(S)$.
  - For $s \in S_P$, the edge $(s,t) \in E$ iff $\delta(s)(t)>0$.
  - $E(s)$ out-going edges from s, and assume $E(s)$ non-empty for all s.

Stochastic Game



Example of stochastic game.

Objective for player 1 is to visit green infinitely often

Strategies

- Strategies are recipe how to move tokens or how to extend plays. Formally, given a history of play (or finite sequence of states), it chooses a probability distribution over out-going edges.
  - σ: $S^* S_1 \rightarrow D(S)$.
  - π: $S^* S_2 \rightarrow D(S)$.

Strategies

- Strategies are recipe how to move tokens or how to extend plays. Formally, given a history of play (or finite sequence of states), it chooses a probability distribution over out-going edges.
  - σ: $S^* S_1 \rightarrow D(S)$.
- History dependent and randomized.
- History independent: depends only current state (memoryless or positional).
  - σ: $S_1 \rightarrow D(S)$
- Deterministic: no randomization (pure strategies).
  - σ: $S^* S_1 \rightarrow S$
- Deterministic and memoryless: no memory and no randomization (pure and memoryless and is the simplest class).
  - σ: $S_1 \rightarrow S$
- Same notations for player 2 strategies π.

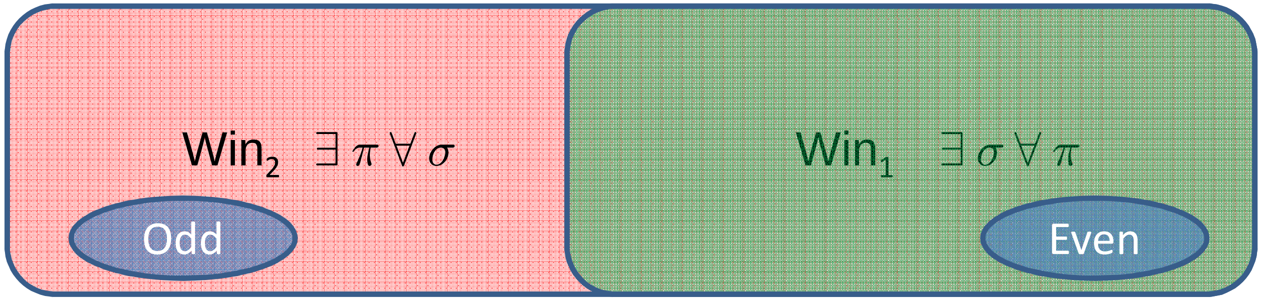
Values in Stochastic Games

- Value at a state for an objective ψ
  - $\mathrm{Val}(\psi)(s) = \sup_{\sigma} \inf_{\pi} \Pr_s^{\sigma,\pi}(\psi)$.
- Qualitative analysis
  - Compute the set of almost-sure (prob 1) winning states (i.e., set of states with value 1).
- Quantitative analysis
  - Compute the value for all states.
- Determinacy: the order of sup inf can be exchanged.

Non-Stochastic Games

- There are no probabilistic states.

- Non-stochastic games with parity objectives
  - Values only 0 or 1.
  - Pure memoryless winning strategies exist.
  - Once a pure memoryless strategy is fixed all cycles winning.

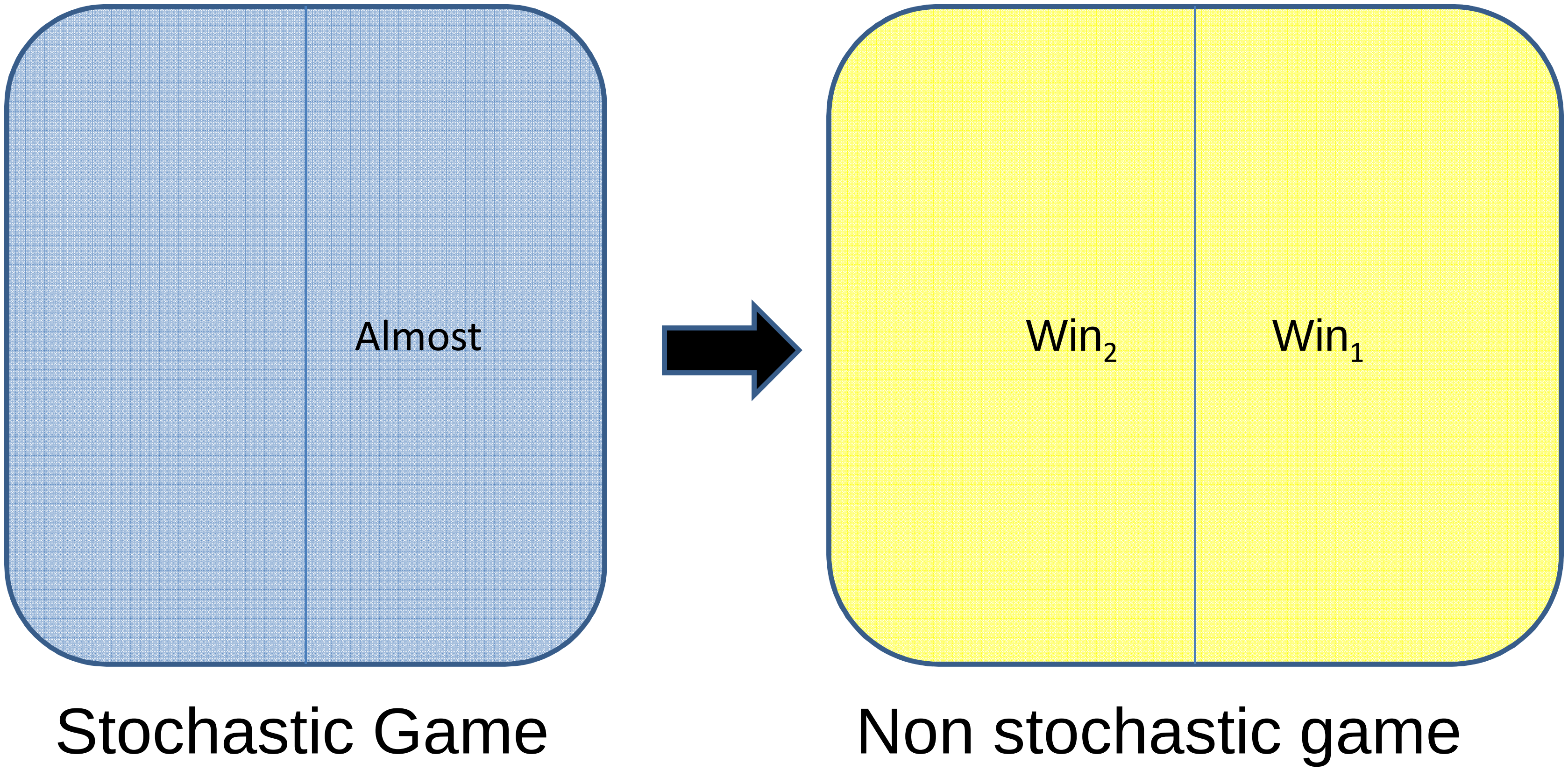


Qualitative and Quantitative Analysis

- Qualitative analysis
  - Reduction to games without probability.
  - Use existence of pure memoryless strategies in games with probability for parity objectives.
  - Show it for Liveness and can be extended to parity.
- Quantitative analysis
  - Combine notion of qualitative and local optimality for quantitative optimality.

