Quantitative objectives in Markov decision processes

Jan Křetínský

IST Austria

Advanced topics in formal methods
June 8, 2015

Quantitative properties of runs

A reward function $r : S \to \mathbb{N}$ maps a run $s_1 s_2 \cdots$ to a sequence of values $v_1 v_2 \cdots = r(s_1) r(s_2) \cdots$

An objective function $f : \mathbb{N}^\omega \to \mathbb{N}$ assigns to it a value

- discounted sum $\sum_{i=1}^{\infty} \lambda^i v_i$
- total reward
  - over finite horizon T: $\sum_{i=1}^{T} v_i$
  - over infinite horizon: $\lim_{T\to\infty} \sum_{i=1}^{T} v_i$
- (limit) average / mean payoff $\lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} v_i$

Mean payoff

Mean payoff (Long-run average reward):

$$v_1 v_2 \cdots = 42\ 1\ 2\ 1\ 2\ 1\ 2 \cdots$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} v_i}{n} = 1.5$$

Mean payoff

Mean payoff (Long-run average reward):

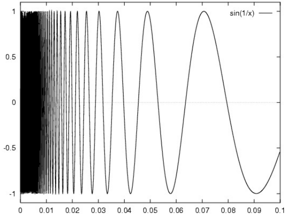
$$v_1 v_2 \cdots = 42\ 1\ 2\ 1\ 2\ 1\ 2 \cdots$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} v_i}{n} = 1.5$$

Limit may not exist:

$$0\ (1)^{10}\ (0)^{1000}\ (1)^{1000000} \cdots$$

$$MP = \liminf_{n\to\infty} \frac{\sum_{i=1}^{n} v_i}{n} = 0$$

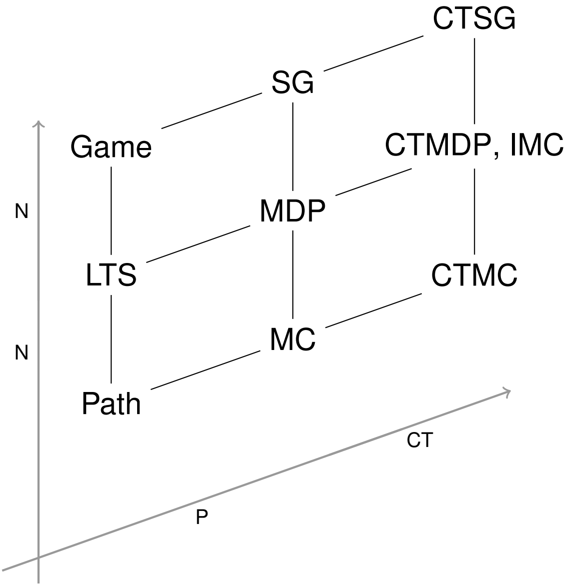


Properties of systems

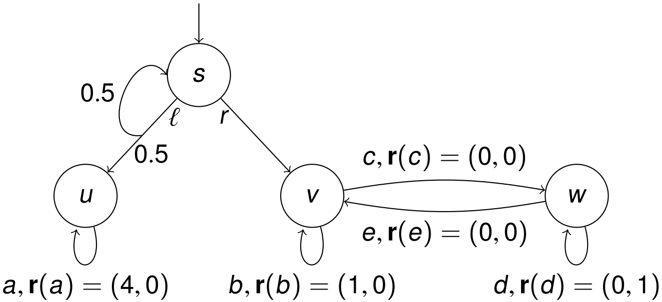
How to combine values of runs into a value of a system?

- $\max(\cdot), \min(\cdot)$
- $\mathbb{E}[\cdot]$
- $\mathbb{P}[\cdot > \textit{value\_threshold}] > \textit{probability\_threshold}$
- combination of objectives
- combinations depending on systems

Systems



Markov decision processes



Example

Money investment

- >0 earning, <0 losing
- maximize expected mean payoff: $\vec{v}_s = \max_\sigma \mathbb{E}_s^\sigma[MP]$

Value iteration (backward induction)

Value vector $\vec{v}$ found by successive approximation
For unichains (every strategy induces a Markov chain with only one recurrent class), extensible to MDPs with constant value vector