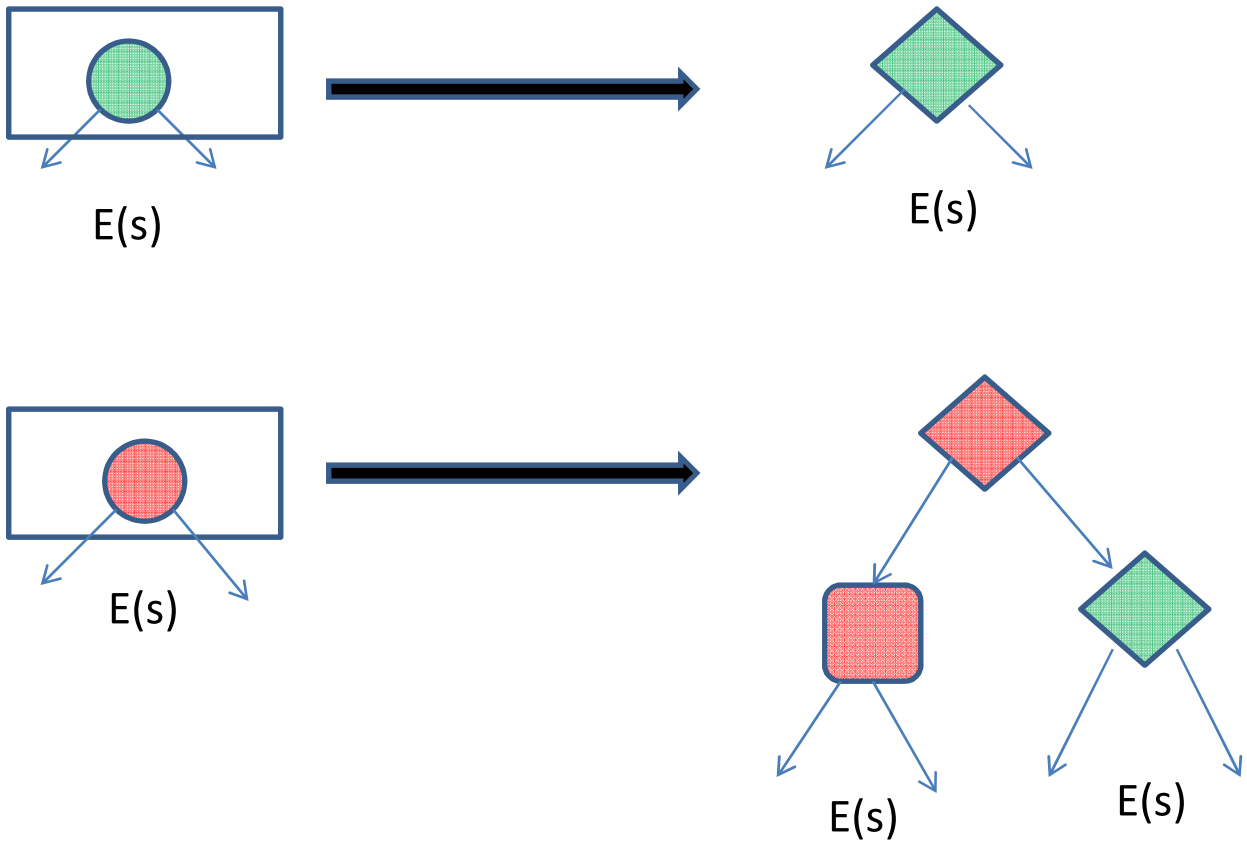
Qualitative Analysis

- Reduction



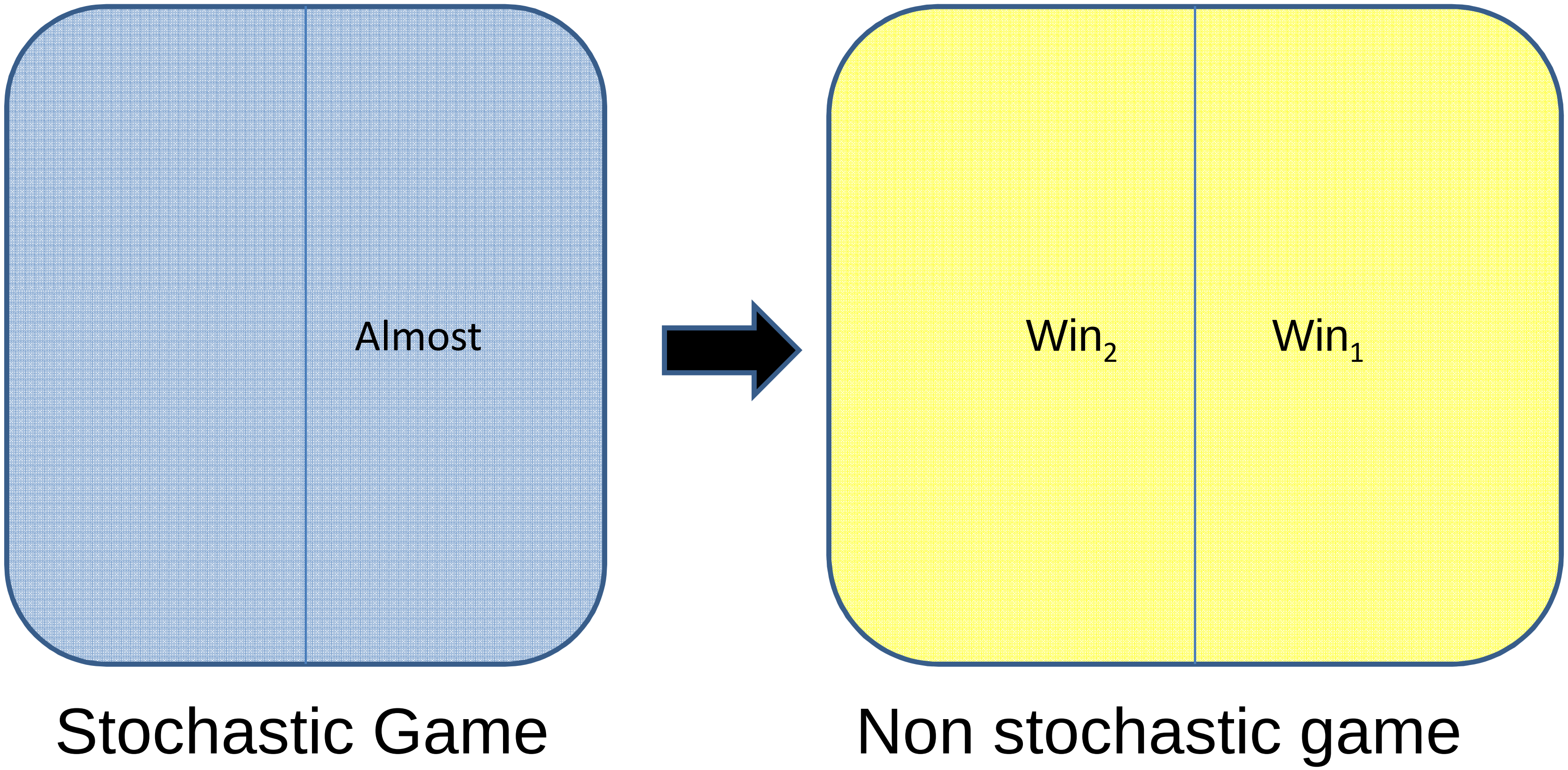
Reduction

- Replace every probabilistic state by two-player gadget. Illustrate it for Liveness.



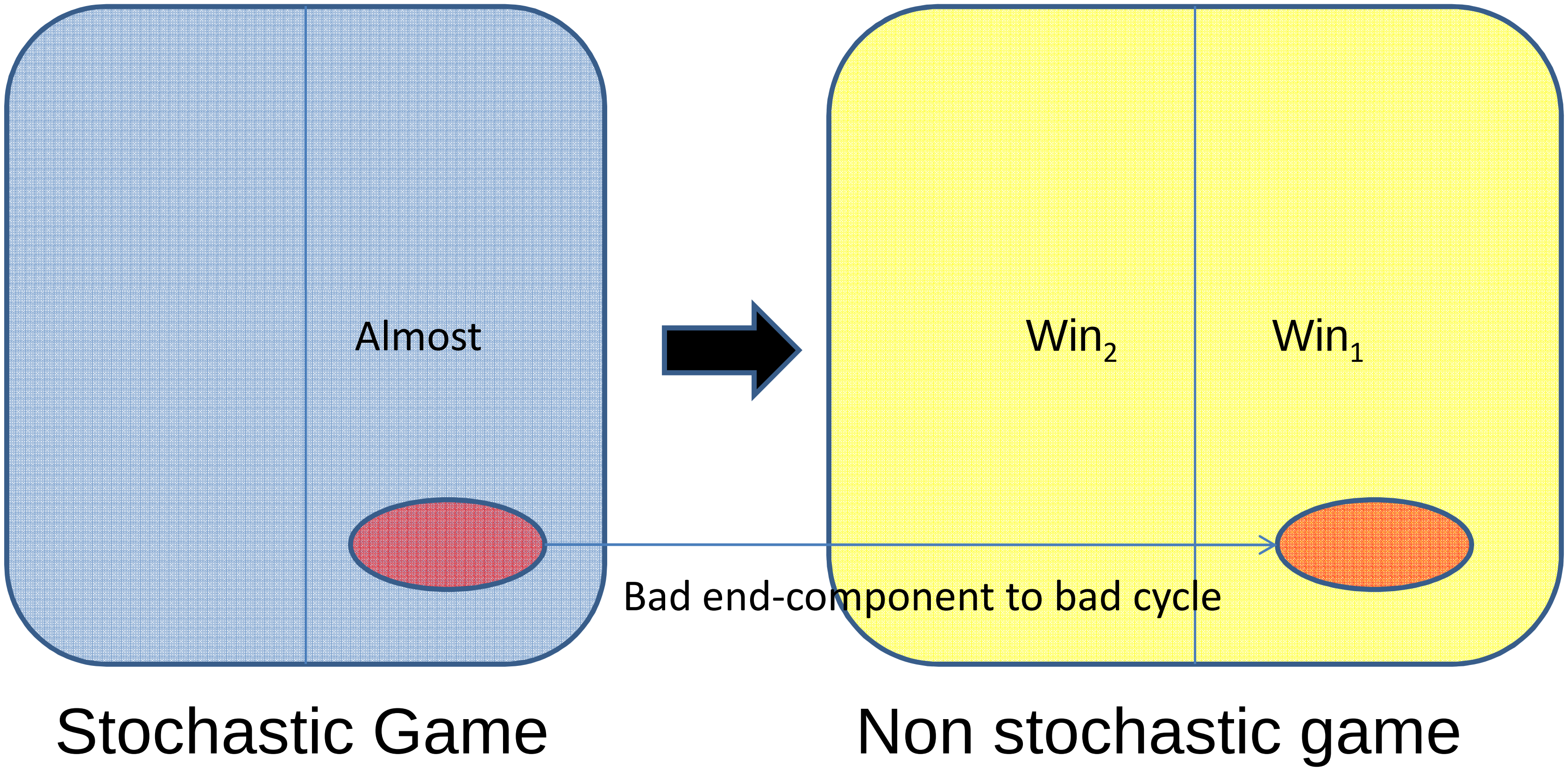
Qualitative Analysis

- Reduction: the end-components are winning.



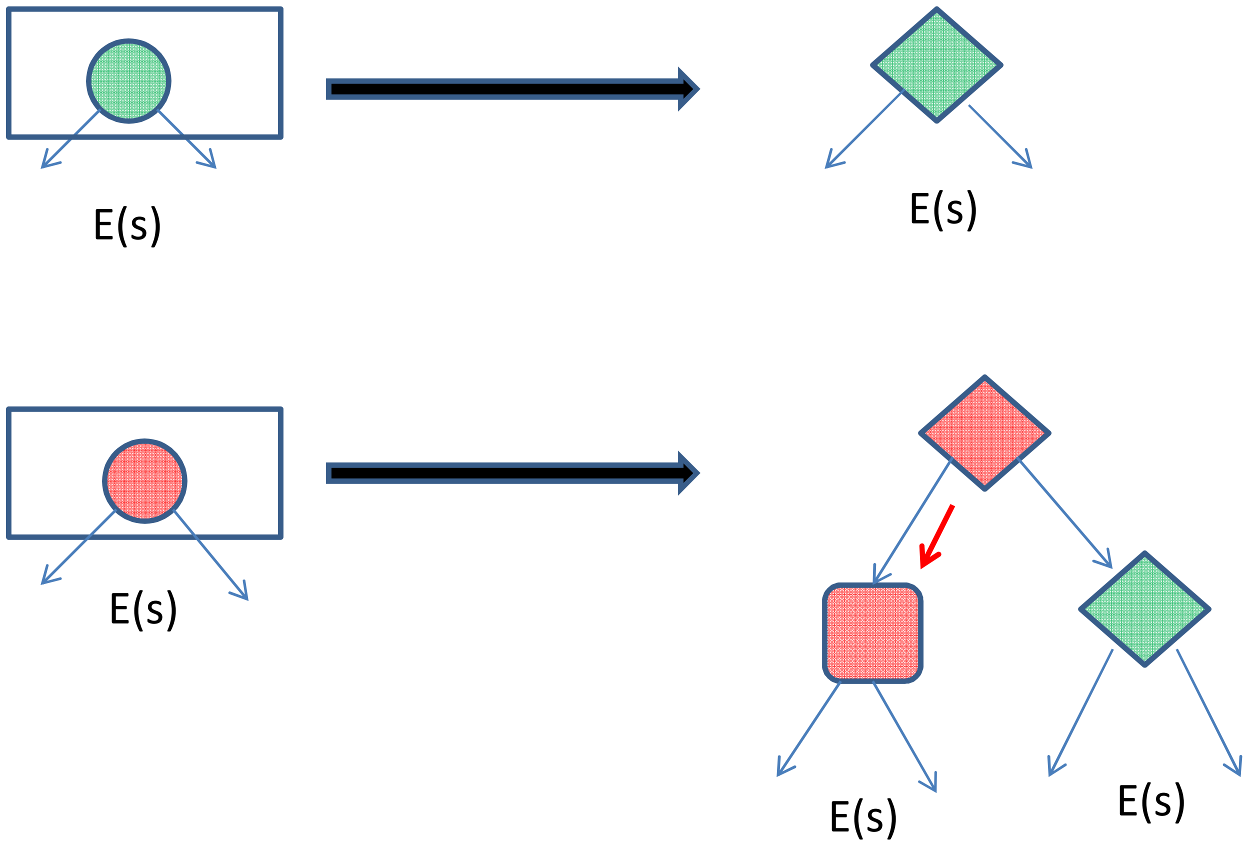
Qualitative Analysis

- Reduction: the end-components are winning.