1. Choose $\varepsilon > 0$, and take $\vec{w} \in \mathbb{R}^{|S|}$ arbitrarily
2. Compute:
   - $q(a) := r(a) + \sum_{s' \in S} \delta(a)(s')\vec{w}_{s'}$, for $s \in S$ and $a \in A(s)$
   - $\vec{u}_s := \max_{a \in A(s)} q(a)$, for $s \in S$, and take f such that $\vec{u}_s = r(f(s)) + \sum_{s' \in S} \delta(f(s), s')\vec{w}_{s'}$
   - $k := \max_{s \in S}(\vec{u}_s - \vec{w}_s)$, $l := \min_{s \in S}(\vec{u}_s - \vec{w}_s)$
3. If $k - l \leq \varepsilon$: f is an ε-optimal strategy and $\frac{k+l}{2}$ is a $\frac{1}{2}\varepsilon$-approximation of the value $\vec{v}$ (Stop)
   Otherwise: $\vec{w} := \vec{u}$ and go to step 2.

$\vec{w}^t$ approximates the optimal total reward in time t
$\vec{w}^t - \vec{w}^{t-1}$, computed as $\vec{u} - \vec{w}$, converges to $\vec{v}$
k and l approximate $\vec{v}$ from above and below, respectively.

Strategy iteration I

Sequence $f^0, f^1, \ldots$ of strategies such that $\vec{v}(f^{t+1}) \geq \vec{v}(f^t)$ and
converging to an optimal strategy

Finitely many strategies $\Rightarrow$ termination

$$\begin{aligned}
\text{for all } s \in S: \quad & \vec{x}_s & &= \textstyle\sum_{s' \in S} \delta(f(s), s')\vec{x}_{s'} \\
\text{for all } s \in S: \quad & \vec{x}_s + \vec{y}_s & &= \textstyle\sum_{s' \in S} \delta(f(s), s')\vec{y}_{s'} + r(f(s)) \\
\text{for all } s \in S: \quad & \vec{y}_s + \vec{z}_s & &= \textstyle\sum_{s' \in S} \delta(f(s), s')\vec{z}_{s'}
\end{aligned} \tag{1}$$

$\vec{x}$ is equal to $\mathbb{E}^f[MP]$
$\vec{y}$ is the difference between total and long-run rewards
$\vec{z}$ is used in the algorithm to prevent cycling

Strategy iteration II

Using $(\vec{x}, \vec{y})$

$$B(s,f) = \left\{ a \in A(s) \;\middle|\; \begin{array}{l} \sum_{s'} \delta(a)(s')\vec{x}_{s'} > \vec{x}_s \text{ or} \\ \sum_{s'} \delta(a)(s')\vec{x}_{s'} = \vec{x}_s \text{ and} \\ r(a) + \sum_{s'} \delta(a)(s')\vec{y}_{s'} > \vec{x}_s + \vec{y}_s \end{array} \right\} \qquad (2)$$

1. Start with any $f \in F$.
2. Determine unique $(\vec{x}$,$\vec{y})$-part in a solution of the linear system (1)
3. For every $s \in S$: determine $B(s,f)$ as defined in (2) using the values $\vec{x}$ and $\vec{y}$ from step 2
4. If $B(s,f) = \emptyset$ for every $s \in S$: go to step 6
   Otherwise: take any $g \neq f$ such that $g(s) \in B(s,f)$ if $g(s) \neq f(s)$
5. $f := g$ and go to step 2
6. f is an average optimal strategy

Linear programming I

$\vec{v}$ the smallest solution of LP, strategy derived from its dual LP

Primary linear program:

$$\begin{aligned}
&\text{Minimize:}\\
&\qquad \sum_{s \in S} \vec{\mu}_s \vec{x}_s\\
&\text{Subject to:}\\
&\qquad \text{for all } s \in S, a \in A(s)\colon \vec{x}_s \geq \sum_{s' \in S} \delta(a)(s')\vec{x}_{s'}\\
&\qquad \text{for all } s \in S, a \in A(s)\colon \vec{x}_s \geq r(a) + \sum_{s' \in S} \delta(a)(s')\vec{y}_{s'} - \vec{y}_s
\end{aligned} \tag{3}$$

where $\vec{\mu}_s > 0$ arbitrarily chosen

Dual linear program:

Maximize:

$$\sum_{a \in A} r(a)\vec{x}_a$$

Subject to:

for all $s \in S$: (4)

$$\vec{\mu}_s + \sum_{a \in A} \delta(a)(s)\vec{y}_a = \sum_{a \in A(s)} \vec{y}_a + \sum_{a \in A(s)} \vec{x}_a$$

for all $s \in S$:

$$\sum_{a \in A} \delta(a)(s)\vec{x}_a = \sum_{a \in A(s)} \vec{x}_a$$

$\vec{x}$: occupation measure in the limit
$\vec{y}_a$: expected number of taking action a during the transient phase

both flows subject to Kirchhof's law

Optimal strategy: f such that

- $\vec{x}_{f(s)} > 0$ if $s \in S_{\vec{x}}$
- $\vec{y}_{f(s)} > 0$ if $s \notin S_{\vec{x}}$

where $S_{\vec{x}} := \{s \in S \mid \sum_{a \in A(s)} \vec{x}_a > 0\}$

Multiple mean payoff

Optimize multiple mean payoffs MP_i, $i \in [n]$, in MDP:

- expectation [BBCFK]

$$\bigwedge_i \mathbb{E}[MP_i] \geq \mathbf{exp}_i$$

- satisfaction (quantiles, percentiles)
  - conjunctive [RRS]

$$\bigwedge_i \mathbb{P}[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$$

  - joint [BBCFK]

$$\mathbb{P}[\bigwedge_i MP_i \geq \mathbf{sat}_i] \geq prob$$

- **conjunctions** thereof [CKK,CR]

[BBCFK] T. Brázdil, V. Brožek, K. Chatterjee, V. Forejt, A. Kučera: *Two Views on Multiple Mean-Payoff Objectives in Markov Decision Processes (LICS'11)*
[RRS] M. Randour, J.-F. Raskin, O. Sankur: *Percentile Queries in Multi-Dimensional Markov Decision Processes (CAV'15)*
[CKK] K. Chatterjee, Z. Komárková, J. Křetínský: *Unifying Two Views on Multiple Mean-Payoff Objectives in Markov Decision Processes (LICS'15)*
[CR] L. Clemente, J.-F. Raskin: *Multidimensional beyond worst-case and almost sure problems for mean-payoff objectives (LICS'15)*

Examples

Example 1: Money investment

- >0 earning, <0 losing
- maximize expected mean payoff $\mathbb{E}[MP]$

Examples

Example 1: Money investment

- > 0 earning, < 0 losing
- maximize expected mean payoff $\mathbb{E}[MP]$
- maximize probability $\mathbb{P}[MP \geq 0]$

Examples

Example 1: Money investment

- >0 earning, <0 losing
- maximize expected mean payoff $\mathbb{E}[MP]$
- maximize probability $\mathbb{P}[MP \geq 0]$
- maximize $\mathbb{E}[MP]$ while ensuring $\mathbb{P}[MP \geq 0] \geq 0.95$

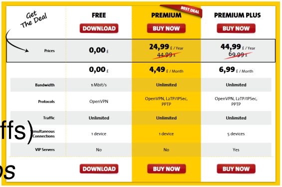
  "risk-averse" strategies

Examples

Example 1: Money investment

- > 0 earning, < 0 losing
- maximize expected mean payoff $\mathbb{E}[MP]$
- maximize probability $\mathbb{P}[MP \geq 0]$
- maximize $\mathbb{E}[MP]$ while ensuring $\mathbb{P}[MP \geq 0] \geq 0.95$

  "risk-averse" strategies

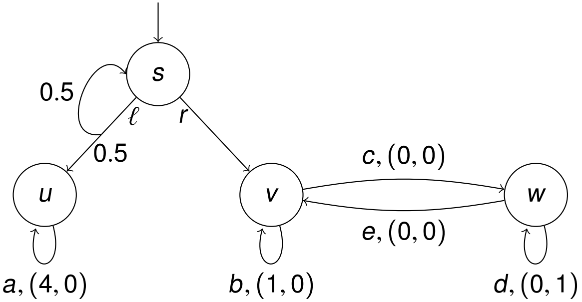


Get The Deal	FREE	PREMIUM	PREMIUM PLUS
	DOWNLOAD	BEST DEAL BUY NOW	BUY NOW
Prices	0,00 £	24,99 £ / Year ~~44,99 £~~	44,99 £ / Year ~~69,99 £~~
	0,00 £	4,49 £ / Month	6,99 £ / Month
Bandwidth	1 Mbit/s	Unlimited	Unlimited
Protocols	OpenVPN	OpenVPN, L2TP/IPSec, PPTP	OpenVPN, L2TP/IPSec, PPTP
Traffic	Unlimited	Unlimited	Unlimited
Simultaneous Connections	1 device	1 device	5 devices
VIP Servers	No	No	Yes
	DOWNLOAD	BUY NOW	BUY NOW