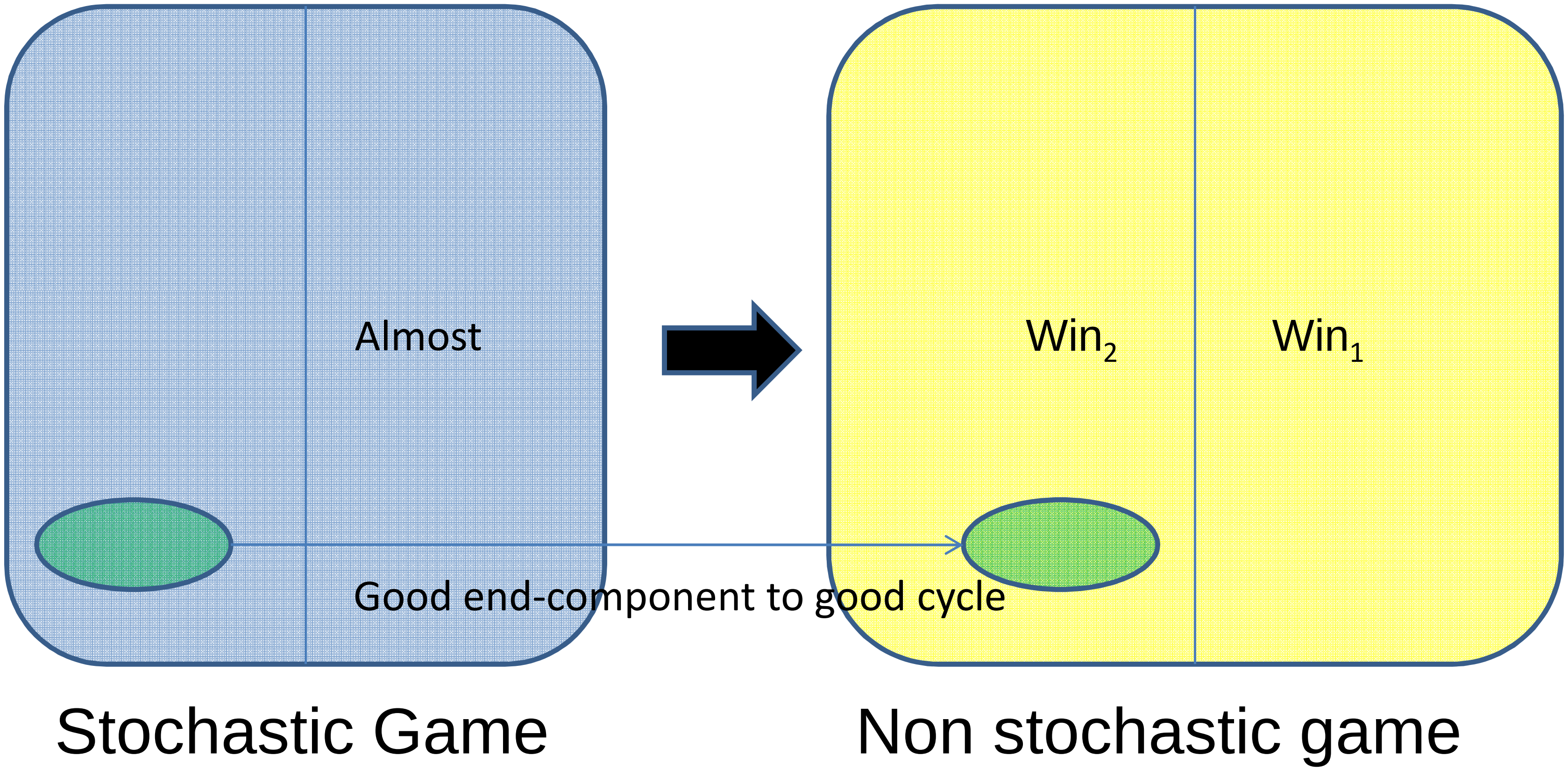
Reduction

- Choice in the gadget



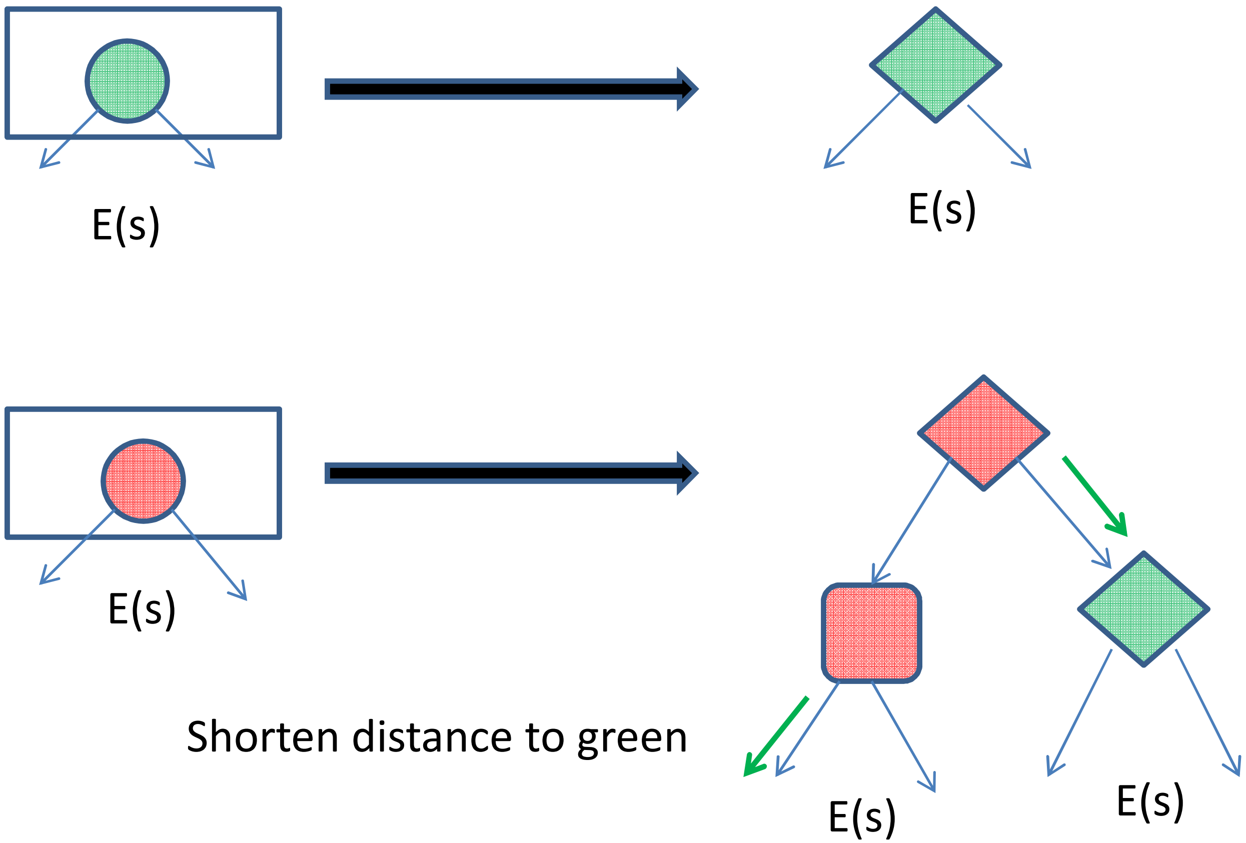
Qualitative Analysis

- Reduction: the end-components are winning.



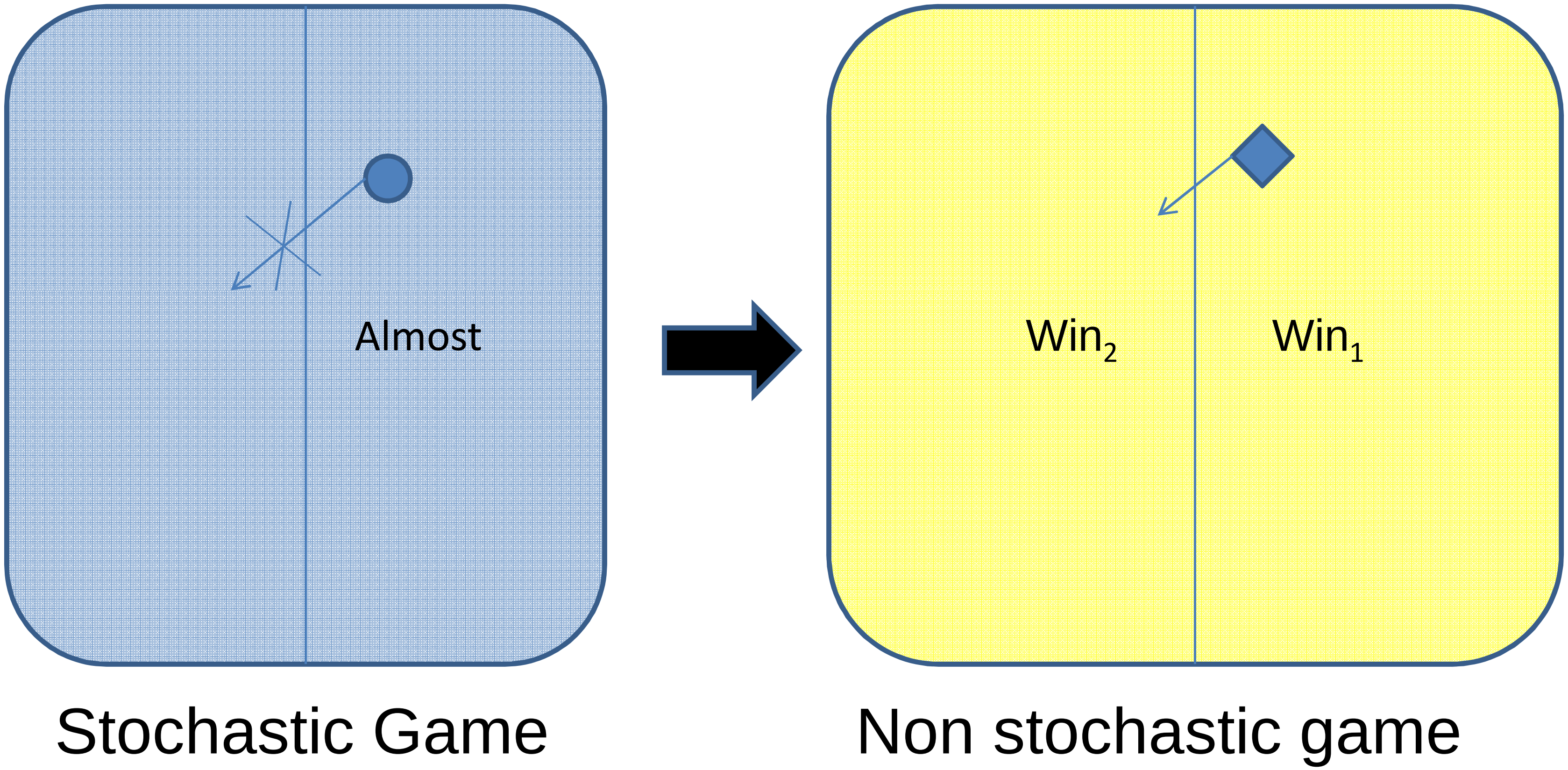
Reduction

- Choice in the gadget



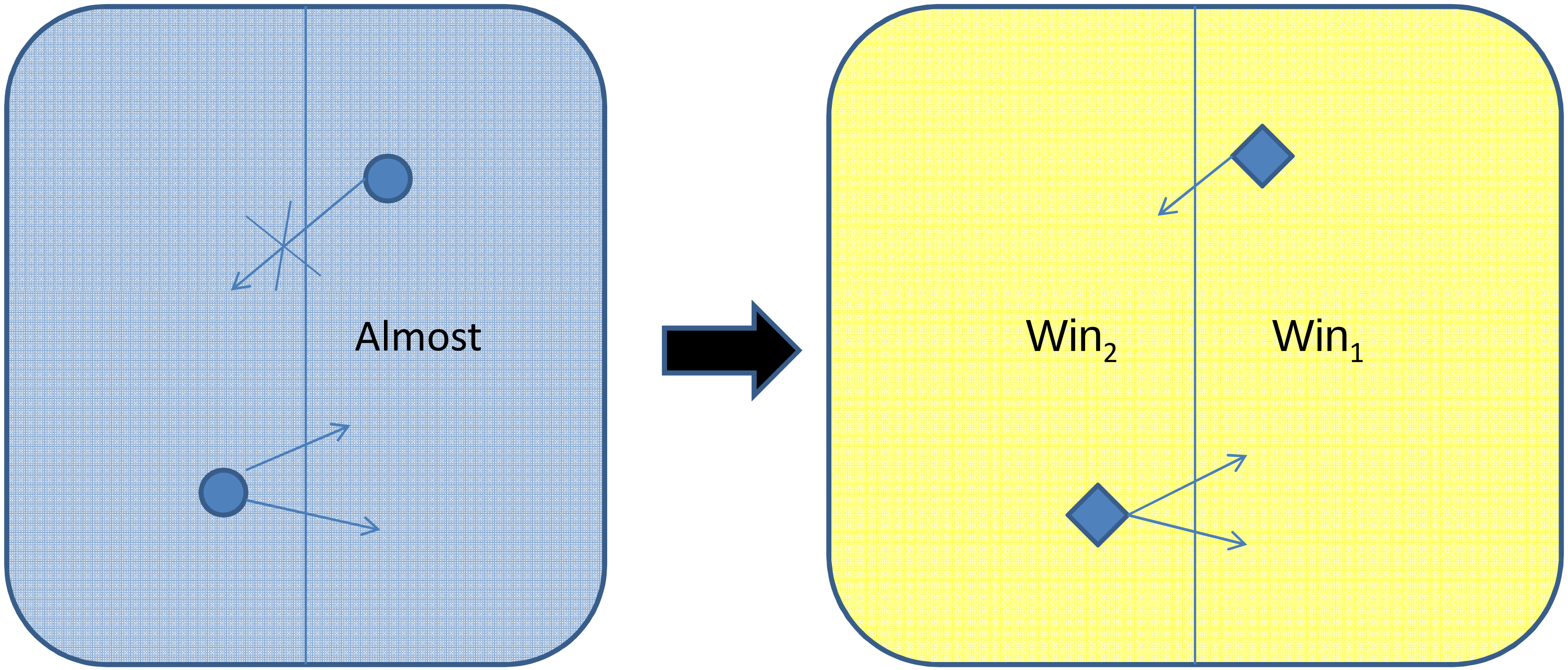
Qualitative Analysis

- Reduction: the end-components are winning.



Qualitative Analysis

- Reduction: the end-components are winning.



Stochastic Game　　　　Non stochastic game

Reduction

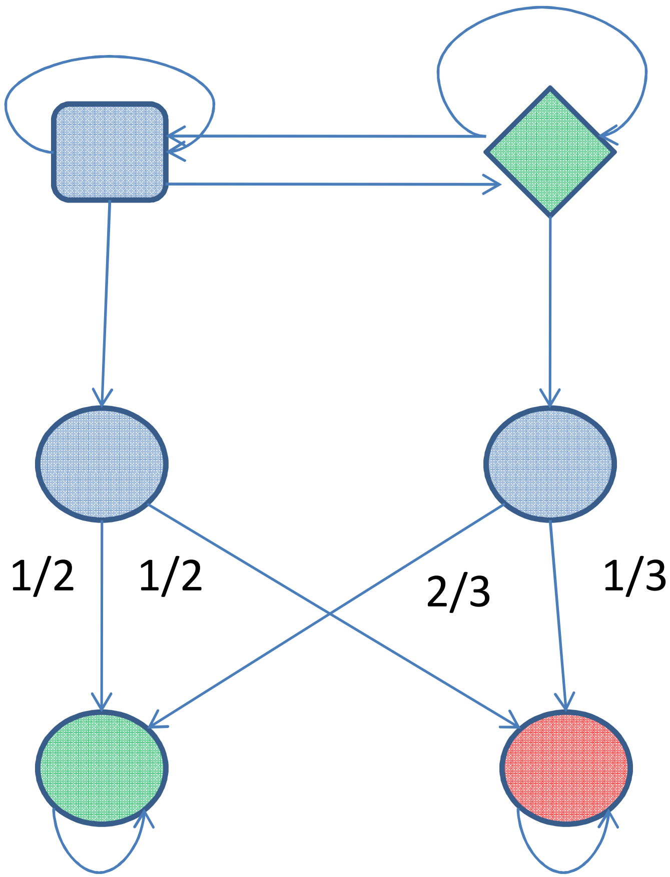
- Gadget based reduction can be extended to parity.

- Qualitative analysis
  - Pure memoryless almost-sure strategies exists.
  - Linear time reduction to non-stochastic games.
  - Same complexity: NP $\cap$ coNP.
  - All algorithms can be used.

Quantitative Analysis

- Unlike MDPs, we cannot do the following:
  - Compute almost-sure winning states.
  - Compute quantitative reachability to almost-sure winning states.
  - We illustrate with an example.

Stochastic Game



Example of stochastic game.

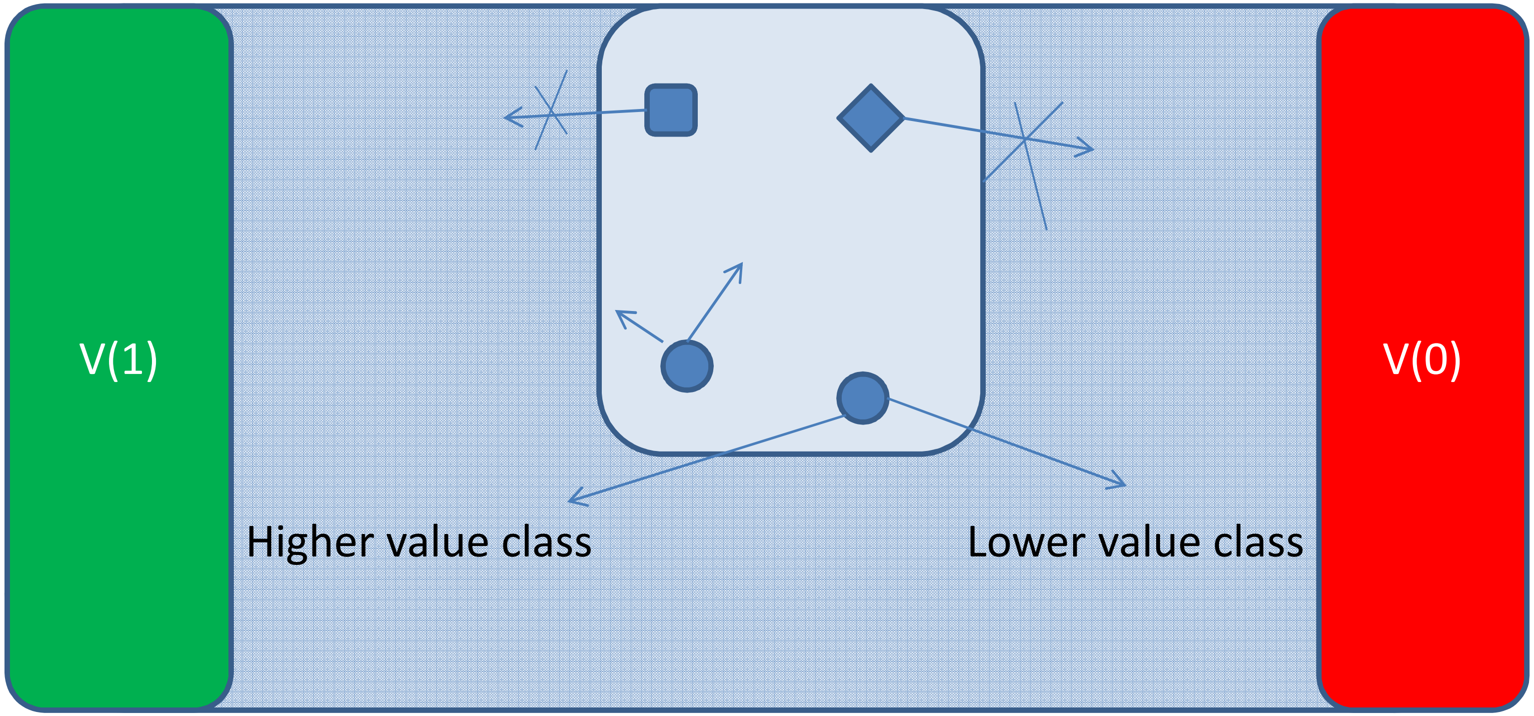
Objective for player 1 is to visit green infinitely often

Cannot ensure to reach green absorbing with prob 2/3.