Example 2: Downloading service (multiple mean payoffs)

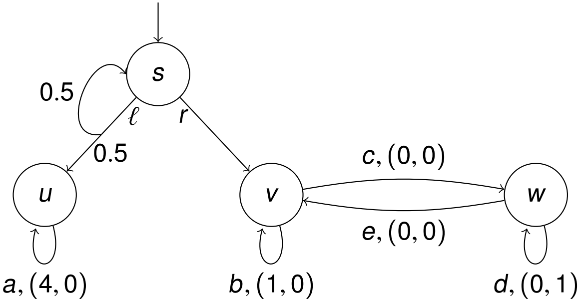
- gratis service: expected throughput $MP_1 = 1 Mbps$
- premium service: $\mathbb{E}[MP_2] = 10 Mbps$ and 95% connections run on $\geq 5 Mbps$; sold at p_2 per Mb
- need to hire MP_3 resources from a cloud each at price p_3
- while satisfying the guarantees, maximize $\mathbb{E}[p_2 \cdot MP_2 - p_3 \cdot MP_3]$

Example



$$\textbf{sat} = (0.5, 0.5), \textbf{prob} = (0.8, 0.8)$$

Example



exp $= (1.1, 0.5)$, **sat** $= (0.5, 0.5)$, **prob** $= (0.8, 0.8)$

Solution: Linear program I

Find strategy σ such that for all $i \in [n]$: $\mathbb{E}^{\sigma}[MP_i] \geq \mathbf{exp}_i$

Linear program [BBCFK]:

3. recurrent flow: for $s \in S$

$$\sum_{a \in A} x_a \cdot \delta(a)(s) = \sum_{a \in A(s)} x_a$$

4. expected rewards:

$$\sum_{a \in A} x_a \cdot \mathbf{r}(a) \geq \mathbf{exp}$$

Solution: Linear program I

Find strategy σ such that for all $i \in [n]$: $\mathbb{E}^{\sigma}[MP_i] \geq \mathbf{exp}_i$

Linear program [BBCFK]:

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_{a \in A} y_a \cdot \delta(a)(s) = \sum_{a \in A(s)} y_a + y_s$$

3. recurrent flow: for $s \in S$

$$\sum_{a \in A} x_a \cdot \delta(a)(s) = \sum_{a \in A(s)} x_a$$

4. expected rewards:

$$\sum_{a \in A} x_a \cdot \mathbf{r}(a) \geq \mathbf{exp}$$

Solution: Linear program I

Find strategy σ such that for all $i \in [n]$: $\mathbb{E}^{\sigma}[MP_i] \geq \mathbf{exp}_i$

Linear program [BBCFK]:

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_{a \in A} y_a \cdot \delta(a)(s) = \sum_{a \in A(s)} y_a + y_s$$

2. probability of switching in a MEC is the frequency of using its actions: for $C \in$ MEC

$$\sum_{s \in C} y_s = \sum_{a \in C} x_a$$

3. recurrent flow: for $s \in S$

$$\sum_{a \in A} x_a \cdot \delta(a)(s) = \sum_{a \in A(s)} x_a$$

4. expected rewards:

$$\sum_{a \in A} x_a \cdot \mathbf{r}(a) \geq \mathbf{exp}$$

Solution: Linear program II

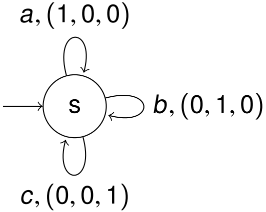
Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^{\sigma}[MP_i] \geq \mathbf{exp}_i$,
- $\mathbb{P}^{\sigma}[MP_i \geq \mathbf{sat}_i] \geq 1$.