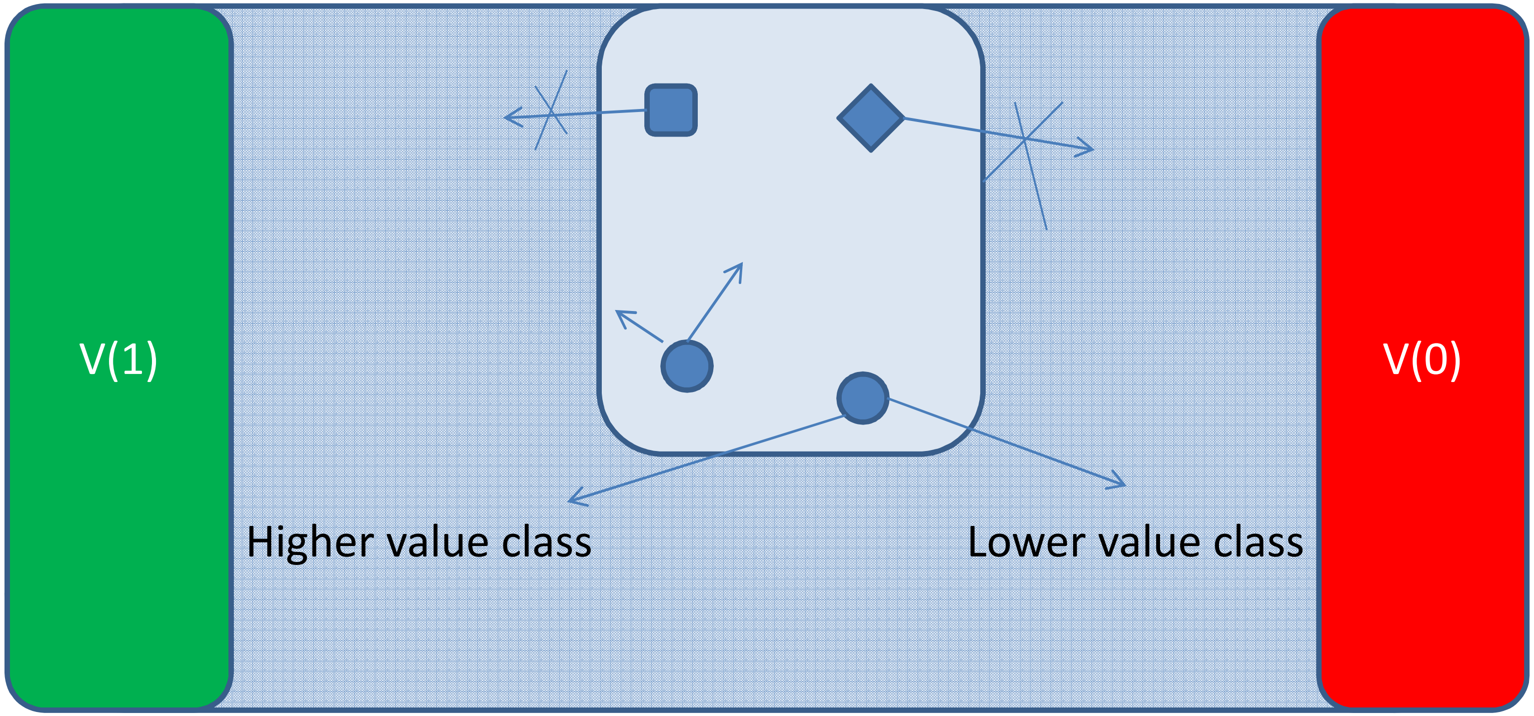
Quantitative Analysis

- Quantitative optimality
  - Local optimality
  - Qualitative optimality
- Value class: the set of states with same value. V(r) is the set of states with value r.

Value Class Property



Value Class: Boundary Probabilistic States



Value Class Reduction

- Remove edges going out to lower value class (local optimality).

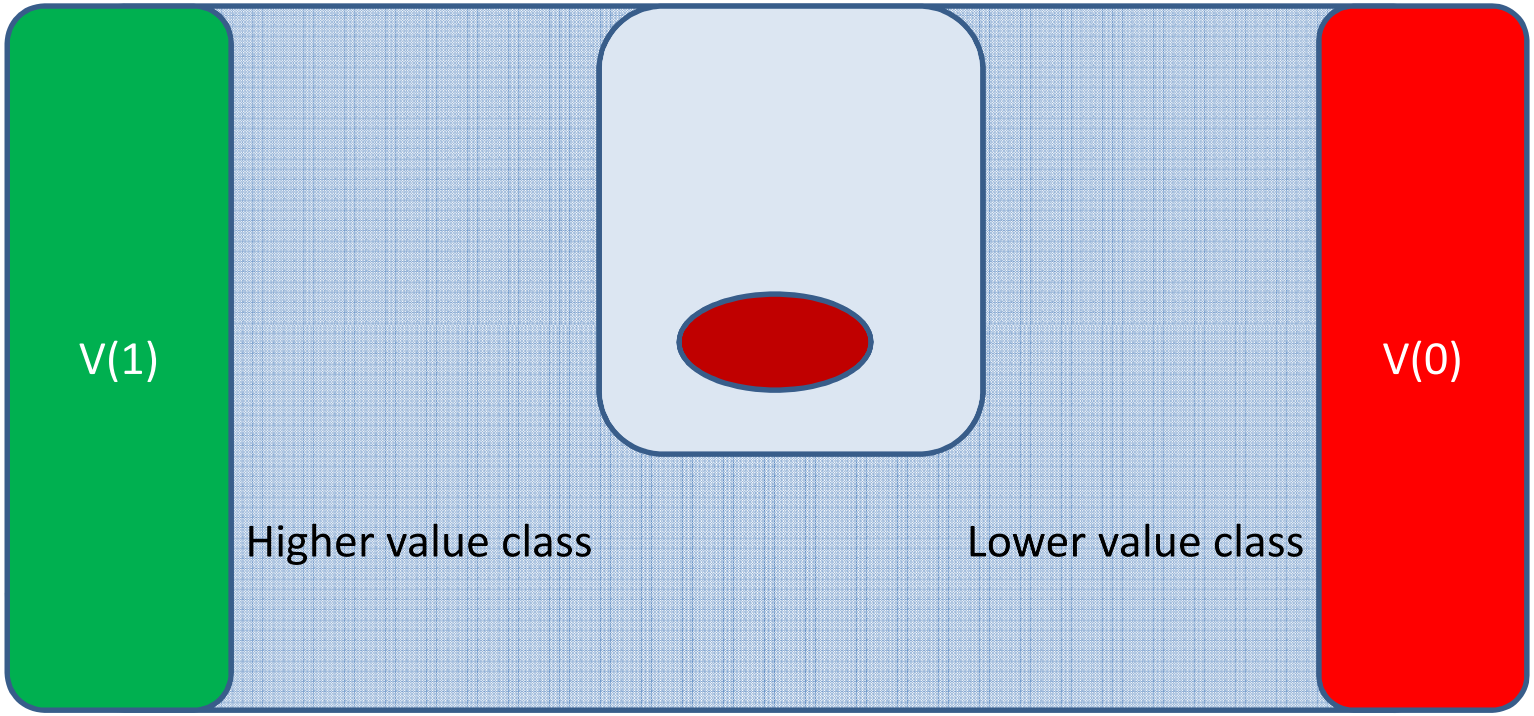
- Change boundary probabilistic states to winning states for player 1.

- Claim: In this sub-game player 1 wins almost-surely everywhere.

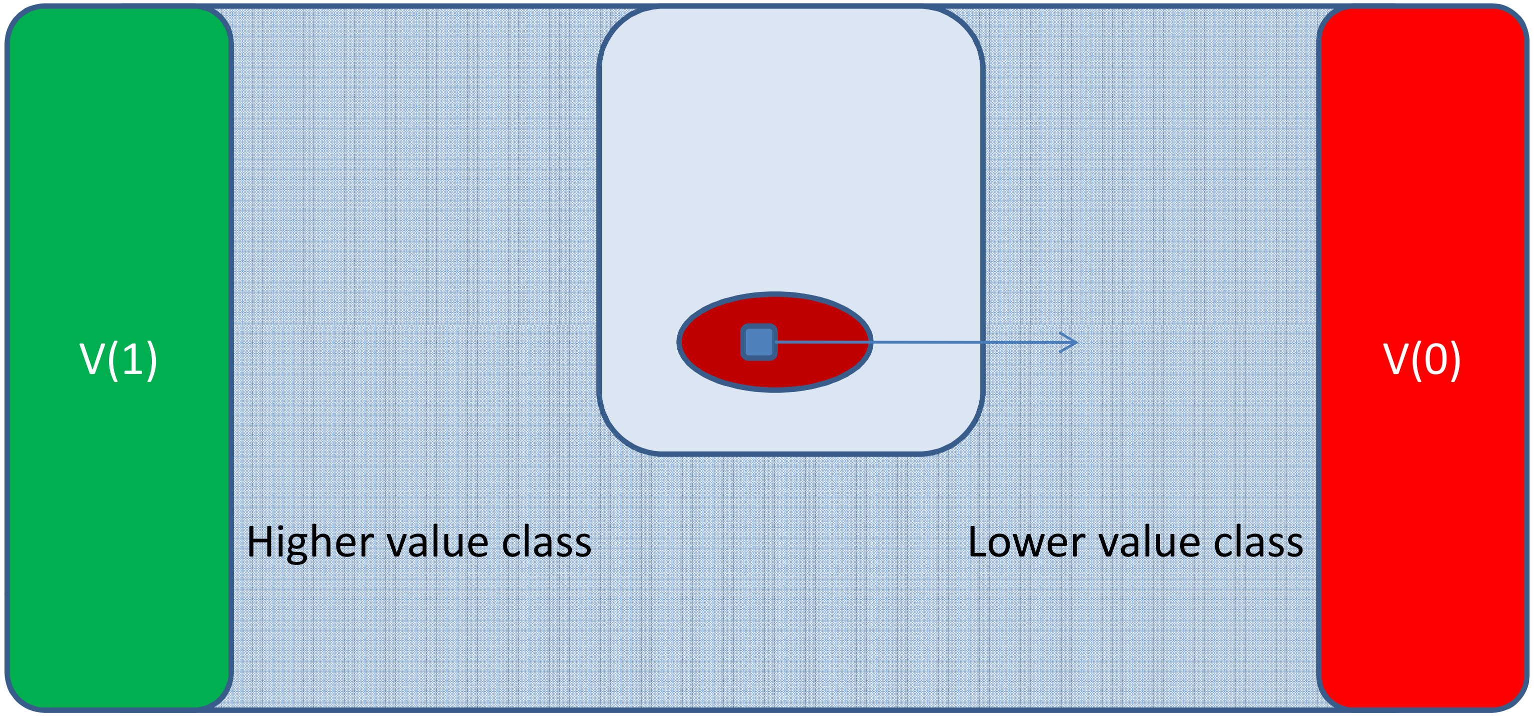
Sub-game Qualitative Optimality

- Claim: Player 1 wins almost-surely.
- Proof: Suppose not.
  - Then player 2 wins with positive probability somewhere.
  - Player 2 wins almost-surely somewhere.
  - Player 1 if stays in the value class loses with probability 1 or else jumps to a lower value class.
  - Contradiction.

Value Class: Boundary Probabilistic States



Value Class: Boundary Probabilistic States



Value Class Property

- In value classes if we assume boundary probabilistic vertices winning for player 1 then player 1 wins almost surely.

- Conditional almost-sure winning strategies.

- Stitching lemma: Compose them to get a optimal strategy.

Stitching Lemma

- Proof idea:
  - If the game stays in some value class player 1 wins with probability 1.
  - Else it leaves the value class through the boundary probabilistic vertex or goes to a higher value class.
  - Invoke sub-martingale Theorem or use results from MDPs.

Quantitative Analysis

- Pure memoryless optimal strategies exist.
- Complexity bound
  - NP $\cap$ coNP.
- Algorithms: Strategy improvement algorithms, uses qualitative algorithms and local optimality.

Stochastic Games Summary

	Reachability	**Liveness**	**Parity**
Qualitative	O(n m)	O(n m)	NP ∩ coNP Linear reduction to non-stochastic parity
Quantitative	NP ∩ coNP	NP ∩ coNP	NP ∩ coNP

Summary and Messages

- Markov chains
  - Qualitative: Linear time algorithm through closed recurrent states (bottom scc's).
  - Quantitative analysis: Linear equalities, Gaussian elimination.
- MDPs
  - Qualitative: Iterative algorithm.
  - Quantitative: Reduction to quantitative reachability using end-components.
  - Quantitative reachability: Linear programming.
- Stochastic games
  - Qualitative: Reduction to non-stochastic games.
  - Quantitative: qualitative and local optimality.

Extensions

- Perfect-information turn-based finite state stochastic games

  - Infinite state games: pushdown games, timed games.
  - Concurrent games: simultaneous interaction.
  - Imperfect-information games.

CONCURRENT GAMES

Games on Graphs

Games on graphs:

1. Turn-based:
   - Chess.
   - Tic-tac-toe.

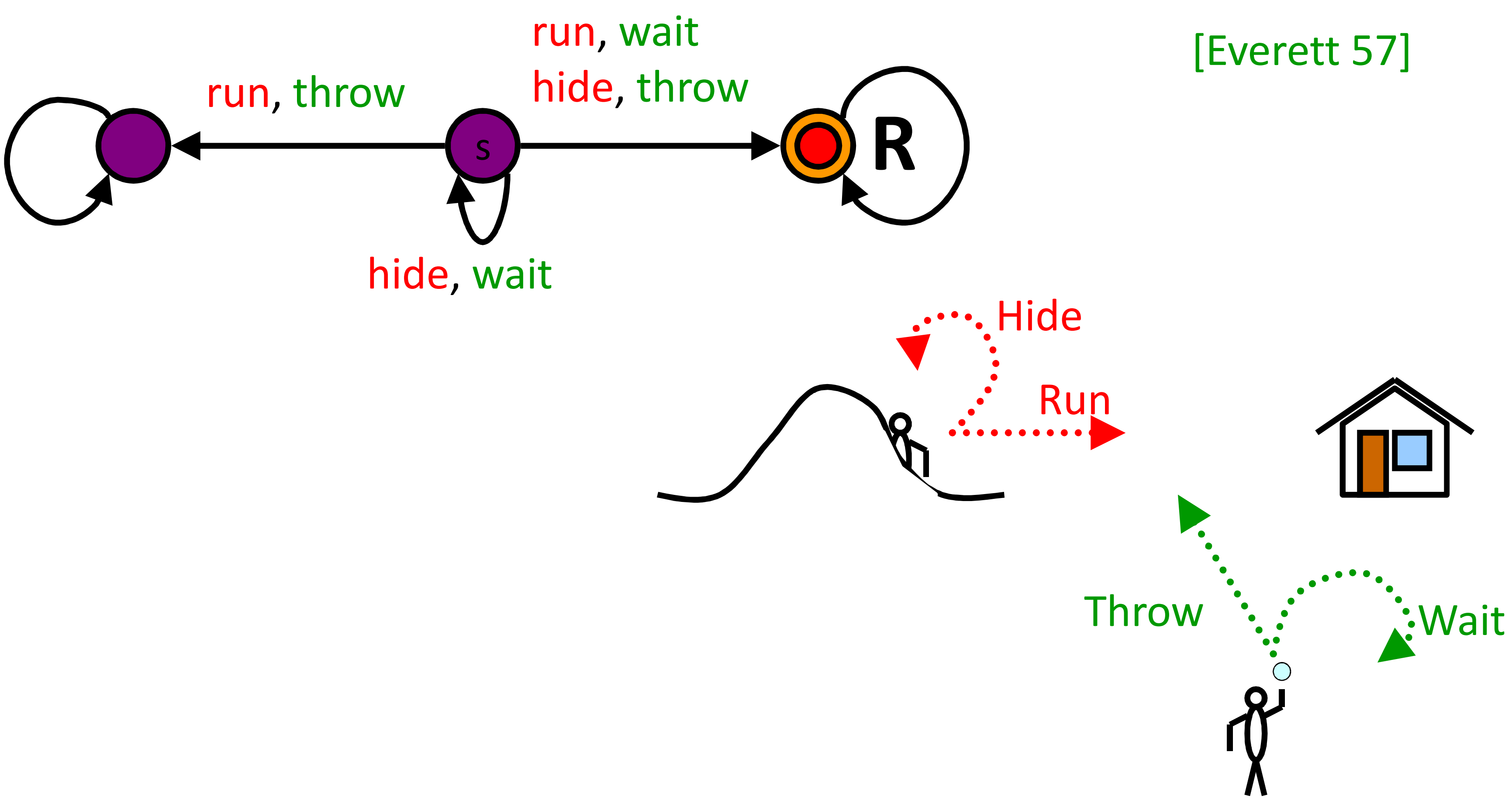
2. Concurrent:
   - Penalty Shoot-out.
   - Rock-paper-scissor.

Concurrent Game Graphs

A concurrent game graph is a tuple G =(S,M,Γ_1,Γ_2,δ)

- S is a finite set of states.
- M is a finite set of moves or actions.
- Γ_i: S $\rightarrow 2^M \setminus \emptyset$ is an action assignment function that assigns the non-empty set Γ_i(s) of actions to player i at s, where i $\in$ {1,2}.
- δ: S $\times$ M $\times$ M $\rightarrow$ Dist(S), is a probabilistic transition function that given a state and actions of both players gives a probability distribution of the next state.

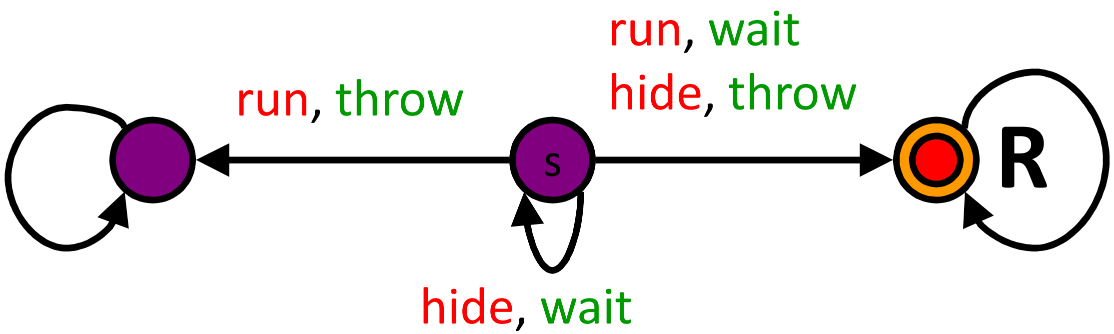
An Example (Deterministic Transition)



[Everett 57]

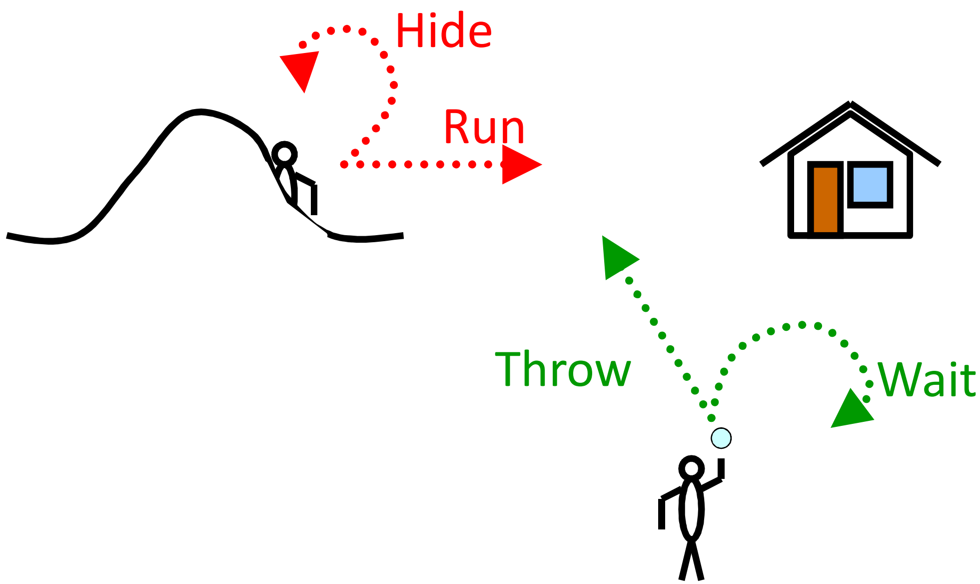
Concurrent reachability games

[Everett 57]