5. almost-sure satisfaction: for $C \in \mathrm{MEC}$

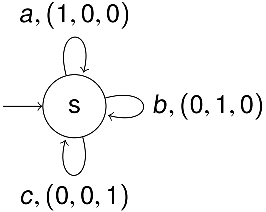
$$\sum_{a \in C} x_a \cdot \mathbf{r}(a) \geq \sum_{a \in C} x_a \cdot \mathbf{sat}$$

Difficulty: Multiple quantitative satisfaction



sat $= (1, 1, 1)$, **prob** $= (1/3, 1/3, 1/3)$

Difficulty: Multiple quantitative satisfaction



sat $= (1, 1, 1)$, **prob** $= (1/3, 1/3, 1/3)$

sat $= (1/2, 1/2, 1/2)$, **prob** $= (2/3, 2/3, 2/3)$

Solution: Linear program III

Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^\sigma[MP_i] \geq \mathbf{exp}_i$
- $\mathbb{P}^\sigma[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$

Idea: Split x_a into $x_{a,N}$ for $N \subseteq [n]$; similarly for y_s

3. MEC flow: for $s \in S$, $N \subseteq [n]$

$$\sum_a x_{a,N} \cdot \delta(a,s) = \sum_{a \in A(s)} x_{a,N}$$

Solution: Linear program III

Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^\sigma[MP_i] \geq \mathbf{exp}_i$
- $\mathbb{P}^\sigma[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$

Idea: Split x_a into $x_{a,N}$ for $N \subseteq [n]$; similarly for y_s

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_a y_a \cdot \delta(a, s) = \sum_{a \in A(s)} y_a + \sum_{N \subseteq [n]} y_{s,N}$$

2. frequencies of actions used in MECs: for $C \in \mathrm{MEC}, N \subseteq [n]$

$$\sum_{s \in C} y_{s,N} = \sum_{a \in C} x_{a,N}$$

3. MEC flow: for $s \in S, N \subseteq [n]$

$$\sum_a x_{a,N} \cdot \delta(a, s) = \sum_{a \in A(s)} x_{a,N}$$

Solution: Linear program III

Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^\sigma[MP_i] \geq \mathbf{exp}_i$
- $\mathbb{P}^\sigma[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$

Idea: Split x_a into $x_{a,N}$ for $N \subseteq [n]$; similarly for y_s

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_a y_a \cdot \delta(a,s) = \sum_{a \in A(s)} y_a + \sum_{N \subseteq [n]} y_{s,N}$$

2. frequencies of actions used in MECs: for $C \in \text{MEC}, N \subseteq [n]$

$$\sum_{s \in C} y_{s,N} = \sum_{a \in C} x_{a,N}$$

3. MEC flow: for $s \in S, N \subseteq [n]$

$$\sum_a x_{a,N} \cdot \delta(a,s) = \sum_{a \in A(s)} x_{a,N}$$

4. expected rewards:

$$\sum_{a \in A} \sum_{N \subseteq [n]} x_{a,N} \cdot \mathbf{r} \geq \mathbf{exp}$$

5. commitment to satisfaction: for $C \in \text{MEC}, \boxed{N \subseteq [n], i \in N}$

$$\sum_{a \in C} x_{a,N} \cdot \mathbf{r}_i(a) \geq \sum_{a \in C} x_{a,N} \cdot \mathbf{sat}_i$$

Solution: Linear program III

Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^\sigma[MP_i] \geq \mathbf{exp}_i$
- $\mathbb{P}^\sigma[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$

Idea: Split x_a into $x_{a,N}$ for $N \subseteq [n]$; similarly for y_s

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_a y_a \cdot \delta(a,s) = \sum_{a \in A(s)} y_a + \sum_{N \subseteq [n]} y_{s,N}$$

2. frequencies of actions used in MECs: for $C \in \text{MEC}$, $N \subseteq [n]$

$$\sum_{s \in C} y_{s,N} = \sum_{a \in C} x_{a,N}$$

3. MEC flow: for $s \in S$, $N \subseteq [n]$

$$\sum_a x_{a,N} \cdot \delta(a,s) = \sum_{a \in A(s)} x_{a,N}$$

4. expected rewards:

$$\sum_{a \in A} \sum_{N \subseteq [n]} x_{a,N} \cdot \mathbf{r} \geq \mathbf{exp}$$

5. commitment to satisfaction: for $C \in \text{MEC}$, $\boxed{N \subseteq [n], i \in N}$

$$\sum_{a \in C} x_{a,N} \cdot \mathbf{r}_i(a) \geq \sum_{a \in C} x_{a,N} \cdot \mathbf{sat}_i$$

6. satisfaction: for $i \in [n]$

$$\sum_{a \in A} \sum_{N \ni i} x_{a,N} \geq \mathbf{prob}_i$$

Solution: Linear program III

Find strategy σ such that for all $i \in [n]$

- $\mathbb{E}^\sigma[MP_i] \geq \mathbf{exp}_i$
- $\mathbb{P}^\sigma[MP_i \geq \mathbf{sat}_i] \geq \mathbf{prob}_i$

Idea: Split x_a into $x_{a,N}$ for $N \subseteq [n]$; similarly for y_s

1. transient flow: for $s \in S$

$$\vec{1}_{s_0}(s) + \sum_a y_a \cdot \delta(a,s) = \sum_{a \in A(s)} y_a + \sum_{N \subseteq [n]} y_{s,N}$$

2. frequencies of actions used in MECs: for $C \in \text{MEC}$, $N \subseteq [n]$

$$\sum_{s \in C} y_{s,N} = \sum_{a \in C} x_{a,N}$$

3. MEC flow: for $s \in S$, $N \subseteq [n]$

$$\sum_a x_{a,N} \cdot \delta(a,s) = \sum_{a \in A(s)} x_{a,N}$$

4. expected rewards:

$$\sum_{a \in A} \sum_{N \subseteq [n]} x_{a,N} \cdot \mathbf{r} \geq \mathbf{exp}$$

5. commitment to satisfaction: for $C \in \text{MEC}$, $\boxed{N \subseteq [n], i \in N}$

$$\sum_{a \in C} x_{a,N} \cdot \mathbf{r}_i(a) \geq \sum_{a \in C} x_{a,N} \cdot \mathbf{sat}_i$$

6. satisfaction: for $i \in [n]$

$$\sum_{a \in A} \sum_{N \ni i} x_{a,N} \geq \mathbf{prob}_i$$

Complexity I

Case	Alg. c.	Witness strat. c.	ε-witness strat. c.
single	$\mathit{poly}(\lvert G\rvert)$	det. 1-mem.	det. 1-mem.
multiple	$\mathit{poly}(\lvert G\rvert, n)$	rand. ∞-mem.	rand. 2-mem.
multiple combined	$\mathit{poly}(\lvert G\rvert, 2^n)$	rand. ∞-mem.	rand. $\leq 2^n$-mem.
	NP-hard		$\geq n$-mem.

Optimization algorithms

- single linear program
  $\Rightarrow$
  can optimize thresholds, linear combinations of expectations etc.
- ε-approximation of Pareto curve
  - polynomial in MDP size
  - polynomial in $1/\varepsilon$
  - exponential in dimension

“conjunctive satisfaction” with “joint satisfaction” is NP-hard:

Complexity II

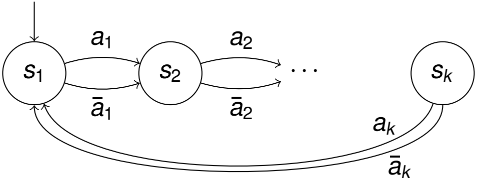
"conjunctive satisfaction" with "joint satisfaction" is NP-hard:

formula φ with clauses $C = \{c_1, \dots, c_k\}$

- $\mathbf{r}_{c_i}(\ell) = \begin{cases} 1 & \text{if } \ell \models c_i \\ 0 & \text{if } \ell \not\models c_i \end{cases}$

- $\mathbf{r}_a(\ell) = \begin{cases} 1 & \text{if } \ell = a_i \\ -1 & \text{if } \ell = \bar{a}_i \\ 0 & \text{otherwise} \end{cases}$

- $\mathbf{r}_{\bar{a}}(\ell) = \begin{cases} -1 & \text{if } \ell = a_i \\ 1 & \text{if } \ell = \bar{a}_i \\ 0 & \text{otherwise} \end{cases}$



$$\mathbb{P}^\sigma[MP_\ell \geq \frac{1}{k}] \geq \frac{1}{2} \qquad \text{for each } \ell \in Ap \cup \overline{Ap}$$

$$\mathbb{P}^\sigma[\bigwedge_{c \in C} MP_c \geq \frac{1}{k}] \geq \frac{1}{2}$$

Conclusion

Summary

- maximizing discounted/total/average reward in MDP/games/stoch.games
- expectation, satisfaction, combinations (risk averse), multiple resources
- value iteration, strategy iteration, linear programming
- feasible and practically useful