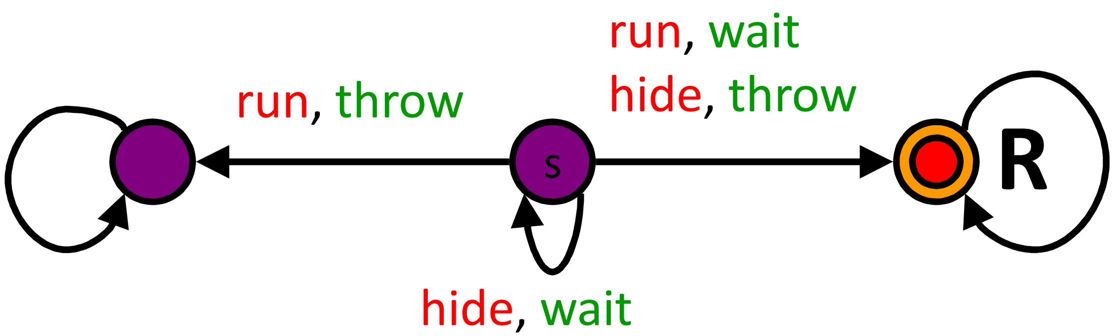
Move	Probability	
run	q	
hide	1-q	(q>0)

Win at s with probability 1-q, for all q > 0.



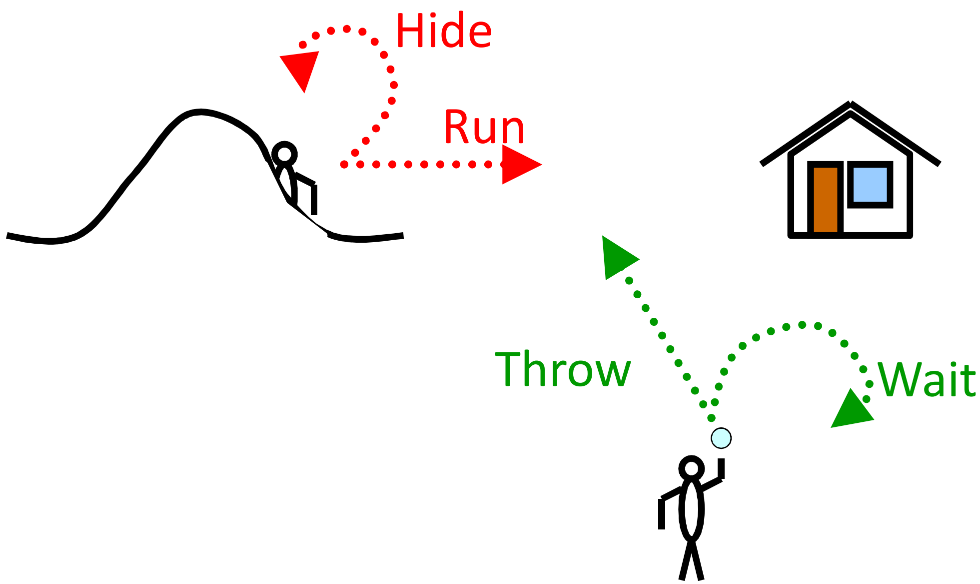
Concurrent reachability games

[Everett 57]



Move	Probability	
run	q	
hide	1-q	(q>0)

Win at s with probability 1-q, for all q > 0.



Player 1 cannot achieve v(s) = 1, only v(s) = 1-q for all q > 0.

Concurrent Games

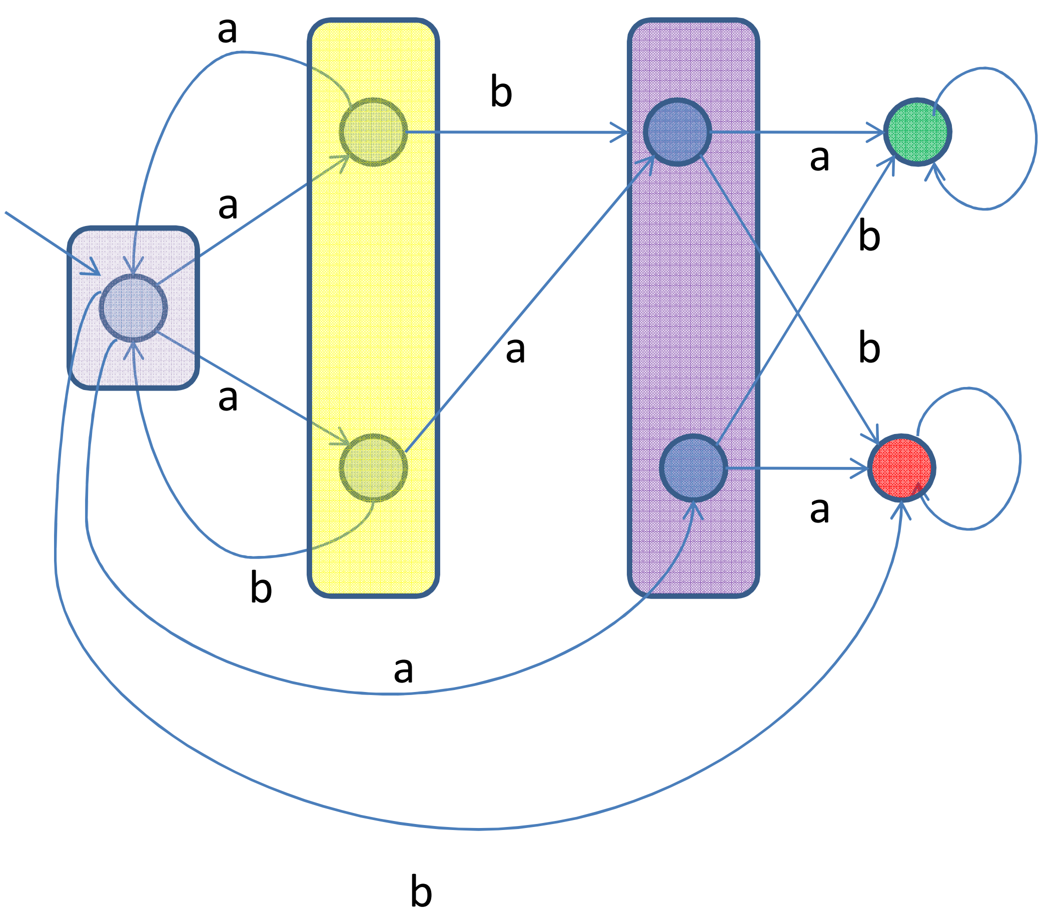
- Strategies
  - Require randomization.
  - May not be optimal.
  - Only ϵ-optimal, for $\epsilon>0$.
  - For liveness requires infinite memory.

- Values can be irrational for concurrent deterministic reachability games.

- Qualitative and quantitative analysis still decidable
  - Qualitative analysis is NP $\cap$ coNP.
  - Quantitative analysis is PSPACE.

PARTIAL-INFORMATION GAMES

Partial-information Games



In starting play a.
In yellow play a and b at random.
In purple:
- if last was yellow then a
- if last was starting, then b.

Requires both randomization and memory

Partial-information Games

- Strategies
  - Require randomization.
  - May not be optimal.
  - Only ϵ-optimal, for ϵ>0.
  - For liveness requires infinite memory.
  - More complicated than concurrent games.

- Quantitative analysis
  - Undecidable.

- Qualitative analysis
  - Reachability, Liveness: EXPTIME-complete.
  - Parity: Undecidable.

Conclusion

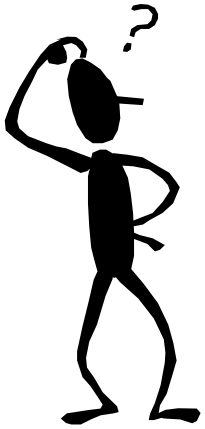
- Perfect-information stochastic games
  - Applications: verification and synthesis of stochastic reactive systems.
  - Markov chains, MDPs and stochastic games with parity objectives.
- Glimpses of the world of games beyond.

References

- Applications and connections:
  - Church 62, Pnueli-Rosner 89, Ramadge-Wonham 87, Courcoubetis-Yannakakis 95, Thomas Handbook 97, and many more.
- Markov chains:
  - Book of Kemeny
- Markov Decision Processes:
  - Book of Filar Vrieze.
  - Probabilistic verification: Courcoubetis-Yannakakis.
  - PhD Thesis of deAlfaro.
- Stochastic games
  - Condon 92, 93.
  - PhD Thesis of Chatterjee

The end

Thank you !



Questions ?